

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**PEKİŞTİRMELİ ÖĞRENME İLE YÜZ GÖRÜNTÜSÜNÜN
YÜKSEK ÇÖZÜNÜRLÜKTE ELDE EDİLMESİ**

DOKTORA TEZİ

Emre ALTINKAYA

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Bilim Dalı

OCAK 2025

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**PEKİŞTİRMELİ ÖĞRENME İLE YÜZ GÖRÜNTÜSÜNÜN
YÜKSEK ÇÖZÜNÜRLÜKTE ELDE EDİLMESİ**

DOKTORA TEZİ

Emre ALTINKAYA

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Bilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Burhan BARAKLI

OCAK 2025

Emre ALTINKAYA tarafından hazırlanan “Pekiştirmeli Öğrenme İle Yüz Görüntüsünün Yüksek Çözünürlükte Elde Edilmesi” adlı tez çalışması 23.01.2025 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalı Elektrik Mühendisliği Bilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	Doç. Dr. Gökçen ÇETİNEL Sakarya Üniversitesi	
Jüri Üyesi :	Dr. Öğr. Üyesi Burhan BARAKLI Sakarya Üniversitesi	
Jüri Üyesi :	Dr. Öğr. Üyesi Tuğrul TAŞCI Sakarya Üniversitesi	
Jüri Üyesi :	Doç. Dr. Ali Furkan KAMANLI Sakarya Uygulamalı Bilimler Üniversitesi	
Jüri Üyesi :	Dr. Öğr. Üyesi Selman HIZAL Sakarya Uygulamalı Bilimler Üniversitesi	



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “Pekiştirmeli Öğrenme İle Yüz Görüntüsünün Yüksek Çözünürlükte Elde Edilmesi” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(...../...../20.....).

Emre ALTINKAYA





Eşime ...



TEŞEKKÜR

Tez çalışmamın hazırlanmasında bana yol gösteren, bilgi ve deneyimlerini esirgmeden paylaştan saygıdeğer danışman hocam Burhan BARAKLI'ya en derin şükranlarımı sunuyorum. Kendisi, sadece bilimsel anlamda değil, kişisel gelişimim konusunda da bana her zaman inançla yaklaştı ve bu inancını her fırsatta hissettirdi. Kendi potansiyelimi görmemi sağlayarak bu süreci en verimli şekilde geçirmeme katkıda bulundu. Hocamın rehberliği, sabrı ve teşviki sayesinde, karşılaştığım zorlukları aşmayı ve bu çalışmayı tamamlamayı başardım.

Bana her anlamda destek olan ve varlığıyla bana güç veren sevgili eşim Elif ZOROĞLU ALTINKAYA'ya sonsuz teşekkürlerimi sunuyorum. Kendisi, bu süreçte gösterdiği sabır, anlayış ve teşvikle en büyük dayanağım olmuştur.

Hayatım boyunca yanımda olan, bana her zaman inanan ve sevgilerini esirgemeyen değerli aileme gönülden teşekkür ederim. Onların varlığı, bu sürecin büyük bir motivasyon kaynağı olmuştur.

Ayrıca, sıcak bir yuva olarak beni bağrılarına basan ve artık kendi ailem gibi gördüğüm eşimin değerli ailesine de bu süreçteki destekleri ve iyi dilekleri için teşekkürlerimi sunarım.

Süreç boyunca yanımda olan, moral veren ve her daim desteğini hissettiren sevgili arkadaşlarıma da teşekkür etmek isterim. Onların dostluğu ve motivasyonu, bu zorlu sürecin daha kolay ve keyifli geçmesini sağladı.

Bu çalışmayı tamamlamamda emeği, sevgisi ve desteği olan herkese şükranlarımı sunarım.

Emre ALTINKAYA



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxv
1. GİRİŞ	1
2. DERİN PEKİŞTİRMELİ ÖĞRENME VE DİKKAT MEKANİZMASI	9
2.1. Süper Çözünürlükte Temel Makine Öğrenimi Yapıları	9
2.2. Pekİştİrmeli Öğrenme ve Derin Pekİştİrmeli Öğrenme	11
2.3. Pekİştİrmeli Öğrenme Algorİtmaları	16
2.4. Tekrarlayan Sinir Ağı (Recurrent Neural Network - RNN)	22
2.5. Görsel Dikkat Mekanizması (Visual Attention Mechanism)	28
3. GÖRÜNTÜ SÜPER ÇÖZÜNÜRLÜĞÜ	31
3.1. Problem Tanımı	31
3.2. Literatür Araştırması	31
3.3. Derin Öğrenme Mekanizmalarına Genel Bakış	43
4. DERİN PEKİŞTİRMELİ ÖĞRENME İLE YÜZ SÜPER ÇÖZÜNÜRLÜĞÜ	47
4.1. Yüz Bölgelerinin Derin Pekİştİrmeli Öğrenme Tabanlı Süper Çözünürlüğü (DRL-SRFR) Ağı (Deep Reinforcement Learning Based Super Resolution of Face Regions (DRL-SRFR) Network)	50
4.1.1. Model detayları	51
4.1.2. Pekİştİrmeli yöntem için durum, aksiyon ve ödülün tanımı	53
4.1.3. FSR yapısı ve RRDB	60
4.1.4. DRL yapısı	64
4.2. Model Eğİtİmi	69
5. DENEYSEL ÇALIŞMALAR VE BULGULAR	73
5.1. Giriş	73
5.2. Veri Setleri ve Uygulama Detayları	74
5.2.1. Eğİtİm ve test veri setleri	74
5.2.2. Uygulama detayları	78
5.3. Performans Sonuçları	80
5.3.1. Değerlendirme ölçütleri	80
5.3.2. Görsel karşılaştırmalar	82
5.3.3. Ablasyon çalışmaları	100
5.4. Optimum Hiperparametrelerin Ayarlanması	107
5.5. Yöntemlerin Karşılaştırılması	111
5.5.1. Genel değerlendirme	111

5.5.2. Yöntemler arası farklılıklar	112
5.5.3. Önerilen modelin gücü	117
6. SONUÇLAR	119
KAYNAKLAR.....	123
ÖZGEÇMİŞ.....	133



KISALTMALAR

ADAM	: Adaptive Moment Estimation (Uyarlanabilir Moment Tahmini)
CNN	: Convolutional Neural Network (Evrişimli Sinir Ağı)
DL	: Deep Learning (Derin Öğrenme)
DRL	: Deep Reinforcement Learning (Derin Pekiştirmeli Öğrenme)
ESRGAN	: Enhanced Super-ResolutionGAN (Geliştirilmiş Süper Çözünürlüklü)
GAN	: Generative Adversarial Network (Üretken Çekişmeli Ağ)
FH	: Face Hallucination (Yüz Halüsinasyonu)
FSR	: Face Super-Resolution (Yüz Süper Çözünürlüğü)
GPU	: Graphics Processing Unit (Grafik İşlem Birimi)
GRU	: Gated Recurrent Unit (Geçitli Tekrarlayan Birim)
GT	: Ground Truth (Hedef Yüksek Çözünürlüklü Görüntü)
HR	: High-Resolution (Yüksek Çözünürlük)
LR	: Low-Resolution (Düşük Çözünürlük)
LSTM	: Long Short-Term Memory (Uzun Kısa Süreli Bellek)
MAE	: Mean Absolute Error (Ortalama Mutlak Hata)
MEAN	: Mean Value (Ortalama Değer)
MSE	: Mean Squared Error (Ortalama Karese Hata)
PSNR	: Peak Signal-to-Noise Ratio (Tepe Sinyal-Gürültü Oranı)
RAM	: Recurrent Attention Mechanism (Tekrarlayan Dikkat Mekanizması)
ReLU	: Rectified Linear Unit (Doğrultulmuş Lineer Ünite)
RGB	: Red, Green, Blue (Kırmızı, Yeşil, Mavi)
RNN	: Recurrent Neural Network (Tekrarlayan Sinir Ağı)
ROI	: Region of Interest (İlgi Bölgesi)
RRDB	: Residual-in-Residual Dense Block (Artık İçinde Artık Yoğun Blok)
SGD	: Stochastic Gradient Descent (Stokastik Gradyan İnişi)
SR	: Super-Resolution (Süper Çözünürlük)
SSIM	: Structural Similarity Index Measure (Yapısal Benzerlik Endeksi)
STD	: Standard Deviation (Standart Sapma)



SİMGELER

α	: Öğrenme Oranı (Learning Rate)
δ	: Delta (Fark Değeri - Difference Value)
ε	: Şekil Değiştirme (Strain)
f	: Frekans (Frequency) [Hz]
γ	: Ödül Fonksiyonu Katsayısı (Reward Function Coefficient)
h	: Gizli Durum (Hidden State)
λ	: Regularization Parametresi (Regularization Parameter)
Π	: Politika Fonksiyonu (Policy Function)
R	: Rezidüel Değer (Residual Value)
$\mathbf{R}$	: Ödül
t	: Zaman [s]
u, v	: Görüntüde seçilen parçaların piksellerinin eksen konum bilgisi
m, n	: Görüntünün eksen konum bilgisi
∇	: Gradyan
ψ	: Aktivasyon Fonksiyonu
Φ	: Potansiyel Fonksiyon
σ	: Keşif ve sömürü dengesini belirleyen parametre



TABLO LİSTESİ

Sayfa

Tablo 2.1. Pekiştirmeli öğrenme örnekleri.	14
Tablo 4.1. DRL ve FSR modellerinin giriş ve çıkışları.....	52
Tablo 4.2. RRDB süreç formülleri.	63
Tablo 5.1. Uygulama parametreleri.....	79
Tablo 5.2. Karşılaştırılan yöntem parametreleri.....	80
Tablo 5.3. PSNR değerlerinin 4 ve 8 ölçek oranına göre karşılaştırılması.	84
Tablo 5.4. SSIM değerlerinin 4 ve 8 ölçek oranına göre karşılaştırılması.	85
Tablo 5.5. Yöntemin hesaplama yükü açısından karşılaştırılması.	94
Tablo 5.6. T 'nin farklı değerlerinin 4 ve 8 ölçekleme faktöründe karşılaştırılması.....	100
Tablo 5.7. Parça boyutu karşılaştırması.	104
Tablo 5.8. Farklı sayıda RRDB katmanının karşılaştırılması(Ölçekleme 4x).....	104



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Makine öğrenmesi üç temel yapı.	10
Şekil 2.2. Pekiştirmeli öğrenme genel bakış.	12
Şekil 2.3. Pekiştirmeli öğrenme döngü ve parametreleri.	13
Şekil 2.4. Derin pekiştirmeli öğrenme yapısı.	15
Şekil 2.5. Pekiştirmeli öğrenme algoritmaları.	17
Şekil 2.6. RNN, LSTM ve GRU mimarileri.	26
Şekil 3.1. Süper çözünürlük için DL tabanlı F modeli nihai çıktı.	31
Şekil 3.2. Artık öğrenme (Residual Learning) yapısı.	43
Şekil 3.3. Yoğun bağlantılar (Dense Connections) yapısı.	44
Şekil 3.4. Çok yollu öğrenme (Multi-Path Learning) yapısı.	44
Şekil 3.5. Dikkat Mekanizmaları (Attention Mechanisms).	45
Şekil 3.6. Dönüştürücü Tabanlı Modeller (Transformer-Based Models).	45
Şekil 4.1. Modelin temel işleyişi.	48
Şekil 4.2. Önerilen yöntemin genel yapısı.	51
Şekil 4.3. Durum modülü giriş ve çıkış akış diyagramı.	54
Şekil 4.4. Ödül fonksiyon girdileri.	58
Şekil 4.5. FSR yapı akışı.	61
Şekil 4.6. $FFSR$ modelindeki iyileştirme ağı.	62
Şekil 4.7. DRL yapısı ve durum oluşturma akışı.	64
Şekil 4.8. Politika için durum katmanını oluşumu.	65
Şekil 4.9. Politika ağıtarafından üretilen skor değerler ve çıkış.	66
Şekil 4.10. Tekrarlayan yapının son üç adımının görsel analizi.	70
Şekil 5.1. CelebA veri setinden bazı örnek görseller.	74
Şekil 5.2. BioID veri setinden bazı örnek görseller.	75
Şekil 5.3. LFW veri setinden bazı örnek görseller.	76
Şekil 5.4. Pubfig veri setinden bazı örnek görseller.	77
Şekil 5.5. Ölçekleme faktörü 8 için LFW veri seti üzeri farklı yöntem çıktıları.. Ölçekleme faktörü 8 için LFW veri seti üzeri farklı yöntem çıktıları.	88
Şekil 5.6. Ölçekleme faktörü 4 için LFW veri seti üzeri farklı yöntem çıktıları.	90
Şekil 5.7. BioID, CelebA, LFW ve PubFig veri kümelerinde ölçek 4 için DRL-SRFR modelinin çıktıları. (a) LR girdileri. (b) Temel gerçek. (c) DRL-SRFR. .	91
Şekil 5.8. DRL-SRFR modelimizin PubFig veri kümeleri üzerinde ölçek 4 için küresel / yerel görsel ve nicel çıktıları.	93
Şekil 5.9. Seçilen parçaların örnek bir gösterimi.	95
Şekil 5.10. Seçilen bir parçanın orijinal, düşük çözünürlük ve model çıktısı halleri.	97
Şekil 5.11. Farklı eğitim epoklarında görsel çıktı ve parça dağılımı (8x).	98
Şekil 5.12. Farklı eğitim epoklarında görsel çıktı ve parça dağılımı (4x).	99
Şekil 5.13. İyileştirme aşamasındaki epoklara bağlı olarak seçilen parça haritası. .	102
Şekil 5.14. Eğitim anında RRDB katman sayısına bağlı ortalama PSNR değerleri.	105
Şekil 5.15. Eğitim anında RRDB katman sayısına bağlı ortalama SSIM değerleri.	106
Şekil 5.16. Hiperparametre değerlerine göre PSNR performans grafiği.	110

Şekil 5.17. SR3'ün U-Net mimarisi.	114
Şekil 5.18. Görüntü restorasyonu için önerilen SwinIR mimarisi.	116



PEKİŞTİRMELİ ÖĞRENME İLE YÜZ GÖRÜNTÜSÜNÜN YÜKSEK ÇÖZÜNÜRLÜKTE ELDE EDİLMESİ

ÖZET

Pekiştirmeli öğrenme ile yüz görüntüsünün yüksek çözünürlükte elde edilmesini amaçlayan çalışma, düşük çözünürlüklü yüz görüntülerinin yüksek çözünürlüklü hale getirilmesi sürecinde derin pekiştirmeli öğrenme (DRL) ve artık içinde artık yoğun bloklar (RRDB) tabanlı bir yöntemi bir araya getiren yenilikçi bir yaklaşım önermektedir. Günümüzde yüz süper çözünürlüğü (FSR), biyometrik güvenlik, yüz tanıma sistemleri, adli bilişim, sağlık sektörü, dijital medya ve eğlence endüstrisi gibi birçok kritik alanda büyük bir öneme sahiptir. Düşük kaliteli yüz görüntülerinin iyileştirilmesi, özellikle güvenlik kameraları, eski fotoğraflar ve düşük çözünürlükte çekilmiş dijital görüntüler gibi çeşitli kaynaklardan elde edilen verilerin analizinde kritik bir gereklilik haline gelmektedir. Mevcut süper çözünürlük teknikleri, büyük ölçüde evrimsel sinir ağları (CNN), üretken çekişmeli ağlar (GAN) ve derin öğrenme tabanlı yöntemlere dayanmaktadır. Ancak yöntemlerin çoğu, yüz görüntülerinin belirli bölgelerinde meydana gelen detay kayıplarını dikkate almadan, tüm görüntüyü eşit şekilde işleyerek iyileştirme yapmaktadır. Böylece yüz görüntülerinin belirli kritik bölgelerinde detay kaybı oluşabilmekte ve genel görsel bütünlük bozulabilmektedir. Önerilen çalışmada yüz görüntülerinin önemli bölgeleri belirlenerek seçilen parçaların iteratif olarak iyileştirilmesini sağlayan, pekiştirmeli öğrenme tabanlı bir dikkat mekanizması ile desteklenmiş süper çözünürlük modeli geliştirilmiştir.

Önerilen model, yüz görüntülerinde hangi bölgelerin iyileştirilmesi gerektiğini öğrenen bir mekanizma kullanarak, süper çözünürlük problemini dinamik bir süreç haline getirmektedir. Geleneksel yöntemler tüm yüz görüntüsüne aynı şekilde yaklaşarak detay iyileştirmeleri yapmaya çalışmaktadır. Önerilen model yüz görüntüsündeki farklı bölgelerin farklı seviyelerde detaylandırılması gerektiğini öğrenen bir karar alma sürecine sahiptir. Modelin dikkat mekanizması, yüz görüntüsünde belirli bölgeleri iteratif olarak seçmekte ve her iterasyonda bu bölgelerin süper çözünürlük işlemiyle yeniden yapılandırılmasını sağlamaktadır. Süreç, belirlenen her yeni yüz bölgesinin iyileştirilmesi ve önceki iterasyonlarda iyileştirilmiş bölgelerle bütünleşerek yüz görüntüsünün küresel yapısını koruması ile devam etmektedir.

Önerilen modelin temel bileşenlerinden biri pekiştirmeli öğrenme tabanlı stratejik parça seçim mekanizmasıdır. Mekanizma yüz görüntüsündeki kritik detayları belirleyerek, modelin hangi bölgelerin daha fazla iyileştirilmesi gerektiğini anlamasını sağlamaktadır. Geleneksel süper çözünürlük yöntemlerinde görüntüler bütün olarak işlenirken, önerilen model dikkat mekanizmasını kullanarak her iterasyonda belirli bir yüz bölgesine odaklanmakta ve iyileştirme işlemi gerçekleştirmektedir. Model her adımda yüz görüntüsündeki eksik detayları analiz edebilme yeteneğine ulaşmaktadır. Model iyileştirilmesi gereken bölgeleri belirler ve RRDB tabanlı bir süper çözünürlük ağı ile bu bölgeleri detaylandırarak görüntüye

geri entegre etmektedir. Modelin buradaki süreci, her iterasyonda yeni bir yüz bölgesinin seçilmesi ve iyileştirilmesi ile devam etmektedir. Amaç iyileştirilmiş yüksek çözünürlüklü bir yüz görüntüsü elde etmektir.

Modelin en büyük yeniliklerinden biri stokastik eylem seçimi mekanizmasıdır. Geleneksel süper çözünürlük yöntemleri deterministik (önceden belirlenmiş) kurallara göre hareket ederek görüntü iyileştirme işlemlerini yürütmektedir. Önerilen model, hangi yüz bölgesinin iyileştirileceğine dair seçimleri stokastik (olasılıklı) bir süreç ile karar vermektedir. Stokastik süreç, modelin her görüntüye özgü en iyi iyileştirme stratejisini öğrenmesine olanak tanımaktadır. Stokastik eylem seçimi sayesinde model, farklı yüz yapılarında en kritik detayları belirleyerek her bir görüntüye özel bir iyileştirme stratejisi oluşturabilmektedir. Böylece, model sabit bir iyileştirme algoritması yerine her yüz görüntüsünün farklı gereksinimlerine göre dinamik olarak şekillenebilen bir süper çözünürlük yöntemi sunmaktadır.

Modelin süper çözünürlük aşaması, artık içinde artık yoğun bloklar (RRDB) tabanlı bir ağ kullanılarak gerçekleştirilmektedir. RRDB yapısı yoğun bağlantılar ve artık öğrenme mekanizmaları sayesinde düşük çözünürlüklü yüz görüntülerindeki eksik detayları etkili bir şekilde geri kazandırmaktadır. Her iterasyonda model, seçilen yüz bölgesinin detaylarını artırarak yüksek çözünürlüklü hale getirmektedir. Ardından seçilen bölgeyi düşük çözünürlüklü yüz görüntüsüne entegre etmektedir. Böylece, yüzün hem genel yapısı korunmakta hem de en önemli detayların iyileştirilmesi sağlanmaktadır. Modelin başarımı yüz görüntüleri veri setlerinden yaygın olarak kullanılan LFW, CelebA, BioID ve PubFig veri setleri üzerinde test edilerek değerlendirilmektedir. Sonuçlar modelin mevcut süper çözünürlük yöntemlerine kıyasla hem görsel hem de sayısal anlamda üstün performans sergilediğini göstermektedir. Özellikle PSNR (Peak Signal-to-Noise Ratio) ve SSIM (Structural Similarity Index Measure) gibi metrikler kullanılarak yapılan analizlerde, modelin yüz görüntülerindeki detayları koruma ve küresel yapıyı iyileştirme konusunda oldukça başarılı olduğu görülmektedir. PSNR, iki görüntü arasındaki hata seviyesini ölçen bir metrik olup, düşük çözünürlüklü görüntü ile yüksek çözünürlüklü tahmin edilen görüntü arasındaki farkın ne kadar küçük olduğunu sayısal olarak ifade etmektedir. Daha yüksek PSNR değerleri daha düşük hata oranını ve dolayısıyla görüntünün yüksek çözünürlüğe daha yakın olduğunu göstermektedir. Ancak PSNR metriği tek başına bir görüntünün kalitesini tam olarak ifade etmekte yetersiz kalabilmektedir. Dolayısıyla süper çözünürlük gibi yüksek düzeyde görsel detay gerektiren uygulamalarda SSIM metriği ile desteklenmesi kritik bir önem taşımaktadır.

SSIM, yalnızca pikseller arasındaki doğrudan farkı değil aynı zamanda görüntüdeki yapısal benzerlikleri, parlaklık, kontrast ve kenar detaylarını da göz önünde bulundurarak görüntü kalitesini değerlendiren bir ölçüttür. İnsan görsel algısı mutlak piksel farklarından ziyade yapısal bilginin korunmasına daha duyarlı olduğu için SSIM metriği süper çözünürlük değerlendirmelerinde en kritik kalite ölçütlerinden biri olarak kabul edilmektedir. Özellikle yüz görüntülerinde gözler, burun, ağız gibi kritik bölgelerin detaylarının korunması süper çözünürlük başarısını belirleyen en önemli faktörlerden biridir.

Çalışmada önerilen model, hem PSNR açısından düşük hata oranına sahip olmasıyla hem de SSIM değeri yüksek çıkacak şekilde yüzün küresel yapısını koruyarak detayları iyileştirmesiyle mevcut süper çözünürlük yöntemlerine kıyasla üstün bir performans sergilemektedir. Özellikle geleneksel yöntemlerle üretilen süper

özünürlüklü yüz görüntülerinde sıka karşılaşılan bulanık kenarlar, yapay detay ekleme ve yüzün doğal yapısının bozulması gibi sorunlar, önerilen modelin dikkat mekanizması ve iteratif süper özünürlük yaklaşımı sayesinde minimize edilmiştir. SSIM değerlendirmeleri modelin yüz bütünlüğünü koruyarak, detayları en iyi şekilde iyileştirdiğini ve insan algısına daha uygun sonuçlar ürettiğini kanıtlamaktadır.

Önerilen alışma yüz süper özünürlüğü alanında pekiştirmeli öğrenme tabanlı süper özünürlük yöntemlerinin uygulanabilirliğini artırarak, biyometrik güvenlik, dijital medya, adli bilişim ve gözetim sistemleri gibi birçok alanda yenilikçi uygulamalar sunmaktadır. Model, yalnızca yerel detayları değil aynı zamanda küresel yapı bütünlüğünü de dikkate alarak daha gerçekçi ve doğal sonuçlar üretmektedir. Özellikle güvenlik sistemleri, eski fotoğrafların iyileştirilmesi, düşük kaliteli güvenlik kamerası görüntülerinin netleştirilmesi ve biyometrik doğrulama gibi alanlarda modelin önemli bir katkı sunacağı öngörülmektedir.

Modelin performansını daha da artırmak için çeşitli geliştirmeler yapılması olasıdır. Öğrenme oranının adaptif olarak ayarlanması, modelin farklı veri setleri için en iyi sonucu vermesine olanak sağlayabilir. Bunun yanı sıra, üretken ekişmeli ağlar (GAN) ve dönüştürücü tabanlı modeller (Transformer-Based Models) gibi farklı derin öğrenme teknikleriyle modelin hibrit bir yapıya entegre edilmesi, süper özünürlük başarısını daha da artırabilir. Ayrıca, daha geniş veri setleri üzerinde testler yapılarak modelin genelleme yeteneği artırılabilir ve gerçek dünya uygulamalarına daha uygun hale getirilebilir.

Sonuç olarak önerilen alışma, yüz süper özünürlüğü problemini özmek için derin pekiştirmeli öğrenme ve artık içinde artık yoğun bloklar (RRDB) tabanlı bir model önermektedir. Stratejik para seçimi mekanizması ile yüz görüntüsünün en kritik bölgeleri belirlenmekte, detaylar iteratif olarak iyileştirilmekte ve küresel yapı bütünlüğü korunmaktadır. Modelin geniş kullanım alanı ve üstün başarımı göz önüne alındığında yüz süper özünürlüğü ve görüntü iyileştirme alanında önemli bir ilerleme sunduğu söylenebilmektedir.



HIGH RESOLUTION FACIAL IMAGE ACQUISITION WITH REINFORCEMENT LEARNING

SUMMARY

This study proposes an innovative approach that combines deep reinforcement learning (DRL) and a residual dense blocks in residuals (RRDB) based method to convert low-resolution face images into high-resolution ones. Nowadays, facial super resolution (FSR) is of great importance in many critical areas such as biometric security, face recognition systems, forensics, healthcare, digital media and entertainment industry. The enhancement of low-quality facial images is becoming a critical requirement, especially in analyzing data from a variety of sources such as surveillance cameras, old photographs, and low-resolution digital images. Existing super-resolution techniques are largely based on convolutional neural networks (CNN), generative adversarial networks (GAN) and deep learning-based methods. However, most of the methods do not take into account the loss of detail in certain regions of the face images and process the whole image equally. Thus, detail loss may occur in certain critical regions of the face images and the overall visual integrity may be compromised. In the proposed work, we develop a super-resolution model supported by a reinforcement learning-based attention mechanism that identifies important regions of facial images and iteratively enhances the selected parts.

The proposed model makes the super-resolution problem a dynamic process by using a mechanism that learns which regions of the face images need to be enhanced. Traditional methods approach the whole face image in the same way and try to improve the detail. The proposed model has a decision-making process that learns that different regions in the face image should be enhanced at different levels of detail. The attention mechanism of the model iteratively selects certain regions in the face image and reconstructs them with a super-resolution process at each iteration. The process continues by enhancing each new face region and integrating it with the regions enhanced in the previous iterations to maintain the global structure of the face image.

One of the key components of the proposed model is a reinforcement learning-based strategic part selection mechanism. By identifying critical details in the face image, the mechanism enables the model to understand which regions need further enhancement. While traditional super-resolution methods process images as a whole, the proposed model uses the attention mechanism to focus on a specific facial region and perform enhancement at each iteration. At each step, the model achieves the ability to analyze the missing details in the face image. The model identifies the regions that need to be improved and integrates them back into the image by detailing these regions with an RRDB-based super-resolution network. The process of the model here continues with the selection and enhancement of a new face region in each iteration. The goal is to obtain an improved high-resolution face image.

One of the major innovations of the model is the stochastic action selection mechanism. Traditional super-resolution methods perform image enhancement by acting according to deterministic (predetermined) rules. The proposed model uses a stochastic (probabilistic) process to decide which facial region to enhance. The stochastic process allows the model to learn the best enhancement strategy for each image. Thanks to the stochastic action selection, the model is able to identify the most critical details in different facial structures and create an enhancement strategy specific to each image. Thus, instead of a fixed enhancement algorithm, the model provides a super-resolution method that can be dynamically shaped according to the different requirements of each face image.

The super-resolution phase of the model is realized using a residual dense blocks within residuals (RRDB) based network. The RRDB structure efficiently recovers missing details in low-resolution face images through dense connections and residual learning mechanisms. In each iteration, the model increases the details of the selected face region to high resolution. It then integrates the selected region into the low-resolution face image. Thus, the overall structure of the face is preserved and the most important details are enhanced. The performance of the model is evaluated by testing it on the widely used face image datasets LFW, CelebA, BioID and PubFig. The results show that the model outperforms existing super-resolution methods both visually and numerically. In particular, the analysis using metrics such as PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) shows that the model is very successful in preserving the details in the face images and improving the global structure. PSNR is a metric that measures the level of error between two images and numerically expresses how small the difference between the low-resolution image and the high-resolution predicted image is. Higher PSNR values indicate a lower error rate and hence the image is closer to high resolution. However, the PSNR metric alone may not be sufficient to fully express the quality of an image. Therefore, in applications that require a high level of visual detail, such as super resolution, it is critical to support it with the SSIM metric.

SSIM is a metric that evaluates image quality by considering not only the direct difference between pixels but also the structural similarities, brightness, contrast and edge details in the image. Since human visual perception is more sensitive to the preservation of structural information rather than absolute pixel differences, the SSIM metric is considered one of the most critical quality measures in super-resolution evaluations. Especially in face images, preserving the details of critical regions such as eyes, nose and mouth is one of the most important factors determining the super-resolution success.

The model proposed in this study shows a superior performance compared to the existing super-resolution methods by having a low error rate in terms of PSNR and improving the details by preserving the spherical structure of the face with a high SSIM value. In particular, problems such as blurred edges, artificial detail addition and distortion of the natural structure of the face, which are frequently encountered in super-resolution face images produced by conventional methods, are minimized thanks to the attention mechanism and iterative super-resolution approach of the proposed model. SSIM evaluations prove that the model preserves the integrity of the face, enhances the details in the best way and produces results that are more suitable for human perception.

The proposed work improves the applicability of reinforcement learning-based super-resolution methods in the field of face super-resolution and offers innovative applications in many areas such as biometric security, digital media, forensics and surveillance systems. The model produces more realistic and natural results by taking into account not only local details but also global structural integrity. The model is expected to make a significant contribution especially in areas such as security systems, enhancement of old photographs, clarification of poor quality security camera images and biometric verification.

Various enhancements are possible to further improve the performance of the model. An adaptive adjustment of the learning rate could allow the model to perform optimally for different datasets. In addition, integrating the model into a hybrid structure with different deep learning techniques such as generative adversarial networks (GAN) and transformer-based models can further improve the super-resolution performance. Furthermore, tests on larger datasets can be conducted to increase the generalization capability of the model and make it more suitable for real-world applications.

In conclusion, the proposed work proposes a model based on deep reinforcement learning and residual dense blocks within residuals (RRDB) to solve the face super-resolution problem. With a strategic part selection mechanism, the most critical regions of the face image are identified and the details are iteratively improved while maintaining the global structure integrity. Considering the wide range of applications and superior performance of the model, it can be said that it offers a significant advance in the field of facial super-resolution and image enhancement.



1. GİRİŞ

Çoklu-ortam aygıtları geçmişten günümüze kadar hiç bu zamanki kadar hızla gelişmemiştir. Dijitalleşmenin de etkisi ile teknolojik araçlar hayatımızın her alanına nüfuz etmektedir. Akıllı telefonlar, bilgisayarlar, tabletler, kameralar ve diğer akıllı cihazlar gündelik yaşamın ayrılmaz bir parçası haline gelmiştir. Bulut tabanlı sistemler ve bireyler arası veri akışımı çok büyük boyutlara ulaşmıştır. Tek tıklamayla dünyanın herhangi bir yerinden veri alınabilmekte, görüntüler izlenebilmekte veya belgeler saniyeler içerisinde aktarılabilmektedir. Teknolojideki gelişmeler, iletişim ve bilgi paylaşımını kolaylaştırır da bazı sorunları beraberinde getirmektedir. Artan veri hacmi nedeni ile bilgi güvenliği ve bütünlüğü konusunda endişeler ve problemler ortaya çıkarmaktadır. Bilgiler çok hızlı bir şekilde paylaşılrken verilerin kaybolma, bozulma ya da izinsiz erişim riski de giderek artmaktadır. Verilerin doğru ve güvenilir bir şekilde saklanması ve işlenmesi zorlaşmaktadır. Özellikle görüntü ve video gibi büyük boyutlu verilerin saklanması ve işlenmesi her zaman mevcut teknolojideki donanımlarla ya da yöntemlerle mümkün olmamaktadır. Verinin sadece fiziksel olarak saklanması yeterli değildir. Aynı zamanda erişilebilir, bütünlüğü korunmuş ve bozulmalara karşı dirençli olması da büyük önem taşımaktadır. Özellikle büyük ölçekli veri akışının yönetimi, gelişmiş algoritmalar ve güçlü donanımlar gerektirmektedir. Veri merkezleri ve bulut tabanlı sistemler bu soruna çözüm üretmek adına yaygınlaşsa da büyük hacimli verilerin işlenmesi ve analiz edilmesi için daha güçlü yöntemlerin geliştirilmesine ihtiyaç duyulmaktadır.

Büyük boyutlu görsel içeriklerin işlenmesi ve saklanması, yalnızca büyük depolama alanları gerektirmekle kalmaz, aynı zamanda yüksek hesaplama gücüne de ihtiyaç duymaktadır. Fakat bazı geleneksel yöntemler ile veri işlenmesi sırasında kayıpların meydana gelmesi kaçınılmazdır. Veri kayıpları özellikle tıp, güvenlik, gözetim ve medya gibi alanlarda kritik bir sorun haline gelmektedir. Verilerin detay kaybı olmadan ya da buna en az şekilde maruz kalarak işlenmesi ve gerektiğinde iyileştirilmesi de büyük önem taşımaktadır. Veri işlemede ortaya çıkan zorluklar bilgisayar işlem kapasitelerinin artmasıyla birlikte daha etkili veri işleme yöntemleri

ile aşılmaya çalışılmaktadır. Makine öğrenmesi, derin öğrenme ve yapay zeka destekli algoritmalar verinin daha hızlı ve verimli işlenmesine olanak tanımaktadır. Aynı zamanda görüntü ve video gibi karmaşık veri türleri üzerinde daha hassas analizler yapılması sağlanmaktadır. Özellikle görsel verilerin işlenmesi ve iyileştirilmesi, bilgi kaybını en aza indirmek ve daha yüksek kaliteli çıktılar elde etmek adına önemli bir araştırma alanı haline gelmiştir.

Görüntü işleme, dijital dünyada bilginin görsel formda saklanmasını, analiz edilmesini ve yorumlanmasını sağlayan temel bir alandır. Günümüzde akıllı telefonlardan güvenlik kameralarına, medikal cihazlardan uydu görüntüleme sistemlerine kadar geniş bir yelpazede kullanılmaktadır. Görüntü işleme alanı insan gözünün algıladığı dünyayı makineler tarafından anlamlandırılabilir hale getirmek istemektedir. Fotoğraflar, videolar, grafikler ve üç boyutlu görseller, çeşitli algoritmalar yardımıyla analiz edilerek daha net, daha anlamlı ve daha kullanılabilir hale getirilmektedir. Özellikle görüntü verilerinin önemli olduğu tıp, güvenlik, uzay araştırmaları ve yapay zeka gibi alanlarda görüntülerin kalitesini artırmak ve detayları daha belirgin hale getirmek büyük bir önem taşımaktadır.

Görüntüler, sahip oldukları detaylarla anlam kazanmaktadır. Bir görselin içerisindeki ince ayrıntılar bazen kritik kararların alınmasını sağlayan önemli bilgiler içerebilmektedir. Örneğin, bir tıbbi taramada görüntüdeki küçük bir anormallik, bir hastalığın erken teşhis edilmesini sağlayabilmektedir. Uydu görüntülerinde yüzey yapılarının incelenmesi, doğal afetlerin tespiti ya da şehir planlaması için hayati bilgiler sağlayabilmektedir. Güvenlik kameralarındaki yüz tanıma sistemleri, görüntüdeki en küçük detaylara bile duyarlıdır ve bu sayede bireylerin kimlikleri doğrulanabilmektedir. Ancak, görüntülerin kalitesini belirleyen en önemli faktörlerden biri, onların içerdiği detayları anlaşılabilmesidir. Düşük kaliteli bir görüntü, içerdiği bilgiyi eksik veya hatalı bir şekilde yansıtabilmektedir. Dolayısıyla, görüntülerdeki detayların korunması ve eksik kalan bilgilerin tamamlanması, görüntü işleme alanında önemli bir araştırma konusu haline gelmektedir.

Görüntü kalitesini belirleyen en temel faktörlerden biri çözünürlüktür. Çözünürlük, bir görüntünün sahip olduğu piksel sayısını ifade etse de, tek başına yüksek piksel sayısı her zaman daha kaliteli bir görüntü anlamına gelmemektedir. Çözünürlük, yalnızca piksel yoğunluğu ile değil aynı zamanda görüntüdeki detayların ne kadar net algılanabildiği, keskinlik, kontrast, renk doğruluğu ve gürültü seviyesi gibi

faktörlerle de doğrudan ilişkilidir. Örneğin, 1920x1080 piksel çözünürlüğe sahip bir görüntü, 1280x720 piksel çözünürlüğe sahip bir görüntüden daha fazla detay içerebilmektedir. Ancak bu yalnızca piksel sayısının artışıyla değil aynı zamanda görüntüdeki detayların belirginliğiyle de ilgilidir. İki farklı görüntü aynı piksel sayısına sahip olabilir, ancak kullanılan sensör kalitesi, görüntü işleme teknikleri ve sıkıştırma oranı gibi faktörler görsel algıyı tamamen farklı hale getirebilmektedir. Eğer yüksek çözünürlüklü bir görüntü aşırı sıkıştırılmışsa, gürültü oranı fazlaysa veya çekim sırasında bulanıklık oluşmuşsa görüntü kalitesi düşebilmekte ve detaylar kaybolabilmektedir. Ayrıca optik çözünürlük ve sinyal-gürültü oranı gibi faktörler de görüntü kalitesini doğrudan etkileyen unsurlar arasında yer almaktadır. Optik çözünürlük, yalnızca dijital olarak üretilen piksellerin değil, kullanılan lens ve sensör gibi fiziksel bileşenlerin de görüntü netliği üzerindeki etkisini ifade etmektedir. Kaliteli bir kamera, ışığı daha iyi yakalayarak ve detayları daha doğru şekilde işleyerek daha yüksek çözünürlüklü ve gerçekçi görüntüler üretebilmektedir. Öte yandan düşük kaliteli bir sensörle çekilmiş bir görüntü, piksel sayısı yüksek olsa bile bulanık ve detaylardan yoksun olabilmektedir. Benzer şekilde, gürültü oranı yüksek olan bir görüntü, daha fazla piksele sahip olsa bile detayları kaybolmuş, bozulmuş veya hatalar içeren bir görüntü haline gelebilmektedir. Dolayısıyla, çözünürlüğün yalnızca piksel sayısı ile tanımlanamayacağı açıktır. Piksel sayısı artırılarak görüntü boyut olarak büyütülebilmekte ancak bu durum kaybolan detayların kendiliğinden geri kazanılmasını sağlamamaktadır. Yüksek kaliteli bir görüntü elde etmek için kaybolan detayların yapay zeka ve görüntü işleme teknikleriyle düzenlenmesi, gürültü seviyesinin azaltılması ve sıkıştırma gibi nedenlerle kaybolan bilgilerin tahmin edilerek onarılması gerekmektedir.

Görüntü verileri kullanılan çalışma alanlarında düşük çözünürlüklü görüntülerin iyileştirilmesi ve daha yüksek çözünürlüğe ulaştırılması büyük bir ihtiyaç haline gelmektedir. Günümüzde birçok farklı teknik kullanılarak düşük çözünürlüklü görüntülerin kalitesi artırılmaya çalışılmaktadır. Geleneksel yöntemler, genellikle bulanıklığa neden olur ve kaybolan detayları geri getirememektedir. Dolayısıyla gelişmiş algoritmalar kullanılarak daha gerçekçi ve yüksek kaliteli görüntüler üretmek, görüntü işleme alanında önemli bir hedef haline gelmektedir. İhtiyaç doğrultusunda, özellikle süper çözünürlük (Super-Resolution, SR) adı verilen teknikler geliştirilmektedir. Süper çözünürlük, düşük çözünürlüklü görüntülerden

yüksek çözünürlüklü görüntü elde etmeyi hedefleyen yaklaşımdır ve geleneksel yöntemlere kıyasla çok daha başarılı sonuçlar elde etmeyi sağlamaktadır. Düşük kaliteli bir görüntü bulanıklık, gürültü veya sıkıştırma artefaktları içerebilmektedir. Süper çözünürlük teknikleri bu tür sorunları minimize ederek görüntüleri daha net ve kullanılabilir hale getirmeyi amaçlamaktadır. Bunun için görüntüde eksik kalan ayrıntılar tahmin edilerek tamamlanmakta ve görsel bütünlük korunmaya çalışılmaktadır. Gelişmiş görüntü işleme yöntemleri kullanılarak eksik bilgilerin anlamlı bir şekilde tamamlanması sağlanabilmektedir.

Süper çözünürlük teknolojisi, yalnızca belirli sektörlerde değil, dijital dünyada genel bir gereklilik haline gelmiştir. Günümüzde düşük çözünürlüklü eski fotoğrafların veya video kayıtlarının iyileştirilmesi, yapay zeka destekli görsel analiz sistemlerinde doğruluk oranının artırılması ve mobil cihazlarda daha yüksek kaliteli görüntü oluşturma süreçleri gibi birçok alanda süper çözünürlük önemli bir çözüm sunmaktadır. Geliştirilen yöntemler, yalnızca yeni nesil uygulamalar için değil, aynı zamanda geçmişte kaydedilmiş düşük kaliteli görüntülerin yeniden anlamlandırılması açısından da büyük bir potansiyel taşımaktadır. Özellikle artan veri işleme kapasitesi ve gelişmiş algoritmalar sayesinde süper çözünürlük teknikleri, giderek daha başarılı hale gelmekte ve görsel verinin daha etkin bir şekilde işlenmesini sağlamaktadır. Yapay zeka tabanlı yöntemlerin yaygınlaşmasıyla, düşük çözünürlüklü görüntülerin yalnızca büyütülmesi değil, aynı zamanda detaylarının korunarak ve eksik bilgilerin anlamlandırılarak iyileştirilmesi mümkün hale gelmektedir. Tıp, güvenlik, medya ve gözetleme sistemleri gibi birçok sektörde süper çözünürlüğün kullanımı yaygınlaşırken, belirli alanlarda daha spesifik çözümler geliştirilmesi büyük bir önem kazanmaktadır. Özellikle yüz tanıma ve kimlik doğrulama sistemlerinin gelişmesiyle birlikte yüz süper çözünürlüğü (Face Super-Resolution, FSR) adı verilen özel bir teknik ortaya çıkmış ve bu alanda yeni araştırmalar hız kazanmıştır.

Yüz süper çözünürlüğü, düşük çözünürlüklü yüz görüntülerini yüksek çözünürlüğe dönüştürerek yüz tanıma sistemlerinin performansını artırmayı amaçlamaktadır. Güvenlik kameralarından elde edilen düşük kaliteli yüz görüntülerinin iyileştirilmesi, adli analizlerde netlik kazandırılması ve biyometrik doğrulama sistemlerinde doğruluk oranının artırılması gibi kritik görevlerde bu teknoloji büyük bir rol oynamaktadır. Geleneksel yüz tanıma sistemleri yüksek kaliteli görüntüler üzerinde çalışmak üzere optimize edilmiş olsa da pratikte düşük çözünürlüklü veya bulanık

görüntüler üzerinde çalışması gerekebilmektedir. Dolayısıyla yüz süper çözünürlüğü teknikleri, kimlik doğrulama ve yüz analizi gibi sistemlerin başarısını doğrudan etkileyerek yapay zeka destekli güvenlik çözümlerinin daha güvenilir hale gelmesini sağlamaktadır. Ayrıca, yüz süper çözünürlüğü yalnızca güvenlik ve biyometri alanlarında değil medya, dijital sanat ve eğlence sektörlerinde de önemli bir yere sahiptir. Eski fotoğrafların veya video kayıtlarının netleştirilmesi, düşük çözünürlükte çekilmiş görsellerin kaliteli hale getirilmesi gibi uygulamalar, yüz süper çözünürlüğünün kullanım alanlarını genişletmektedir. Yapay zeka tabanlı gelişmiş teknikler sayesinde yüz detayları daha gerçekçi bir şekilde geri kazanılmakta ve insan algısına en uygun sonuçlar elde edilmektedir. Süper çözünürlük teknolojisinin yüz görüntülerine uyarlanmasıyla ortaya çıkan bu yenilikçi yaklaşımlar hem bireysel hem de kurumsal kullanımlar için büyük faydalar sunmaktadır.

Yüz süper çözünürlüğü literatürde yüz halüsinasyonu (Face Hallucination, FH) olarak adlandırılan bir kavramla yakından ilişkilidir. Yüz halüsinasyonu, düşük kaliteli veya eksik bilgileri içeren yüz görüntülerinin yapay zeka ve görüntü işleme teknikleri kullanılarak yüksek çözünürlüklü ve detaylı hale getirilmesini ifade etmektedir. Süper çözünürlük teknikleri genellikle yalnızca var olan pikselleri analiz ederek görüntüleri iyileştirmeye çalışırken, yüz halüsinasyonu yöntemleri yüzün yapısal özelliklerine dair eksik bilgileri de tahmin ederek görsel bütünlüğü tamamlamayı hedeflemektedir. Yüz halüsinasyonu, özellikle kimlik doğrulama, adli analizler ve biyometrik güvenlik uygulamalarında büyük bir avantaj sağlamaktadır. Güvenlik kameraları gibi düşük kaliteli görüntü kaynakları, çoğu zaman yüz tanıma sistemleri için yetersiz olabilmektedir. Yüz halüsinasyonu teknikleri, eksik veya belirsiz alanları tahmin ederek yüz tanıma süreçlerini daha güvenilir hale getirebilmektedir. Bunun yanı sıra yüz halüsinasyonu sanal gerçeklik, dijital sanat ve medya sektörlerinde de yaygın olarak kullanılmaktadır. Eski veya düşük çözünürlüklü fotoğrafların restore edilmesi, yüz ifadesi detaylarının kaybolmadan iyileştirilmesi gibi uygulamalar, bu teknolojinin potansiyelini ortaya koymaktadır.

FSR olarak da bilinen yüz halüsinasyonu insan algısının yanı sıra bilgisayar tabanlı yüz tanıma ve analiz sistemlerini geliştirmeyi amaçlamaktadır. Uzaktan veya zayıf ışık koşullarında çekilen görüntülerde yüz özelliklerini daha belirgin hale getirerek özellikle güvenlik alanında yüz tanıma sistemlerinin doğruluğunu artırmaktadır.

FSR'den türetilen yüz görüntüleri, suçluların kimlik tespitini kolaylaştırmak için görgü tanığı çizimlerinin fotoğraflara dönüştürülmesinde de etkili olmaktadır. Ayrıca yüz hizalama, yüz ayrıştırma ve yüz algılama gibi yüzle ilgili çeşitli alt görevlerde de önemli bir rol oynar [1-4]. Baker ve Kanade ilk olarak FSR problemini tanımlamış ve görüntülerdeki yüksek frekanslı bileşenleri çıkarmak için Bayesian tabanlı bir istatistiksel yöntem sunmaktadır [5].

Yüz süper çözünürlüğü ve yüz halüsinasyonu süreçleri, görüntü işleme alanında uzun yıllardır geliştirilen makine öğrenimi ve derin öğrenme yöntemleriyle büyük bir ilerleme kaydetmiştir. Geleneksel yöntemler, yüz görüntülerinin istatistiksel özelliklerini kullanarak eksik detayları tahmin etmeye çalışırken, modern derin öğrenme tabanlı yaklaşımlar, büyük veri kümeleri üzerinde eğitilmiş Konvolüsyonel Sinir Ağları (CNN) ve Üretken Çekişmeli Ağlar (GAN) gibi modeller aracılığıyla daha gerçekçi ve yüksek kaliteli yüz görüntüleri üretmeyi mümkün kılmaktadır. CNN tabanlı modeller düşük çözünürlüklü yüz görüntülerinin yüksek çözünürlüklü karşılıklarını öğrenerek dönüşüm sürecini otomatik hale getirirken, GAN tabanlı modeller ise üretilen görüntülerin daha doğal ve gerçekçi olmasını sağlayarak insan algısına daha yakın sonuçlar üretmektedir. Ancak bu yöntemler çoğu zaman sabit stratejilere dayanmakta ve tüm yüz görüntülerine aynı iyileştirme sürecini uygulamaktadır. Oysa yüz görüntülerinde farklı bölgeler farklı seviyelerde detay kaybına uğrayabilmektedir. Bu noktada, pekiştirmeli öğrenme (Reinforcement Learning, RL) tabanlı yöntemler, yüz süper çözünürlüğü ve yüz halüsinasyonu süreçlerine daha dinamik ve adaptif bir yaklaşım kazandırmaktadır. Geleneksel derin öğrenme modelleri büyük miktarda eğitim verisi üzerinden belirli bir dönüşüm fonksiyonunu öğrenirken, pekiştirmeli öğrenme tabanlı sistemler hangi bölgelerin daha fazla iyileştirilmesi gerektiğini öğrenerek stratejik kararlar alabilmektedir. Böylece yüz görüntüsünün kritik detaylara sahip bölgeleri (örneğin gözler, burun ve ağız çevresi) daha yüksek hassasiyetle iyileştirilirken daha az önemli alanlar için gereksiz hesaplama yükü önlenebilmektedir. Sonraki bölümlerde pekiştirmeli öğrenmenin yüz süper çözünürlüğü bağlamında nasıl çalıştığı, karar alma mekanizması ve stratejik parça seçimi gibi kavramlar detaylı bir şekilde ele alınacaktır.

Çalışma kapsamında, yüz süper çözünürlüğü ve yüz halüsinasyonu süreçlerini daha etkili hale getirmek için derin pekiştirmeli öğrenme (Deep Reinforcement Learning,

DRL) ve artık yoğun bloklar (Residual-in-Residual Dense Block, RRDB) tabanlı bir yaklaşım önerilmektedir. Geleneksel süper çözünürlük yöntemleri, genellikle düşük çözünürlüklü görüntülerden yüksek çözünürlüklü çıktılar elde etmek için sabit dönüşüm modelleri veya doğrudan öğrenme tabanlı yöntemler kullanmaktadır. Ancak yüz görüntülerinde detay kaybı yaşandığında sabit büyütme yöntemleri yeterli olmamakta ve yüzün karakteristik özellikleri kaybolabilmektedir. Tez kapsamında geliştirilen yöntem, pekiştirmeli öğrenme destekli stratejik parça seçimi ile yüz süper çözünürlüğünü optimize etmeyi amaçlamaktadır. Önerilen model, yüz görüntüsü üzerinde hangi bölgelerin daha fazla iyileştirilmesi gerektiğini öğrenen bir dikkat mekanizması ile çalışmaktadır. DRL tabanlı parça seçimi sayesinde, düşük çözünürlüklü bir yüz görüntüsünde en kritik bölgeler belirlenmekte ve her adımda iteratif olarak iyileştirilmektedir. Seçilen bölgelere uygulanan RRDB tabanlı süper çözünürlük modeli, detayların daha etkin bir şekilde geri kazanılmasını sağlarken, yüzün genel yapısının korunmasını da garanti etmektedir. Yaklaşım yalnızca düşük kaliteli yüz görüntülerinin iyileştirilmesini değil aynı zamanda yüz halüsinasyonu sürecini de destekleyerek eksik yüz detaylarının en iyi şekilde oluşturulmasını hedeflemektedir. Çalışma kapsamında önerilen yöntem, yüz süper çözünürlüğü ve yüz tanıma sistemleri başta olmak üzere güvenlik, adli bilişim, tıbbi görüntüleme ve medya sektörlerinde uygulanabilecek yenilikçi bir çözüm sunmaktadır.

Çalışmanın 2. kısmında modelin temellerinden olan derin pekiştirmeli öğrenme ve dikkat mekanizmaları konuları ele alınacaktır. 3. kısmında görüntü süper çözünürlüğü problem tanımı yapılmakta ve ardından literatür taraması verilmektedir. 4. kısımda önerilen yöntem anlatılmaktadır. 5. kısımda elde edilen deneysel sonuçlar gösterilmekte ve yorumlanmaktadır. 6. kısımda ise sonuçlar ele alınmaktadır.



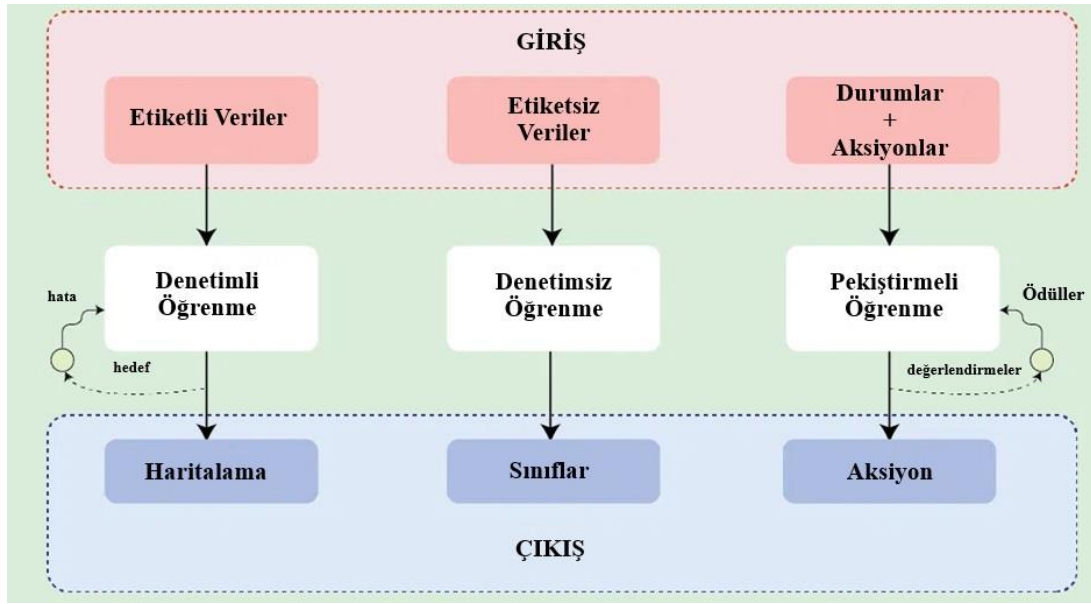
2. DERİN PEKİŞTİRMELİ ÖĞRENME VE DİKKAT MEKANİZMASI

Geliştirilen derin öğrenme tabanlı yüz süper çözünürlüğü metodu, pekiştirmeli öğrenme ve kendini tekrarlayan sinir ağları tabanlıdır. Ayrıca görsel dikkat mekanizması da kullanılmaktadır. Dolayısıyla aşağıdaki kısımda, çalışmanın rahat anlaşılabilmesi için gereken alt yapı sunulmaktadır. İlk kısımda temel makine öğrenme yöntemlerine kısaca değinilmektedir. Ardından pekiştirmeli öğrenmenin temelleri anlatılmaktadır. Okuyucunun verilen alt yapıyı inceledikten sonra diğer kısımlara geçmesi tavsiye edilmektedir.

2.1. Süper Çözünürlükte Temel Makine Öğrenimi Yapıları

Süper çözünürlük alanında makine öğrenimi yöntemleri, görüntüdeki eksik detayları tamamlayarak daha net ve ayrıntılı sonuçlar elde etmede kritik bir rol oynamaktadır. Makine öğrenimi, bir bilgisayar programının önceden belirlenmiş kurallar olmadan, veri aracılığıyla öğrenmesini ve performansını geliştirmesini sağlayan bir yaklaşımdır. Yapay zekanın temel yapılarından biri olan makine öğrenimi, veri bilimi ve görüntü işleme alanlarında giderek önem kazanmaktadır. Süper çözünürlükte kullanılan makine öğrenimi algoritmaları veri örneklerinden öğrenilen bilgilere dayanarak düşük çözünürlüklü görüntülerden yüksek çözünürlüklü tahminler oluşturmaktadır.

Makine öğrenimi süreçleri, kullanılan veri türüne ve öğrenme yöntemine bağlı olarak farklı kategorilere ayrılmaktadır. Süper çözünürlük alanında en yaygın kullanılan yaklaşımlar denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme yöntemleridir. Denetimli öğrenme, düşük ve yüksek çözünürlüklü görüntü çiftlerinden modelin öğrenmesini sağlayarak yeni düşük çözünürlüklü girişler için yüksek çözünürlüklü tahminler üretmektedir. Denetimsiz öğrenme ise etiketlenmiş veri gerektirmeden düşük çözünürlüklü görüntüler arasındaki örüntüleri keşfederek çözünürlüğü artırmayı amaçlamaktadır. Pekiştirmeli öğrenme ise ödül mekanizması ile modelin iteratif olarak en iyi stratejiyi öğrenmesine yardımcı olur ve görüntüde odaklanması gereken kritik bölgeleri belirleyerek süper çözünürlüğü optimize etmektedir. Şekil 2.1’de bu üç temel öğrenme paradigması gösterilmektedir.



Şekil 2.1. Makine öğrenmesi üç temel yapı.

Denetimli öğrenme, modelin giriş verilerini hedef çıktılarla eşleştiren bir fonksiyon öğrenmesini sağlayan bir makine öğrenimi yöntemidir. Öğrenme sürecinde her giriş verisine karşılık gelen doğru çıktının bulunduğu etiketli veri kümeleri kullanılarak model eğitilmektedir. Model (x_i, y_i) çiftlerinden oluşan bir veri kümesi üzerinde tüm $i=1, 2, \dots, N$ örnekleri için $f: X \rightarrow Y$ fonksiyonunu öğrenmeye çalışmaktadır. Fonksiyon, modelin giriş verisine dayanarak en doğru tahmini üretmesini amaçlamaktadır. Modelin başarımı tahmin edilen çıktı $f(x_i)$ ile gerçek değer y_i arasındaki farkı en aza indirmek için optimize edilen bir kayıp fonksiyonu ile değerlendirilmektedir [6].

Denetimsiz öğrenme, modelin giriş verileri arasındaki gizli yapıları ve örüntüleri keşfetmesini sağlayan bir makine öğrenimi yöntemidir. Öğrenme sürecinde etiketli veri yerine yalnızca giriş verileri kullanılarak model eğitilmektedir. Denetimsiz öğrenmede hedef N adet girdi (x_i) üzerinden $f: X \rightarrow Y$ fonksiyonunu öğrenmektir; ancak burada y_i (çıktı) model tarafından keşfedilir ve etiketli veriler değildir. Burada modelin amacı verileri anlamlı gruplara ayırmak veya veri dağılımını temel alarak yeni temsiller oluşturmaktır. Modelin başarımı belirlenen veri yapılarının tutarlılığı veya kümeleme gibi görevlerdeki performansı ile değerlendirilmektedir. Denetimsiz öğrenme etiketsiz büyük veri kümeleri üzerinde çalışırken tercih sebebi olabilmektedir. Ancak bu öğrenme türünde doğru sonuca ulaşma garantisi olmadığından sonuçlar genellikle keşif niteliğindedir ve elde edilen sonuçların doğruluğu insan müdahalesine veya ek değerlendirmelere dayanabilmektedir.

Pekiştirmeli öğrenme, denetimli ve denetimsiz öğrenmeden farklı bir teknikte eğitim süreci kullanmaktadır. Pekiştirmeli öğrenme, bir ajanın çevresiyle etkileşim kurarak deneyim yoluyla öğrenmesini sağlayan bir makine öğrenimi yöntemidir [7]. Öğrenme sürecinde ajan belirli bir ortam içinde hareket eder ve gerçekleştirdiği her eylemin sonucunda bir ödül veya ceza alarak stratejisini geliştirmektedir. Amaç, ajanı en yüksek toplam ödülü elde edecek şekilde yönlendirerek en iyi kararları öğrenmesini sağlamaktır. Denetimli öğrenmede, doğru eylemler veya sonuçlar modele verilir ve model giriş verilerine dayanarak öğrenme gerçekleştirmektedir. Ancak pekiştirmeli öğrenmede doğru eylemler ajana sunulmamakta; bunun yerine, ajan deneme-yanılma yoluyla kendi stratejisini geliştirmektedir. Denetimsiz öğrenmeden farkı ise ajanın belirli bir davranışı öğrenmeye zorlanmasıdır. Ajan her adımda çevre tarafından değerlendirilmekte ve ödüllendirilmekte. Denetimsiz öğrenmede bir hedef veya yönlendirme olmadan verilerin yapısını keşfetme amacı varken, pekiştirmeli öğrenmede ajan her etkileşiminde bir ödül fonksiyonuna göre yönlendirilmektedir. Ödül fonksiyon yapısına göre her adımda bir değerlendirme almaktadır.

Pekiştirmeli öğrenme yapısında bahsi geçen parametreler ve ifadeler (ajan, ödül, durum, çevre) ile birlikte yapı detayları sonraki başlıkta anlatılmaktadır. Ayrıca derin pekiştirmeli öğrenme yapısından da yine aynı başlık altında bahsedilmektedir.

2.2. Pekiştirmeli Öğrenme ve Derin Pekiştirmeli Öğrenme

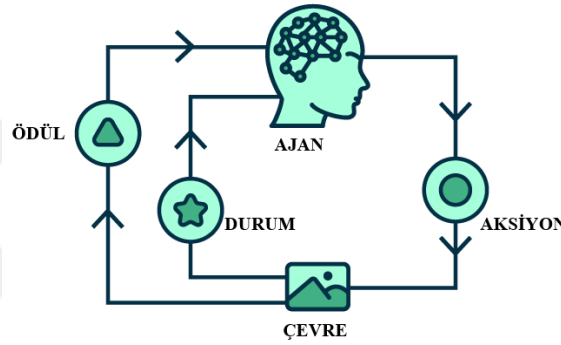
Pekiştirmeli öğrenme genellikle Markov Karar Süreci (Markov Decision Process - MDP) modeli ile tanımlanmaktadır [8]. Pekiştirmeli öğrenmede ajan, bir durumda bulunur ve bir aksiyon gerçekleştirmektedir. Aksiyon sonucunda ajan yeni bir duruma geçer ve bir ödül almaktadır. Ajanın hedefi, uzun vadeli ödülü maksimize edecek en iyi stratejiyi öğrenmektir. Pekiştirmeli öğrenme, özellikle otonom sistemlerde, oyun oynayan yapay zeka algoritmalarında ve robotikte yaygın olarak kullanılmaktadır. MDP'nin temel bileşenleri şunlardır:

Durumlar (States): Ajanın bulunduğu mevcut durumlardır. Her durum, ajanın çevresiyle ilgili bilgileri içermektedir.

Aksiyonlar (Actions): Ajanın belirli bir durumda gerçekleştirebileceği olası hareketlerdir.

Ödül (Reward): Ajanın bir eylem gerçekleştirdiğinde aldığı geri bildirimdir. Ödül, ajanın belirli bir durumda yaptığı hareketin iyi ya da kötü olduğunu göstermektedir.

Çevre (Environment): Ajanın etkileşimde bulunduğu ve kararlarının sonuçlarını gözlemlediği dış ortamdır. Çevre, ajan tarafından gerçekleştirilen her aksiyona karşılık olarak yeni bir durum üretir ve belirli bir ödül sağlamaktadır. Ajan, çevreden aldığı geri bildirimlerle stratejisini geliştirir ve gelecekteki adımlarını optimize etmektedir. Etkileşim süreci, ajan ile çevre arasındaki dinamik ilişkiyi şekillendirir ve öğrenme sürecinin temelini oluşturmaktadır.

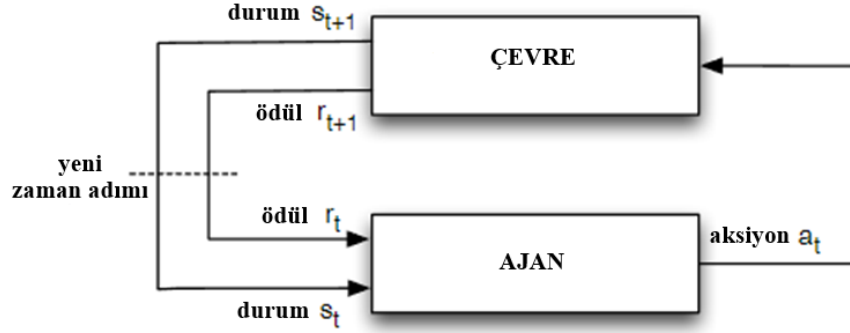


Şekil 2.2. Pekiştirmeli öğrenme genel bakış.

Temsili yapı olarak Şekil 2.2 incelendiğinde pekiştirmeli öğrenme, bir ajanı (öğrenen varlık) çevrede nasıl hareket etmesi gerektiğini öğretmek için zamanla toplam ödülü maksimize etmeye dayanmaktadır [9]. Ajanın her hareketi, bir ödül fonksiyonuna göre ya ödüllendirilmekte ya da cezalandırılmaktadır. Bu nedenle, ödül fonksiyonu ajana ne öğretileceğinin bir temsili olarak kabul edilmektedir. Ajan, çevreyi farklı durumlarda çeşitli eylemler gerçekleştirerek keşfeder ve süreç boyunca deneme-yanılma yöntemiyle tecrübe kazanmaktadır. Daha sonra öğrendiği tecrübeleri kullanarak anlık ve gelecekteki ödülleri göz önünde bulundurarak en yüksek ödülü elde etmeye çalışmaktadır.

Ajan terimi girdileri algılayabilen ve bunlara çıktılarla tepki verebilen her türlü program, algoritma, robot veya başka bir sistem için kullanılabilir. Ajanın kullandığı girdiler ise durum, kullandığı çıktıları ise aksiyon olarak adlandırılmaktadır. Ajanın davranışı durumlardan aksiyonları haritalanması politika olarak adlandırılır ve ajana geri bildirim yoluyla öğretilmektedir. Politikanın öğretilmesi için kullanılan geri bildirim ise ödül olarak adlandırılır ve ajana ödüle göre hareket etmesini sağlayacak bir strateji kurulmaktadır. Durumlar, fiziksel bir makinenin gerçek durumlarını (örneğin açık/kapalı), sanal bir ortamın dahili bir

temsili ya da soyut kavramlar olabilmektedir. Örneğin bir ajan “bu nesnenin ne olduğunu bilmiyorum” durumunda olabilir. Aynı durum aksiyonlar için de geçerlidir: Bunlar çok somut düşük seviyeli motor komutları ya da “yukarı yönde git” gibi yüksek kavramlar olabilmektedir. Çevre, ajanın çıktılarını (aksiyonlarını) girdi olarak kullanır ve çıktılarını durum olarak ajana geri bildirimler verecektir.



Şekil 2.3. Pekiştirmeli öğrenme döngü ve parametreleri.

Şekil 2.3, ajan ve çevre arasındaki etkileşim döngüsünü göstermektedir. Ajan çevreden mevcut bir durum bilgisi (s_t) alır ve bu duruma uygun bir aksiyon (a_t) seçerek karşılık vermektedir. Seçilen aksiyon, çevrede bir değişikliğe yol açarak yeni bir duruma (s_{t+1}) ve bir ödüle (r_{t+1}) neden olur. Çevrenin, ajan için herhangi bir “niyeti” veya “planı” yoktur; yalnızca pasif bir şekilde ajanın aksiyonlarına göre tepkiler vermektedir. Bu durumda, ajan ödüllerin kaynağından çok miktarına odaklanır. Ajan, yalnızca elde ettiği ödüllerden performansını değerlendirir ve ödülleri bazen gecikmeli olarak geldiği için öğrenme süreci daha zorlu olabilmektedir. Ajanın bu sınırlı bilgiyle en iyi politikayı nasıl öğrenip geliştireceği, pekiştirmeli öğrenmenin cevaplamaya çalıştığı temel sorudur.

Pekiştirmeli öğrenme aşağıda verilen insan hayatından örnekler ile anlatılmaya çalışılmaktadır. İlk örnekte bir çocuğun bisiklet sürmeyi öğrenme süreci, deneme-yanılma yoluyla denge sağlamayı ve doğru hareketleri öğrenmeyi içermektedir. Çocuk, ilk başta denge kurmakta zorlanır ve düşebilir. Ancak her yeni denemede dengede kalmayı başararak ödüllendirilmektedir. Süreçte her düşüş veya başarı, çocuğun denge kurma stratejisini geliştirmesine katkıda bulunmaktadır. İkinci örnekte yeni bir işe başlayan bir çalışanın müşteri ilişkilerini geliştirmesi anlatılmaktadır. Çalışan, müşterilerle etkileşimlerinde hangi yaklaşımların olumlu sonuç verdiğini deneyimleyerek öğrenmektedir. Başlangıçta müşteri taleplerini anlamakta zorlanabilmektedir. Ancak zamanla müşteri memnuniyetini artıran

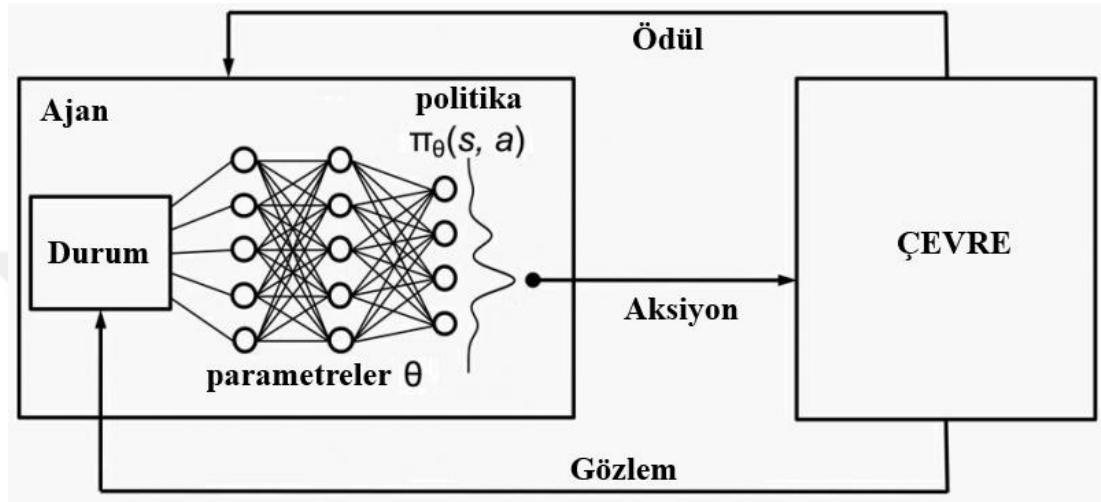
stratejiler geliştirir ve uyguladıkça ödüllerle (örneğin, müşteri memnuniyeti) pekiştirilir. Bu sayede çalışan, müşteri ilişkilerini daha etkili bir şekilde yönetmeyi öğrenmektedir. Üçüncü ve son örnekte tez çalışmasından yola çıkarak, derin pekiştirmeli öğrenme (DRL) ile düşük çözünürlüklü yüz görüntülerini iyileştirmek hedeflenmektedir. Öğrenme sürecinde model yüzün belirli bölgelerini stratejik olarak seçip iyileştirerek daha net ve yüksek çözünürlüklü bir görüntü elde etmeyi amaçlamaktadır. Model her seçtiği ve iyileştirdiği parça için ödüller alarak en iyi stratejiyi öğrenir ve yüz genelinde daha kaliteli bir sonuç elde etmeyi sağlamaktadır. Bahsi geçen örneklerin paradigmaları ve olay örgüsü Tablo 2.1’de verilmektedir.

Tablo 2.1. Pekiştirmeli öğrenme örnekleri.

Parametre	Olay 1: Çocuğun Bisiklet Sürmeyi Öğrenmesi	Olay 2: Çalışanın Müşteri İlişkilerini Geliştirmesi	Olay 3: Tez (Yüz Görüntüleri İyileştirme)
Ajan	Çocuk	Çalışan	Model
Çevre	Yol, bisiklet, çevredeki engeller	İş ortamı, müşteri talepleri, şirket politikaları	Yüz görüntüsü ve özellikleri
Durum	Çocuğun bisiklet üzerindeki mevcut durumu	Müşteriyle o anki etkileşim	Modelin üzerinde çalıştığı mevcut yüz durumu
Aksiyon	Pedalları çevirmek, direksiyonu yönlendirmek	Müşteriyle etkileşim kurmak, stratejileri değiştirmek	Yüzdeki belirli bir parçayı seçip iyileştirmek
Ödül	Düşmemek veya uzun süre dengede kalmak	Müşteri memnuniyeti sağlanması	İyileştirilen parçanın hedef parça ile benzerliği, MSE ve perceptual loss
Politika	Dengeyi sağlamak için geliştirilen strateji	Müşteriyle daha etkili etkileşim kurma stratejisi	Hangi parçaların iyileştirileceğine dair strateji
Sürecin Gelişimi	Çocuk zamanla dengede kalmayı öğrenir	Çalışan deneyimle müşteriyi memnun edecek stratejileri öğrenir	Model, süreç boyunca en iyi iyileştirme stratejisini öğrenir

Derin pekiştirmeli öğrenme, geleneksel pekiştirmeli öğrenmenin derin sinir ağları ile güçlendirilmiş bir versiyonudur. Geleneksel pekiştirmeli öğrenmede, bir ajan

çevresiyle etkileşime girerek aldığı ödüller doğrultusunda optimal bir strateji öğrenmeye çalışmaktadır. Ancak klasik yöntemler yüksek boyutlu ve karmaşık ortamlar için verimli sonuçlar alamayabilmektedir. Derin pekiştirmeli öğrenme, sinir ağları aracılığıyla büyük ve yüksek boyutlu durum uzaylarını daha verimli temsil edebilir ve karmaşık görevlerde daha başarılı öğrenme sağlayabilmektedir. Derin pekiştirmeli öğrenmenin temel yapısı Şekil 2.4’te gösterilmektedir.



Şekil 2.4. Derin pekiştirmeli öğrenme yapısı

Derin pekiştirmeli öğrenmenin klasik pekiştirmeli öğrenmeden temel farkı, ajanın karar mekanizmasını doğrudan öğrenmesine olanak tanıyan derin sinir ağları kullanmasıdır. Geleneksel pekiştirmeli öğrenme genellikle tablo tabanlı yöntemler veya doğrusal fonksiyon kestirimi ile çalışırken, derin pekiştirmeli öğrenme karmaşık örüntüleri öğrenebilir ve genel performansı artırabilmektedir. Özellikle büyük veri ortamlarında ve sürekli aksiyon alanlarına sahip problemlerde derin sinir ağları sayesinde daha ölçeklenebilir ve esnek bir öğrenme süreci elde edilmektedir.

Derin pekiştirmeli öğrenmenin başlıca avantajları arasında yüksek boyutlu veri kümeleri ile çalışabilme, karmaşık çevresel dinamikleri öğrenebilme ve insan seviyesinde karar mekanizmaları geliştirebilme yeteneği bulunmaktadır.

Pekiştirmeli öğrenme yapılarında ajanın amacı zamanla ödülleri maksimize edecek en iyi politika fonksiyonunu (π) öğrenmektir. Süreç, durum-eylem-ödül döngüsü olarak bilinir ve ajan bu döngüde deneyim kazandıkça daha iyi stratejiler öğrenmektedir.

Pekiştirmeli öğrenme, ajanın stratejisini temsil eden bir politika fonksiyonu $\pi: S \rightarrow A$ öğrenme sürecidir. Bu fonksiyon, girdileri (durumlar) $s \in S$ 'den çıktılara (eylemler)

$a \in A$ 'ya eşlemektedir. Çevre, her bir zaman adımında s_t durumunu üretir ve mevcut durum s_t verildiğinde ajan bir aksiyon a_t ile tepki vermektedir. Ajan, bir politika $\pi(a_t|s_t)$ doğrultusunda hareket eder; bu s_t durumunda bir aksiyon a_t alma olasılığını temsil etmektedir (deterministik bir çevrede $\pi(s_t) = a_t$ olur). Ajan aksiyonu a_t aldıktan sonra çevre, yeni bir durum s_{t+1} ve bu durumun ne kadar iyi olduğunu belirten bir ödül r_{t+1} sağlamaktadır.

Öğrenme süreci, olası tüm durumlar için toplam ödülü temsil eden değer fonksiyonu $V_\pi(s)$ 'nin maksimize edilmesiyle gerçekleştirilmektedir. Bu bağlamda, denetimsiz öğrenmeye benzemektedir. Ancak denetimsiz öğrenmenin aksine, $V_\pi(s)$ fonksiyonu sabit tanımlı yapıda değildir. Ajan çevre ile etkileşime girerek dolayısıyla olası tüm durumlarda olası tüm eylemleri gerçekleştirerek bu fonksiyonu da öğrenmektedir.

Maksimize edilmek istenilen ödüller: $\max \sum r_t$ şeklinde hesaplanmaktadır. Kümülatif (genellikle indirimli) ödüllerin toplamı, getiri (return) R_t olarak adlandırılır (Denklem 2.1):

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^T \gamma^k r_{t+k+1} \quad (2.1)$$

Ayrıca, V -fonksiyonu ve Q -fonksiyonlarını denklem 2.2 ve 2.3'teki gibi tanımlanmaktadır:

$$V_\pi(s) = \mathbb{E}_\pi[R_t | S_t = s] \quad (2.2)$$

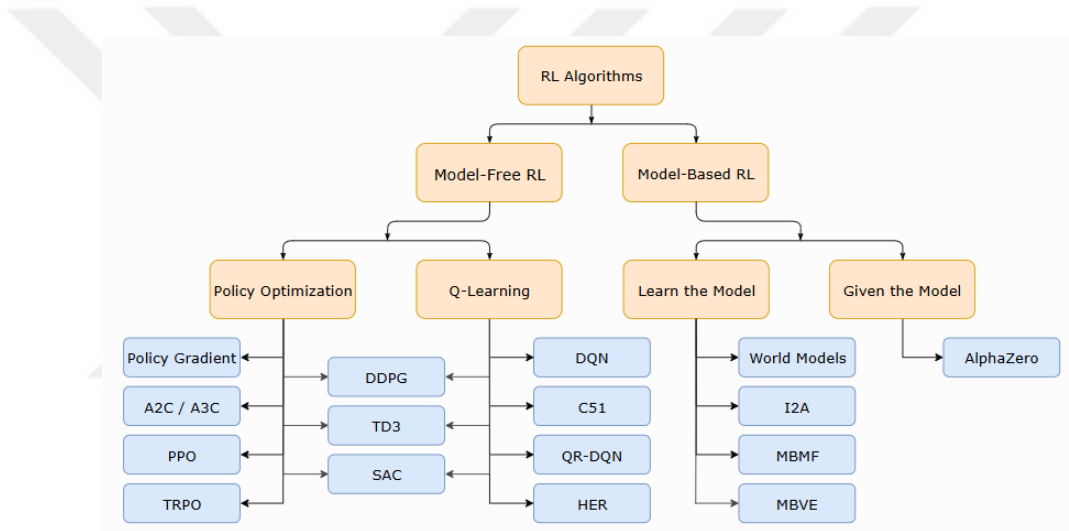
$$Q_\pi(s, a) = \mathbb{E}_\pi[R_t | S_t = s, A_t = a] \quad (2.3)$$

Burada $V_\pi(s)$, belirli bir durumda beklenen getiri değerini ifade ederken, $Q_\pi(s, a)$, belirli bir durumda belirli bir aksiyonu aldıktan sonra beklenen getiri değerini göstermektedir. ($\mathbb{E}$ =beklenen değer, γ = indirim oranını (discount factor) $[0, 1]$ aralığında bir sabittir ve gelecekteki ödüllerin ne kadar önemli olduğunu belirlemektedir.)

2.3. Pekiştirmeli Öğrenme Algoritmaları

Pekiştirmeli öğrenme, Richard Bellman'ın dinamik programlama çalışmalarıyla temelleri atılan ve ardından Arthur Samuel'in oyun oynama sistemleriyle

popülerleşen bir makine öğrenmesi alanıdır [10, 11]. Kullanılan algoritmalar, ajanın çevreyi nasıl modelleyeceğine ve aksiyonlarını nasıl seçeceğine bağlı olarak çeşitli kategorilere ayrılmaktadır (Şekil 2.5). Genel olarak, pekiştirmeli öğrenme algoritmaları iki ana grupta incelenir: Model Tabanlı (Model-Based) ve Model Tabansız (Model-Free) yöntemler [9]. Her iki yaklaşım da ajanın ödülleri ve deneyimleri aracılığıyla çevresini anlamasına ve kararlar almasına olanak tanımaktadır. Model tabanlı algoritmalar, çevrenin dinamiklerini modelleyerek ajanın ileriye yönelik adımlarını simüle etmesine imkan tanırken; model tabansız algoritmalar, ajanın doğrudan deneyimlerinden öğrenerek hareket etmesini sağlamaktadır. Bu bölümde, iki ana kategoriye temel bir bakış sağlanmakta ve çalışmamızda kullandığımız model incelenecektir.



Şekil 2.5. Pekiştirmeli öğrenme algoritmaları.

Kategoriler, ajanın çevresini nasıl öğrendiğine ve aksiyonlarını nasıl seçtiğine dayanan farklı veri işleme yaklaşımlarını temsil etmektedir. Her iki ana kategori de alt dallarında farklı yöntemler sunarak çeşitli problem çözme yaklaşımlarına imkan tanımaktadırlar. Aşağıda iki model tipi daha ayrıntılı incelenecektir.

- **Model Tabanlı (Model-Based) Pekiştirmeli Öğrenme:**

Model Tabanlı yöntemlerde ajan, çevrenin dinamiklerini öğrenmeye veya modellemeye çalışmaktadır. Çevre modelini öğrenmek, ajanın gelecekteki durumları ve ödülleri tahmin etmesine yardımcı olmaktadır. Model tabanlı yöntemler, verileri daha verimli kullanmaktadır ve çevre hakkında bilgi sahibi olduğundan daha hızlı sonuçlar elde edilmektedir. Ajan, geçiş olasılıklarını ($P(s'|s,a)$) ve ödül fonksiyonunu ($R(s,a)$) tahmin ederek çevreyi modellemeye çalışmaktadır. Model

sayesinde gelecekteki adımların sonuçları simüle edilir ve optimal politikalar daha hızlı bir şekilde öğrenilebilmektedir. Model tabanlı yöntemler, genellikle veri toplamanın zor olduğu senaryolarda (örneğin, tıbbi deneylerde) tercih edilmektedir [12].

- **Model Tabansız (Model-Free) Pekiştirmeli Öğrenme:**

Model tabansız yöntemlerde ajan, çevre hakkında herhangi bir model öğrenmemektedir. Yalnızca deneyimlerinden ve tecrübelerden doğrudan öğrenmektedir. Model tabansız yöntemler, verinin bol ve erişimin kolay olduğu ancak hesaplama kapasitesinin sınırlı olduğu durumlarda daha uygun bir seçimdir [13]. Model tabansız yöntemler kendi içinde Değer Tabanlı (Value-Based / Q-learning) ve Politika tabanlı yöntemler olmak üzere iki ana kategoriye ayrılmaktadır.

Değer tabanlı yöntemler, pekiştirmeli öğrenmede ajanın her durum (veya durum-eylem çifti) için bir değer fonksiyonu öğrenmesini hedeflemektedir. Fonksiyon, ajan için en yüksek değeri sağlayacak aksiyonu seçmesine yardımcı olmaktadır. Değer tabanlı yöntemlerin temelinde Q-öğrenme algoritması bulunmaktadır. Q-öğrenme, her durum-eylem çifti için gelecekte beklenen ödülleri temsil eden bir Q-değer fonksiyonu öğrenmekte ve bu değerleri güncelleyerek ajanın en uygun stratejiyi öğrenmesini sağlamaktadır [8]. Geleneksel Q-öğrenme yöntemleri, küçük durum-eylem uzaylarında etkili olsa da büyük ve karmaşık çevrelerde verimliliği sınırlı kalmaktadır. Oluşabilecek bu sorunu çözmek için derin öğrenme ile entegre edilmiş pekiştirmeli öğrenme yöntemleri geliştirilmiştir. Örneğin, Derin Q Ağları (Deep Q-Networks - DQN), Q-öğrenmeyi derin sinir ağlarıyla birleştirerek büyük durum-eylem uzaylarında etkili sonuçlar elde etmeyi mümkün kılarak özellikle karmaşık görevlerde ajanın uzun vadeli stratejiler geliştirmesine olanak tanımaktadır [14].

Politika (Policy) tabanlı veya doğrudan pekiştirmeli öğrenme yöntemlerde ajan, durumlardan aksiyonlara doğrudan bir politika fonksiyonu öğrenmektedir. Politika arama, parametrelili bir politika fonksiyonu aracılığıyla aksiyon seçimini doğrudan optimize etmektedir. Politika tabanlı yöntem, büyük ve sürekli aksiyon uzaylarında etkili olmaktadır. Ajan, belirli bir durumdayken en iyi aksiyonu seçmek için doğrudan bir politika öğrenmektedir. Politika gradyanı (Policy Gradient) algoritması ve PPO (Proksimal Politika Optimizasyonu) gibi algoritmalar politika tabanlı yaklaşıma örnek algoritmalar [15].

Önerilen çalışmada, Politika Tabanlı (Policy-Based) yöntemler üzerinde durulacaktır. Politika tabanlı yöntemler değer tabanlı yöntemlerden farklı olarak, her durumda optimal aksiyonu seçmek için doğrudan bir politika öğrenmektedirler. Bu yaklaşım, büyük ve karmaşık durum-aksiyon uzaylarında etkili olup, çevrenin karmaşık olduğu durumlarda daha iyi sonuçlar verebilmektedir. Pekiştirmeli öğrenme, hem politika tabanlı yöntemlerle hem de derin öğrenme tabanlı yaklaşımlarla gerçekleştirilebilmektedir. Derin pekiştirmeli öğrenme, derin sinir ağları ile pekiştirmeli öğrenmeyi birleştirerek daha karmaşık çevrelerde daha başarılı ajanlar oluşmasını sağlamaktadır.

Stokastik politikalar, bir aksiyonun belirli bir durumda seçilme olasılığını verirken, deterministik politikalar doğrudan bir aksiyon belirlemektedir. Bir politika, durum s verildiğinde aksiyon a 'yı seçme olasılığını $\pi(a|s)$ ile tanımlanır ve bu politika, parametrelerle ifade edildiğinde $\pi(a|s; \theta)$ şeklinde gösterilmektedir. Politika tabanlı yöntemler, politikanın doğrudan optimizasyonu için kullanılmakta ve bu optimizasyon genellikle politika gradyanı algoritmaları ile gerçekleştirilmektedir. Örneğin politika gradyan modelleri, politikanın parametrelerini maksimum ödül sağlayacak şekilde güncellemek için gradyan inişi yöntemi kullanmaktadır. Politika tabanlı yöntemlerin popüler örnekleri şunlardır:

Reinforce Algoritması: Temel politika gradyan algoritmalarından biri olan reinforce, ajanın her bir aksiyonunu ödüllendirme veya cezalandırma yoluyla politikasını güncellemektedir.

Proximal Policy Optimization (PPO) [15]: Politikayı daha kararlı ve güvenilir hale getirmek için çeşitli kısıtlamalarla politikayı güncellemektedir.

Advantage Actor-Critic (A2C/A3C) [16]: Hem aktör (actor) hem de eleştirici (critic) yapılarını kullanarak politikanın ve değer fonksiyonunun aynı anda optimize edilmesini sağlamaktadır.

Politika Gradyanı (Policy Gradient) Teoremi: Politika gradyanı yöntemlerinde hedef, ajanın toplam beklenen ödülünü en büyük hale getirecek şekilde politikayı optimize etmektir [9, 17, 18]. Politikayı güncellemek için politika gradyanı teoremi kullanılmaktadır. Beklenen toplam ödül $J(\theta)$, parametreler θ ile tanımlanan politika $\pi(a|s; \theta)$ 'dan türetilmektedir. Önerilen çalışmada reinforce algoritması ve politika gradyanı teoremi kullanılmaktadır.

$$J(\theta) = \mathbb{E}_{\pi_\theta} \left[\sum_{t=0}^T R_t \right] \quad (2.4)$$

Denklem 2.4'te R_t her adımdaki ödölü temsil etmektedir. Politika gradyan teoremi, $J(\theta)$ 'nin gradyanını aşağıdaki şekilde denklem 2.5'te tanımlar:

$$\nabla_\theta J(\theta) = \mathbb{E}_{\pi_\theta} [\nabla_\theta \log \pi_\theta(a|s) \cdot \hat{Q}(s, a)] \quad (2.5)$$

burada $\hat{Q}(s, a)$, belirli bir durumda belirli bir aksiyonu seçmenin beklenen getirisini ifade etmektedir. Formül, ajanın politika parametrelerini güncellemesi için gerekli olan gradyanı sağlamaktadır. $\nabla_\theta \log \pi_\theta(a|s)$ terimi, ajanın seçtiği aksiyonun olasılığının logaritmasının gradyanı olarak ajanın politikayı nasıl güncelleyeceğini belirlemektedir.

Reinforce algoritmasının teorisi, temel bir politika gradyan yöntemidir ve Monte Carlo yöntemini kullanarak beklenen ödülleri hesaplamaktadır [9]. Her eğitim döngüsü sonunda, ajanın ödülleri temel alınarak politika güncellenmektedir. Algoritmanın güncelleme kuralı denklem 2.6'da verilmektedir.

$$\theta \leftarrow \theta + \alpha \nabla_\theta \log \pi_\theta(a_t|s_t) \cdot G_t \quad (2.6)$$

burada α öğrenme oranı, denklem 2.7'de gösterilen G_t ise t anından itibaren gelecekteki toplam ödölü temsil etmektedir. Beklenen getiri olarak düşünülebilir. ($\hat{Q}(s, a) = G_t$) :

$$G_t = \sum_{k=t}^T \gamma^{k-t} R_k \quad (2.7)$$

burada γ indirim faktörü ve R_k k'nci adımdaki ödöldür.

Politika gradyan teoreminde baseline kavramı vardır. Baseline, politika gradyanı yöntemlerinde öğrenme sürecini hızlandırmak ve daha düşük varyansla politika güncellemeleri yapmak için kullanılmaktadır. Baseline, bir referans değer olarak kullanılır ve ajanın politika gradyanı hesaplamasında beklenen getiriyi (return) azaltarak güncellemelerin varyansını düşürmeye yardımcı olmaktadır. Baseline, gradyan tahminlerini daha kararlı hale getirir ve öğrenme sürecini hızlandırmaktadır.

Baseline kullanmak, ödülün mutlak değerinden ziyade, ödülün beklenen değere göre ne kadar iyi olduğunu vurgulamaktadır.

Politika gradyanı teoremi, baseline $b(s)$ eklenerek denklem 2.8'deki gibi yazılabilir:

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \cdot (G_t - b(s_t))] \quad (2.8)$$

burada $b(s_t)$, s_t durumuna bağlı bir baseline değeridir.

Denklem 2.8'deki formülde, $b(s_t)$ 'nin eklenmesi, gradyanın beklenen değerini değiştirmez ancak varyansını azaltmaktadır. Özetle baseline kullanımı, öğrenme sürecini hızlandırırken aynı zamanda güncellemelerin tutarlılığını artırmaktadır.

En yaygın kullanılan baseline, duruma bağlı değer fonksiyonu $V(s)$ 'dir. $V(s)$, belirli bir durumda ajanın beklenen getirilerini hesaplamaktadır ve bu değer, politikadan bağımsız bir tahminle öğrenilebilmektedir. $V(s)$, duruma bağlı ortalama ödül temsil ettiği için politikadaki güncellemeleri normalize ederek daha etkili hale getirmektedir.

Baselinenin varyans azaltıcı etkisi denklem 2.9'da gösterilmektedir:

$$\mathbb{E}_{\pi_{\theta}} [\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \cdot (G_t - V(s_t))] \quad (2.9)$$

burada G_t gelecekteki toplam ödül ve $V(s_t)$, ajanın o anki durumda beklediği getirideğerini ifade etmektedir. En iyi baseline, hesaplamayı değiştirmeden varyansı en çok azaltan baseline'dır. Matematiksel olarak, en iyi baseline $b^*(s)$ şu şekilde denklem 2.10'daki hali ile hesaplanabilir:

$$b^*(s) = \mathbb{E}_{\pi_{\theta}} [G_t | s_t = s] \quad (2.10)$$

burada belirli bir durum için beklenen getiri ifade edilmektedir.

Williams'ın reinforce algoritması [18] ve Sutton ve Barto'nun "Reinforcement Learning: An Introduction" kitabında [9] açıklanan politika gradyanı yöntemleri, baseline kullanımının matematiksel temellerini detaylandırmaktadır.

Sonuç olarak, baseline kullanımı, politika tabanlı pekiştirmeli öğrenme algoritmalarında kritik bir rol oynar ve ajanın öğrenme sürecinin daha verimli hale gelmesini sağlamaktadır.

Bir sonraki başlıkta modelde kullanılan tekrarlı yapıları daha iyi anlayabilmek için Tekrarlayan Sinir Ağı yapılarından bahsedilecektir.

2.4. Tekrarlayan Sinir Ağı (Recurrent Neural Network - RNN)

Zaman serileri ve ardışık verilerin işlenmesi, birçok uygulama alanında kritik bir öneme sahiptir. Metin işleme, konuşma tanıma, zaman serisi analizi ve video işleme gibi ardışık bilgilere dayalı problemler, veriler arasındaki ardışık ilişkileri doğru bir şekilde modelleyen ve öğrenebilen yapıların kullanımını gerektirmektedir. Bu nedenle Tekrarlayan Sinir Ağları, ardışık bilgileri işlemeye uygun yapılarıyla öne çıkmaktadır. RNN'ler, geçmişteki bilgileri mevcut duruma entegre ederek modelleme yeteneği sunar ve bu da onları ardışık verilerin yer aldığı birçok uygulamada etkili bir çözüm haline getirmektedir [19, 20].

Önerilen yöntemde yüz görüntülerinin sıralı parçalarını seçip bu parçaların çözünürlüğünü artırmak için kullanılan metodoloji RNN'lerin sunduğu ardışık modelleme avantajını kullanmaktadır. Çalışmadaki model, RNN yapılarının ardışık işlem kabiliyetinden faydalananarak her bir adımda önemli yüz bölgelerini tespit eder ve bu bölgelerin çözünürlüğünü artırmaktadır. Her adımda seçilen parça çözünürlük açısından iyileştirilir ve daha önceki adımlardan gelen bilgi kullanılarak görüntünün genel kalitesi korunmaktadır. Süreç, tekrarlayan bir dikkat mekanizması aracılığıyla yüzün detaylı bölgelerini iyileştirir ve nihai görüntünün global ve yerel kalitesini artırmaktadır. RNN yapı sayesinde görüntü üzerinde seçilen parçaların, geçmişteki bilgilerle birleştirilerek daha geniş alanlara odaklanması ve her bir parçanın çözünürlüğünün artırılması sağlanmaktadır. Aşağıda RNN'lerin çalışma prensipleri, matematiksel temelleri ve kullanım alanlarına detaylı bir şekilde değinilmektedir.

Bir RNN'deki her bir gizli katman önceki zamandaki gizli durum bilgisi ile birlikte güncellenmektedir. Yapı, RNN'nin girişteki ardışık verilere bağlı olarak durumsal ilişkileri öğrenmesini sağlamaktadır. Denklem 2.11'de, RNN'nin gizli durumu h_t , önceki gizli durum h_{t-1} ve mevcut giriş x_t kullanılarak matematiksel hesaplaması verilmektedir.

$$h_t = f(W_h h_{t-1} + W_x x_t + b_h) \quad (2.11)$$

Burada:

h_t , zaman t anındaki gizli durumu temsil etmektedir.

x_t , zaman t anındaki giriş vektörüdür.

W_h , önceki gizli durumlardan gelen ağırlık matrisidir.

W_x , girişten gelen ağırlık matrisidir.

b_h , bias terimidir.

f , genellikle tanh veya ReLU gibi bir aktivasyon fonksiyonudur.

RNN' in çıktısı ise gizli durumdan hesaplanarak denklem 2.12'deki gibi elde edilmektedir:

$$y_t = c(W_y h_t + b_y) \quad (2.12)$$

Burada:

y_t , zaman t anındaki çıkış vektörüdür,

W_y , gizli durumdan çıkışa olan ağırlık matrisidir,

b_y , çıkış katmanındaki bias terimidir,

c , çıkış fonksiyonu olup, probleme bağlı olarak softmax veya sigmoid gibi fonksiyonlar olabilmektedir.

Zaman Üzerinde Geriye Yayılım (Backpropagation Through Time - BPTT), ardışık veri işlemede ve bu verilerle model eğitmede önemli bir rol oynamaktadır. Önerilen çalışmada, görüntülerin sıralı parça seçimleri ve bu parçaların çözünürlüğünün artırılması gibi ardışık süreçlerin etkin bir şekilde yönetilmesi için RNN'lerin döngüsel yapıları ve bu yapılar için kullanılan BPTT algoritmasının anlaşılması gerekmektedir. Görüntü işleme sürecinde, modelin her adımda geçmiş bilgilere dayanarak gelecekteki adımları stratejik olarak seçmesi gerektiği için, BPTT'nin hata sinyallerini zaman boyunca nasıl geri yaydığı ve ağırlıkları nasıl güncellediği konuları kritik öneme sahip olmaktadır. Bu nedenle, BPTT'nin temel prensiplerini ve RNN'lerin eğitimi sırasında karşılaşılan zorlukları anlamak, çalışmanın metodolojisinin daha iyi kavranmasına olanak sağlamaktadır.

Geriye Yayılım (Backpropagation), yapay sinir ağlarının eğitimi için kullanılan temel bir öğrenme algoritmasıdır. Algoritma, kayıp fonksiyonunun çıktılarına ve ağırlık parametrelerine olan türevini alarak ağırlıkları günceller ve ağırlık doğruluğunu

artırmayı amaçlamaktadır. Ancak Tekrarlayan Sinir Ağları gibi zaman boyutunda bağlantılı olan sinir ağlarının eğitimi, standart geriye yayılım algoritması ile doğrudan gerçekleştirilemez. Bunun yerine zaman boyutundaki bağımlılıkları dikkate almak için “Zaman Üzerinde Geriye Yayılım” adı verilen bir uzantı kullanılmaktadır.

RNN’ler, hata sinyalini zaman boyunca geri yayarak öğrenim gerçekleştirmektedir. Bu işlem, Zaman Üzerinde Geriye Yayılım olarak bilinir. BPTT, RNN’de ortaya çıkan zaman bağımlılıklarını hesaba katarak, ağırlıkların güncellenmesini sağlamaktadır. Zaman boyunca ardışık durumların birbirine bağlı olması nedeniyle hata sinyali, her bir zamandaki tüm gizli durumlar üzerinden yayılmaktadır.

Sonuç olarak RNN yapılarında ardışık bağımlılık, modelin geçmiş bilgileri hafızasında tutarak uzun dizileri öğrenmesine olanak sağlamaktadır. BPTT, ağırlık çıktıları ile gerçek değerler arasındaki hata sinyalini zaman boyunca geri yayarak ağırlıkların tüm zaman adımlarında güncellenmektedir.

BPTT’de, RNN’nin kayıp fonksiyonu (loss function) tüm zaman adımlarında birikmiş hataların toplamı olarak ifade edilmektedir. Örneğin, T zaman adımına sahip bir ardışık veri için toplam kayıp, denklem 2.13’teki gibi tanımlanmaktadır:

$$\mathcal{L} = \sum_{t=1}^T \mathcal{L}_t(y_t, \hat{y}_t) \quad (2.13)$$

Burada:

$\mathcal{L}_t$, her bir zaman adımındaki kaybı temsil eder,

y_t , gerçek çıktıyı,

$\hat{y}_t$, modelin tahmin ettiği çıktıyı ifade eder.

BPTT algoritması, bu toplam kaybın ağırlık parametrelerine göre türevini hesaplamaktadır. RNN’lerdeki gizli durum h_t , denklem 2.11’deki gibi önceki gizli durum h_{t-1} ve mevcut giriş x_t ile hesaplanmaktadır.

BPTT’de, her bir zaman adımındaki kayıp, yalnızca o zamandaki parametrelerden değil, aynı zamanda geçmiş zamandaki gizli durumlar aracılığıyla önceki adımlardaki parametrelerden de etkilenmektedir. Dolayısıyla her bir parametre

güncellemesi için gradyan hesaplanırken zaman boyutundaki tüm geçmiş durumlar dikkate alınmaktadır.

Zaman boyutunda hesaplanan toplam gradyan aşağıda denklem 2.14'teki gibi yazılabilmektedir:

$$\frac{\partial \mathcal{L}}{\partial W_h} = \sum_{t=1}^T \frac{\partial \mathcal{L}_t}{\partial h_t} \frac{\partial h_t}{\partial W_h} \quad (2.14)$$

Türev, zincir kuralı kullanılarak her bir zaman adımına geri doğru yayılmaktadır. RNN'de her bir gizli durum bir önceki duruma bağlı olduğu için her bir zamandaki hata sinyalinin, ağıın parametreleri üzerinde kümülatif bir etkiye sahip olması sağlanmaktadır.

Süreç boyunca zaman bağımlılığı nedeniyle “kaybolan gradyan” veya “patlayan gradyan” problemleri ortaya çıkabilmektedir. Bu sorunları gidermek için LSTM (Long Short-Term Memory) ve GRU (Gated Recurrent Unit) gibi daha karmaşık hücre yapıları kullanılmaktadır.

Kaybolan Gradyan Problemi: Gradyanlar zaman adımlarında geriye doğru yayılırken zincir kuralı nedeniyle değerleri hızla küçülebilmektedir. Bu durum, ağıın daha eski bilgileri öğrenmesini zorlaştırmaktadır. Çünkü gradyanların değeri çok küçük olduğunda parametre güncellemeleri etkisiz hale gelmektedir.

Patlayan Gradyan Problemi: Bazı durumlarda, gradyanlar geriye yayılırken çok büyük değerlere ulaşabilir ve bu da öğrenme sürecini kararsız hale getirmektedir. Oluşan problemi çözmek için, gradyanların bir sınır değeriyle kesildiği “gradyan kesme” (gradient clipping) yöntemi kullanılabilmektedir.

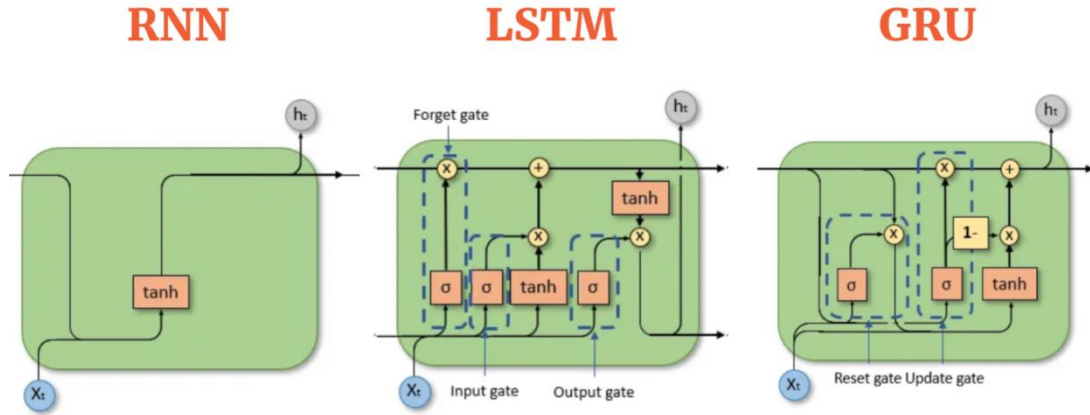
BPTT algoritmasının temel adımları şu şekildedir:

İleri Yayılım (Forward Pass): Girdiler tüm zaman adımlarında ileri doğru iletilir ve her bir zaman adımındaki çıktılar ve gizli durumlar hesaplanmaktadır.

Kayıp Hesaplama (Loss Calculation): Her bir zaman adımındaki tahmin edilen çıktılar ile gerçek çıktılar arasındaki kayıp hesaplanır ve toplam kayıp elde edilmektedir.

Geriye Yayılım (Backward Pass): Toplam kaybın her bir parametreye göre türevi, zincir kuralı ile geçmiş tüm zaman adımlarını kapsayacak şekilde hesaplanmaktadır. Gradyanlar, tüm zaman adımları üzerinden birikimli olarak geri yayılmaktadır.

Ağırlık Güncellemesi (Weight Update): Elde edilen gradyanlar kullanılarak ağırlık güncellenir ve bu süreç her bir öğrenme döngüsünde (epoch) tekrarlanmaktadır.



Şekil 2.6. RNN, LSTM ve GRU mimarileri.

BPTT, standart RNN'ler için etkin bir öğrenme algoritmasıdır. Ancak, kaybolan gradyan problemleri gibi zorluklarla başa çıkmak için LSTM ve GRU gibi gelişmiş RNN yapıları geliştirilmiştir. Bu yapılar, uzun vadeli bağımlılıkları daha verimli öğrenebilmek için özel kapılar kullanarak, bilgi akışını daha etkin bir şekilde kontrol etmektedir. Şekil 2.6'da mimarileri verilen yapıların matematiksel ifadelerine aşağıda değinilmektedir [21, 22].

- **Uzun Kısa Süreli Hafıza (Long Short-Term Memory - LSTM):**

LSTM, klasik RNN'lerdeki kaybolan gradyan problemini çözmek amacıyla geliştirilmiş özel bir RNN türüdür. LSTM hücresi, giriş kapısı, unutma kapısı ve çıkış kapısı gibi yapıları kullanarak hangi bilgilerin saklanacağını veya unutulacağını kontrol etmektedir. Bu sayede, model uzun süreli bağımlılıkları daha iyi öğrenmektedir. LSTM hücresinin temel denklemleri denklem 2.15, 2.16, 2.17, 2.18 ve 2.19'daki gibidir.

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (2.15)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (2.16)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (2.17)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \quad (2.18)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (2.19)$$

Burada:

i_t , giriş kapısıdır ve hangi yeni bilgilerin hücre durumuna ekleneceğini belirler.

f_t , unutma kapısıdır ve hücre durumundaki hangi bilgilerin unutulacağını belirler.

o_t , çıkış kapısıdır ve hücre durumundan hangi bilgilerin çıkış olarak aktarılacağını belirler.

c_t , hücre durumudur.

h_t , gizli durumdur.

- **Kapılı Tekrarlayan Birim (Gated Recurrent Unit - GRU):**

GRU, LSTM'ye benzer şekilde uzun vadeli bağımlılıkları öğrenmek amacıyla geliştirilmiş başka bir RNN türüdür. GRU'da, LSTM'nin aksine tek bir hücre durumu vardır ve GRU yapısı daha az parametre içermektedir. GRU'nun temel matematiksel ifadeleri denklem 2.20, 2.21, 2.22 ve 2.23'teki şekildedir:

$$z_t = \sigma(W_{zx}x_t + W_{zh}h_{t-1} + b_z) \quad (2.20)$$

$$r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r) \quad (2.21)$$

$$\tilde{h}_t = \tanh(W_{hx}x_t + r_t \cdot (W_{hh}h_{t-1} + b_h)) \quad (2.22)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (2.23)$$

Burada:

z_t , güncelleme kapısıdır ve önceki durumun ne kadarını yeni durumla değiştireceğini belirlemektedir.

r_t , reset kapısıdır ve önceki durumun ne kadarının göz ardı edileceğini belirler.

h_t , zaman t anındaki gizli durumdur.

Güncelleme ve reset kapıları sayesinde GRU, daha hızlı ve daha verimli bir şekilde uzun vadeli bağımlılıkları öğrenebilmektedir.

Bir sonraki kısımda dikkat mekanizma yapısından bahsedilmektedir.

2.5. Görsel Dikkat Mekanizması (Visual Attention Mechanism)

İnsanlar, geniş ve ayrıntılı sahnelerde tüm veriyi aynı anda işlemlemek yerine, dikkati belirli bölgelere yönlendirerek seçici bir algılama gerçekleştirmektedir. Görsel dikkat mekanizmaları, özellikle nesne segmentasyonu, nesne tanıma, görüntü başlığı oluşturma, insan eylemi tanıma ve görsel soru cevaplama gibi bilgisayarla görme uygulamalarında yoğun olarak kullanılmaktadır. Önerilen çalışmada bir görüntünün görsel olarak analizi yapıp, değiştirilmesi ya da iyileştirilmesi gereken kısımlara odaklanmasını sağlayacak bir görsel dikkat mekanizması temel alınmaktadır [23].

Tekrarlayan görsel dikkat ise, bir görsel sahnenin belirli alanlarına tekrarlayan bir dikkat süreciyle odaklanmayı amaçlayan bir dikkat mekanizmasıdır [24]. Yaklaşım, görüntünün önemli bölgelerini kademeli olarak analiz etmek için tekrarlayıcı (recurrent) bir yapı kullanılmaktadır. Her adımda, dikkat mekanizması önceki durumlara dayalı olarak görüntünün farklı bir bölgesine odaklanır ve böylece sahnenin bütünsel bir algısını oluşturmaktadır.

Derin öğrenmedeki gelişmelerle birlikte, görsel dikkat mekanizmalarının verimli şekilde uygulanması, çeşitli görevlerde model performansını önemli ölçüde artırmaktadır. Görsel dikkat modelleri, temel olarak Yumuşak Dikkat (Soft Attention) ve Sert Dikkat olarak iki ana kategoriye ayrılmaktadır [23]. Bu kategoriler, dikkat mekanizmasının görsel sahneye nasıl odaklandığını ve seçici olarak nasıl işlem yaptığını farklı şekillerde ele almaktadır.

Yumuşak dikkat mekanizması, görüntünün tüm bölgelerine sürekli olarak odaklanmak yerine, her bir bölgeye belirli bir ağırlık atayarak çalışan bir yapıdır [25]. Ağırlıklar, modelin görüntüde belirli bölgelerine daha fazla dikkat vermesini sağlamaktadır. Yumuşak dikkat mekanizmasında, tüm görüntü üzerinde bir odaklanma ağırlığı hesaplanır ve bu ağırlıklar bir softmax fonksiyonu kullanılarak normalize edilmektedir. Model dikkat verdiği her bir konum için bir olasılık dağılımı oluşturmaktadır.

Sert dikkat mekanizması, yumuşak dikkatten farklı olarak sahnenin yalnızca belirli bir bölgesine odaklanmaktadır. Model, her adımda tek bir konumu seçmekte ve o konuma ilişkin özellikleri işlemektedir [26]. Sert dikkat mekanizması, bir “Monte Carlo” yaklaşımıyla çalışır ve genellikle pekiştirmeli öğrenme algoritmalarıyla, özellikle reinforce algoritmasıyla birlikte uygulanmaktadır. Sert dikkat yapısı, dikkatin dağılımını sınırlandırarak modelin yalnızca belirli alanlarda dikkat harcamasını sağlamaktadır.

Önerilen çalışmada sert dikkat (Hard Attention) kullanılmaktadır. Sert dikkat yaklaşımı, seçiciliği ve işlem verimliliğini artırarak modelin daha etkili sonuçlar üretmesine olanak tanımaktadır. Sert dikkat yönteminin, özellikle derin öğrenme modellerinde stratejik görsel analiz için uygun bir yapı sunduğu, yapılan deneyler ve araştırmalarla gösterilmektedir.

Önerilen yöntemdeki işlevini daha iyi anlayabilmek için sert dikkat mekanizmasının matematiksel alt yapısı aşağıda incelenmektedir.

Sert dikkat yapısında seçilen konuma yönelik bir ödül sinyali alınmaktadır. Örneğin, görüntü tanıma gibi görevlerde model, seçilen konumdaki bilgi ile doğru tahmin yaparsa pozitif bir ödül alır, aksi durumda negatif bir ödül verilmektedir. Ödül sinyali reinforce algoritması ile hesaplanmakta ve bu ödüller, modelin gelecekte daha verimli dikkat seçimi yapabilmesi için politika öğrenmesine yardımcı olmaktadır.

Amaç Fonksiyonu ve Gradyan Hesaplama: Sert dikkat mekanizmasında amaç, görsel bilginin sadece ilgili bölgelere odaklanarak işlenmesini sağlamak ve ödülleri maksimize etmektir. Amaca yönelik olarak her bir adımda modelin aldığı ödüllerin toplamını maksimize edecek bir amaç fonksiyonu tanımlanmaktadır. Ancak sert dikkatin doğası gereği gradyanlar doğrudan hesaplanamaz. Çünkü konum seçimi rastgeledir. Bundan dolayı Monte Carlo örneklemeleri ile gradyan tahmini yapılmaktadır. Amaç fonksiyonu $L(\theta)$ denklem 2.24'deki gibi tanımlanabilir:

$$L(\theta) = \mathbb{E}_{l \sim p(l|\theta)} \left[\sum_{t=1}^T R_t \right] \quad (2.24)$$

Burada:

θ , model parametrelerini ifade etmektedir.

R_t , zaman t anında alınan ödül temsil etmektedir.

Amaç, reinforce algoritması ile elde edilen bu fonksiyonun gradyanlarını tahmin ederek modelin parametrelerini güncellemektir.

Reinforce Algoritması ile Güncelleme: Sert dikkat mekanizmasında, reinforce algoritması kullanılarak politika öğrenimi yapılmaktadır. Algoritma, her bir adımda gözlemlenen ödüllere göre politika güncellemelerini gerçekleştirmektedir. Amaç fonksiyonunun gradyanı, Monte Carlo örneklemeleri ile hesaplanır:

$$\nabla_{\theta} L(\theta) \approx \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \nabla_{\theta} \log p(l_t | \theta) R_t \quad (2.25)$$

Denklem 2.25’de verilen Monte Carlo örnekleme ifadesi model parametreleri θ için gradyanı vermekte ve her bir dikkat seçimi için ödül sinyalleri kullanılarak modelin öğrenme süreci optimize edilmektedir.

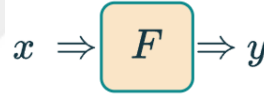
Tekrarlayan görsel dikkat yapısı, önerilen çalışma kapsamında ele alınan pekiştirmeli öğrenme ve dikkat mekanizmalarıyla uyumlu bir şekilde görsel sahnelerde belirli alanların iyileştirilmesi ve seçimi gibi süreçlerde etkili bir çözüm sunmaktadır. Görsel dikkat sürecinde, model her adımda bir görsel bölgeyi seçer ve bu seçimi bir sonraki dikkat alanını optimize etmek için kullanmaktadır. Model tekrarlayan bir dikkat döngüsü içinde çalışarak, görsel sahnenin en bilgilendirici alanlarına odaklanmaktadır.

Çalışmanın 3. kısmında görüntü süper çözünürlüğü hakkında temel problem tanımı verilmekte ve ardından çalışmayı ilgilendiren literatür hakkında bilgilendirme sunulmaktadır.

3. GÖRÜNTÜ SÜPER ÇÖZÜNÜRLÜĞÜ

3.1. Problem Tanımı

Süper çözünürlük yöntemleri, düşük çözünürlüklü görüntülerden yüksek çözünürlüklü versiyonlarını elde etmeyi amaçlayan çeşitli teknikler içermektedir. DL tabanlı FSR yöntemlerinde amaç, Şekil 3.1’de görüldüğü gibi düşük çözünürlüklü bir x giriş yüz görüntüsünü, yüksek çözünürlüklü bir y hedef yüz görüntüsüne dönüştürmektir.



Şekil 3.1. Süper çözünürlük için DL tabanlı F modeli nihai çıktı.

Çözünürlük artırma amacıyla x görüntüsü, model parametresi θ olan DL yöntemi $\mathcal{F}$ ile işlenerek hedef yüz görüntüsünün tahmin eden $\hat{x}$ yüz görüntüsü elde edilmektedir. Temel işlem yapısı denklem 3.1’de verilmiştir.

$$\hat{x} = \mathcal{F}(x | \theta) \quad (3.1)$$

Süper çözünürlük alanında geliştirilen yaklaşımlar, farklı veri işleme stratejileri ve modelleme yöntemleri kullanarak çözünürlük artırma performansını iyileştirmeye çalışmaktadır.

Bir sonraki başlıkta, konuya daha derinlemesine bakabilmek için literatürdeki çalışmalar incelenerek mevcut yöntemler ve elde edilen ilerlemeler değerlendirilecektir.

3.2. Literatür Araştırması

Çözünürlüğü artırma alanı; yüz tanıma [27], güvenlik sistemleri [28], tıbbi görüntüleme [29] ve afet yardımı [30] gibi çeşitli uygulama alanlarında önemli bir yere sahiptir. Çözünürlük alanındaki çalışmalar, özellikle yüz görüntüleri gibi detay açısından kritik alanlarda büyük önem taşımaktadır. Yüz süper çözünürlük yöntemleri genel olarak dört ana kategoriye ayrılmaktadır: enterpolasyon tabanlı

yöntemler, istatistiksel yöntemler, rekonstrüksiyon tabanlı yöntemler ve öğrenme tabanlı yöntemler [31]. Her bir yapı, düşük çözünürlüklü yüz görüntülerini yüksek çözünürlüklü hale getirmek için farklı stratejiler kullanmaktadır.

Enterpolasyon tabanlı yöntemler, FSR alanındaki en eski tekniklerden biri olup düşük çözünürlüklü görüntülerde eksik piksel değerlerini tahmin etmeye çalışmaktadır. Enterpolasyon yöntemleri, bilinen veri noktaları arasındaki eksik değerleri hesaplayarak yeni piksel değerleri oluşturmaktadır. Enterpolasyon yöntemlerinin en yaygın kullanılanları arasında bicubic ve bilinear enterpolasyon yer almaktadır [32-34]. Bicubic enterpolasyon, komşu 16 pikselin değerlerini kullanarak yeni bir piksel değeri tahmin ederken, bilinear enterpolasyon daha basit bir formül kullanarak sadece 4 komşu pikselden veri almaktadır [33]. Yöntemler, temel matematiksel modeller üzerine kuruludur ve hızlı hesaplama avantajına sahiptir. Ancak bu yöntemler yüz gibi detay odaklı alanlarda yetersiz kalmaktadır. Çünkü kullanılan teknikler yüksek frekanslı detayları geri kazanmakta başarısızdır ve bulanık görüntülerle sonuçlanmaktadır [34].

Enterpolasyon tabanlı yöntemlerin önemli bir kısıtı, yöntemlerin yüksek çözünürlük elde etmek için tahmin süreçlerini kullanmasıdır. Tahmin yöntemleri, yüz özelliklerinin bozulmasına ve detayların kaybolmasına neden olmaktadır. Özellikle düşük çözünürlüklü bir görüntünün 4x veya 8x ölçeklendirilmesi gibi durumlarda, enterpolasyon yöntemleri önemli derecede bulanıklıklar oluşturmaktadır. Dolayısıyla hem insan algısında hem de yüz tanıma algoritmalarında başarısız sonuçlara yol açmaktadır [31]. Özetle enterpolasyon yöntemleri, basit ve hızlı bir çözüm sunmalarına rağmen yüksek çözünürlüklü ve detaylı yüz görüntüleri üretme konusunda yetersiz kalmaktadır [31-34].

İstatistiksel yöntemler, düşük çözünürlüklü ve yüksek çözünürlüklü görüntüler arasındaki ilişkileri modellemeye çalışarak çözünürlüğü artırmayı hedeflemektedir. Görüntülerdeki yüksek frekanslı bileşenleri geri kazanmak için geliştirilmişlerdir ve yüz gibi karmaşık yapıları modellemeye çalışmaktadır. Özellikle piksel dağılımını modelleyerek yüksek çözünürlüklü görüntüler üretme konusunda başarı sağlamaktadırlar. İlk istatistiksel FSR yaklaşımı Baker ve Kanade tarafından, bayesian tabanlı bir çerçevede tanıtılmıştır ve yüz görüntülerindeki yüksek frekanslı detayların geri kazanılmasına odaklanmaktadır [5]. C. Liu ve ark., yüz görüntülerini yüksek çözünürlüklü hale getirmek için iki aşamalı bir yaklaşım sunmaktadır. İlk

aşamada, global parametrelerle tüm yüzün genel yapısı oluşturulur, ardından yerel, parametrik olmayan bir model ile daha ince detaylar eklenmektedir. Bu iki modelin birleşimi, düşük çözünürlüklü yüz görüntülerinden yüksek çözünürlüklü yüzlerin üretilmesinde önemli bir adım olarak kabul edilmektedir. Bahsi geçen çalışma, istatistiksel modelleme ile yüz yapılarının daha doğru şekilde elde edilmesini sağlamaktadır [35]. X.Wang ve ark., yüz süper çözünürlüğü için eigentransformation adı verilen bir yöntem önermektedir. Yöntem, yüzlerin düşük çözünürlüklü görüntülerinden yüksek çözünürlüklü versiyonlarını üretmek için yüzlerin temel bileşen analizini (PCA) kullanmaktadır. Yüz verisi, bu temel bileşenler uzayına projelendirilir ve ardından yüksek çözünürlüklü uzaya geri dönüştürülmektedir. Eigentransformation tabanlı yöntem, yüz görüntülerinde genel yapı ve detayların korunmasını sağlayan istatistiksel bir yaklaşımdır [36]. S. T. Roweis ve ark., yerel doğrusal gömme (LLE) yöntemini tanıtarak, yüz görüntülerini düşük çözünürlükten yüksek çözünürlüğe çıkarırken boyut indirgeme tekniklerini kullanmaktadır. LLE, her bir veri noktasını kendi komşularına göre analiz ederek, yüz görüntüsünün daha yüksek çözünürlüklü versiyonunu oluşturmak için komşu ilişkileri korumaktadır. LLE yaklaşımı, özellikle verilerin karmaşık ve çok boyutlu olduğu durumlarda yüz süper çözünürlüğü için etkili bir çözüm sunmaktadır [37]. H. Chang ve ark., düşük çözünürlüklü görüntülerin çözünürlüğünü artırmak için komşuluk gömme (neighbor embedding) yöntemi kullanmaktadır. Yöntem, her bir düşük çözünürlüklü görüntü parçasını, eğitim verilerindeki yüksek çözünürlüklü karşılıkları ile ilişkilendirerek süper çözünürlük sağlamaktadır. Komşu gömme yaklaşımı, düşük çözünürlüklü parçaların yüksek çözünürlüklü karşılıklarını bulmada başarılı bir yöntem olarak öne çıkar ve yüksek doğrulukta yüz süper çözünürlüğü sonuçları üretmektedir [38]. Bazı istatistiksel yaklaşımlar, özellikle farklı uzamsal dönüşümler kullanarak varyans ana bileşenini çıkarma yöntemlerini benimsemektedir. Y. Zhuang ve ark., yüz görüntülerindeki detayların iyileştirilmesi için LPH (local position and harmony) süper çözünürlük ve komşuluk rekonstrüksiyonu yöntemini kullanmaktadır. Yöntem, düşük çözünürlüklü görüntülerdeki kalan hataları düzeltmek ve daha net detaylar sağlamak amacıyla komşu piksellerle uyum içinde çalışmaktadır. LPH, düşük çözünürlüklü yüzlerin daha doğru ve detaylı bir şekilde yeniden yapılandırılmasını sağlamaktadır [39]. W. Liu ve ark., yüz çözünürlüğünü artırmak için komşu kombinasyonu ve dönüşümü adı verilen bir yöntem önermektedirler. Komşuluk tabanlı bu yöntem, düşük çözünürlüklü yüz görüntülerini yüksek çözünürlüklü hale

getirmek için komşu görüntülerdeki bilgiyi birleştirmekte ve dönüştürmektedir. Yaklaşım, yüz detaylarını daha iyi yakalayıp süper çözünürlük elde etmek amacıyla veri transformasyonunu kullanmaktadır [40]. İstatistiksel yöntemler, özellikle basit yüz yapılarında başarılı olabilirken karmaşık yüz yapıları ve detaylar içeren görüntülerde yeterince esnek olamamaktadır. Sabit parametrelerle çalışmak, özellikle yüz yapısının karmaşıklığını yeterince yansıtamayan bir problem olarak öne çıkmaktadır. Yüz görüntülerinde göz, burun, ağız gibi farklı bölgelerde değişkenlik gösteren detaylar sabit parametreler kullanıldığında doğru bir şekilde modellenemeyebilir [35-40]. İstatistiksel yöntemlerin bazılarında varyans temel bileşeni farklı uzamsal dönüşümlerde kullanılırken, diğerlerinde komşuluklar arasındaki yerel ve küresel ilişkileri çıkarmak veya piksellerin istatistiksel dağılımını modellemek için kullanılmaktadır. İstatistiksel çıkarım, yüz konturunu ve görüntünün genel görünümünü korurken yüksek frekanslı bileşenlerin vurgusunu artırmak için de kullanılmaktadır. Sabit parametreler veya sınırlı modelleme yetenekleri nedeniyle istatistiksel yöntemler özellikle yüz yapılarının ve varyasyonlarının karmaşık olduğu durumlarda yeterince esnek değildir. Ayrıntıları yakalamada zorluklar yaşamaktadır. Ayrıca yöntem parametrelerinin sabit değerlerde kullanılması, yüzün bazı özelliklerini yeterince yansıtmadığından elde edilen görüntülerde bulanık veya gerçekçi olmayan bölgelerin oluşmasına neden olabilmektedir. İstatistiksel yöntemlere kıyasla daha esnek teknikler geliştirilmiştir. Bunlardan ilki, hedef görüntünün bulanıklaştırılmış ve alt örneklenmiş halini düşük çözünürlüklü görüntüye yakın hale getirmeye çalışan yeniden yapılandırma (Rekonstrüksiyon) tekniğidir.

Rekonstrüksiyon tabanlı yöntemler, enterpolasyon ve istatistiksel yöntemlerin sınırlamalarını aşmayı amaçlayarak yüksek frekanslı detayların ve kenarların geri kazanılmasına odaklanmaktadır. Rekonstrüksiyon tabanlı yöntemler, görüntüdeki kayıp detayları yeniden oluşturmak için çeşitli stratejiler kullanmaktadır. Özellikle kenar koruma ve gradyan profillemeye teknikleri ile bu yöntemler, yüz yapısındaki ince detayları geri kazandırmaya çalışmaktadır [41]. Rekonstrüksiyon tabanlı yöntemlerin ana hedefi, düşük çözünürlüklü görüntülerdeki gürültüyü azaltırken, yüksek frekanslı detayları vurgulamak ve keskinliği artırmaktır. Rekonstrüksiyon, kenar bulanıklaştırma ve gradyan profili oluşturma yoluyla keskin ayrıntılara odaklanmayı [42-44] veya düşük frekanslı bölgelerdeki gürültüyü gidererek keskin ayrıntıları

korumayı amaçlamaktadır [45, 46]. J. Sun ve ark., görüntü süper çözünürlüğünde gradyan profili önceliği (gradient profile prior) kullanarak yüksek çözünürlüklü görüntüler elde etmeye odaklanmaktadır. Gradyan profilinin, bir görüntüdeki kenarların netliğini ve detaylarını artırmada önemli bir rol oynadığı vurgulanmaktadır. Yöntem, özellikle kenar detaylarını iyileştirerek görüntüdeki keskinlik kaybını azaltmayı hedeflemektedir. Sonuçlar, gradyan profil önceliğinin kenar bulanıklaştırmayı minimuma indirmede etkili olduğunu göstermektedir [42]. S. Dai ve ark., renkli görüntü süper çözünürlüğü için kenar yumuşaklığına dayalı bir öncelik olan SoftCuts yöntemini tanıtmışlardır. Yöntem, kenar bulanıklaştırmayı azaltarak daha keskin kenarların elde edilmesini sağlarken yumuşak kenar geçişleri oluşturmaya yönelik bir algoritma kullanmaktadır. Renkli görüntülerde hem renk doğruluğunu korurken hem de kenar detaylarını iyileştirme üzerine odaklanılmıştır. SoftCuts yaklaşımı, görüntülerin daha pürüzsüz ve doğal görünmesini sağlarken keskin detayların korunmasına yardımcı olmaktadır [43]. Q. Yan ve ark., tek bir görüntüden yüksek çözünürlüklü bir görüntü üretmek için gradyan profilinin keskinliği üzerine odaklanmaktadır. Gradyan profilleri, kenarların ne kadar keskin olduğunu belirlemede kullanılır ve bu yöntem, özellikle kenar keskinliğini artırarak daha net görüntüler elde etmeyi amaçlamaktadır. Yüksek çözünürlüklü görüntülerde, gradyan profili keskinliği görüntünün netliği ve detayları üzerinde önemli bir etkiye sahiptir. Yöntem, kenarların daha fazla belirgin olmasını sağlarken, bulanıklığı en aza indirmektedir [44]. A. Marquina ve ark., toplam varyasyon (TV) düzenlemesi ve Bregman iterasyonu kullanarak görüntü süper çözünürlüğü elde etmeye odaklanmıştır. TV düzenlemesi, görüntülerdeki gürültüyü ve istenmeyen artefaktları ortadan kaldırarak daha net görüntüler elde edilmesini sağlamaktadır. Teknik, özellikle düşük frekanslı bölgelerde gürültüyü giderip, kenarları koruyarak keskin detaylar sağlamada başarılı olmaktadır. Bregman iterasyonu ise çözünürlük artırma işlemi sırasında dengeyi sağlamak için kullanılmaktadır [45]. A. Buades ve ark., görüntüdeki gürültüyü azaltmak için yerel olmayan bir algoritma geliştirmişlerdir. Yöntem, görüntünün çeşitli bölgelerindeki benzerlikleri analiz ederek düşük frekanslı bölgelerdeki gürültüyü azaltmaktadır. Gürültü giderme işlemi özellikle görüntüdeki ince detayları korumak amacıyla yerel ve global bilgiler kullanılarak yapılmaktadır. Yöntem, keskin detayları koruyarak gürültüsüz ve yüksek kaliteli görüntüler üretir [46]. Birden fazla görüntüden hareket algılamayı birleştiren veya bazı tahmin yöntemleri kullanarak görüntüleri birleştiren yöntemler vardır [47-50]. Yeniden

yapılandırma tekniğinde keskin detaylara odaklanması nedeniyle çözüm uzayı sınırlı olmaktadır. Ayrıca, ölçek faktörü arttığında performans hızla düşmektedir. R. C. Hardie ve ark., eksik örnekleme yapılmış görüntülerin bir dizi halinde kullanılarak yüksek çözünürlüklü görüntü tahmini yapılmasını hedeflemektedir. Maksimum A Posteriori (MAP) kayıt yöntemi kullanılarak, birden fazla düşük çözünürlüklü görüntü hizalanır ve yüksek çözünürlüklü bir görüntü elde edilmektedir. Yaklaşım, hareketli veya sabit sahnelerden alınan görüntüler arasında en uygun hizalamayı sağlarken eksik verilerin tamamlanmasında da başarılı olabilmektedir. Özellikle birden fazla görüntü kullanılarak çözünürlük artırma için kullanılan bu yöntem, verimli sonuçlar sunmaktadır [47]. S. Farsiu ve ark., çoklu çerçeve süper çözünürlük (multiframe super-resolution) için hızlı ve sağlam bir algoritma geliştirmiştir. Yöntem, birden fazla düşük çözünürlüklü görüntüyü analiz ederek gürültü ve bulanıklık gibi sorunları ortadan kaldırmak ve yüksek çözünürlüklü bir görüntü üretmek üzerine kuruludur. Kayıp verilerin hızlı bir şekilde geri kazanılması için optimizasyon teknikleri kullanılmaktadır. Algoritma, hem hesaplama açısından verimlidir hem de gürültüye karşı dayanıklıdır [48]. M. Irani ve ark., görüntü iyileştirme için hareket analizi kullanarak çözünürlük artırma ve perdeleme gibi zorluklarla başa çıkmayı amaçlamaktadır. Yazarlar, bir görüntü sahnesindeki farklı hareketleri analiz ederek örtüşen veya şeffaf nesnelerin neden olduğu bilgi kaybını telafi etmeye çalışmaktadırlar. Hareket bilgisi, düşük çözünürlüklü görüntülerden yüksek çözünürlüklü sonuçlar üretmek için kritik bir rol oynamaktadır. Yöntem, çoklu çerçeve süper çözünürlük tekniklerinin geliştirilmesinde öncü olmaktadır [49]. Z. Lin ve ark., yeniden yapılandırma temelli süper çözünürlük algoritmalarını incelemektedir. Özellikle, görüntülerin çeviri hareketine tabi olduğu senaryolarda çözünürlük artırma tekniklerinin ne ölçüde başarılı olabileceği analiz edilmektedir. Yeniden yapılandırma yöntemlerinin belirli sınırlamalarla karşı karşıya kalabileceği ve çözünürlüğü artırma kapasitesinin sınırlı olduğu gösterilmektedir. Çalışma, yeniden yapılandırma tabanlı yöntemlerin performansını iyileştirmek için rehber niteliğindedir [50]. Rekonstrüksiyon yöntemleri çoklu görüntülerin birleştirilmesi ya da hareket tespiti gibi teknikler kullanarak, yüz detaylarını korumayı ve iyileştirmeyi amaçlamaktadır [28]. Örneğin, birden fazla düşük çözünürlüklü görüntüyü birleştirerek hareketli sahnelerde daha net yüz görüntüleri elde edilebilmektedir. Ancak bu yöntemler, özellikle büyük ölçeklendirme faktörlerinde performans kaybı yaşayabilmektedir. Ölçeklendirme faktörü arttıkça rekonstrüksiyon tabanlı

yöntemlerin çözüm uzayı daralmakta ve görüntülerin keskinliği azalmaktadır. Bu nedenle, rekonstrüksiyon tabanlı yöntemler SR problemlerinde belirli bir noktaya kadar etkili olabilirken daha yüksek ölçeklendirme gerektiren uygulamalarda performansları sınırlı kalmaktadır [51]. Rekonstrüksiyon tabanlı yöntemler, yüz görüntülerinde belirli bir keskinlik ve detay sağlasa da yüksek çözünürlüklü ve doğal görünen yüz görüntüleri üretme konusunda sınırlı yeteneklere sahip olabilmektedirler.

SR alanındaki son araştırmalar öğrenme tabanlı yöntemlere odaklanmaktadır [52-56]. H. Huang ve ark., yüz tanıma için süper çözünürlük yöntemi geliştirmiş ve doğrusal olmayan haritalamalar kullanılarak uyumlu özellikler üzerinde çözünürlük artırma gerçekleştirmişlerdir. Yazarlar, düşük çözünürlüklü yüz görüntülerindeki temel yüz özelliklerini yüksek çözünürlükte yeniden yapılandırmak için doğrusal olmayan bir dönüşüm modeli önermektedir. Yöntem, yüz tanıma sistemlerinde düşük çözünürlüklü görüntülerin etkili bir şekilde kullanılmasına olanak tanımaktadır. Sonuçlar, yaklaşımın doğruluğu artırarak yüz tanıma performansını iyileştirdiğini göstermektedir [52]. K. Nguyen ve ark., Gabor tabanlı özellikleri kullanarak yüz ve iris tanıma için özellik tabanlı bir süper çözünürlük çerçevesi sunmaktadır. Gabor filtreleri, biyometrik sistemlerde sıkça kullanılan yüz ve iris özelliklerini çıkarmak için kullanılmaktadır. Çalışmada, özellikle düşük çözünürlüklü biyometrik görüntüler üzerinde performansın artırılması için süper çözünürlük teknikleri uygulanmaktadır. Çözüm, düşük çözünürlüklü biyometrik verilerin yüksek doğrulukla tanınmasını sağlamak amacıyla çalışmaktadır [53]. L.Wu ve ark., yüz görüntülerinin öğrenme tabanlı bir süper çözünürlük algoritmasıyla hızlı bir şekilde yeniden yapılandırılmasını hedeflemektedir. Algoritma, düşük çözünürlüklü yüz görüntülerinden yüksek çözünürlüklü görüntüler elde etmek için verimli ve hızlı bir yöntem sunmaktadır. Çalışma, yüz tanıma sistemlerinde düşük çözünürlüklü görüntülerin işleme sürecini hızlandırmaya yönelik bir çözüm sunmaktadır. Sonuçlar yaklaşımın hızlı ve etkili olduğunu göstermektedir [54]. W. W. W. Zou ve ark., düşük ve yüksek çözünürlüklü yüz görüntüleri arasındaki ilişkiyi kernel uzayında öğrenmeye yönelik bir yöntem geliştirmiştir. Yöntem, düşük çözünürlüklü görüntülerden yüksek çözünürlüklü karşılıklarını tahmin etmek için kernel tabanlı bir dönüşüm kullanmaktadır. Yüz görüntülerinin süper çözünürlüğü için kullanılan bu yöntem, düşük çözünürlüklü görüntülerdeki özelliklerin yüksek çözünürlükte doğru

şekilde yeniden yapılandırılmasına olanak tanımaktadır. Kernel tabanlı yaklaşım, verimli ve doğru bir yüz süper çözünürlük çözümü sunmaktadır [55]. Y. He ve ark., tek kare yüz görüntülerinden süper çözünürlük elde etmek için öğrenme tabanlı bir yaklaşım geliştirmiştir. Model, düşük çözünürlüklü bir yüz görüntüsünden yüksek çözünürlüklü bir versiyon oluşturmak için veri tabanında öğrenilen örüntüleri kullanmaktadır. Yöntem özellikle yüz tanıma gibi biyometrik uygulamalar için tek kare üzerinden yüksek kaliteli yüz görüntüsü üretme üzerine odaklanmaktadır [56].

Son yıllarda popüler olan derin öğrenme (DL) ve konvolüsyonel sinir ağları (CNN) kullanılarak çözünürlük alanında büyük ilerlemeler kaydedilmektedir [57]. Öğrenme tabanlı yöntemler, düşük çözünürlüklü ve yüksek çözünürlüklü görüntü çiftleri arasında bir öğrenme süreci gerçekleştirmekte ve eğitim verisi üzerinden modelin öğrenmesi sağlanmaktadır. Öğrenme tabanlı SR yöntemleri, geleneksel yöntemlere kıyasla daha fazla detay yakalama ve görüntüdeki ince ayrıntıları daha doğru bir şekilde geri getirme potansiyeline sahiptirler [58]. CNN tabanlı yaklaşımlar, düşük çözünürlüklü görüntülerde kaybolan detayları geri kazanarak daha doğal ve yüksek çözünürlüklü yüz görüntüleri oluşturmaktadır. Görüntü özellikleriyle çalışan makine öğrenmesi yöntemlerine özellik tabanlı yöntemler denilmektedir. Makine öğrenmesi kullanılarak elde edilen görüntülerde yüksek frekanslı detaylara odaklanmada sınırlar vardır. Oluşabilecek sorunların üstesinden gelmek ve daha etkili sonuçlar elde etmek için DL modelleri önerilmiştir [59].

DL tabanlı SR çalışmaları üç kategoriye ayrılmaktadır: yerel tabanlı yöntemler, global tabanlı yöntemler ve yerel-global tabanlı yöntemler. Yerel tabanlı kategoride, görüntü veya görüntünün bir bloğu derin ağ modeline girdi olarak verilmektedir [7, 60-66]. Yerel yaklaşımda, bölge tabanlı derin evrimsel ağlar kullanılarak yüz çözünürlüğü artırma (face hallucination) işlemi gerçekleştirilebilmektedir. Örneğin, Lu ve ark., bölge tabanlı derin evrimsel ağlar (CNN) kullanarak yüz görüntülerinde belirli bölgelere odaklanmış ve bu bölgelerde çözünürlüğü artırmayı hedeflemişlerdir [60]. Daha sonra aynı araştırma grubu, bölgesel dikkat mekanizmaları (split-attention) ile yüz çözünürlüğünü artırmayı sağlayan bir yöntem geliştirmiştir [61]. Diğer taraftan, Cheng ve ark. ve Shi ve ark. tarafından sunulan yaklaşımlar yüz çözünürlüğü artırmada kimlik korunmasına ve dikkat mekanizmalarına dayalı derin pekiştirmeli öğrenme teknikleri kullanmaktadır. Gelişen yeni teknikler, özellikle yüzün belirli alanlarına odaklanarak hem düşük hem de yüksek çözünürlüklü

görüntüler arasındaki kimlik tutarlılığını korumaya çalışmaktadırlar [7, 62]. Cao ve ark. tarafından geliştirilen bir başka model, yüz çözünürlüğü artırmada dikkat tabanlı derin pekiştirmeli öğrenmeyi kullanarak belirli bölgelere odaklanıp daha detaylı bir çözünürlük artırma gerçekleştirmektedir [63].

Yerel ve global yaklaşımları birleştiren yöntemlerde, görüntünün farklı bölgeleri dikkat mekanizmalarıyla işlenirken yüzün genel yapısı da korunmaktadır. Yerel-global yöntemlerde, hem bölgesel ayrıntılar iyileştirilirken hem de yüzün global yapısındaki bütünlük sağlanmaktadır. Pan ve ark. tarafından geliştirilen çift evrişimli sinir ağı (Dual CNN), düşük seviye görsel görevlerde yerel ve global detayları bir araya getirerek başarılı sonuçlar elde etmiştir [64]. Benzer şekilde, Dong ve ark. tarafından önerilen derin Wiener dekonvolüsyon ağı (DWDN), hem yerel detayları hem de genel yapıyı dikkate alarak bulanık görüntülerin iyileştirilmesini amaçlamaktadır [65, 66].

Bir görüntüyü daha küçük bloklara bölerek uygulanan SR yöntemleri global bilgi kaybına, bulanıklaşmaya ve blokların kenarlarında artefaktlara neden olmaktadır. Görüntünün ana yapısını korumak için global tabanlı yaklaşımlar geliştirilmiştir. Zhou ve arkadaşları, yüz görüntülerini doğal koşullarında işleyerek tüm yüzün global yapısını korumayı amaçlayan bir yüz halüsinasyon yaklaşımı sunmaktadırlar [67]. Huang ve arkadaşları ise çok ölçekli yüz süper çözünürlüğü elde etmek amacıyla dalgacık tabanlı bir CNN kullanarak, global yapıyı korurken detaylı yüz özelliklerini iyileştiren bir yöntem geliştirmişlerdir [68].

CNN yöntemlerine ek olarak eğitimden önce hazırlanan ek yüz bilgilerini ve haritalarını içeren yöntemler de bulunmaktadır. Chen ve ark. tarafından geliştirilen FSRNet, yüz çözünürlüğünü artırmada yüz önceliklerini öğrenmeye dayalı uçtan uca bir yaklaşım sunmaktadır [69]. Ma ve ark., dikkatli iyileşme ve işaret tahmini arasında denge sağlayarak, yüz süper çözünürlüğünü adım adım iyileştiren bir yöntem önermişlerdir [70].

Ayrıca, hem frekans detaylarını hem de genel yapıyı korumak için uzamsal ve frekans bilgilerini birleştiren yöntemler de önerilmiştir. Wang ve ark., yüz süper çözünürlüğü için uzamsal ve frekans karşılıklı öğrenme mekanizmasını geliştirmişlerdir [71]. Dastmalchi ve Aghaeinia, kimlik korumayı hedefleyen bir

dalgacık entegreli çekışmeli ađ tabanlı yöntem sunarak, çok düşük çözünürlüklü yüz görüntülerinde yüksek kalitede sonuçlar elde etmeyi başarmaktadırlar [72].

Yerel veya global tabanlı yöntemlerin eksiklikleri, mesafe ve katmanlar arası ilişkileri daha iyi kullanan yerel-global tabanlı yöntemlerle giderilmiştir. Ledig ve ark. tarafından geliştirilen SRGAN, gerçekçi görünümlü tek görüntü süper çözünürlüğü elde etmek için üretken çekışmeli ađ (GAN) kullanmaktadır [73]. Zhang ve ark., residual dense network (RDN) yapısını kullanarak görüntü süper çözünürlüğünde yerel ve global bilgileri entegre eden başarılı bir yöntem önermektedir [74]. Ayrıca, Xu ve ark., residual dense network yapısını temel alarak daha yoğun bir bilgi akışını sağlayan Dense ByNet modelini geliştirmişlerdir [75]. Wang ve ark., ESRGAN ile hem düşük seviyeli hem de yüksek detaylı görsel bilgilerin geri kazanılmasını sağlayan bir yapıyı sunmaktadır [76].

Güncel çalışmalarda, [77] RRDB (Residual-in-Residual Dense Block) modeline dayalı görüntü süper çözünürlük algoritmaları yüksek performans sunmakta ve genel yapının korunması konusunda etkili sonuçlar üretmektedir. Chen ve arkadaşları, bu modeli iki boyutlu RRDB ve kenar iyileştirme sağlayan bir GAN yapısı ile geliştirerek, özellikle uydu görüntülerinin süper çözünürlüğünde başarılı sonuçlar elde etmektedirler [78]. RRDB, görüntülerin çözünürlüğünü artırırken aynı zamanda insan algısına daha yakın sonuçlar üretmektedir [31]. RRDB'nin avantajı, düşük çözünürlüklü yüz görüntülerini yüksek detaylı ve doğal görünen yüz görüntülerine dönüştürme yeteneğidir. RRDB yapısı özellikle göz, burun ve ağız gibi yüzün kritik bölgelerinde daha fazla detay sağlar ve bu da yüz tanıma gibi uygulamalarda önemli bir avantaj sunmaktadır.

Son zamanlarda derin üretken modeller ortaya çıkmaktadır. Gerçekçi veri örneklerini (ör. görüntü, ses, metin) öğrenme ve üretme yetenekleri nedeniyle stil aktarımı, veri boyama, görüntü oluşturma ve süper çözünürlük gibi birçok zorlu görev için derin üretken modeller tercih edilmektedirler. Yöntemlerin en belirgin özelliđi, verideki dağılımı öğrenerek iki aşamalı bir süreçle yeni benzer örnekler üretmesidir [79]. Süper çözünürlük alanında yüksek kaliteli görüntüler üreten üretken çekışmeli ađ [69, 73, 76, 80], normalize edici akışlar [81, 82], otoregresif modeller [83, 84] ve difüzyon tabanlı [79, 85] yöntemlere örnek olarak verilebilir.

GAN'lar, üreteç ve ayrıştırıcı olmak üzere iki bileşenden oluşan modellerdir. Üreteç, düşük çözünürlüklü verilerden yüksek çözünürlüklü örnekler üretirken ayrıştırıcı, bu verilerin gerçek mi sahte mi olduğunu anlamaya çalışmaktadır. Goodfellow ve ark. tarafından önerilen çalışma, GAN'ların bu özelliğini vurgulamakta ve süper çözünürlük gibi zorlu görevlerde kullanılabilirliğini göstermektedir [80]. Ancak, GAN'ların en büyük sorunlarından biri optimizasyon kararsızlığı ve mod çökmesi (çeşitlilik kaybı) problemidir. NF (Normalize Edici Akışlar) ise verinin dağılımını öğrenerek tersinir dönüşümlerle yeni örnekler üretmektedir. Yöntem, daha hızlı eğitim süreci sunmasına rağmen ürettiği örnekler bazen optimal kalitenin altında kalabilmektedir. Yu ve ark., NF yönteminde dalgacık tabanlı bir yaklaşımla eğitimi hızlandırmayı başarmışlardır [81], Kingma ve Dhariwal ise Glow modelini sunarak tersinir 1×1 konvolüsyonlarla yüksek kaliteli görüntüler üretmektedir [82].

Otoregresif Modeller ise her adımda bir sonraki örneği tahmin ederek veri üretmektedir. A. van den Oord ve ark. tarafından geliştirilen WaveNet, otoregresif yapıların ses ve görüntü üretimi konusundaki potansiyelini göstermektedir [83]. PixelCNN ise, otoregresif modeller arasında yer alır ve görüntüleri piksel bazında işleyerek süper çözünürlük gibi görevlerde başarılı sonuçlar elde etmektedir [84]. Otoregresif modellerin en büyük sınırlaması ise yüksek hesaplama maliyeti ve yavaş üretim hızıdır.

Son zamanlarda difüzyon tabanlı yöntemlerin popülerliği artmaktadır. İteratif yapıları nedeniyle yüksek hesaplama maliyetleri ve uzun eğitim süreleri gerektirmektedir. Saharia ve ark., difüzyon tabanlı süper çözünürlük yöntemlerini kullanarak tekrarlamalı iyileştirme yoluyla yüksek kaliteli görüntüler üretmektedir [79]. Ancak bu modellerin belirsizlik azaltma süreci bazen yüz görüntülerinin doğal görünümünü etkileyen yapay etkilere yol açabilmektedir. Noroozi ve ark. tarafından geliştirilen “stable diffusion” bu alanda ölçeklendirilmiş dönüşüm yöntemiyle hızlı ve etkili çözümler sunarak yapay etiket sorunlarını en aza indirmeyi amaçlamaktadır [85]. Kısıtlamalarına rağmen, difüzyon tabanlı yöntemler detay ve doğruluk gerektiren problemlerin çözümünde öne çıkmakta ve SR uygulamalarında daha yaygın olarak kullanılması beklenmektedir [79].

Her bir derin üretken model türünün belirli avantajları olduğu gibi bazı zayıf yönleri de bulunmaktadır. Örneğin difüzyon tabanlı modeller insan algısına hitap eden yüksek kaliteli ve gerçekçi görüntüler üretebilse de çıktıların hedef görüntünün

gerçekliğini karşılayıp karşılamadığı konusunda belirsizlikleri bulunmaktadır. Corvi ve arkadaşları, GAN'lar ve difüzyon modelleri tarafından üretilen sentetik görüntülerin ilginç özelliklerini araştırmış ve bu görüntülerin gerçekçilik seviyeleri üzerinde durmuşlardır [86]. Bununla birlikte, difüzyon yöntemlerinin algısal performanslarına rağmen süper çözünürlük karşılaştırma metriklerinde beklediği gibi performans göstermediği de gözlemlenmektedir [87]. Ayrıca, bu tür sentetik görüntülerin tespit edilmesi ve doğruluklarının ölçülmesi üzerine çalışmalar da devam etmektedir [88].

DL alanında, bilgisayarla görme problemlerindeki son gelişmelerden biri dikkat mekanizmalarıdır [23]. Kısım 2'de ayrıntılı şekilde bahsedilen dikkat mekanizmasının literatür taraması bu kısımda incelenmektedir.

DL'deki dikkat fikri en alakalı verilere odaklanmayı amaçlamaktadır. Hassanin ve ark. tarafından yapılan kapsamlı bir inceleme dikkat mekanizmalarının görüntü işleme ve derin öğrenmede oynadığı rolü ortaya koymaktadır [23]. Dikkat temelli çalışmaların iki temel yapısına kısım 2'de değilmişti. Bunlar yumuşak dikkat ve sert dikkat yapılarıdır.

Yumuşak dikkat mekanizması çoğunlukla Softmax ve sigmoid gibi farklılaştırılabilir yumuşak geçiş fonksiyonlarını kullanmaktadır. Böylece, eğitim süresi boyunca model kendini sürekli ve verimli bir şekilde geliştirebilmektedir. Ancak yumuşak dikkat yöntemleri yüksek miktarda girdi verisi ve hesaplama karmaşıklığına sahiptirler [7, 23, 62, 63]. Cheng ve ark., derin pekiştirmeli öğrenme kullanarak kimlik koruyucu yüz çözünürlüğü artırmada yumuşak dikkat mekanizmasının önemini vurgulamaktadırlar [62]. Shi ve ark., dikkatli sıralama optimizasyonunu kullanarak yüz çözünürlüğü artırma problemini çözmek için yumuşak dikkati ve DRL'yi birleştirmektedirler [7].

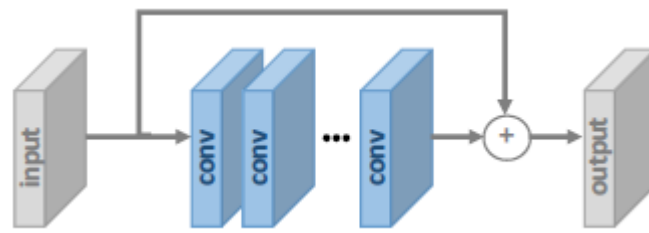
Sert dikkatte türevlenebilir fonksiyonlar yerine varyasyonel yöntemler veya pekiştirmeli öğrenme yöntemleri kullanılmaktadır [23, 89-91]. Cao ve ark., dikkat farkındalıklı yüz halüsinasyon yöntemlerinde sert dikkati ve derin pekiştirmeli öğrenmeyi birleştirerek başarılı sonuçlar elde etmişlerdir [63]. Sert dikkat yöntemlerinde daha düşük hesaplama maliyetleri sağlanırken verinin en kritik özelliklerine odaklanmak mümkündür. Ancak, RL yöntemlerinin optimizasyonu ve eğitim süreçlerinde zorluklar yaşanabilmektedir [89-91]. Dosovitskiy ve ark.,

görüntü tanımda transformers tabanlı bir yaklaşım önererek sert dikkat mekanizmalarının görsel görevlerde nasıl kullanıldığını göstermektedirler [89]. Lu ve ark., tek görüntü süper çözünürlüğü için transformer tabanlı dikkat modelleri kullanarak hem yumuşak hem de sert dikkat yöntemlerini birleştirmişlerdir [90]. Ayrıca, You ve ark., sert dikkat için Gaussian tabanlı dikkat mekanizmaları kullanarak doğal dil işleme alanında başarılı sonuçlar elde edebilmektedirler [91].

3.3. Derin Öğrenme Mekanizmalarına Genel Bakış

Derin öğrenme modellerinin performansını artırmak ve daha karmaşık problemleri çözmek için farklı yapısal mekanizmalar geliştirilmiştir. Mekanizmalar, öğrenme sürecini iyileştirmek modelin derinleşmesini sağlamak ve bilgi akışını optimize etmek amacıyla tasarlanmıştır. Görüntü süper çözünürlüğü uygulamalarında kullanılan yaygın tekniklerin temel yapılarına aşağıda değinilecektir.

- Artık Öğrenme (Residual Learning)

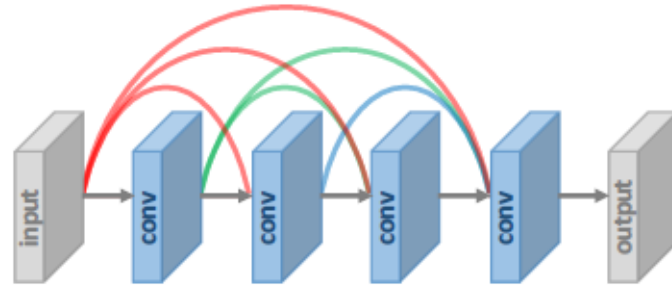


Şekil 3.2. Artık öğrenme (Residual Learning) yapısı.

Artık öğrenme, derin sinir ağlarında gradyan kaybolma (vanishing gradient) ve gradyan patlaması (exploding gradient) gibi problemlerin üstesinden gelmek için kullanılan bir yapıdır. Yöntem, giriş verisini doğrudan çıktı katmanına ileten bir atlama bağlantısı (skip connection) ekleyerek, modelin öğrenmesi gereken bilgiyi sadeleştirmesine yardımcı olmaktadır [92].

Şekil 3.2’de görüldüğü gibi, geleneksel evrişim katmanlarının (conv) üzerine eklenen doğrudan bağlantı, giriş verisini son çıktıyla toplayarak ağırlık yükünü azaltmaktadır. Böylece derin katmanlarda kaybolan gradyanlar daha iyi taşınır ve modelin öğrenme süreci hızlanmaktadır. Özellikle ResNet (Residual Network) gibi derin öğrenme modellerinde yaygın olarak kullanılan artık öğrenme mekanizması daha derin ağların eğitilmesini kolaylaştırarak yüksek doğrulukta sonuçlar elde edilmesini sağlamaktadır [73].

- Yoğun Bağlantılar (Dense Connections)

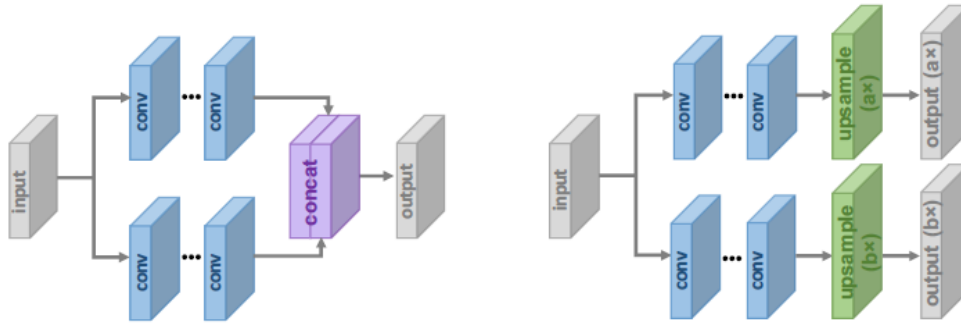


Şekil 3.3. Yoğun bağlantılar (Dense Connections) yapısı.

Yoğun bağlantılar, DenseNet (Densely Connected Convolutional Networks) gibi derin öğrenme modellerinde kullanılan bir bağlantı mekanizmasıdır [74, 77]. Yoğun bağlantı yapısı, her katmandaki çıktının sonraki tüm katmanlara giriş olarak iletilmesini sağlamaktadır. Böylece, derin sinir ağlarında bilgi akışı artırılır ve parametre verimliliği optimize edilmektedir.

Şekil 3.3'te görüldüğü gibi, her evrişim (conv) katmanının çıktısı yalnızca bir sonraki katmana değil ondan sonra gelen tüm katmanlara da iletilmektedir. Farklı renklerle gösterilen bağlantılardan görüldüğü üzere her katmanın daha önceki katmanlardan gelen özellikleri doğrudan kullanması sağlanmaktadır.

- Çok Yollu Öğrenme (Multi-Path Learning)



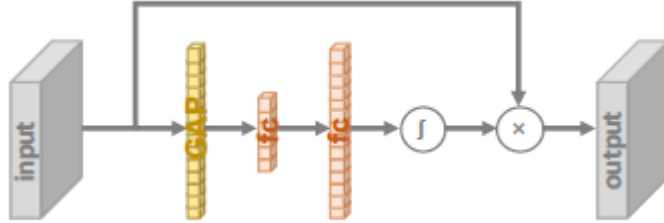
Şekil 3.4. Çok yollu öğrenme (Multi-Path Learning) yapısı.

Çok yollu öğrenme (Multi-Path Learning), derin sinir ağlarında bilgi akışını çeşitlendirmek ve farklı özellikleri aynı anda öğrenmek amacıyla kullanılan bir yöntemdir. Çok yollu öğrenme yapısı, modelin aynı giriş verisini farklı yollardan işlemeye olanak tanıyarak belirli bir göreve özgü daha zengin ve kapsamlı özellik haritaları oluşturmasını sağlamaktadır.

Şekil 3.4'te gösterildiği gibi giriş verisi birden fazla paralel evrişimli yol (conv) üzerinden işlenmektedir. Soldaki yapı her yolun çıktılarını birleştiren

bir concat (birleştirme) işlemi içerirken sağdaki yapı farklı ölçeklerde çıktılar üretmek için upsample (yukarı örnekleme) katmanları içermektedir. Çok yönlü öğrenme yöntemine örnek olarak LapSRN modeli [93], hem düşük hem de yüksek boyutlu uzaylarda bilgi akışını sağlamaktadır.

- Dikkat Mekanizmaları (Attention Mechanisms)

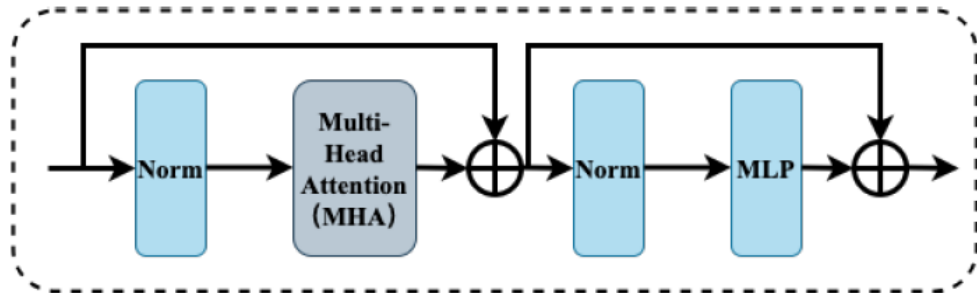


Şekil 3.5. Dikkat Mekanizmaları (Attention Mechanisms).

Dikkat mekanizmaları, derin öğrenme modellerinde belirli özellikleri ön plana çıkararak bilgi işleme sürecini iyileştiren yapılar arasında yer almaktadır. Dikkat mekanizmaları, giriş verisinin farklı bölümlerine dinamik olarak ağırlık vererek, önemli bilgilere odaklanılmasını sağlar ve modelin gereksiz bilgileri göz ardı etmesine yardımcı olmaktadır.

Şekil 3.5'teki yapı Kanal Dikkat Mekanizması (Channel Attention Mechanism) olarak adlandırılan bir dikkat mekanizmasına aittir [63]. Mekanizma, giriş özellik haritalarının kanallar arası önem derecelerini belirlemek için kullanılmaktadır. Kanal dikkati ve uzamsal dikkat [94] gibi mekanizmalar, ağın hangi bilgiyi daha detaylı işlemesi gerektiğini öğrenmesine olanak tanımaktadır.

- Dönüştürücü Tabanlı Modeller (Transformer-Based Models)



Şekil 3.6. Dönüştürücü Tabanlı Modeller (Transformer-Based Models).

Dönüştürücü tabanlı modeller, derin öğrenmede uzun vadeli bağımlılıkları yakalayabilen ve sekans tabanlı veriler üzerinde üstün performans gösteren bir mimari türüdür. Transformer mimarisi özellikle doğal dil işleme (NLP), bilgisayarla

görü (CV) ve görüntü süper çözünürlük gibi alanlarda yaygın olarak kullanılmaktadır.

Şekil 3.6’da gösterilen yapı, Transformer Bloğunun temel bileşenlerinden oluşmaktadır. Yapı, çoklu başlık dikkat (Multi-Head Attention - MHA) ve tam bağlantılı katman (MLP - Multi-Layer Perceptron) bileşenlerini içeren, normalizasyon (Norm) ve artık bağlantılar (Residual Connections) ile optimize edilmiş bir sistemdir. Dönüştürücü tabanlı yapı, girdinin dikkat mekanizması üzerinden geçirilerek önemli özellikleri vurgulamasını ve ardından tam bağlantılı bir ağ ile derin öğrenme işlemlerinin gerçekleştirilmesini sağlamaktadır. Transformer modelleri örnek SwinIR [95] gibi modeller, hem küresel hem de yerel özellikleri öğrenebilmektedir.



4. DERİN PEKİŞTİRMELİ ÖĞRENME İLE YÜZ SÜPER ÇÖZÜNÜRLÜĞÜ

Son yıllarda FSR üzerine yapılan çalışmalar derin öğrenme tabanlı yöntemlerin başarısını göstermiş olsa da bazı önemli sınırlamaları da beraberinde getirmektedir. Literatürde yer alan mevcut yöntemler yüz görüntülerindeki detayları iyileştirmede başarılı olsa da bazı yönlerden eksik kalmaktadır [74, 96]:

Bölgesel Detay Kaybı: Mevcut yöntemler [60, 69, 71], tüm yüz görüntüsünü aynı ölçüde iyileştirmeye çalışırken, yüzün farklı bölgelerinde aynı düzeyde başarı gösterememektedir. Örneğin göz, burun, ağız gibi yüzün önemli bölgelerindeki detayların korunması kritik öneme sahipken mevcut yöntemler bu bölgelerde istenen iyileştirmeyi tam anlamıyla sağlayamamaktadır. Özellikle yüksek frekanslı detayların kaybolması yüz tanıma başarısını olumsuz etkileyen bir faktördür.

Genelleme Problemi: Literatürdeki bazı FSR yöntemleri [1, 7, 62], eğitim aşamasında belirli veri setleri üzerinden eğitildiklerinden farklı yüz yapıları ve ifadeler içeren görüntülere genelleme yapmakta zorlanmaktadır. Farklı yaş grupları, etnik kökenler veya yüz ifadeleri gibi varyasyonlar mevcut yöntemlerin performansını sınırlayan faktörlerdir. Genelleme kapasitesi düşük olan modeller, farklı ortamlarda alınan yüz görüntülerinde yeterli performansı sağlayamamaktadır.

Zaman ve Hesaplama Gücü Gereksinimi: Bazı derin öğrenme tabanlı FSR modelleri [58, 93, 97], çok katmanlı ağ yapıları nedeniyle yüksek hesaplama gücü ve uzun işlem süresi gerektirmektedir. Uzun işlem süresi uygulamalarda yöntemlerin kullanılabilirliğini kısıtlamaktadır. Ayrıca yüksek hesaplama gerektiren modeller daha düşük işlem gücüne sahip cihazlarda kullanılamamakta veya performans kayıplarına yol açmaktadır.

Önerilen çalışmanın temel motivasyonu, yukarıda bahsedilmekte olan mevcut FSR yöntemlerinin eksik kaldığı alanlarda iyileştirmeler yaparak daha başarılı ve etkili bir yüz süper çözünürlük yöntemi sunmaktır. Yöntem, derin pekiştirmeli öğrenme tabanlı bir parça seçimi stratejisi ile yüz görüntüsündeki belirli bölgeleri iteratif olarak seçerek görüntüyü iyileştirmeyi amaçlamaktadır. Böylece yüzün önemli bölgelerindeki detayları koruyarak daha yüksek bir doğruluk oranı sağlanmaktadır.

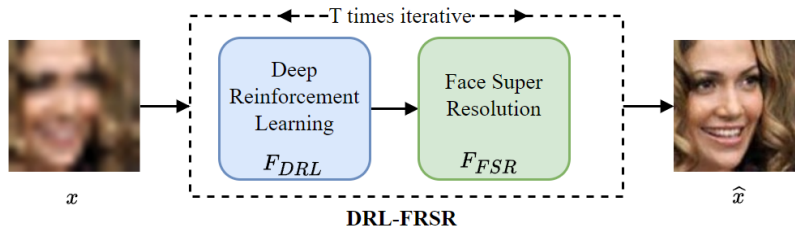
Çalışma aynı zamanda RRDB tabanlı bir yapı kullanarak hem yerel seçimler ile hem de yüz genelinde daha gerçekçi ve detaylı iyileştirmeler yapmaktadır.

Modelde sürecin temelini, derin pekiştirmeli öğrenme tabanlı bir dikkat mekanizması oluşturmaktadır. Model düşük çözünürlüklü bir yüz görüntüsünden sıralı bir şekilde belirli sayıda parça seçer. Her adımda seçilen parça yüzün yüksek çözünürlüklü bölgeleriyle uyumlu olacak şekilde iyileştirilmekte ve görüntüye gömülmektedir. Yöntem, FSR problemini Kısmen Gözlemlenebilir Markov Karar Süreci (Partially Observable Markov Decision Process, POMDP) olarak ele alır. Yaklaşım her adımda mevcut gözlemlerden öğrenerek ve geçmiş bilgilerden faydalanarak iyileştirilmesi gereken bölgeleri belirlemektedir. Süreç, modelin hem geçmiş kararları analiz edebilmesini hem de yeni bölgeler üzerinde daha bilinçli seçimler yapabilmesini sağlamaktadır.

Sürecin Genel Akışı:

- Durum (State): Her bir iterasyon, yüz görüntüsünün mevcut durumunu analiz eder ve seçilen parçalar üzerinden bilgi çıkarılmaktadır.
- Aksiyon (Action): Model, dikkat mekanizmasını kullanarak bir sonraki adımda iyileştirilecek parça bölgesini belirlemektedir.
- Ödül Mekanizması: Yüzün genel yapısına ve detayların iyileştirilmesine dayalı bir ödül sistemi kullanılarak, modelin her iterasyonda daha doğru parça seçimleri yapması sağlanmaktadır.

İteratif süreç, yüz görüntüsündeki her bir bölgenin belirli bir sıra ile iyileştirilmesi ve görüntüye entegre edilmesiyle sonuçlanmaktadır. Modelin öğrenme süreci, geçmiş bilgilerle sürekli güncellenerek görüntünün genel çözünürlüğünü artırır ve aynı zamanda kritik yüz bölgelerinin detaylarını iyileştirmektedir.



Şekil 4.1. Modelin temel işleyişi.

Şekil 4.1, önerilen “Yüz Bölgelerinin Derin Pekiştirmeli Öğrenme Tabanlı Süper Çözünürlüğü (DRL-FRSR)” yönteminin genel işleyişini göstermektedir. Modelin ana

yapısı, düşük çözünürlüklü yüz görüntüsünden (giriş görüntüsü, x) yüksek çözünürlüklü tahmin edilen yüz görüntüsünü ($\hat{x}$) elde etmeyi amaçlamaktadır. Süreç, iteratif bir yapıda çalışan iki ana bileşen ile gerçekleştirilmektedir.

Derin Pekiştirmeli Öğrenme Modülü ($\mathcal{F}_{DRL}$): Modül, yüz görüntüsündeki kritik bölgeleri (parçaları) belirleyerek her adımda hangi bölgenin iyileştirileceğini seçmektedir. Pekiştirmeli öğrenme tabanlı modül, önceki adımlarda iyileştirilen parçaların bilgilerini kullanarak yüz genelinde tutarlılığı sağlarken yüksek çözünürlüklü bölgelerin belirginliğini artırmaktadır.

Yüz Süper Çözünürlük Modülü ($\mathcal{F}_{FSR}$): Seçilen parça, süper çözünürlük modülü tarafından iyileştirilmektedir. Modül, DRL tarafından belirlenen bölgeye odaklanarak detayları artırır ve iyileştirilmiş parça yüz görüntüsüne entegre edilir. Böylece yüz genelindeki çözünürlük artışı sağlanmaktadır.

$\mathcal{F}_{DRL}$ ve $\mathcal{F}_{FSR}$ modülleri, T kez yinelemeli olarak çalışarak, düşük çözünürlüklü yüz görüntüsünden adım adım yüksek çözünürlüklü bir görüntü oluşturmaktadır. İteratif yapı, modelin yüz genelindeki ince detayları daha iyi öğrenmesine ve yüksek çözünürlüklü görüntüde tutarlılığı korumasına yardımcı olur.

Yöntemde, T tane bölge (parça) belirlenmektedir. Her adımda bir parça iyileştirilerek ve görüntüye gömülerek yeni görüntü $\hat{x}$ oluşturulmaktadır. Parça iyileştirme işlemi aşağıdaki şekilde ifade edilebilir.

$$\hat{x} = \mathcal{F}_{FSR}((\mathcal{F}_{DRL}(x | \theta_r)) | \theta_h) \quad (4.1)$$

Denklem 4.1’de, $\mathcal{F}_{DRL}$ ve $\mathcal{F}_{FSR}$ sırasıyla DRL modelini ve çözünürlüğü arttıran modeli temsil etmektedir. θ_r ve θ_h ise modellerin parametrelerini belirtmektedir. Yüz bölgelerinin seçiminde DRL kullanıldığından ve seçilen bölge SR yöntemi ile iyileştirildiğinden önerilen yöntem “Yüz Bölgelerinin Derin Pekiştirmeli Öğrenme Tabanlı Süper Çözünürlüğü” olarak isimlendirilmiştir.

$\mathcal{F}_{DRL}$ modelinde, yüz görüntüsüne dikkat tabanlı bir yaklaşım uygulanmaktadır. Modelin çıktısı, dikkat edilmesi (iyileştirilmesi) gereken muhtemel konum ve bölgedir. Seçilen dikkat bölgesi, artık bağlantıları ve yoğun katmanları içeren $\mathcal{F}_{FSR}$ modeli ile iyileştirilmektedir.

$\mathcal{F}_{DRL}$ ve $\mathcal{F}_{FSR}$ modellerinde bazı zorluklar bulunmaktadır. Detayları kaybolmuş bölgeleri seçme ve yüzdeki diğer hedef bölgeleri belirleme temel zorluklardan

birisidir. Ayrıca, $\mathcal{F}_{DRL}$ için ödül mekanizmasının nasıl oluşturulacağı model için büyük önem teşkil etmektedir. Bir yüz görüntüsünde ayrıntıları iyileştirilecek birçok bölge bulunabilmektedir. Amaca yönelik iteratif bir dikkat modeli ile T sayıda bölge seçilmektedir. Her adımda yüzün yüksek frekanslı bölgelerinin ve detayların ortaya çıkarılması hedeflenmektedir. İteratif şekilde kullanılan dikkat modellerinde geçmiş bilgilere ihtiyaç bulunmaktadır. Dolayısıyla, iteratif adımlarda iyileştirilmiş bölgelerin ve görüntülerin saklanması ayrıca geçmiş model deneyimlerinden yararlanılması, sürecin tekrarlı yapısında karşılaşılan başka bir zorluktur. İyileştirmede kullanılan SR yönteminde, RRDB modelinin uygulanışı modifiye edilmektedir [77]. RRDB’de birçok katman bulunmaktadır ve modelin girişi bir görüntüdür. Ancak, önerilen yöntemde SR modelinin girişi iteratif yapıdan elde edilen iyileştirilmiş görüntü ve her adımda yeniden elde edilen görüntü üzeri seçilen parça görüntüdür. Model girişinin değiştirilmesi nedeniyle RRDB katmanlı ağ yapısındaki artık öğrenme yapısı da güncellenmektedir. RRDB yapısında bulunan ayırıcı ağ önerilen çalışmada kullanılmamaktadır. Ödül mekanizması ile iyileştirme ağı optimizasyonu yapılmaktadır. $\mathcal{F}_{DRL}$ model çıktısı, $\mathcal{F}_{FSR}$ modelinin girişi olduğundan model parametrelerin güncellenmesi için kesintisiz bir geriye yayılımın sağlanması gerekmektedir. Ayrıca iteratif yapıda T adım sayısı, seçilecek bölgelerin boyutları, artık yoğun bloklarının katman sayısı, filtre boyutları, bağlantı yoğunluğu parametrelerinin belirlenebilmesi ayrı zorlukları beraberinde getirmektedir.

Çalışmanın bu kısmında önce DRL-SRFR model tanıtımı yapılmakta ve ayrıntılarıyla gösterilmektedir. Sonrasında ise model eğitim süreci ayrıntıları ile verilmektedir.

4.1. Yüz Bölgelerinin Derin Pekiştirmeli Öğrenme Tabanlı Süper Çözünürlüğü (DRL-SRFR) Ağı (Deep Reinforcement Learning Based Super Resolution of Face Regions (DRL-SRFR) Network)

Önerilen modelde daha önceki DL-tabanlı FSR yöntemlerinden farklı olarak, SR problemi kısmen gözlemlenebilir markov karar süreci olarak ele alınmaktadır. $\mathcal{F}_{DRL}$ modeli, tekrarlayan dikkat modeli üzerine kuruludur. Tekrarlayan dikkat modelinde (RAM), gradyan-tabanlı bir tekrarlayan politika ağı kullanılmaktadır. Her adım t ’de, karar verici ortamın mevcut durumu (şu anki gözlem) hakkında kısmi bir gözlem almaktadır. Mevcut duruma bağlı olarak, karar verici, iyileştirilmesi gereken parçaları seçmektedir.

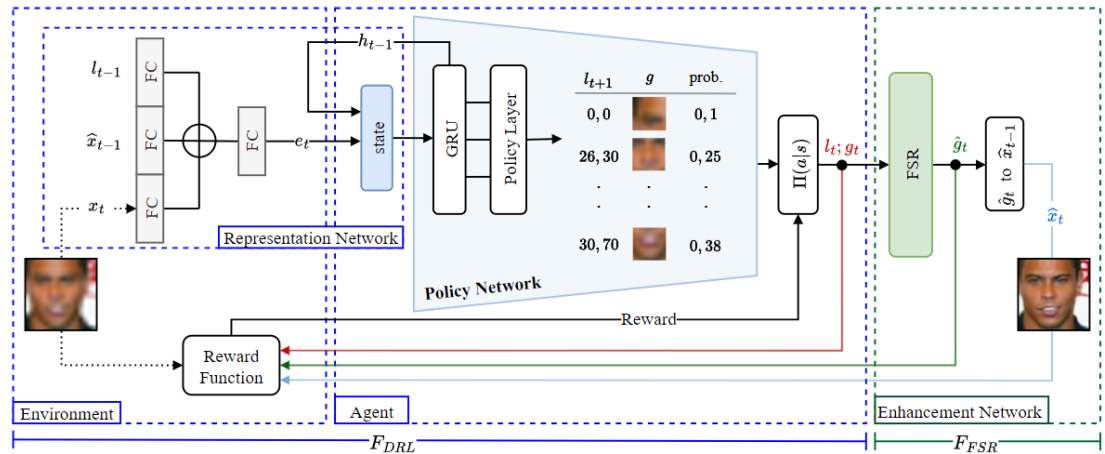
$\mathcal{F}_{FSR}$ modelinde düşük çözünürlüklü girdilerle tutarlı olan, ince ayrıntıları vurgulayan, yoğun bağlantılarla gradyan akışını iyileştiren ve hiyerarşik ve ağ yapılı kalıntı yapısını kullanan bir SR yöntemi oluşturulmuştur. DRL-SRFR’de, gözlemler her bir iteratif süreç sonunda elde edilir ve her durum bir sonraki politikanın girdisi olmaktadır.

Politika katmanı, yüz görüntüsündeki bir konumu ve bir yüz parçasını dikkatlice belirlemektedir. Yüz parçası $\mathcal{F}_{FSR}$ modeli ile iyileştirilir ve önceki adımlardan elde edilen iyileştirilmiş görüntüye gömülmektedir. Seçilen tüm parçalar işlendiğinde yüksek çözünürlüklü yüz görüntüsü elde edilmektedir. Sürecin her adımında ödül mekanizması çalıştırılır, geri yayılım algoritması ve biriken ödüller yardımıyla $\mathcal{F}_{DRL}$ ve $\mathcal{F}_{FSR}$ modellerine ait parametreler güncellenmektedir.

4.1.1. Model detayları

Çoğu FSR ve halüsinasyon yöntemi yüz görüntüsündeki pikseller arası korelasyonu kullanmaz ve yüz özelliklerini bağımsız olarak ele almaktadır. Önerilen model sıralı öğrenme yoluyla yüksek çözünürlüklü yüz görüntüleri üretir, görüntünün bozulmuş kısımlarına odaklanır ve iyileştirilen bölgeler aracılığıyla bölgeler arası ilişkileri kullanmaktadır.

Şekil 4.2’de gösterildiği gibi önerilen yöntem üç bölümden oluşur. İlk bölüm, mevcut durumu ve ödülleri politika ağına sunan bir ajan ağı ve ödül fonksiyonudur. İkinci bölüm, konum ve parça bilgisi üreten karar verici (politika ağı) olarak görev yapmaktadır. Son bölüm, seçilen parçanın iyileştirilmesinden sorumlu kısımdan oluşmaktadır.



Şekil 4.2. Önerilen yöntemin genel yapısı.

Yöntemin genel yapısında politika ağı dikkat edilecek parçaları belirlerken, ödül fonksiyonu $\mathcal{F}_{DRL}$ 'nin sıralı eğitimini yönlendirmektedir. $\mathcal{F}_{FSR}$ ise derin RRDB katmanları ile parçaları iyileştirerek yüz görüntüsüne eklemektedir. Ajan ağı, doğrusal katmanlar ile iteratif süreçte elde edilen gözlemlerin özelliklerini çıkarmaktadır. $\mathcal{F}_{DRL}$ politika ağı, GRU ağı ve bir konum üretici içermektedir. GRU, geçmiş bilgiden elde edilen bilgiler saklanmakta ve yeni konumları belirlemek için kullanılmaktadır. $\mathcal{F}_{FSR}$ 'de RRDB yöntemi, artık bağlantılar kullanarak uzun menzilli katman ilişkileri ile uygulanmaktadır. Detaylar ileriki bölümlerde verilmektedir.

Notasyonlar: $\mathcal{F}_{FSR}$ yönteminin amacı, iteratif süreçteki adımlarla bir görüntü $x \in \mathbb{R}^{h,w}$ için pozisyonları $l = \{l_1, l_2, \dots, l_T\} \in \mathbb{R}^2$ ve bu pozisyonlardaki parçaları $g = \{g_1, g_2, \dots, g_T\} \in \mathbb{R}^{u,v}$ sırasıyla belirlemektir. Burada h ve w , görüntünün boyutlarını, u ve v , görüntüden seçilen parça boyutlarını ifade etmektedir. Belirlenen parçalar her adımda iyileştirilmiş parçalar $\hat{g} = \hat{g}_1, \hat{g}_2, \dots, \hat{g}_T$ elde etmek için işlenir ve iyileştirilmiş bir görüntü $\hat{x} = \hat{x}_1, \hat{x}_2, \dots, \hat{x}_T$, her bir adımda aynı konumda parçayı önceki görüntüye gömerek oluşturulmaktadır. İterasyon T 'deki görüntü çıktı görüntüsüdür. Her bir adım için işlem aşağıdaki denklemlerle ifade edilmektedir. Tablo 4.1, iteratif süreçte üretilen ara yapılar ve çıktıları göstermektedir.

Tablo 4.1. DRL ve FSR modellerinin giriş ve çıkışları.

Adım (t)	DRL giriş→durum	DRL çıkış→	FSR girdi→	FSR çıktı→	Gömme $\hat{g}$ 'yi $\hat{x}$ 'e
1	$x, \hat{x}_0, l_0, h_0$	l_1, g_1, h_1	g_1	$\hat{g}_1$	$\hat{x}_1$
2	$x, \hat{x}_1, l_1, h_1$	l_2, g_2, h_2	g_2	$\hat{g}_2$	$\hat{x}_2$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$T - 1$	$x, \hat{x}_{T-2}, l_{T-2}, h_{T-2}$	$l_{T-1}, g_{T-1}, h_{T-1}$	g_{T-1}	$\hat{g}_{T-1}$	$\hat{x}_{T-1}$
T	$x, \hat{x}_{T-1}, l_{T-1}, h_{T-1}$	l_T, g_T, h_T	g_T	$\hat{g}_T$	$\hat{x}_T$

İteratif politika katmanı $f_{\pi}(s_{t-1}; \theta_r)$, $t - 1$ adımında durum s_{t-1} için denklem 4.2’de formül de görülen bir konum bilgisi $l_t = \{m, n\}$ üretir ve m, n konum bilgisi ile denklem 4.3’teki gibi g_t yüz görüntüsünden bir parça seçilmektedir.

$$l_t = \{m, n\} = f_{\pi}(s_{t-1}; \theta_r) \quad (4.2)$$

Parça g_t , $\mathcal{F}_{FSR}$ ağı ile iyileştirilerek iyileştirilmiş $\hat{g}$ parçası elde edilmektedir.

$$\hat{g}_t = \mathcal{F}_{FSR}(g, x, \hat{x}_{t-1}; \theta_h) \quad (4.3)$$

Her adımda $\hat{g}_t$, önceki adımda iyileştirilen yüz görüntüsüne $\hat{x}_{t-1}$ gömülür. Denklem 4.4’te, süreç benzer şekilde ilerler ve T adımının sonunda nihai bir görüntü $\hat{x}$ elde edilmektedir.

$$\hat{x}_t = \hat{x}_{t-1} \oplus \hat{g}_t \text{ where } t = 1, 2, \dots, T \quad (4.4)$$

Burada $\oplus$, $\hat{g}_t$ ’nin önceki iyileştirilmiş görüntüye gömülmesini ifade etmektedir. İşlem sırasında alt bölge $\hat{g}$, $\hat{x}_{t-1}$ ’den belirlenmektedir.

Bir sonraki başlıkta modelin pekiştirmeli öğrenme kısmı ayrıntılı şekilde ele alınacaktır.

4.1.2. Pekiştirmeli yöntem için durum, aksiyon ve ödülün tanımı

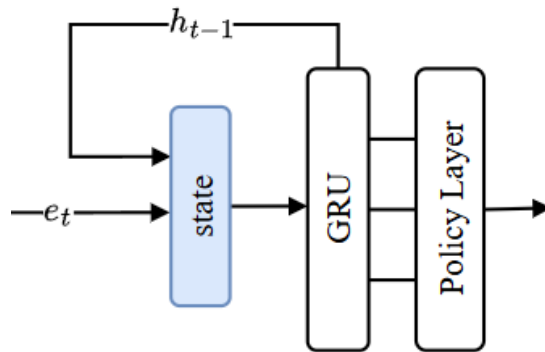
Tekrarlayan dikkat modeli (RAM), özellikle belirli hedeflere ulaşmayı amaçlayan faaliyetler bağlamında görsel bir ortamda dikkati anlamak ve kullanmak için tasarlanmış bir hesaplama çerçevesidir. Teorik olarak, bilişsel bilim ve yapay zeka alanlarındaki sıralı karar verme süreçleri paradigması üzerine kuruludur. RAM, dağınıklıkla başa çıkma ve büyük giriş görüntülerine ölçeklendirme konusunda CNN’den daha iyi performans göstermektedir. Dolayısıyla RAM, karmaşık bir görsel ortamda sınırlı hesaplama kaynaklarıyla gerçek dünya uygulamalarında daha verimli ve etkili olmaktadır. RAM’de karar verici çevresini gözlemlemek için sınırlı bant genişliğine sahip bir ağ kullanmaktadır. Diğer bir deyişle sadece belirli bir çözünürlükte veya belirli bir alanı kapsayabilen bir girdi (durum) bulunmaktadır. Karar verici bir zaman adımında tüm çevresini değil, sadece odaklandığı kısımları ele alarak aksiyonları belirlemektedir. Dolayısıyla karar verici zaman içinde bilgiyi entegre etmek zorundadır. Alınan aksiyonlara karşılık çevreden geri bildirim ile ödüller oluşturulmaktadır. Çevreyle etkileşiminin sonuçlarını yansıtan ödül, anlık

veya belirli iterasyon sonucu toplam şekilde elde edilebilmektedir. Süreç içerisindeki ana amaç iteratif süreç sonundaki toplam ödülü en iyi değere ulaştırmaktır. Aşağıda $\mathcal{F}_{DRL}$ 'deki RAM mekanizması için durum, aksiyon ve ödülleri tanımlanmaktadır. θ öğrenme parametreleri hem doğrusallaştırma katmanlarında hem de modellerde farklı indis ile belirtilmeliyken karmaşıklığı önlemek adına her fonksiyon için indis kullanılmadığına dikkat ediniz.

Durum: Her bir parça, önceki adımların yanlış seçim konumlarını belirleyebilmelidir. Dolayısıyla bir durum, önceki konum bilgisini l_{t-1} ve önceki yerleştirme bilgisi h_{t-1} 'yi yansıtmalıdır. Bahsi geçen amaç doğrultusunda dikkat mekanizması eğitim sürecinde doğru tercihleri ve konumları öğrenmektedir. Ayrıca düşük çözünürlüklü bir görüntü x 'de her adımda yeni bir konum seçmek ve görüntünün bütünsel özelliklerini incelemek için durum girdisine dahil edilmektedir. Geliştirilmiş görüntü $\hat{x}_{t-1}$ 'nin durumda ardışık olarak eklenmesi, geçmişin etkisini korur ve hem yerel hem de global iyileşmelere katkıda bulunmaktadır. Verilen amaçlara yönelik olarak; x , $\hat{x}_{t-1}$ ve l_{t-1} giriş verileri temsilci ağı $f_s(x, \hat{x}_{t-1}, l_{t-1}; \theta)$ ile e_t vektörüne dönüştürülmektedir. Ardından s_t durumu h_{t-1} 'in e_t 'ye eklenmesiyle denklem 4.5'teki gibi elde edilmektedir.

$$s_t = (e_t, h_{t-1}) \quad (4.5)$$

İteratif sürecin başlangıcında iyileştirilmiş görüntü $\hat{x}$, orijinal düşük çözünürlüklü görüntü x 'e atanır ve başlangıç konumu görüntünün merkezi olarak belirlenmektedir. f_s , girdi parametrelerini doğrusal hale getiren bir temsil ağıdır. Politika ağı, dikkat bölgesinin seçilme olasılığını hesaplamak için durumu kullanmaktadır. Şekil 4.3'te durum ve politika ağı arası akış gösterilmektedir.



Şekil 4.3. Durum modülü giriş ve çıkış akış diyagramı.

Durumun bileşenleri aşağıdaki şekilde hesaplanmaktadır:

h_t , t 'inci iterasyondaki gizli durum bilgisi ve h_{t-1} bir önceki iterasyondaki gizli durum bilgisi ile denklem 4.6'daki gibi hesaplanmaktadır.

$$h_t = GRU(h_{t-1}, e_t) \quad (4.6)$$

Yeni gözlemlerden çıkarılan özellikler (e_t), mevcut iterasyondaki giriş (x), bir önceki iyileştirilmiş görüntü ($\hat{x}_{t-1}$) ve bir önceki aksiyon (l_{t-1}) üzerinden denklem 4.7'deki gibi hesaplanmaktadır.

$$e_t = f_s(x, \hat{x}_{t-1}, l_{t-1}) \quad (4.7)$$

Durumun iteratif güncellenmesi denklem 4.5 üzerinden hesaplanır ve her iterasyon sonunda h_t ile güncellenmektedir.

İlk iterasyon ($t = 0$):

- Başlangıç durumunda:

$$h_0 = 0$$

- Yeni gözlemler:

$$e_0 = f_s(x, x, \text{Merkez Parça})$$

- Başlangıç durumu:

$$s_t = (e_0, h_0)$$

Sonraki iterasyonlar ($t > 0$):

- Yeni gözlemler:

$$e_t = f_s(x, \hat{x}_{t-1}, l_{t-1})$$

- Güncel durum:

$$s_t = (e_t, h_{t-1})$$

Gizli Durumun Güncellenmesi:

Gizli durum (h_t), s_t 'ye bağlı olarak denklem 4.8'de verilen GRU yapısı ile güncellenmektedir.

$$h_t = GRU(h_{t-1}, e_t) \quad (4.8)$$

Eylem: Politika ağının ana görevi iyileştirilecek yeni parçanın konumunu belirlemektir. Durum s_t için parçalarla ilgili tüm konum olasılıkları (u, v) hesaplanmaktadır. Olasılıklardan l_t için yeni parça konumu ana görüntü $\hat{x}$ 'den seçilmektedir. Süreç boyunca bir sonraki konum belirlenirken politika ağı önceki durumla birlikte mevcut durumu değerlendirmekte ve iyileştirme ağına odaklanacak konumu belirlemektedir.

Bir eylem GRU ve doğrusal bir ağ tarafından gerçekleştirilmektedir. RAM yöntemleri gibi bir GRU bellektir ve iki durum çıkarır: geçmiş durum ve bir çıktı. İlk durum başlangıç durumu ikincisi ise geçmiş bilgilerdir. GRU önceki deneyimler ve kullanıcı izlenimlerine dayalı öğrenme sağlamaktadır. Böylece her durumda yeni konum doğru bir şekilde güncellenebilmektedir. Tüm olası konumlar $l_t = (m, n | m \leq h, n \leq w)$ olarak tanımlanmaktadır. Konum, bir rastgele değişken olarak ele alınır ve denklem 4.9 ve 4.10'da temel işlemi verilen olasılık kütle fonksiyonu ile tanımlanmaktadır.

$$\mathcal{P}(a|s, \theta) = \mathcal{P}(l_t|s_t, \theta) = \mathcal{P}((m, n)|s_t, \theta) \quad (4.9)$$

$$\mathcal{P}((m, n)|s_t, \theta) = \frac{\exp(f_\pi(s_t, (m, n)))}{\sum_{(m', n') \in l_t} \exp(f_\pi(s_t, (m', n')))} \quad (4.10)$$

Burada:

$l_t = (m, n)$: Seçilen parçanın konum bilgisi.

s_t : Mevcut durum bilgisi.

θ : Politika ağının parametreleri.

$f_\pi(s_t, (m, n))$: Durum bilgisi ve aksiyon üzerinde politika ağı tarafından üretilen skor.

$l_t = (m, n | m \leq h, n \leq w)$: Tüm olası konumların kümesi.

$\sum_{(m', n') \in l_t} \exp(f_\pi(s_t, (m', n')))$: Tüm olası konumlar üzerindeki toplam değer olmaktadır.

Denklem 4.10, her bir konum için bir skor hesaplar ve bu skorları normalize ederek toplamı 1 olan bir olasılık dağılımı oluşturmaktadır.

Rastgele konum seçimi:

Konum (l_t) olasılık dağılımından denklem 4.11’de verildiği gibi rastgele örneklenmektedir.

$$l_t \sim \mathcal{P}(l_t|s_t, \theta) \quad (4.11)$$

Burada yüksek olasılıklı konumlar daha sık seçilmektedir.

Denklem 4.12 politika ağı ile konum seçimini göstermektedir.

$$f_\pi(s_t, (m, n)) = Wh_t + b \quad (4.12)$$

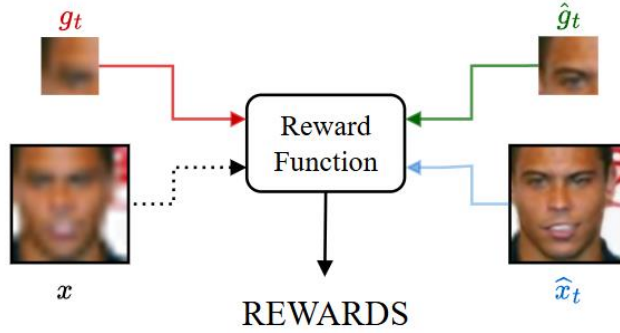
W, b : Politika ağının öğrenilebilir parametrelerini temsil etmektedir.

Skorlar olasılık kütle fonksiyonu ile normalize edilerek olasılığa dönüştürülmektedir. İlgili formül denklem 4.13’te verilmektedir.

$$\mathcal{P}(l_t|s_t, \theta) = \frac{\exp(Wh_t + b)}{\sum_{(m', n') \in l_t} \exp(Wh_t + b)} \quad (4.13)$$

Ödül: DRL yöntemlerinde karar vericiyi doğru eylemlere yönlendirebilmek amacıyla ödül fonksiyonları kullanılmaktadır. Karar verici, yüksek ödülle sonuçlanan eylemleri tekrar etmeye meyilli hale gelirken, düşük ödüllü eylemlerden mümkün mertebe kaçınma eğiliminde olmaktadır. Önerilen çalışmada görüntünün parçalarını ve görüntünün tamamını iyileştirecek şekilde yerel-global ödül fonksiyonları kullanılmaktadır. Ödül fonksiyon yapısında dört adet ödül fonksiyonu ve bir görüntünün “İyileştirilmiş piksel sayısı/Tüm piksel sayısı” oranını kullanan yardımcı bir ödül parametresi E_{pc} kullanılmaktadır. Son çalışmalar yalnız piksel bazında ödüller kullanmanın yapısal farklılıkların göz ardı edilmesine ve algısal kalitenin düşük olmasına neden olduğunu göstermektedir [76]. ESRGAN yönteminde sunulan algısal kayıp fonksiyonu, görüntünün yapısal benzerliğini ön planda tutarak algısal kalitenin artırılmasını sağlamaktadır. Önerilen yöntemde de algısal kayıp fonksiyonu ödül amaçlı olarak kullanılmaktadır. Parça bölgeleri üzerinden iki adet yerel ödül değeri, görüntünün tamamından iki adet global ödül değeri ile toplam ödül belirlenmektedir. Yerel ödüller her adımdaki parça $\hat{g}_t$ ile hedef parça g_t^y üzerinden oluşturulmaktadır. Modelde düşük çözünürlüklü görüntüden seçilen g_t parçasının hedef görüntüdeki karşılığı, hedef parça g_t^y belirlenmektedir. İki parça üzerinden, $MSE(\hat{g}_t, g_t^y)$ değeri ile $VGG(\hat{g}_t), VGG(g_t^y)$ ’den elde edilen özellikler arasındaki

L2-Norm değeri ile iki ödül hesaplanmaktadır. (VGG19, 19 katmanlı bir CNN olup 2014 ImageNet yarışmasında iyi performans göstermektedir. VGG19'un özellik katmanlarından çıkarılan özellikler, içerik kaybını hesaplamak için kullanılmaktadır.)



Şekil 4.4. Ödül fonksiyon girdileri.

Şekil 4.4, ödül fonksiyonunun girişlerini göstermektedir. Ödül fonksiyonu, modelin hem yerel hem de global düzeydeki iyileştirme performansını değerlendirmek amacıyla farklı kaynaklardan gelen bilgileri birleştirmektedir. Girdiler ödül fonksiyonuna verilerek modelin iyileştirme sürecinde aldığı kararların doğruluğu ölçülmektedir. Nihai ödül, bu girdilere dayalı olarak hesaplanmakta ve modelin öğrenme sürecine geri besleme olarak iletilmektedir.

Yerel Ödüller (R_{yerel}):

Yerel ödüller modelin her iterasyonda model sonucu elde ettiği parçaların ($\hat{g}_t$) hedef parçalar (g_t^y) ile benzerliğini ölçmektedir. Bu ödüller iki ana bileşene ayrılmaktadır:

1. Piksel Bazlı Yerel Ödül ($R_{yerel,piksel} = rl$): Seçilen parçaların hedef parçalarla piksel seviyesindeki farkını değerlendirmekte ve genellikle MSE (Mean Squared Error) tabanlı olmaktadır.
2. Algısal Yerel Ödül ($R_{yerel,algısal} = rl_{per}$): Seçilen parçaların hedef parçalarla yüksek seviyeli özellik benzerliğini ölçmektedir. Özellik uzayı, önceden eğitilmiş bir ağ (VGG19) üzerinden çıkarılmakta ve bu özellikler arasındaki fark ödül olarak hesaplanmaktadır.

Global ödüller ise iteratif süreç sonunda oluşturulmaktadır. Oluşan son görüntü $\hat{x}$ ve hedef görüntü y olmak üzere, global ödüller $MSE(\hat{x}, y)$ değeri ile $VGG(\hat{x}), VGG(y)$ den elde edilen özellikler arasındaki L2-Norm değerinden hesaplanmaktadır.

Global Ödüller (R_{global}):

Global ödülleri iyileştirilmiş tüm görüntünün ($\hat{x}$) hedef görüntü (y) ile benzerliğini ölçmektedir. Global ödüller de iki bileşene ayrılmaktadır:

1. Piksel Bazlı Global Ödül ($R_{global,piksel} = rx$): Nihai görüntünün hedef görüntüyle olan piksel seviyesindeki farkı değerlendirmektedir.
2. Algısal Global Ödül ($R_{global,algısal} = rx_{per}$): Nihai görüntünün hedef görüntüyle olan yüksek seviyeli özellik benzerliğini ölçmektedir. Özellik uzayı yine VGG19 gibi bir model üzerinden elde edilmektedir.

Ödüllerin Birleştirilmesi:

Toplam ödül fonksiyonu yukarıda belirtilen yerel ve global ödüllerin birleştirilmesiyle elde edilmektedir. Fonksiyon modelin hem yerel detayları hem de genel bütünlüğü iyileştirme başarısını ölçmektedir. Ödül fonksiyonu denklem 4.14'teki gibi ifade edilmektedir.

$$R_t = \alpha \cdot R_{yerel} + \beta \cdot R_{global} \quad (4.14)$$

Burada:

α : Yerel ödüllerin ağırlık katsayısı.

β : Global ödüllerin ağırlık katsayısı.

Yerel ve global ödüller kendi içlerinde denklem 4.15 ve 4.16'daki gibi birleştirilmektedir:

$$R_{yerel} = \lambda_1 R_{yerel,piksel} + \lambda_2 R_{yerel,algısal} \quad (4.15)$$

$$R_{global} = \lambda_3 R_{global,piksel} + \lambda_4 R_{global,algısal} \quad (4.16)$$

Burada $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ katsayıları, ödül bileşenlerinin göreceli önemini belirlemektedir.

Piksel bazlı ödüller genellikle MSE (Mean Squared Error) üzerinden tanımlanmaktadır ve ölçekleri sabittir. Dolayısıyla bu tür ödüller genellikle doğrudan toplam ödüle dahil edilmektedir. MSE tabanlı ödüllerin etkisi belirli bir ölçekte olduğu için ek bir ağırlıklandırmaya (λ) gerek duymamaktadır. Bazı ödüllerde λ 'nın olmaması o bileşenin temel ödül metriklerinden biri olarak görülmesinden kaynaklanmaktadır. Piksel bazlı ödüller doğrudan hesaplandığı için temel metrik olarak kabul edilmekte ve genel katsayılar ile denge kurulmaktadır.

E_{pc} parametresi ise global iyileştirmeye fayda sağlamaktadır. Ne kadar çok piksel değerinin iyileştirilmesine katkı sağlanmışlığının bir göstergesidir. Bir döngü boyunca seçilen tüm piksellere 1 değeri atanmakta diğerleri ise 0 değeri olarak belirlenmektedir. E_{pc} için oluşan (0;1) haritası 5. kısımda ablasyon çalışmalarında verilmektedir. Söz konusu yerel ve global ödüllere ilişkin genel ifadeler sırasıyla denklem 4.17, 4.18, 4.19 ve 4.20’de aşağıda verilmektedir:

$$rl = \frac{1}{T \times c \times u \times v} \sum_{t=1}^T (g_t^y - \hat{g}_t)^2 \quad (4.17)$$

$$rl_{per} = \lambda \frac{1}{T \times c \times u \times v} \sum_{t=1}^T \|\phi^{vgg} \hat{g}_t - \phi^{vgg} g_t^y\|_2^2 \quad (4.18)$$

$$rx = \frac{1}{c \times h \times w} (\hat{x} - y)^2 \times E_{pc} \quad (4.19)$$

$$rx_{per} = \lambda \frac{1}{c \times h \times w} \|\phi^{vgg} \hat{x} - \phi^{vgg} y\|_2^2 \times E_{pc} \quad (4.20)$$

ϕ^{vgg} fonksiyonu için VGG19 modeli kullanılarak özellikler çıkarılmaktadır. Ön eğitilmiş VGG19 modelinin son kısmındaki adaptif ortalama havuzlama ve sınıflandırma kaldırılarak özellik katmanları kullanılmaktadır. Ayrıca VGG19 modelinin bütün parametreleri dondurularak eğitilmiş model parametreleri üzerinden özellikler belirlenmektedir.

Toplam ödül R , denklem 4.21’deki gibi oluşturulmakta ve karar vericiyi eğitmek için kullanılmaktadır.

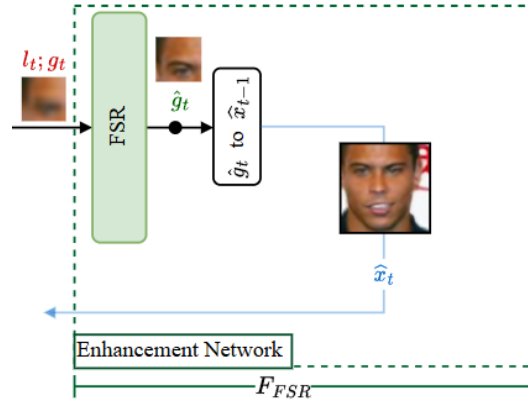
$$R = -(rl + rl_{per} + rx + rx_{per}) \quad (4.21)$$

Bir sonraki başlıkta FSR yapısı ve modifiye edilen RRDB yapısı incelenmektedir.

4.1.3. FSR yapısı ve RRDB

SR yapılarında görüntü detaylarını açığa çıkartmak ve global yapıyı korumak için derin ağlar önerilmektedir [96]. $\mathcal{F}_{DRL}$ ’de tekrarlı dikkat mekanizmasından elde edilen parça artık bağlantılar, hiyerarşik özellikler ve derin yoğun katmanlar ile iyileştirilmektedir. Klasik CNN tabanlı yöntemler görüntü özelliklerinin çıkarılması işleminde bazı sınırlamalara sahiptir. Şekil 4.6’da da gösterilen RDB yapısı daha ayrıntılı görüntü özellikleri elde etmek ve detayların daha iyi korunmasını sağlamak

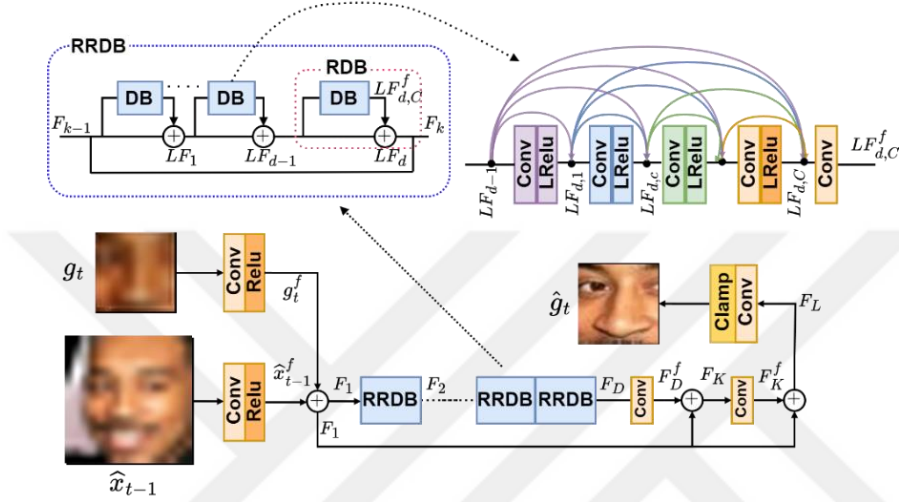
için geliştirilmiştir [74]. RDB yapıları artık bağlantılar ve yoğun bağlantılardan oluşmaktadır. ESRGAN yönteminde RDB yapılarının birleştirilmesiyle RDB'ye göre daha karmaşık ve verimli olan RRDB yapıları geliştirilmiştir. RRDB yönteminde, RDB'nin temel özellikleri korunurken tekrarlayan iç içe yapı benimsenmektedir. Bir RRDB bloğu içerisinde birden fazla RDB bloğunu içermektedir. Bahsedilen iç içe yapı özellikle detayların ve dokuların korunması açısından son derece etkili olmaktadır [58, 77]. Aynı zamanda RRDB ile gradyanların derin ağ katmanlarında daha etkili bir şekilde hesaplanması sağlanmaktadır. Böylece gradyan kaybolması problemine de bir çözüm getirilmektedir. Önerilen RRDB yapımızda ise ESRGAN yöntemindeki gibi bir giriş görüntü kullanılması ile birlikte çıkış katmanındaki veride bu katmana giriş olarak verilmektedir. İteratif süreç boyunca oluşan bir önceki iyileştirilmiş görüntü $\hat{x}$ ve tekrarlı dikkat mekanizmandan elde edilen parça g ile RRDB yapısı Şekil 4.6'daki gibi oluşturulmaktadır. Önerilen yapının önemli bir üstünlüğü gradyan akışını kesintisiz bir şekilde sağlamasıdır. Bununla birlikte farklı amaçtaki $\mathcal{F}_{DRL}$ ve $\mathcal{F}_{FSR}$ model yapıları arasında bir köprü gerekmektedir. $(t - 1)$. adımdan elde edilen iyileştirilmiş yüz görüntüsü ile yeni bölgenin entegrasyonu modeller arasında köprü vazifesi görmektedir. Böylece modellere ait ağ parametreleri geri yayılım algoritması ile güncellenebilmektedir.



Şekil 4.5. FSR yapı akışı.

Şekil 4.5, önerilen FSR yönteminin yapı akışını görsel olarak ifade etmektedir. FSR yapısının ilk adımında düşük çözünürlüklü giriş görüntüsü (x) üzerinden belirli bir yüz parçası (g_t) seçilir. Seçilen yüz parçası Şekil 4.6'da verilen RRDB tabanlı süper çözünürlük modeli ile işlenmektedir. Süreçte parça, daha yüksek çözünürlüğe ulaştırılarak detayların iyileştirilmesi sağlanmaktadır.

Bir RRDB’de çok sayıda RDB katmanı yer almaktadır. Ayrıca RRDB ve RDB yapılarında artık bağlantılar bulunmaktadır. Böylece ağın başlangıç katmanlarında yüz görüntüsünün kenar, doku ve renk gibi temel öğelerinin özellikleri belirirken ağın derinleşmesiyle birlikte daha karmaşık ve soyut özellikler ortaya çıkmaktadır. Sonuç olarak düşük seviyeli ve yüksek seviyeli özellikler derin katmanlarda işlenerek yüz özellikleri korunmaktadır.



Şekil 4.6. $\mathcal{F}_{FSR}$ modelindeki iyileştirme ağı.

Şekil 4.6’da bir t anı için modelin FSR yapısı gösterilmektedir. $\hat{x}_{t-1}$ bir önceki adımda iyileştirilmiş yüz görüntüsü, g_t ise $\mathcal{F}_{DRL}$ ’de belirlenen parça bölgesidir. $\hat{x}_{t-1}$ ve g_t ’den özellikler çıkarılarak RRDB yapılarından geçirilmektedir. RRDB yapısındaki RDB’de artık bağlantılar kullanılmaktadır. RDB’de yine artık bağlantılar ile özellikler belirlenmektedir. $\hat{x}_{t-1} \in R^{c,h,w}$ ile $g_t \in R^{c,u,v}$ parçasına konvolüsyon ve ReLU aktivasyonu uygulanarak sırasıyla $\hat{x}_{t-1}^f \in R^{c,u,v}$ ve $g_t^f \in R^{c,u,v}$ sığ özellik vektörleri elde edilmektedir. Her iki vektör birleştirilerek $F_1 \in R^{c,u,v}$ ile belirtilen derin RRDB’ye ait giriş vektörü oluşturulmaktadır. Giriş vektöründen hiyerarşik özellikler derin RRDB katmanlarında kullanılmaktadır. Sıralı RRDB yapısının sonuna iki noktada artık bağlantı eklenmektedir. Son konvolüsyon katmanları ile $\hat{g}_t$ iyileştirilmiş parçası elde edilmektedir. Bir RRDB’de D tane RDB bulunmaktadır. Her RDB yapısında residual bağlantılar kullanılan özellik çıkarımları mevcuttur.

Tablo 4.2’de Şekil 4.6’daki işlemler ve formüller özetlenmektedir. RDB bloğu haricinde diğer katmanların konvolüsyon ağırlıkları W ile temsil edilmiş ve basitlik olması açısından bias terimleri verilmemektedir. Hem RDB hem derin RRDB

yapısında sürekli bellek yapısı ve yerel özellik füzyonu bulunmaktadır. σ , ReLU aktivasyon fonksiyonunu ve RDB’de $W_{d,c}$ ve $W_{d,LF}$ ise sırasıyla c’inci konvolüsyon katmanı ile son katmanın ağırlıklarını temsil etmektedir. Clamp fonksiyonu son konvolüsyon katman çıkışını $\{0,1\}$ arasına ölçekleyerek $\hat{g}_t$ üretmektedir.

Tablo 4.2. RRDB süreç formülleri.

	Süreç	Formül
RDB	Sürekli bellek mekanizması	$LF_{d,c} = \sigma(W_{d,c}[LF_{d-1}, LF_{d,1}, \dots, LF_{d,c-1}])$
	Yerel spesiyalite lerin füzyonu	$LF_{d,c}^f = W_{d,f}([LF_{d-1}, LF_{d,1}, \dots, F_{d,c}, \dots, F_{d,c}])$
	Yerel artık öğrenme	$LF_d = LF_{d-1} + LF_{d,c}^f$
Derin RRDB	RRDB Blok gösterimi	$F_k = F_{k-1} + LF_d$
	Sürekli bellek mekanizması	$F_D = RRDB(D \text{ times } \dots RRDB(RRDB(F_1)))$
	Yerel spesiyalite lerin füzyonu	$F_D^f = W_{D,f}(F_D)$
Derin RRDB’ nin artık bağlantısı	Sıralı artık öğrenme	$F_K = F_1 + F_D^f$
		$F_K^f = W_{K,f}(F_K)$
		$F_L = F_1 + F_K^f$
Derin RRDB’ ni Giriş	g_t ve $\hat{x}_{t-1}$ ’ in sığ özelliklerinin kombinasyonu	$F_1 = g_t^f + \hat{x}_{t-1}^f$
Girişlerin Sığ Özellikleri	g_t ve $\hat{x}_{t-1}$ ’ in özellik çıkarımı	$g_t^f = \sigma(W \times g_t)$
		$\hat{x}_{t-1}^f = \sigma(W \times \hat{x}_{t-1})$

RDB yoğun bir blok (DB) yapısına sahiptir. Özellikler DB içinde sürekli bir bellek yapısı oluşturularak birleştirilmektedir. RDB yapısı, DB’nin girişinin artık bağlantısı ve DB’nin çıkışının artık bağlantısı toplanarak oluşturulmaktadır. Sıralı RDB’lerde RRDB yapısı ilk RDB girişinin artık bağlantısı ile oluşturulmaktadır. RRDB’ler için ayrıca sürekli bir bellek mekanizması kurulur ve özellik füzyonları gerçekleştirilmektedir. $\hat{x}_{t-1}$ ve g_t , hem yüzey hem de içsel ve global özellikleri kullanarak geliştirilmiş bir $\hat{g}_t$ parçası elde etmek için doğrusal katmanlara ve aktivasyon fonksiyonuna sahip derin RRDB ağına girdi olarak verilmektedir.

FSR’de son adımda iyileştirilen parça (g_t) modelin bir önceki adımda elde edilen iyileştirilmiş görüntüsüne ($\hat{x}_{t-1}$) gömülmektedir. Yapılan işlem, tüm yüz görüntüsünün kademeli olarak yüksek çözünürlüklü bir forma dönüştürülmesini sağlamaktadır. Her iterasyonda yeni bir yüz parçası seçilir, iyileştirilir ve görüntüye eklenmektedir. Böylece hem yerel detaylar hem de yüzün genel yapısı kademeli olarak iyileştirilmektedir.

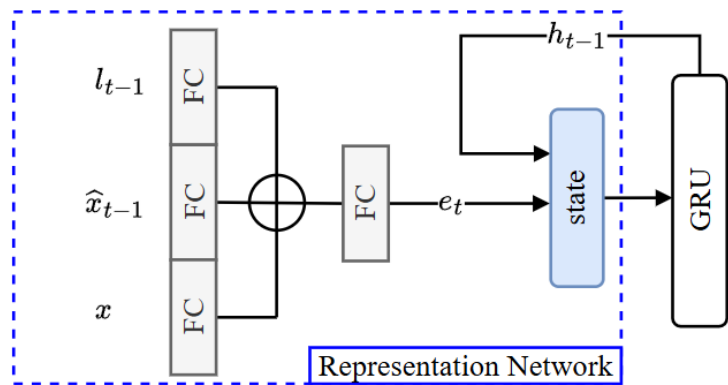
Şekil 4.7’de gösterilmekte olan yapıda hem yerel bölgelerdeki detayların iyileştirilmesini hem de global bütünlüğün korunmasını hedeflemektedir. RRDB’nin sunduğu derin yapılar yüksek kaliteli ve gerçekçi yüz görüntüleri üretmek için kritik bir rol oynamaktadır.

Bir sonraki başlıkta DRL yapısı daha detaylı anlatılmaktadır.

4.1.4. DRL yapısı

Önerilen modelde DL yapısındaki durumların temsili ve politika ağı için RAM’in yapısı aşağıda detaylandırılmaktadır.

Şekil 4.7, önerilen DRL yapısında durum ($state=s_t$) oluşturma ve modelin öğrenme sürecindeki temel bileşenleri içeren akış şemasını göstermektedir. Yapının bu kısmında modelin iteratif bir süreçle iyileştirme kararlarını vermesini sağlayan temel birimler olan temsil ağı (Representation Network), durum vektörü (State) ve GRU (Gated Recurrent Unit) yapıları ve ilişkileri gösterilmektedir.



Şekil 4.7. DRL yapısı ve durum oluşturma akışı.

Temsil ağı, giriş bileşenlerini işleyerek durumu oluşturan bir gözlem vektörü (e_t) çıkarmaktadır. Giriş bileşenlerinin her biri bir dizi tam bağlantılı katman (FC) aracılığıyla işlenmektedir.

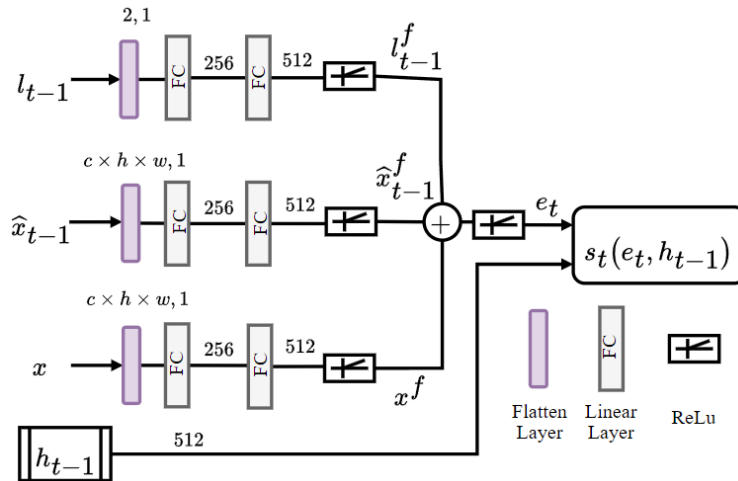
Gözlem vektörü (e_t), bir önceki iterasyonun gizli durumu (h_{t-1}) ile birlikte modelin mevcut durumu (s_t)'yi oluşturmaktadır. Birleşim modelin geçmiş bilgileri ve yeni gözlemleri birlikte değerlendirmesini sağlamaktadır.

GRU, geçmiş durumu (h_{t-1}) ve yeni gözlemi (e_t) işleyerek güncellenmiş gizli durumu (h_t) üretmektedir. Mekanizma modelin geçmiş aksiyonların etkilerini unutmadan yeni duruma uygun bir şekilde öğrenme sürecine devam edilmesini sağlamaktadır. Şekil 4.7, DRL yapısında durum oluşturma sürecini detaylandırmakla birlikte sürecin modelin öğrenme mekanizmasındaki temel rolünü vurgulamaktadır. Temsil ağı, giriş bileşenlerinden anlamlı bir gözlem (e_t) çıkarırken, GRU birimi modelin karar verme süreçlerini sürekli iyileştirmek için bu gözlemi işlemekte ve hafızalı yapı kurmaktadır.

- **Durum Temsili**

Temsil ağı Şekil 4.8'de gösterilmektedir. Bir özellik çıkarıcı görevi görür ve birkaç bağlı doğrusal katmandan oluşmaktadır. Temsil ağı geçmiş bilgilerden h_{t-1} ve yeni gözlem özelliklerinden (e_t) oluşturmaktadır. Geçmiş bilgiler GRU katmanından çıkarılırken yeni gözlem denklem 4.22'deki gibi sıralı doğrusal katmanlar ve aktivasyon fonksiyonları tarafından oluşturulmaktadır.

$$e_t = f_s(x, \hat{x}_{t-1}, l_{t-1}) = x^f + \hat{x}_{t-1}^f + l_{t-1}^f \quad (4.22)$$



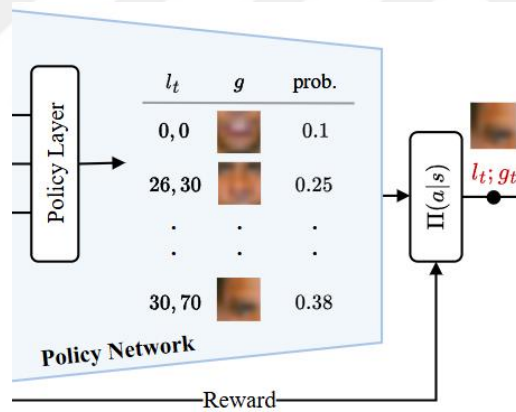
Şekil 4.8. Politika için durum katmanı oluşumu.

Modelde bir durum, e ve h_{t-1} 'den oluşmaktadır. Konum ve $(t - 1)$ adımından oluşturulan görüntü ile orijinal görüntü, doğrusallaştırma katmanlarından özellikler

çıkarılarak birleştirilmektedir. Denklem 4.22’de üst indeksi f olan terimler matematiksel olarak, x , $\hat{x}_{t-1}$ ve l_{t-1} girişlerine iki ardışık doğrusal katman uygulanarak oluşturulmaktadır. Konum l_{t-1} , 2×1 boyutlu vektördür. x ve $\hat{x}$ tek sütunlu bir vektör haline getirilmektedir. Ardından ilk doğrusallaştırma katmanında 256 ve sonraki katmanda 512 boyutunda vektörler elde edilmekte ve e ’yi elde etmek için toplanmaktadır. Durumun diğer bileşeni olan h_{t-1} , GRU yapısındaki geçmiş bilgisidir ve 512 boyutunda bir vektördür.

- **Politika Ağı**

Yinelemeli bağlantılar, yinelemeli süreci oluşturmak ve geçmişle ilgili bilgileri sonraki anlara aktarmak için kullanılmaktadır. Politika ağı GRU katmanı ve politika katmanlarından oluşmaktadır. GRU katmanı durum (s_t) özelliklerini alır ve bunları bir iç duruma (bellek) dönüştürmektedir. Böylece dahili durum zaman bilgisi ve önceki adımların birikimi saklanmaktadır. Politika katmanı, iç durumu belirli bir politika olasılığına dönüştürmektedir. Politika olasılığı, modelin bir sonraki adımda hangi eylemi gerçekleştirmesi gerektiğini belirleyen normal dağılımıdır.



Şekil 4.9. Politika ağı tarafından üretilen skor değerler ve çıkış.

Şekil 4.9, politika ağının karar alma sürecini ve ürettiği çıktıların görsel temsiliyle sistemin işleyişini detaylandırmaktadır. Politika ağı, girdi uzayındaki farklı adaylar için l_t , g ve “prob.” (olasılık) değerlerini hesaplayarak her bir adayın eylem seçimindeki uygunluğunu olasılıksal bir şekilde değerlendirmektedir. Süreçte, politika katmanı (Policy Layer) tarafından her bir aday için belirlenen olasılıklar, sistemin belirli bir eylemi seçmesi amacıyla kullanılmaktadır. Şeklin sağ tarafında yer alan politika fonksiyonu $\Pi(a|s)$, politika ağının seçtiği çıktıyı (bir görsel parça) ifade etmektedir. Seçim, olasılığı en yüksek olan aday üzerinden

gerçekleştirilmektedir. Ayrıca modelin öğrenme sürecini iyileştirmek amacıyla ödül mekanizması ile geri bildirim sağlanmaktadır. Politika yapısı, genellikle pekiştirmeli öğrenme bağlamında kullanılarak modelin hedeflenen görev doğrultusunda optimal stratejiler geliştirmesine olanak tanımaktadır.

İlerleyen kısımda model eğitimi anlatılırken Şekil 4.10'da politika ağının ayrıntıları gösterilmektedir. Bir GRU bellek hücresi $s_t = (e_t, h_{t-1})$ değerini bir eylem grubunun olasılığına dönüştürmektedir. Hafıza hücresi, önceki hücre hafıza birimi h_{t-1} ile mevcut adımdaki giriş vektörü e_t 'yi bütünleştirmektedir. Verilen bir durum (e_t, h_{t-1}) için GRU katmanı $h_t = \text{GRU}(e_t, h_{t-1})$, ağdaki birimlerin aktivasyonlarını denklem 4.23 ile yinelemeli olarak hesaplayarak bir çıkış dizisi h_t hesaplamaktadır.

$$h_t = \tanh(W_{si}e_t + b_{si} + W_{hi}h_{t-1} + b_{hi}) \quad (4.23)$$

Burada t , e_t , h_t , W_{si} , W_{hi} ve b sırasıyla t adımını, giriş vektörünü, gizli vektörü, e_t girişine karşılık gelen ağırlık matrisini, yeni gizli duruma karşılık gelen gizli vektörü ve bias terimlerini ifade etmektedir. e_t 'deki konum başlangıç değerleri $l_0 = [-1, 1]$ 'dir. Diğer adımlarda tekdüze bir dağılım fonksiyonu ile seçilirler. İlk adımda, $\hat{x} = x$ olarak alınmaktadır. Böylece ilk iyileştirilmiş görüntü aslında düşük çözünürlüklü görüntüdür. Diğer e_t durumlarındaki $\hat{x}_{t-1}$ parametresi, $(t - 1)$ zamanında alt bölgesi geliştirilmiş olan görüntüdür. Sıralı öğrenme sürecinde önceki adımlardan gelen geliştirilmiş görüntü bilgileri ve konum bilgileri geçmiş eylem bilgilerinde saklanmaktadır. Bundan dolayı yeni parça seçimi geçmiş bilgilerden etkilenir ve önceki eylemlerin neden olduğu hataları düzeltmek için ayarlanmaktadır. GRU için başlangıç değerleri $h_0 \in \mathbb{R}^{512} = 0$ olarak kullanılmaktadır.

Politika katmanı karar verme sürecini modellemek için kullanılmaktadır. Sürecin her adımında hangi eylemin gerçekleştirileceğini belirleyen bir politika fonksiyonu tanımlanmaktadır. Bir sinir ağından oluşan politika fonksiyonunun çıktısı, eylemlerin olasılık dağılımıdır. Politika ortalama değer (μ) ve standart sapma (σ) parametreleri ile tanımlanan bir Gauss dağılımından örnekleme yapılarak uygulanmaktadır. Gauss dağılımından yeni bir konum stokastik bir süreçle tahmin edilmektedir. Böylece amaç, gelecekteki aşamaları tahmin etmek için mevcut durumu kullanmak ile daha önce denenmemiş eylemlerde bulunmak arasında bir denge kurmaktır.

Ortalama değer ve standart sapma, RL algoritmalarında kritik istatistiksel parametrelerdir. Ortalama değer bir veri setinin merkezi eğilimini gösterirken, standart sapma bu verinin ortalamadan ne kadar saptığını ölçmekte ve dağılımın yayılımını belirlemektedir. Normal bir dağılımda, ortalama değer dağılımın merkezini oluşturur ve en olası eylemi temsil ederken, standart sapma dağılımın genişliğini belirlemektedir. Belirsizliği ve keşif sürecini yönetmektedir. Stokastik süreçlerde, ortalama değer ve standart sapmanın kombinasyonu karar vericinin çevre hakkında bilgi edinerek ödülleri optimize etmesini sağlamaktadır. Standart sapma yüksek olduğunda karar verici daha geniş bir eylem yelpazesi üzerinde örneklemeye yapmaya ve test edilmemiş bir karar vermeye teşvik edilirken, düşük bir standart sapma karar vericinin mevcut durumda onaylanmış en iyi eylemi gerçekleştirmesine olanak tanımaktadır. İki parametre arasındaki denge, karar vericinin hem kısa vadede ödülleri maksimize etmesini hem de uzun vadede genelleştirilebilir ve uyarlanabilir stratejiler geliştirmesini sağlamaktadır. Ortalama değer ve standart sapmanın dinamik olarak ayarlanması, karar vericinin değişen çevresel koşullara uyum sağlamasına ve performansı optimize etmesine de olanak tanımaktadır. Sonuç olarak, RL ve stokastik süreçlerde ortalama değer ve standart sapmanın kullanılması, öğrenme sürecini daha verimli hale getirerek karar verici için esnek bir öğrenme deneyimi sağlamaktadır.

μ , iki doğrusal katman ve bir aktivasyon fonksiyonu kullanılarak $\tanh(W_2 \text{ReLU}(W_1 h_t + b_1) + b_2)$ ilişkisinden hesaplanmaktadır. Standart sapma, $\sigma = \exp(W_\sigma h_t + b_\sigma)$ ile tanımlanan yeniden parametrelendirme yaklaşımından hesaplanmaktadır. Burada $W_1, W_2, W_\sigma, b_1, b_2, b_\sigma$ doğrusal katmanların ağırlık parametreleridir. Hesaplanan parametreler ile denklem 4.24 ile tanımlanan olasılık yoğunluk fonksiyonu oluşturulmaktadır.

$$P(l_t | \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(l_t - \mu)^2}{2\sigma^2}\right) \quad (4.24)$$

Burada l_t belirli bir durumda karar verici tarafından üretilen eylemin bir göstergesidir ve $P(l_t | \mu, \sigma)$ model tarafından örneklenen eylemlerin olasılığını vermektedir. Hesaplanan olasılık dağılımından rastgele bir konum noktası seçilmektedir. Böylece, karar vericinin mevcut durumu keşfetmesine ve belirsizlik altında karar vermesine olanak tanıyan stokastik bir davranış katmanı oluşturulmaktadır.

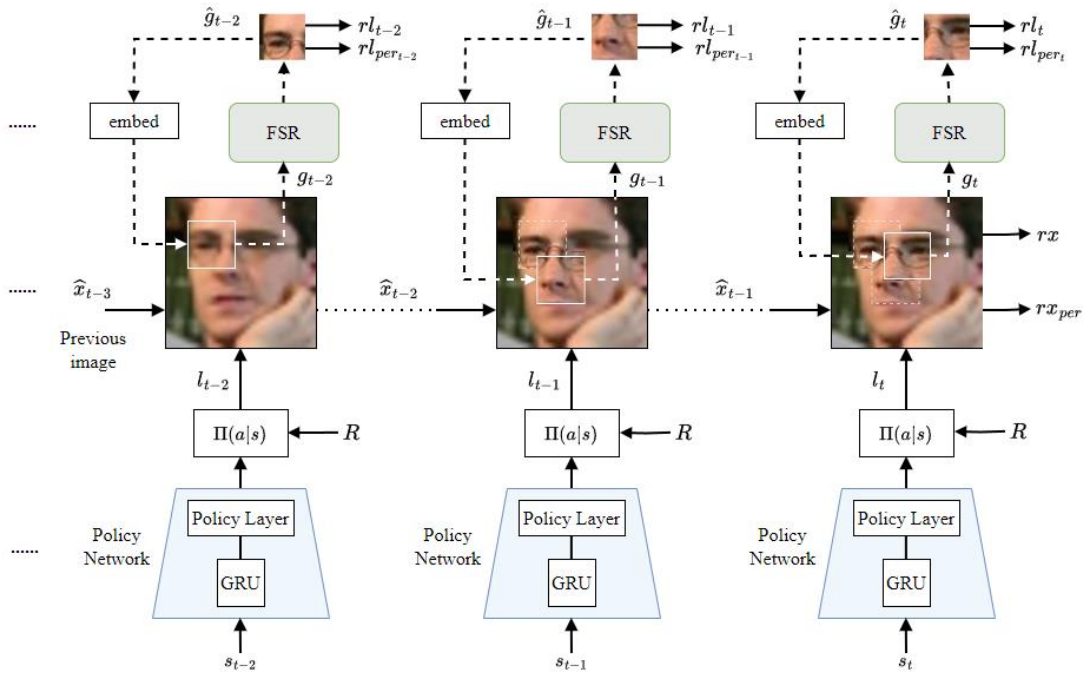
4.2. Model Eğitimi

Önerilen yöntemin $\mathcal{F}_{DRL}$ kısmında, belirsiz ortamlarda konum seçimini verimli bir şekilde uygulamak için POMDP modeli kullanılmaktadır. POMDP'nin amacı belirsizlik altında en iyi yer seçim sırasını belirlemektir. POMDP modelinde tam olarak gözlemlenemeyen durumlarda karar veren bir ajan veya sistem mevcuttur. Ajanın mevcut bilgileri ve geçmiş deneyimlerini kullanarak çevresi hakkında bir içsel olasılıksal temsile sahip olmasına inanç durumu denilmektedir. Dolayısıyla karar vericinin tam olarak gözlemlenemeyen durumlarda en iyi eylemleri seçebilmesi için inanç durumu kullanılarak bir karar verme stratejisi geliştirilmelidir. İnanç durumu, karar vericinin mevcut durumda olma olasılıklarını temsil eder ve olasılıklar, alınan gözlemler ve gerçekleştirilen eylemlerle sürekli olarak güncellenmektedir. İnanç durumu, karar vericinin politikayı belirlemesi ve politika doğrultusunda en iyi eylemleri seçmesi için kritik bir rol oynamaktadır. Ancak karmaşık ve belirsiz ortamlar için en iyi politikayı öğrenmek geleneksel yöntemlerle oldukça zordur. Zorluk karşısında RL yöntemleri devreye girmektedir. Gradyan tabanlı bir Monte Carlo Policy Gradient yöntemi olan reinforce algoritması, inanç durumuna göre en iyi politikayı belirlemek için kullanılmaktadır. Reinforce algoritması politika parametrelerini ödül sinyallerine göre güncellemektedir. Böylece karar verici zaman içinde en yüksek toplam ödülü elde etmeyi öğrenmektedir. Süreçte, karar vericinin belirsizlik altında en iyi kararları vermesini sağlamak için inanç durumları kullanılmaktadır.

Şekil 4.10, sıralı sürecin üç adımını göstermektedir. Model her t adımında çevre s_{t+1} ile gösterilen bir durum üretmektedir. Durumlar arasındaki geçişler, önceki tüm $l_{1:t}$ seçimlerine ve önceki durumlara ($s_{1:t}$) bağlı olan bilinmeyen bir $P(s_{t+1}|l_{1:t}, s_{1:t})$ fonksiyonu tarafından yönetilmektedir. Politika ağı, anlık eylemler üzerindeki bir dağılımı geçmiş etkileşimlerin geçmişine eşleyen stokastik bir politika $\Pi((l_t)|s_{1:t}, \theta)$ öğrenir $h_{1:t} = s_1, l_1, s_2, l_2, \dots, s_{t-1}, l_{t-1}, s_t, l_t$. Bir t anında, l_t seçimi yalnızca s_t anındaki gözlemle etkileşime girer ve ödül sinyalleri bir önceki döngü sırasında hesaplanmaktadır.

Şekil 4.10'da önceki görüntüden alınan parça FSR yapısı ile iyileştirilir ve gelen görüntünün içine gömülmektedir. Ödüller hem parçadan hem de elde edilen görüntüden hesaplanmaktadır. Son adımda geliştirilmiş görüntü çıktı olarak alınmaktadır.

Şekil 4.10 incelendiğinde bir s_t durumu için l_t konumundaki g_t dikkat parçası olarak seçilmektedir. Seçilen parça $\mathcal{F}_{FSR}$ modeli ile iyileştirilerek $\hat{g}_t$ elde edilmiş ve bir önceki iyileştirilmiş yüz görüntüsü $\hat{x}_{t-1}$ 'e gömülmektedir. Ödül mekanizması parça için çalıştırılarak iki yerel ödül (rl , rl_{per}) hesaplanmaktadır. Son adım T 'de oluşan son yüz görüntüsü ve hedef görüntü üzerinden iki global ödül (rx ve rx_{per}) eklenmektedir. Ödül süreci içindeki E_{pc} sayesinde yüzün tüm bölgelerine ulaşılması hedeflenirken, iyileştirilmesi gereken bozuk parçalara $\mathcal{F}_{DRL}$ model eğitimi ile daha çok uğranması amaçlanmaktadır.



Şekil 4.10. Tekrarlayan yapının son üç adımının görsel analizi.

Beklenen toplam ödül $J(\theta)$, politika parametresi θ tarafından hesaplanan toplam olasılık yoğunluk fonksiyonu $p(s_{1:T}; \theta)$ altında r_t toplamalarının beklenen değeri olarak ifade edilmektedir.

$$J(\theta) = \mathbb{E}_{\pi_{\theta}} \left[\sum_{t=0}^T R_t \right] = \mathbb{E}_{p(s_{1:T}; \theta)} \left[\sum_{t=1}^T R_t \right] \quad (4.25)$$

$J(\theta)$ 'yı maksimize etmek için beklenen ödülü maksimize eden θ parametreleri hesaplanmalıdır. Monte-Carlo Policy Gradient kuralı ile beklenen değer üzerinde gradyan alınarak politika parametrelerinin toplam ödül R üzerindeki etkisi analiz edilmelidir [9]. Beklenen değer için logaritmik türev kuralı uygulanarak

denklem 4.25 oluşturulmaktadır. Böylece, ajan en yüksek toplam ödül R 'yi elde etmek üzere eğitilmektedir.

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{(s_{1:T}; \theta)} \left[\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \left[\sum_{t=1}^T R_t - b_t \right] \right] \quad (4.26)$$

Denklem 4.26'da b_t , eylemden bağımsız olarak duruma bağlı olan bir tabandır. Taban, ilgili batchdeki (eğitim yığın boyutu) değerin bir fonksiyonudur. Taban ödüllendirici konumları seçme olasılığını artırır ve tersini seçme olasılığını azaltmaktadır. Sonuç olarak her bir yığın ve parça sayısı üzerinden beklenen toplam ödülün gradyanları hesaplanmaktadır. M yığın sayısı olarak düşünülürse, nihai gradyan kuralı denklem 4.27'de verilmiştir ve reinforce kuralı olarak da bilinmektedir.

$$\nabla_{\theta} J(\theta) = \frac{1}{M} \sum_{i=1}^M \sum_{t=1}^T \nabla_{\theta} \log \pi(a_t^i | s_{1:t}^i; \theta) (R_t^i - b_t) \quad (4.27)$$

Mevcut politikayı kullanarak amaç, $s_{1:T}$ etkileşim dizilerinin örneklerini elde etmek ve ardından θ parametresini ayarlayarak ve düşük ödüllü eylemleri seçmekten kaçınarak yüksek toplam ödüllü eylemleri seçip eylemlerin logaritmik olasılığını artırmaktır.

Buraya kadar anlatılan eğitim süreçleri RL kapsamında olmaktadır. Ödül fonksiyonlarının ayrıntısı yukarıdaki bölümlerde verilmektedir. Fakat gradyan kaybolma olayının çözümünü anlatabilmek ve olayın daha iyi anlaşılması için ödül kısmı kısaca hatırlatılarak devam edilecektir.

Ortalama karesel hata (MSE), RRDB ağındaki yerel bölgelerin eğitim süreçleri ve RRDB ağ parametrelerinin güncellenmesi için kullanılmaktadır. Kayıp fonksiyonu olarak iyileştirilmiş görüntü ile gerçek görüntü arasındaki fark kullanılmaktadır. Parçalar için kayıp fonksiyonu ve global kayıp fonksiyonu MSE kullanılarak hesaplanmaktadır. i 'inci parça için kayıp fonksiyonu $L_i(g) = \frac{1}{n_i} (\hat{g}_i - g_i^y)^2$ olarak hesaplanmaktadır. Burada $n_i = c \times u \times v$ 'dir. T parça için toplam kayıp fonksiyonu denklem 4.28 olarak verilir.

$$L(g) = \sum_{i=1}^T L_i(g) \quad (4.28)$$

T adımının sonunda, küresel kayıp fonksiyonu denklem 4.29 ile hesaplanmaktadır.

$$L_G = \frac{1}{k_i} (\hat{x} - y)^2 \quad (4.29)$$

Burada $k_i = c \times h \times w$ görüntüdeki piksel sayısıdır. $\mathcal{F}_{FSR}$ modelinde sadece $\hat{g}$ 'nin RRDB ile iyileştirilmesi görüntüye geri gömülmesi, global kayıp fonksiyonunun $\hat{x}$ ile y arasında olması nedeniyle gradyanların kaybolmasına neden olmaktadır. Sorunu çözmek için gradyan akışını sağlayacak bir ilişki kurularak $\hat{g}$ 'nin oluşumunda $\hat{x}_{t-1}$ modele dahil edilmektedir. FSR yapısındaki RRDB girişine sadece seçilen g parçası değil aynı zamanda $\hat{x}_{t-1}$ de uygulanmaktadır. Dolayısıyla $\hat{x}$ ve y arasındaki kayıp fonksiyonu $\mathcal{F}_{FSR}$ parametrelerinin güncellenmesini sağlamaktadır. Yöntemin optimum parametrelerini belirlemek için kullanılan nihai kayıp fonksiyonu denklem 4.30'da hesaplanmaktadır.

$$L = L(g) + L_G \quad (4.30)$$

Nihai kayıp fonksiyonunun model parametresi θ 'ya göre türevini hesaplarırken zincir kuralı uygulanmakta ve her bir $L(g)$ ve L_G için oluşturulan gradyanlar denklem 4.31 kullanılarak hesaplanmaktadır.

$$\frac{\partial L}{\partial(\hat{g}, \hat{x})} = \frac{2}{n_i T} \sum_{i=1}^T (\hat{g}_i - g_i^y) + \frac{2}{k_i} (\hat{x} - y) \quad (4.31)$$

Bir sonraki bölüm olan kısım 5'te önerilen çalışmanın deneysel sonuçları anlatılmaktadır.

5. DENEYSEL ÇALIŞMALAR VE BULGULAR

5.1. Giriş

Önerilen modelin üstünlüğünü göstermek için çeşitli kıyaslama deneyleri yapılmaktadır. Çalışmanın bu kısmında ilk olarak eğitim verileri ve uygulama detayları tanıtılmaktadır. Ardından önerilen DRL-SRFR modeli literatürdeki diğer yeni yöntemler ile kapsamlı bir şekilde karşılaştırılmaktadır. Karşılaştırmalarda SR ve FH'deki standart performans ölçütleri kullanılmaktadır. Özellikle önerilen modelin farklı parametrelere bağlı performansı ayrıntılı olarak analiz edilmekte ve önceki yöntemlere göre avantajları vurgulanmaktadır.

Deneylerde, modelin performansını farklı veri kümeleri üzerinde değerlendirmek için çeşitli analizler yapılmaktadır. Analizler, modelin genel uygulanabilirliğini ve çeşitli koşullara uyum sağlama yeteneğini değerlendirmeyi amaçlamaktadır. Deneysel verilerde modeldeki her bir bileşenin katkısını daha da aydınlatmak için ayrıntılı ablasyon çalışmaları yapılmıştır.

Önerilen yöntem SRCNN [98], VDSR [99], SRGAN [73], IPFH [62], Attention-FH [7], SFMNet [71], ESRGAN [76], SR3 [79] ve SwinIR [95] ile karşılaştırılmaktadır. Yöntemler altı gruba ayrılabilir: (i) genel görüntü süper-çözünürlüğü: SRCNN ve VDSR; (ii) yüz halüsinasyonu: SFMNet; (iii) Difüzyon: SR3; (iv) Dönüştürücü: SwinIR; (v) üretken çekişmeli öğrenme: SRGAN ve ESRGAN; (vi) RL: IPFH ve Attention-FH. Birinci gruptaki yöntemler normal görüntü restorasyonunu ele alırken, ikinci ile altıncı gruptaki algoritmalar yüz görüntüsü restorasyonu için yaygın olarak uygulanmaktadır. Üçüncü grup, gürültüyü kademeli olarak kaldırarak yüksek kaliteli görüntüler üreten difüzyon tabanlı yöntemi içermektedir. Dördüncü grup, çoklu katmanlardaki dikkat mekanizmalarını kullanarak yüksek çözünürlüklü görüntüler elde eden transformatör tabanlı yöntemle ilgilidir. Beşinci gruptaki yöntemler görüntü yeniden yapılandırmada yaygın olarak kullanılmaktadır ve etkileyici sonuçlar elde etmektedirler. IPFH ve Attention-FH dışındaki diğer yöntemler tarafımızca uygulaması yapılarak sonuçlar elde edilmektedir. Karşılaştırılan yöntemler eğilirken ilgili makalelerde verilen eğitim parametreleri

kullanılmıştır. Tüm deneyler, NVIDIA 3090 24GB GPU, Ryzen 5 5600 CPU ve 32GB RAM ile gerçekleştirilmiş olup, performans analizleri standart değerlendirme ölçütleri olan PSNR ve SSIM üzerinden yapılmaktadır.

5.2. Veri Setleri ve Uygulama Detayları

5.2.1. Eğitim ve test veri setleri

Çalışmada, önerilen modelin performansını değerlendirmek ve farklı koşullarda nasıl davrandığını analiz etmek amacıyla dört farklı veri seti kullanılmaktadır: CelebA [100], BioID [101], LFW [102] ve PubFig [103]. İlgili veri setleri, hem çeşitlilik hem de yüz görüntüleri üzerinde yapılan çalışmaların literatürdeki yaygınlığı nedeniyle seçilmiştir. Her veri setinin özellikleri, kullanım amaçları ve seçilme nedenleri aşağıda detaylı bir şekilde açıklanmaktadır.

CelebA Veri Seti



Şekil 5.1. CelebA veri setinden bazı örnek görseller.

Şekil 5.1’de CelebA veri setinden bazı örnek görseller sunulmaktadır. CelebFaces Attributes Dataset (CelebA), geniş kapsamlı ve yüksek kaliteli yüz görüntülerinden oluşan bir veri kümesidir. Veri seti, toplamda 202.599 yüz görüntüsü içermekte ve bu görüntüler 10.177 farklı bireyden alınmıştır. Çalışmada bu veri setinden 40.000 görüntü eğitim ve 4.000 görüntü test için kullanılmaktadır. CelebA’nın öne çıkan özellikleri şunlardır:

- Zengin Etiketleme: Her yüz görüntüsü için yüz pozisyonları, cinsiyet, yüz özellikleri (örneğin, gözlük takıp takmadığı, saç modeli, vb.) gibi 40 farklı öznitelik bulunmaktadır.

- Doğal Çeşitlilik: Görüntüler farklı ışık koşulları, yüz ifadeleri ve açılar gibi geniş bir çeşitliliği barındırmaktadır.
- Büyüklük: Veri setinin büyük olması, derin öğrenme modelleri için yeterli çeşitliliği sağlamaktadır.

Kullanım Amacı: CelebA veri seti, modelin geniş çaplı ve karmaşık veri üzerinde nasıl performans gösterdiğini analiz etmek için kullanılmaktadır. Veri setinin farklı yüz özelliklerini içermesi modelin genelleme yeteneğini test etmek için uygundur.

BioID Veri Seti



Şekil 5.2. BioID veri setinden bazı örnek görseller.

Şekil 5.2’de BioID veri setinden bazı örnek görseller sunulmaktadır. BioID, daha çok biyometrik analizler için kullanılan, laboratuvar ortamında çekilmiş 1.521 gri tonlamalı yüz görüntüsü içeren bir veri setidir. Çalışmada BioID veri setinden 1.293 görüntü eğitim ve 228 görüntü test için kullanılmaktadır. BioID’nin dikkat çeken özellikleri şunlardır:

- Homojenlik: Laboratuvar ortamında çekildiği için arka planlar, ışık koşulları ve yüz açıları büyük ölçüde sabit olmaktadır.
- Yüksek Kalite: Görüntülerin sabit bir standartta olması, modellerin düşük varyanslı veri üzerinde nasıl performans gösterdiğini anlamayı sağlamaktadır.
- Biyometrik Odaklılık: Özellikle göz, burun ve ağız gibi biyometrik noktalar üzerinde yoğunlaşmaktadır.

Kullanım Amacı: BioID veri seti, modelin daha az karmaşık ve kontrollü koşullardaki performansını değerlendirmek için idealdir. Veri seti, modelin

laboratuvar ortamı görseller üzerinde detaylı yüz özelliklerini nasıl işlediğini görmek için kullanılmaktadır.

LFW Veri Seti



Şekil 5.3. LFW veri setinden bazı örnek görseller.

Şekil 5.3'te LFW veri setinden bazı örnek görseller sunulmaktadır. Labeled Faces in the Wild (LFW), anlık görüntülerden alınmış 5.749 farklı bireye ait toplamda 13.233 yüz görüntüsü içermektedir. Çalışmada, LFW veri setinden 11.248 görüntü eğitim ve 1.985 görüntü test için kullanılmaktadır. LFW'nin öne çıkan özellikleri şunlardır:

- Anlık Görüntü Koşulları: Görüntüler, anlık doğal ortamlarda çekilmiş olup farklı ışık koşullarını ve yüz ifadelerini içermektedir.
- Tanıma Problemleri: Veri seti, farklı ifadeler içerdiği için yüz özelliklerinin hassas bir şekilde analiz edilmesini sağlamaktadır.
- Çeşitlilik: Farklı pozlar, farklı yaş grupları ve farklı yüz özelliklerini içererek geniş bir çeşitlilik sunmaktadır.

Kullanım Amacı: LFW veri seti, modelin anlık doğal koşullarındaki görüntü kullanımındaki performansı değerlendirmek için önemli bir veri seti olmaktadır.

Şekil 5.4'te Pubfig veri setinden bazı örnek görseller sunulmaktadır. Public Figures Dataset (PubFig), web ortamından toplanmış toplamda 58.797 yüz görüntüsü içermektedir. Çalışmada, PubFig veri setinden 12.000 görüntü eğitim ve 2.000 görüntü test için kullanılmaktadır. PubFig veri seti şu özelliklere sahiptir:

- Web Tabanlı Görüntüler: Görüntüler, farklı platformlardan toplanmış olup geniş bir çeşitlilik sunmaktadır.

- Doğal Hatalar: Veri setinde bulanıklık, kötü ışıklandırma ve düşük çözünürlük gibi doğal sorunlar bulunmaktadır.
- Çeşitli Figürler: Veri setinde, farklı yaş, cinsiyet ve etnik kökenlerden bireylerin görüntüleri yer almaktadır.

PubFig Veri Seti



Şekil 5.4. Pubfig veri setinden bazı örnek görseller.

Kullanım Amacı: PubFig veri seti, modelin düşük kaliteli ve karmaşık veri üzerindeki dayanıklılığını ölçmek için tercih edilmektedir. Web’den toplanan görüntülerin çeşitliliği, modelin genelleme kabiliyetini test etmek için kritik bir rol oynamaktadır.

Yukarıdaki veri setleri şu nedenlerle seçilmiştir:

1. Çeşitlilik: Farklı pozlar, ışıklandırmalar, yaş grupları ve yüz ifadeleri sayesinde modelin genelleme kapasitesi test edilebilmektedir.
2. Boyut: Büyük veri setleri, derin öğrenme modelleri için yeterli miktarda eğitim verisi sağlayarak daha etkili sonuçlar elde edilmesine olanak tanımaktadır.
3. Kontrollü ve Doğal Koşullar: Laboratuvar ortamından doğal koşullara kadar geniş bir spektrumda veri sunarak modelin her iki durumda da performansı incelenmektedir.
4. Literatür Uyumu: Bu veri setleri, yüz süper çözünürlüğü ve yüz tanıma çalışmaları için literatürde yaygın olarak kullanıldığı için karşılaştırma yapılmasına olanak sağlamaktadır.

Sonuç olarak çalışmadaki dört veri seti, önerilen modelin hem genelleme kabiliyetini hem de spesifik görevlerdeki başarısını değerlendirmek için kapsamlı bir veri havuzu olarak kullanılmaktadırlar. Kullanılan veri setlerinin özellikleri, modelin hem yerel detayları hem de yüzün genel yapısal bütünlüğünü etkili bir şekilde geliştirme kapasitesini ortaya koymada olanak sağlamaktadır.

5.2.2. Uygulama detayları

Üzerinde çalışılan tüm yöntemlerde ve veri kümelerinde, RGB formatındaki görüntüler merkez bölge ortalanarak (128,128) olarak kırpılmıştır. Eğitim yığın boyutu 20'dir. Doğru ve güvenilir kıyaslama için, tüm yöntemler ön eğitimde başka veri kümeleri kullanılmadan yalnızca ilgili eğitim kümesi üzerinde eğitilmektedir. Önerilen yöntem, farklı durumları değerlendirmek ve modelin yeterliliğini anlamak için 4 ve 8 ölçeklendirme faktörleriyle değerlendirilmektedir. Ayrıca giriş görüntüleri [0, 1] aralığında normalize edilmektedir. Model hızı, yeterli alan kapsamı ve doğruluk arasında bir denge sağlamak için politika ağı için T değeri 15 olarak ayarlanmıştır. DRL-SRFR modeli, temel öğrenme oranı 10^{-3} , momentum terimi 0,9 olan ADAM gradyan inişi kullanılarak eğitilmektedir. Her 15 epokta bir öğrenme oranı 0,5 azaltılmaktadır. σ değeri 0.25, RDB blok sayısı 3 ve RRDB blok sayısı 8 olarak ayarlanmıştır. Epok sayısı 150 ve yığın boyutu 10 olarak seçilmiştir. Verilen parametrelere dayanarak yöntemin çalışması için gereken donanım RAM miktarı yaklaşık 6 GB'tır. Bir fikir vermesi açısından, LFW veri kümesi için eğitim süresi bir epok için 925 saniyedir. RRDB bloklarının sayısı arttıkça eğitim süresi de artmaktadır. Örneğin 12 bloklulu bir yapıda eğitim süresi 1392 saniyedir. Parametreler tablo 5.1'de listelenmektedir.

Karşılaştırılan yöntemler aşağıdaki parametreler kullanılarak uygulanmıştır:

ESRGAN: blok sayısı 23, yığın boyutu 16, epok sayısı 1000; SWINIR: pencere boyutu 8, transfer blok sayısı 6, gömülü boyut 180, num kafa sayısı 6, çok katmanlı algılayıcı katman boyutu 2, yığın boyutu 8, epok sayısı yöntemde verilen yaklaşık performans değerine kadar devam etmiştir. SR3: iç kanal sayısı 64, kanal genişleme faktörü [1,2,4,8,8,8], dikkat mekanizmasının uygulandığı çözünürlük 16, her çözünürlük seviyesindeki artık blok sayısı 0.2, modelin eğitim sürecinde gürültü ekleme ve çıkarma işlemleri 2000 kez tekrarlanmış, gürültü ekleme ve çıkarma oranı

beta değeri 10^{-6} 'dan 10^{-2} 'ye kadar doğrusal olarak artırılıp azaltılmış, toplam iterasyon sayısı maksimum 500000'dir.

Tablo 5.1. Uygulama parametreleri.

Parametre	Değer
Görüntü Formatı	RGB
Kırpma Boyutu	(128,128)
Yığın Boyutu (Batch Size)	20
Ön Eğitim	Sadece ilgili eğitim verisi kullanılmıştır
Ölçeklendirme Faktörleri	4x, 8x
Normalizasyon	[0, 1] aralığında
Politika Ağı T Değeri	15
Optimizasyon	ADAM
İlk Öğrenme Oranı	10^{-3}
Momentum	0.9
İçsel Azaltma Faktörü (öğrenme oranı)	Her 15 epokta 0.5
σ Değeri	0.25
RDB Blok Sayısı	3
RRDB Blok Sayısı	8
Epok Sayısı	150
Yığın Boyutu	10
RAM Gereksinimi	6 GB

SFMNET iki dallı bir yapıya sahiptir, frekans dalı ve uzamsal dal işlemlerinde yukarı ve aşağı özellik çıkarımı için 8 katmanlı bir yapı kullanılmış ve kaynaştırılmıştır. Batch boyutu 8 ve eğitim sırasında epok sayısı 400'dür. IPH ve Attention-FH: Resmi bir kod sayfası ve başka bir kod uygulaması olmadığı için makalelerinde verilen değerler ve görüntüler kullanılmaktadır. SRCNN: görüntüler önce 33×33 parçalar kullanılarak 50000 epok ve yığın boyutu 16 ile eğitilmiş, özellik katmanını doğrusal olmayan bir katman takip etmiş ve SR görüntüsü son konvolüsyon katmanında yeniden oluşturulmuştur. VDSR: toplam 20 konvolüsyon katmanı vardır ve her katman 3×3 boyutunda filtreler kullanılmaktadır. Her konvolüsyon katmanı 64 filtre

içerir, batch boyutu 16 olarak ayarlanmakta ve model 100 epok için eğitilmektedir. Karşılaştırılan yöntem parametreleri aşağıda tablo 5.2’de verilmektedir.

Tablo 5.2. Karşılaştırılan yöntem parametreleri.

Yöntem	Parametreler
ESRGAN	Blok Sayısı: 23, Yığın Boyutu: 16, Epok: 1000
SWINIR	Pencere Boyutu: 8, Transfer Blok Sayısı: 6, Gömülü Boyut: 180, Num Kafa Sayısı: 6, MLP Katman Boyutu: 2, Yığın Boyutu: 8
SR3	İç Kanal Sayısı: 64, Kanal Genişleme Faktörü: [1,2,4,8,8,8], Dikkat Mekanizması: Çözünürlük: 16, Artık Blok Sayısı: 0.2, Gürültü Ekleyip Çıkarma: 2000 kez, Beta: 10^{-6} ’dan 10^{-2} ’ye, Maksimum İterasyon: 500000
SFMNET	Frekans ve Uzamsal Dal: 8 katman, Yığın Boyutu: 8, Epok: 400
IPH ve Attention-FH	Resmi kod bulunmadığı için makale verileri kullanılmıştır
SRCNN	Parça Boyutu: 33×33 , Epok: 50000, Yığın Boyutu: 16
VDSR	Konvolüyon Katman Sayısı: 20, Filtre Boyutu: 3×3 , Filtre Sayısı: 64, Yığın Boyutu: 16, Epok: 100

5.3. Performans Sonuçları

5.3.1. Değerlendirme ölçütleri

Önerilen DRL-SRFR modelinin performansını değerlendirmek için literatürde yaygın olarak kullanılan Tepe Sinyal-Gürültü Oranı (PSNR) ve Yapısal Benzerlik İndeksi (SSIM) görsel değerlendirme metrikleri tercih edilmektedir. Metrikler, yüz süper çözünürlük modellerinin ürettikleri görüntülerin kalite ve doğruluğunu hem matematiksel hem de algısal açıdan değerlendirmek için temel araçlardır. PSNR ve SSIM’in matematiksel ifadeleri aşağıda denklem 5.1 ve 5.2’de verilmektedir [7, 62, 73]:

PSNR

PSNR, orijinal görüntü ile model tarafından üretilen görüntü arasındaki farkı ölçmek için kullanılmaktadır. Matematiksel ifade şu şekildedir:

$$PSNR(\hat{x}, y) = 10 \times \log \left(\frac{255^2}{\|\hat{x} - y\|_2^2} \right) \quad (5.1)$$

Burada:

- $\hat{x}$: Model tarafından üretilen yüksek çözünürlüklü görüntü.
- y : Orijinal yüksek çözünürlüklü hedef görüntü.
- $\|\hat{x} - y\|_2^2$: İki görüntü arasındaki ortalama karesel hata (MSE).

Daha yüksek PSNR değerleri, modelin ürettiği görüntünün hedef görüntüye daha yakın olduğunu ifade etmektedir. Ancak PSNR yalnızca piksel yoğunluklarındaki farklılıkları değerlendirdiği için insan gözünün algısal özelliklerini tam olarak yansıtmayabilmektedir. Dolayısıyla PSNR sonuçları dikkatli bir şekilde yorumlanmalıdır.

SSIM

SSIM (Structural Similarity Index), iki görüntü arasındaki yapısal benzerliği ölçmek için kullanılmaktadır. SSIM, insan görsel algısına daha yakın bir metrik olarak bilinmekte ve genellikle PSNR'yi tamamlayıcı bir ölçüt olarak değerlendirilmektedir. Matematiksel ifade şu şekildedir:

$$SSIM(\hat{x}, y) = \frac{(2\mu_{\hat{x}}\mu_y + C_1)(2\sigma_{\hat{x}y} + C_2)}{(\mu_{\hat{x}}^2 + \mu_y^2 + C_1)(\sigma_{\hat{x}}^2 + \sigma_y^2 + C_2)} \quad (5.2)$$

Burada:

- $\mu_{\hat{x}}, \mu_y$: $\hat{x}$ ve y görüntülerinin piksel değerlerinin ortalamaları.
- $\sigma_{\hat{x}}^2, \sigma_y^2$: $\hat{x}$ ve y görüntülerinin varyansları.
- $\sigma_{\hat{x}y}$: $\hat{x}$ ve y görüntüleri arasındaki kovaryans.
- C_1 ve C_2 : Küçük sabitler olup, sıfıra bölünmeyi önlemek için kullanılır.

SSIM değeri 0 ile 1 arasında değişmektedir. 1'e yaklaşan bir SSIM değeri, iki görüntü arasında yüksek benzerliği ifade ederken, 0'a yaklaşan bir değer düşük benzerliği göstermektedir. SSIM, özellikle yüz görüntülerinde, dokuların ve ince detayların korunmasını değerlendirmek için oldukça faydalıdır.

Önerilen yöntem diğer mevcut yöntemlerle görsel ve anlamsal olarak karşılaştırılmaktadır. Karşılaştırmalarda PSNR ve SSIM değerleri, yüz görüntülerinin

artan çözünürlüğü ile model çıktı sonuçlarını değerlendirmek için tablolar halinde sunulmaktadır. Farklı veri kümeleri için çeşitli yöntemler arasında görsel karşılaştırmalar da yapılmıştır. Attention-FH modelinin eğitim modeli dosyaları mevcut olmadığından karşılaştırma amacıyla ilgili makaledeki hazır görüntüler kullanılmaktadır.

Önerilen DRL-SRFR modeli, hem görsel hem de sayısal olarak literatürdeki mevcut yöntemlerle kapsamlı bir şekilde karşılaştırılmaktadır. Performans değerlendirmesi yukarıda ayrıntılarına değindiğimiz veri kümeleri üzerinde gerçekleştirilmektedir:

- BioID: Kontrollü laboratuvar koşullarında çekilmiş görüntüler.
- CelebA: Gerçek dünya yüz ifadelerini ve özelliklerini içeren geniş ölçekli bir veri seti.
- LFW: Gerçek dünyadan alınmış, çeşitli açılarda ve koşullarda çekilmiş yüz görüntüleri.
- PubFig: Web tabanlı çeşitli yüz görüntülerini içeren bir veri seti.

Karşılaştırmalar ilgili veri setleri üzerinde farklı ölçekleme faktörleri (4x ve 8x) ile yapılmakta ve PSNR ile SSIM değerleri tablolar ve grafikler kullanılarak sunulmaktadır. Tablo 5.3 ve Tablo 5.4'te farklı veri kümeleri üzerindeki PSNR ve SSIM değerlerinin detaylı bir analizi yer almaktadır.

5.3.2. Görsel karşılaştırmalar

SR'nin yüz görüntülerine uygulanması birkaç kritik faktöre dikkat edilmesini gerektirmektedir. İlk olarak algoritmalar yüksek çözünürlüklü görüntünün cilt tonlarını, dokularını ve yüz özelliklerini koruyarak doğal ve gerçekçi bir görünüme sahip olmasını sağlamalıdır. Benzer şekilde gözler, dudaklar ve saç telleri gibi önemli ayrıntıları doğru bir şekilde yeniden yapılandırarak ayrıntıları geliştirmeye özel önem verilmelidir. İkinci olarak orijinal renkleri korumak ve renk sapmalarını önlemek için renk doğruluğu önemlidir. Üçüncü olarak çeşitli görüntü senaryolarında SR işlemlerini hızlı ve etkili bir şekilde gerçekleştirme becerisi çok önemlidir. Dördüncü olarak, modelin yüz detaylarını anlayabilmesi için eğitim verileri çeşitli örnekler içermelidir.

Ayrıca detay seviyesi de önemli bir faktördür; özellikle ağız, gözler, burun, saç modeli gibi yüzün ana bileşenleriyle ilgili detaylar kullanım amacına bağlı olarak farklı önceliklere sahip olmaktadır. Jestler, kaş hareketleri, kulaklar ve aksesuarlar

gibi yerel detaylar da genel yüz ifadesinin belirlenmesinde etkili olabilmektedir. Ancak sadece teknik detaylara odaklanmak yerine insan gözünün bu detayları nasıl algıladığını anlamak da önemlidir. Algısal koruma düşünüldüğünde global yapının incelenmesi önemlidir. Yüz SR işlemi sonucunda elde edilen görüntü genel yapısı bakımından algısal bütünlüğünü korumalıdır. Küresel yapının dikkate alınması, yüz özelliklerinin sadece yerel detaylarla elde edilemeyecek doğallığını ve orantılılığının yakalanmasını sağlamaktadır. Güvenlik gibi bazı uygulamalarda ayrıntılar önemli olabilirken, başka bir bağlamda yüzün genel formu ve estetiği öncelikli olabilmektedir. Sonuç olarak PSNR ve SSIM gibi teknik metriklere ek olarak, insan algısını anlamak ve kullanım amacına göre öncelikleri belirlemek, yüz SR teknolojilerinin başarılı bir şekilde değerlendirilmesinde kritik bir rol oynamaktadır.

Sayısal ve görsel karşılaştırma sonuçlarından, model davranışının ölçekleme faktörüne ve veri setlerine bağlı olarak değiştiği görülmektedir. Ölçekleme faktörü 4x için mevcut yöntemler ve önerilen yöntem benzer sonuçlar vermektedir. Ayrıca, üretici ağ tabanlı yöntemler bazen görüntüde bulunmayan ek yapaylıklar oluşturmaktadır. Üretici ağ yapılarında oluşan yapaylıklar görsel algıyı iyileştirse de, hem kişisel özelliklerin korunmamasına hem de karşılaştırma metriklerinde bozulmaya neden olmaktadır. Ölçekleme faktörü 8x olduğunda yapaylıklar çıktıda daha belirgin hale gelmektedir. Görsel algı açısından sonuçlar, insanların orijinal hallerine kıyasla farklı yönleri bakması veya yüz ifadelerinin değişmesi gibi olumsuz bir izlenim bırakmaktadır.

Tablo 5.3. PSNR değerlerinin 4 ve 8 ölçek oranına göre karşılaştırılması.

Veri Kümesi	Ölçek	Bicubic	SR		Face Hallucination	Diffusion	Transformer	GAN		Reinforcement Learning		DRL-SRFR
			SRCNN	VDSR	SFMNet	SR3	SwinIR	ESRGAN	SRGAN	IPFH	Attention-FH	
BioID	x 4	24.59	25.99	29.56	31.79	31.67	30.59	31.66	28.29	-	<u>33.38</u>	34.12
	x 8	20.24	21.15	23.12	26.50	28.29	25.17	<u>28.35</u>	24.05	26.59	27.81	28.86
CelebA	x 4	25.76	26.42	28.08	29.25	31.59	29.89	<u>30.75</u>	28.44	-	30.58	30.42
	x 8	21.84	22.05	23.96	24.80	30.04	23.41	<u>28.50</u>	24.61	26.42	26.14	27.26
LFW	x 4	26.79	28.13	31.73	35.01	32.68	33.69	33.05	32.23	-	32.93	<u>34.29</u>
	x 8	21.92	23.22	24.07	28.15	27.35	26.56	<u>28.49</u>	25.99	24.93	27.81	28.66
PubFig	x 4	24.76	25.21	28.05	33.04	34.20	32.80	32.70	27.44	-	28.87	<u>33.23</u>
	x 8	20.75	21.31	21.94	27.72	28.64	26.42	<u>29.01</u>	23.45	-	23.59	29.30

Tablo 5.4. SSIM değerlerinin 4 ve 8 ölçek oranına göre karşılaştırılması.

Veri Kümesi	Ölçek	Bicubic	SR		Face Hallucination	Diffusion	Transformer	GAN		Reinforcement Learning		DRL-SRFR
			SRCNN	VDSR	SFMNet	SR3	SwinIR	ESRGAN	SRGAN	IPFH	Attention-FH	
BioID	x 4	0.8056	0.8268	0.8627	0.8971	<u>0.9518</u>	0.8860	0.8568	0.8630	-	0.9418	0.9524
	x 8	0.6189	0.6015	0.6601	0.7174	0.7895	0.7011	0.6359	0.6881	0,78	<u>0.8568</u>	0.8627
CelebA	x 4	0.7672	0.7634	0.8075	0.8670	0.7902	0.8747	0.7819	0.8300	-	<u>0.8711</u>	0.8705
	x 8	0.6107	0.6120	0.6665	0.7001	0.6324	0.6662	0.5318	0.6753	0,80	<u>0.7441</u>	0.7640
LFW	x 4	0.8469	0.8599	0.8945	0.9353	0.9011	0.9215	0.8915	0.9266	-	<u>0.9418</u>	0.9573
	x 8	0.6712	0.6907	0.7074	0.7859	0.7857	0.7433	0.6679	0.7381	0,69	<u>0.8568</u>	0.8643
PubFig	x 4	0.7600	0.7708	0.8262	<u>0.9262</u>	0.9113	0.9141	0.8827	0.8197	-	0.8587	0.9297
	x 8	0.5782	0.5819	0.5822	<u>0.7758</u>	0.7057	0.7507	0.6582	0.6747	-	0.6805	0.8341

Önerilen model hem lokal hem de tüm görüntüde başarılı çıktılar üretmiş olup görsel algı performansı açısından rekabetçi bir yöntemdir. Sonuçlardan, önerilen yöntemin bakış, yüz ifadesi ve duygu aktarımı açısından yüz ifadesi açısından kaliteli çıktılar ürettiği görülmektedir. Ayrıca önerilen yöntem nicel ve özellikle SSIM değerlendirmelerinde daha iyi sonuçlar üretmektedir.

Tablo 5.3 ve Tablo 5.4, BioID, CelebA, LFW ve PubFig veri kümeleri üzerinde farklı ölçekleme faktörlerinde (4x / 8x) PSNR ve SSIM metrikleri açısından önerilen yöntemin karşılaştırılan yöntemlere göre üstünlüğünü göstermektedir. Farklı modellerin farklı veri kümeleri üzerinde yapılan deneysel testlerinde bazı sonuçlar öne çıkmaktadır. Klasik FSR yöntemlerinin sonuçlarının diğer yöntemlere kıyasla oldukça düşük olduğu görülmektedir. Diğer yöntemler arasında çeşitli açılardan rekabet söz konusudur. SFMNet modeli 4x ölçekleme faktöründe tatmin edici sonuçlar verirken 8x ölçekleme faktöründe zayıf sonuçlar vermektedir.

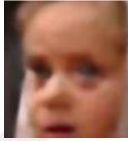
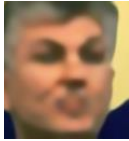
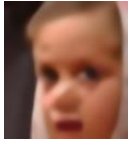
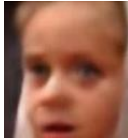
ESRGAN modeli PSNR açısından Tablo 5.3'teki en iyi ikinci yöntemdir, ancak SSIM açısından Tablo 5.4'teki çoğu yöntemden daha kötü sonuç vermektedir. SwinIR modeli iyi sonuçlar üretmesine rağmen transformatör modeli eğitimi için gereken çok sayıda veri kümesi nedeniyle diğer modellerin gerisinde kalmaktadır. Difüzyon tabanlı SR3 yöntemi, üretken doğası nedeniyle büyük veri kümelerinde daha iyi sonuçlar üretmektedir. Ancak SR3 küçük veri kümelerinde önerilen yöntem kadar iyi performans göstermemektedir. DRL-SRFR, küçük bir veri kümesi ile eğitildiğinde dahi yerel detay çıkarımı ve global koruma açısından iyi sonuçlar vermektedir. Veri kümesi sayısı arttıkça DRL-SRFR yönteminin metrik ve algısal performansı da artmaktadır.

Veri kümesi boyutuna bağlı olarak performanstaki değişkenlik, en az veriye sahip BioID ile en çok veriye sahip CelebA'nın sonuçları arasındaki önemli farktan anlaşılmaktadır. Ancak SSIM metriği açısından DRL-SRFR veri kümesi boyutundan bağımsız olarak en iyi sonuçları vermektedir. İkinci en iyi yöntemle karşılaştırıldığında DRL-SRFR, BioID veri kümesi için 4x ölçekleme faktöründe ortalama 0,74 dB PSNR kazancı elde etmektedir.

Model CelebA veri kümesi için 1,17 dB, LFW veri kümesi için 0,72 dB ve PubFig veri kümesi için 0,97 dB farkla üçüncü en iyi sonucu vermektedir. PSNR ölçütü açısından 8x ölçekleme faktöründe DRL-SRFR yöntemi, ikinci en yakın modele

kıyasla BioID, LFW ve PubFig veri kümelerinde sırasıyla ortalama 0,51 dB, 0,17 dB ve 0,29 dB kazanç sağlamaktadır. CelebA veri kümesinde ise 2,78 dB'lik farkla üçüncü sırada yer almaktadır. Benzer şekilde, 4x ölçekleme faktörü için DRL-SRFR yöntemi, en iyi ikinci modele kıyasla BioID, LFW ve PubFig veri kümeleri için sırasıyla 0,0006, 0,0155 ve 0,0035 ortalama SSIM kazançları sağlamaktadır. Sadece CelebA veri kümesinde 0,0039 SSIM kazanç farkı ile üçüncü en iyi sonucu vermektedir. 8x ölçekleme faktörlü SSIM analizine göre DRL-SRFR, ikinci en iyi modele kıyasla ortalama 0,0059, 0,0199, 0,0075 ve 0,0583 SSIM kazanç farkları üreterek diğer tüm yöntemlerden daha iyi performans göstermektedir.

Önerilen yöntemin diğer yöntemlere kıyasla yerel detayları ve küresel görünümü korumadaki performansı aşağıda Şekil 5.5'te görülebilmektedir. Rakamlar ve sayısal değerler genel olarak değerlendirildiğinde Attention-FH ve DRL-SRFR'nin daha iyi sonuçlar verdiği görülmektedir. Diğer yöntemler ya çok daha düşük niceliksel değerler ya da görsel algısı zayıf (bulanık ve yapay çıktı) görüntüler üretmektedir. Şekil 5.5'te, a sütunundaki kişinin gözlükleri diğer yöntemlerde neredeyse algılanamazken, DRL-SRFR ile SR3 modelinde açıkça görülebilmektedir. Sayısal ölçütlere göre ikinci en iyi sonucu veren ESRGAN yönteminin çıktısıyla karşılaştırıldığında DRL-SRFR bütünsel ve görsel algı açısından orijinal görüntüye daha yakın bir çıktı vermektedir. SR3 modelinin görsel algısı yüksek olmasına rağmen sağ gözün kısılması ve sol gözün daha açık olması orijinal görüntüden sapmaya neden olmuştur. Attention-FH çıktısında yüzde şaşkınlık ifadesine yakın bir ifade algılanırken orijinal görüntüde böyle bir algı yoktur. Sütun b'deki kişi analiz edildiğinde, diğer tüm yöntemler orijinalden uzak göz ve yüz yapıları üretmektedir. Buna karşılık DRL-SRFR yöntemi Attention-FH'ye kıyasla göz bölgesinin çözünürlüğü daha düşük bir görüntü üretmektedir. Buna rağmen genel göz yapısı ve yüz ifadesi orijinali daha çok anımsatmaktadır. SR3 yönteminde ise görsel algı yüksek olsa da kişinin bakışlarının farklı yönlere kayması, göz, diş ve ağızdaki yapay oluşumlar orijinal görüntüden uzaklaşılmasına neden olmuştur. c sütununda göz ve kaş bölgesindeki siyahlık diğer yöntemlerin çıktılarında daha yoğun ve karışık ya da SR3 yönteminde daha beyaz tonlu göz bebeği yapay oluşumları gözlemlenirken DRL-SRFR çıktısında daha ayırık ve net, dudak yapısı ve burun detayları orijinale daha yakındır.

	a	b	c	d	e	f
Original						
Bicubic						
SRCNN						
VDSR						
SR3						
SRGAN						
SFMNet						
SwinIR						
Attention-FH						
RL-SRFR (Önerilen model)						

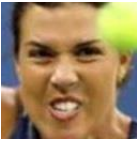
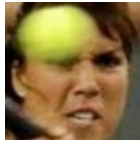
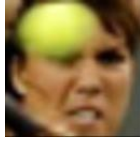
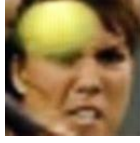
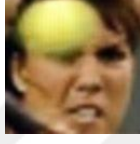
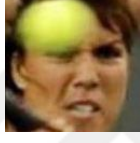
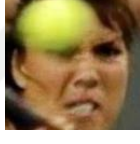
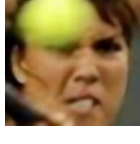
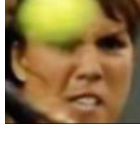
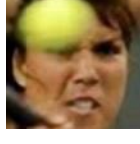
Şekil 5.5. Ölçekleme faktörü 8 için LFW veri seti üzeri farklı yöntem çıktıları..
Ölçekleme faktörü 8 için LFW veri seti üzeri farklı yöntem çıktıları.

d sütununda SR3 modelinde göz torbalarında oluşan çizgiler farklılaşmaya neden olmaktadır.

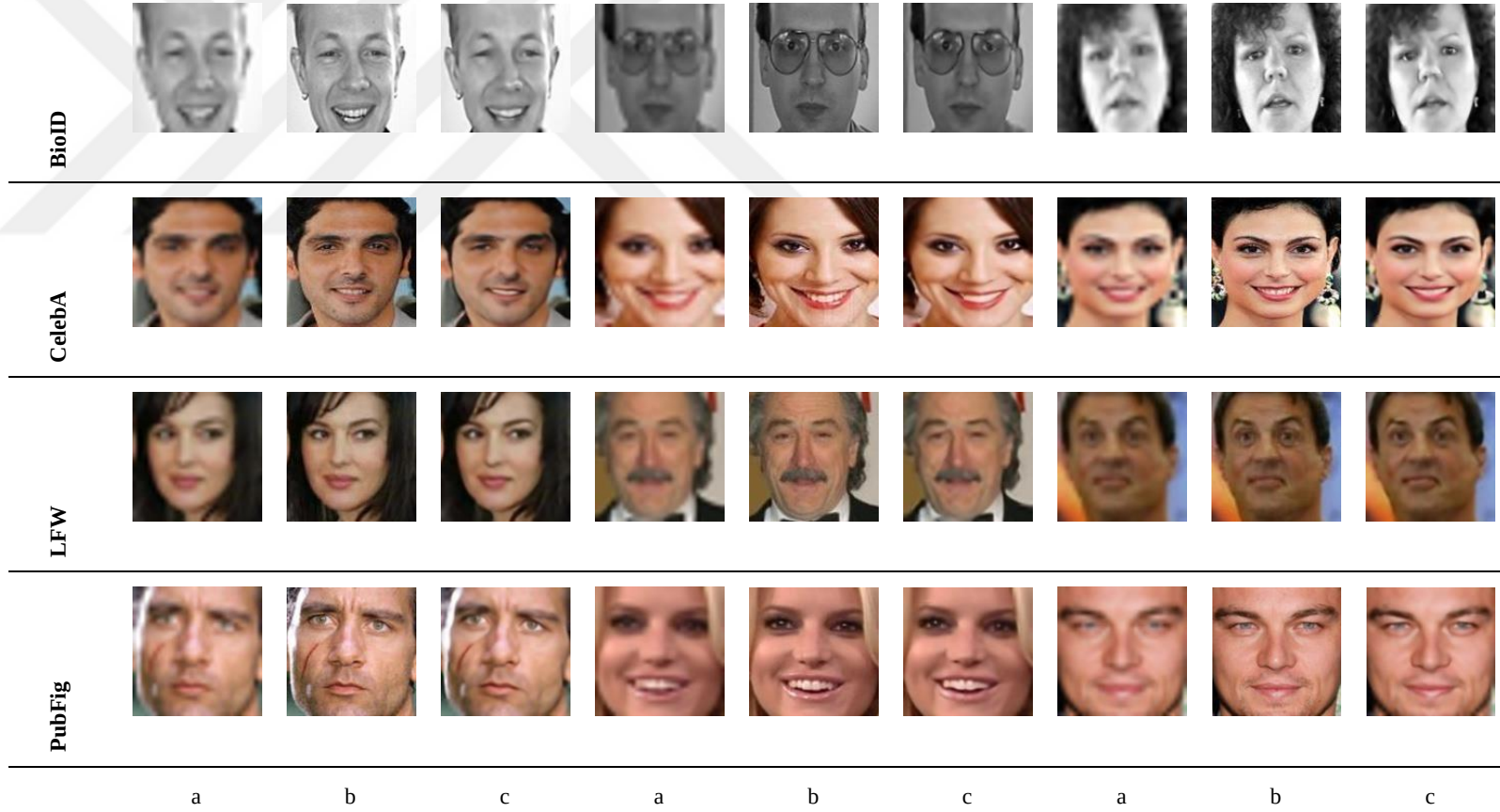
Göz bölgesinde önerilen model diğer yöntemlere göre daha iyi performans gösterirken ağız bölgesinde aynı performansı gösterememiştir. Ancak global yapı orijinale daha yakındır. e sütununda, diğer yöntemler çözünürlüğü artırmaya çalışırken yüzün uzunluğu ve genişliği orijinalere kıyasla önemli ölçüde değişmiştir. Attention-FH modelinde yerel göz çözünürlüğü iyileştirilmiş ve göz belirginleştirilmiş ancak orijinal yapı korunmamaktadır. SR3 modelinde ise görsel algı iyi olmasına rağmen gözlerin kısılması nedeniyle farklı bir duygu aktarılmaktadır. DRL-SRFR çıktısında göz çözünürlüğü daha düşük görünse de çıktı göz bölgesi, genel yüz özellikleri ve büyüklük açısından orijinale yakındır. f sütununda DRL-SRFR'ye yakın en iyi sonucu veren Attention-FH çıktısında ise sağ gözün çözünürlüğü daha iyi görünmekle birlikte orijinalde olmayan şaşkın bir ifade oluşmuştur. Sol gözde de bakış yönünde bir değişiklik vardır. SR3 modelinde ise alındaki çizgi gibi detaylar tüm yöntemlere göre daha nettir ancak sağ gözdeki yapısal değişiklik nedeniyle kişinin bakış hissi değişmiştir. Önerilen yöntemin çıktısında ise ifade orijinale yakın ve göz yönleri doğrudur. Ayrıca burun ile ağız arasındaki bölge incelendiğinde yüz hatlarının daha net ve orijinale daha yakın olduğu görülmektedir.

Şekil 5.6, Attention-FH, SR3, SFMNet, SFMNet, ESRGAN ve DRL-SRFR dışındaki yöntemlerin sağladığı çözünürlüklerin LFW veri kümesinde 4x ölçekleme faktörü için önemli ölçüde düşük olduğunu göstermektedir. Attention-FH, SR3 ve ESRGAN ikinci en iyi sonuçları verse de 8x ölçekleme faktöründe küresel olarak görsel algıda iyi sonuçlar sağlayabilmektedir. Sonuçlar DRL-SRFR yönteminin küresel görsel algı ve nicel değerler açısından iyi performans gösterdiğini ve yerel ayrıntıların ve genel görünümün korunmasında etkileyici bir başarı elde ettiğini göstermektedir.

Şekil 5.7, 4 ölçekleme faktörü için önerilen yöntem çıktısının orijinal görüntüye algısal yakınlığını göstermektedir. BioID veri setinde, siyah ve beyaz kontrast değerlerinde yerel detaylar ve global yapı korunmaktadır. BioID veri kümesindeki ilk kişi için ağız yapısı korunmuş ve aynı gülümseme algısı yakalanmıştır. Benzer şekilde ikinci kişide gözlükteki parlama detayları başarılı bir şekilde yakalanmıştır ve global benzerlik oldukça yakındır. Üçüncü kişide, saç kıvrımı detayı tespit edilebilecek kadar belirgindir ve genel benzerlik orijinaline oldukça yakındır.

	a	b	c	d	e	f
Original						
Bicubic						
SRCNN						
VDSR						
SR3						
ESRGAN						
SFMNet						
SwinIR						
Attention-FH						
DRL-SRFR (Önerilen model)						

Şekil 5.6. Ölçekleme faktörü 4 için LFW veri seti üzeri farklı yöntem çıktıları

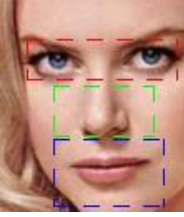
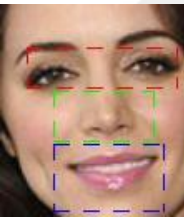
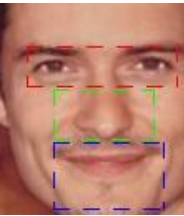


Şekil 5.7. BioID, CelebA, LFW ve PubFig veri kümelerinde ölçek 4 için DRL-SRFR modelinin çıktıları. (a) LR girdileri. (b) Temel gerçek. (c) DRL-SRFR.

Şekil 5.7 CelebA veri kümesinde, birinci kişinin yüz ifadesi aynı şekilde başarıyla yakalanmıştır ve kirli sakal detayları dikkat çekicidir. İkinci kişide gülümsemenin neden olduğu gamze çizgileri oldukça başarılı bir şekilde oluşturulurken genel yüz ifadeleri korunmaktadır. Üçüncü kişide ise aksesuar olarak küpe detayları ve kaş yapısı detayları başarılı bir şekilde oluşturulmaktadır.

Şekil 5.7 LFW veri setinde birinci kişide saç ayrımı detayı ve dudak yapısının şekil benzerliği öne çıkmaktadır. İkinci kişide ise bıyık olmasına rağmen yüzün genel ifadesi korunmuş ve çene kenarlarındaki çukurluklar eşit derecede belirgin gözükmemektedir. Üçüncü kişide ise yüzdeki şaşkın ifadenin korunduğu ve ağız yapısındaki kişisel duruşun aynı şekilde oluşturulabildiği dikkat çekmektedir. Son olarak, PubFig veri kümesinde, birinci kişide yüzdeki yara izi orijinaline önemli ölçüde benzemektedir. İkinci kişide dişlerin diziliminin çukurlaştırıldığı ve düz alanların orijinale yakın hale getirildiği dikkat çekmekte. Üçüncü kişide, yanaktaki çizgiler ve göz torbaları, bakışlardaki duygusal ifadenin korunmasıyla birlikte orijinaline çok yakın bir görsel sunmaktadır.

Şekil 5.8, orijinal yüksek çözünürlüklü (GT) ve düşük çözünürlüklü (LR) görüntülerin ve bunların çeşitli süper çözünürlük yöntemleriyle (ESRGAN, SwinIR, SR3 ve DRL-SRFR) geliştirilmiş versiyonlarının bir karşılaştırmasını sunmaktadır. Seçilen bireylerin görüntüleri PubFig veri setinden alınmıştır. Görüntülerin altındaki satırda yer alan değerler, yüz görüntülerinin farklı yerel bölgelerini (gözler, burun, ağız) ifade etmekte ve yerel bölgelerin iyileştirilmiş parçaları ile GT görüntüsünden oluşturulan parçalardan hesaplanan PSNR ve SSIM değerlerini göstermektedir. İlk yüz için DRL-SRFR yöntemi genel görüntüde 33,52 PSNR ve 0,924 SSIM ile diğer yöntemlerden (SwinIR hariç) daha iyi performans göstermekte ve özellikle ağız bölgesinde çok iyi sayısal değerler vermektedir. Ancak modelleri karşılaştırmak için sadece nicel değerlendirme ya da görsel algı kullanılmamalıdır. Örneğin en yüksek niceliksel değeri veren SwinIR, görsel algı açısından SR3 ve bizim modelimizden daha kötü görüntüler üretmektedir. Öte yandan, ESRGAN modeli nicel değerler ve görsel algı açısından olumlu sonuçlar üretememiştir. İkinci yüz için modelimiz genel görüntüde 34,45 PSNR ve 0,932 SSIM değerleri ile diğer yöntemlerden daha iyi performans göstermiştir. Ağız, burun ve göz kısımlarının nicel değerlendirmelerinde modelimiz diğer yöntemlerden daha iyi performans göstermiştir. Üçüncü yüz için,

GT	LR	ESRGAN	SwinIR	SR3	DRL-SRFR
					
	Görüntü	31.27 – 0.874	33.79 – 0.931	33.19 – 0.918	33.52 – 0.924
	Ağız	30.84 – 0.882	32.94 – 0.931	32.70 – 0.922	32.72 – 0.905
	Burun	30.64 – 0.888	34.80 – 0.930	34.12 – 0.922	35.28 – 0.944
	Gözler	29.65 – 0.837	30.85 – 0.886	30.68 – 0.925	30.76 – 0.887
					
	Görüntü	31.13 – 0.862	34.37 – 0.929	33.19 – 0.906	34.45 – 0.932
	Ağız	30.47 – 0.810	33.03 – 0.882	32.72 – 0.877	33.08 – 0.872
	Burun	31.82 – 0.890	35.86 – 0.923	35.04 – 0.919	35.95 – 0.933
	Gözler	29.69 – 0.791	31.33 – 0.873	30.75 – 0.868	31.90 – 0.925
					
	Görüntü	32.08 – 0.889	35.29 – 0.939	34.28 – 0.917	35.27 – 0.943
	Ağız	31.29 – 0.841	33.86 – 0.897	33.41 – 0.884	34.45 – 0.912
	Burun	32.30 – 0.883	36.35 – 0.940	34.90 – 0.919	36.7 – 0.943
	Gözler	30.25 – 0.863	31.76 – 0.873	31.87 – 0.922	32.43 – 0.933

Şekil 5.8. DRL-SRFR modelimizin PubFig veri kümeleri üzerinde ölçek 4 için küresel / yerel görsel ve nicel çıktıları.

genel görüntüde 35,27 PSNR ve 0,943 SSIM değerleri ile yüksek performans elde edilmektedir. Önerilen model tüm yerel bölgelerin PSNR ve SSIM değerlendirmelerinde diğer modellerden daha iyi performans göstermektedir. Şekil 5.8'deki bulgular, özellikle yüzün önemli bölgelerinde daha yüksek PSNR ve SSIM değerleri sağlayarak DRL-SRFR'nin görsel iyileştirme performansındaki üstünlüğünü göstermektedir. Özetle önerilen yöntem detay ve yapı kalitesini koruyarak keskin ve detaylı yüz görüntüleri elde edilmesini sağlamaktadır.

4x ölçekleme faktörü ve yığın boyutu = 1 için kullanılan parametreler dikkate alındığında, çeşitli yöntemler için hesaplama yükü ve parametre sayıları Tablo 5.5'te verilmektedir. FLOPs (Floating Point Operations per Second) bir modelin çıkarım sırasında gerçekleştirdiği kayan nokta işlemlerinin sayısını ifade ederken, hesaplama yükü GMAC (Giga Multiply-Accumulate Operations) modelin bir girdiyi işlemek için ihtiyaç duyduğu milyarlarca çarpma ve toplama işlemlerinin toplam sayısını ölçmektedir. Önerilen yöntemin parametre sayısı, flop sayısı ve hesaplama yükü mevcut SR çalışmalarına kıyasla düşüktür. Bir görüntüyü oluşturma süresi difüzyon hariç tüm yöntemler için benzerdir. Difüzyon yönteminde bir görüntü 61 saniyede oluşturulabilmektedir.

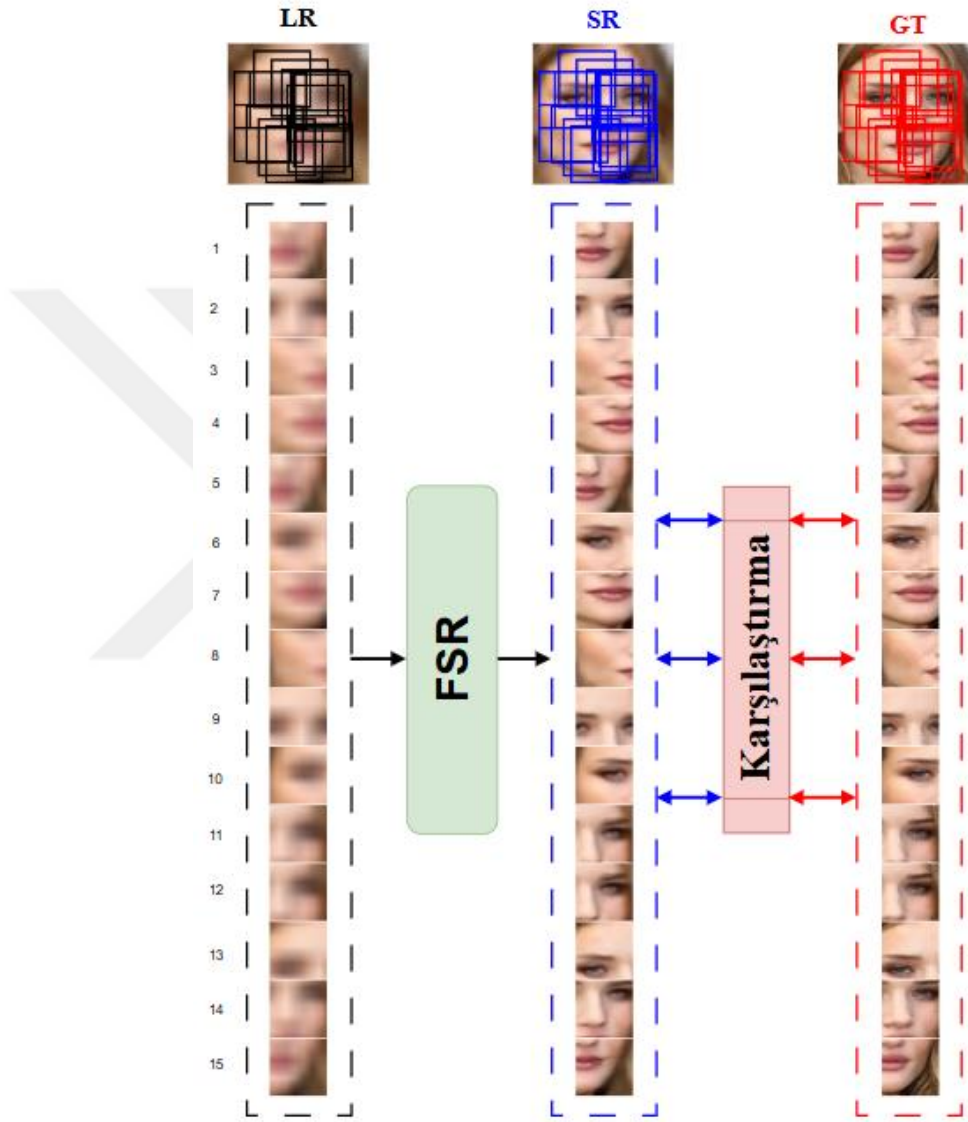
Tablo 5.5. Yöntemin hesaplama yükü açısından karşılaştırılması.

Model	TFlops	Parametre Sayısı	Test zamanı (Sn)	Hesaplama yükü (GMAC)
ESRGAN	84,1	43,242,820	0.99	42.06
SWINIR	60,3	11,900,199	1.39	30.17
SR3	91,7	97,807,491	61	45.89
DRL-SRFR	26,7	16,579,942	1.55	13.35

Aşağıdaki Şekil 5.9'da, önerilen DRL-SRFR yönteminin yüz görüntülerinin süper çözünürlüğündeki performansı detaylı bir şekilde gösterilmektedir. Görsel üç temel aşamayı içermektedir: giriş (LR – 8x düşük çözünürlüklü görüntü), çıkış (SR - model tarafından üretilen yüksek çözünürlüklü görüntü) ve orijinal (GT - hedef yüksek çözünürlüklü görüntü).

Şekil 5.9'daki görselde siyah dikdörtgenlerle işaretlenmiş alanlar, modelin iteratif süreçte seçtiği parçaları göstermektedir. Seçilen parçalar özellikle yüzün kritik

bölgelerini, örneğin göz, burun ve ağız gibi detayların daha iyi iyileştirilebilmesi için seçilmektedir. Düşük çözünürlüklü görüntüde bu alanlar bulanık, detayları kaybolmuş bir halde görülmektedir. Örneğin, göz çevresinde detaylar tamamen silik ve netlik kaybı mevcuttur. Süreçte DRL modülü yüzün hangi bölgelerinin iyileştirilmeye daha çok ihtiyaç duyduğunu belirlemektedir.



Şekil 5.9. Seçilen parçaların örnek bir gösterimi.

Şekil 5.9’da çıkış görüntüsünde mavi dikdörtgenlerle işaretlenmiş bölgeler, DRL tarafından seçilen parçaların FSR modülü ile iyileştirilmiş halini temsil etmektedir. Örneğin:

Gözler: Giriş görüntüsüne kıyasla göz çevresindeki detaylar çok daha net bir şekilde elde edilmiştir. Göz bebeği ve kirpikler belirgin hale gelmiş, bulanıklık azalmıştır.

Burun: Burun çevresindeki ton geçişleri ve kenar hatları daha doğal ve keskin bir görünüm kazanmıştır.

Ağız: Dudakların hatları ve çevresindeki detaylar orijinal yapıya daha yakın bir şekilde modellenmiştir. Ağız kenarlarında giriş görüntüsüne kıyasla önemli bir detay artışı görülmektedir.

Çıkış görüntüsü, hem yerel detayların iyileştirilmesi hem de genel yüz yapısının korunması açısından başarılı bir sonuç göstermektedir.

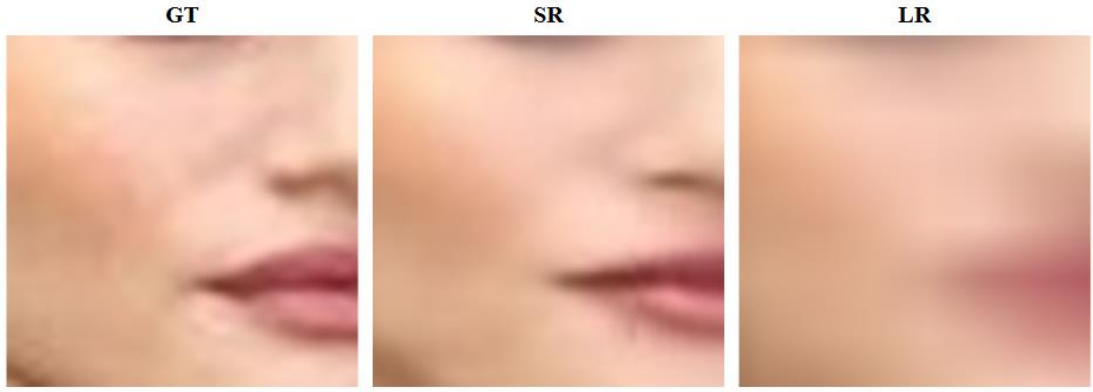
Şekil 5.9'da orijinal görüntüde kırmızı dikdörtgenlerle işaretlenmiş alanlar, giriş ve çıkış görüntüleri ile karşılaştırma için referans olarak kullanılmaktadır. Orijinal görüntüde gözler, burun ve ağız gibi bölgelerde tüm detaylar belirgin ve net bir şekilde görülebilmektedir. Çıkış görüntüsü ile karşılaştırıldığında önerilen modelin bu detaylara oldukça yakın sonuçlar ürettiği gözlemlenmektedir.

Örneğin:

Göz çevresindeki ince çizgiler ve göz bebeği detayları, model çıktısında orijinal ile benzer bir kaliteye sahiptir.

Dudaklar ve burun kenarlarındaki doku yapıları korunmuş, giriş görüntüsüne göre algısal olarak çok daha yakın bir sonuç elde edilmiştir.

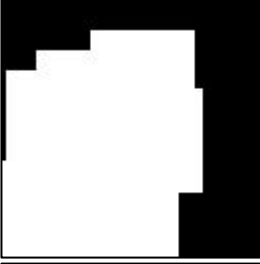
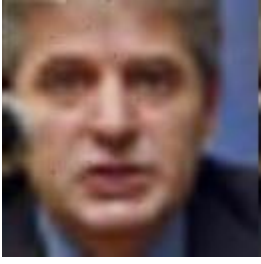
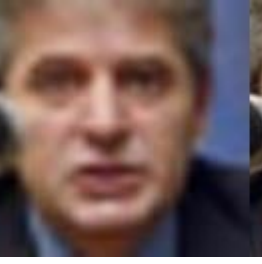
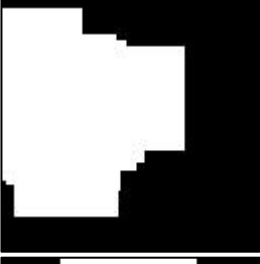
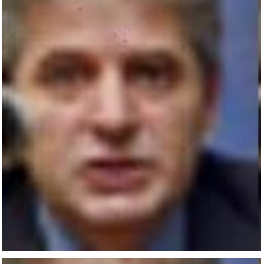
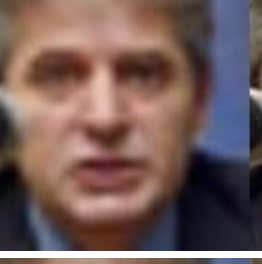
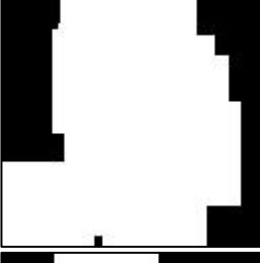
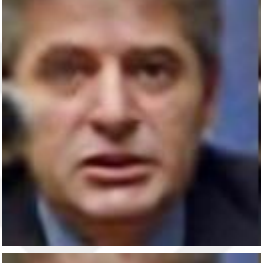
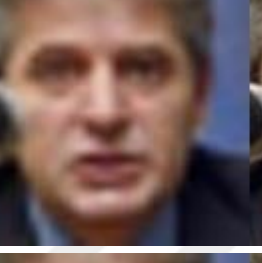
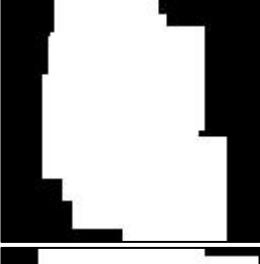
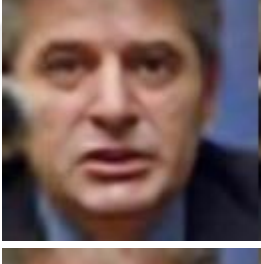
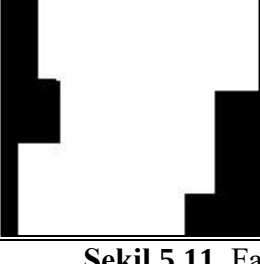
Şekil 5.9'da dikkat çeken bir diğer önemli husus, özellikle 11 ve 12 numaralı parçaların düşük çözünürlüklü (LR) versiyonlarında neredeyse ayırt edilemez derecede bulanık olmasıdır. İlgili parçalar, giriş görüntüsünde birbirine oldukça yakın bölgelerden seçilmiş olmasına rağmen önerilen model belirsizliklerin üstesinden gelerek her bir parçayı doğru bir şekilde ayırt edebilmektedir. Bu durum, modelin iteratif seçim mekanizmasının ve öğrenme sürecinin başarısını ortaya koymaktadır. SR çıktısında, seçilen parçalar daha net bir şekilde oluşturulmuş ve göz, burun ya da ağız gibi kritik bölgelerin detayları, hedef görüntüye oldukça yakın bir şekilde yeniden yapılandırılmıştır. Modelin görüntüdeki belirsizlikleri etkili bir şekilde işleyebilmesi, dikkat mekanizmasının ve DRL tabanlı parça seçiminin üstünlüğünü göstermekte ve hem yerel detaylarda hem de genel yüz bütünlüğünde yüksek kaliteli çıktılar sunmasını sağlamaktadır. Elde edilen başarı, önerilen modelin karmaşık ve düşük çözünürlüklü girişlerden dahi anlamlı detaylar çıkarma yeteneğinin bir kanıtıdır.



Şekil 5.10. Seçilen bir parçanın orijinal, düşük çözünürlük ve model çıktısı halleri.

Şekil 5.10’da, DRL-SRFR modelinin düşük çözünürlüklü (LR), süper çözünürlük uygulanmış (SR) ve hedef yüksek çözünürlüklü (GT) parçalar üzerindeki etkisi büyütülmüş olarak gösterilmektedir. Görseller, modelin kritik bölgelerdeki detay iyileştirme performansını ve görsel kaliteyi artırma yeteneğini ortaya koymaktadır. Düşük çözünürlüklü (8x - LR) görüntüde, ciddi derecede bulanıklık ve detay kaybı gözlemlenmektedir. Özellikle dudakların kenar hatları silik olup, dolgunluk ve kıvrımlar gibi önemli detaylar tamamen kaybolmuştur. Ayrıca cilt yüzeyinde homojen olmayan renk geçişleri ve doğal olmayan ton farklılıkları dikkat çekmektedir. Farklılıklar düşük çözünürlüklü giriş görüntüsünün orijinal detaylara olan uzaklığını net bir şekilde göstermekte ve süper çözünürlük sürecinin zorluğunu ortaya koymaktadır. SR görüntüde ise giriş görüntüsüne kıyasla belirgin bir iyileşme sağlanmaktadır. Dudak kenarları daha keskin ve belirgin hale gelirken, dudakların dolgunluğu ve kıvrımları geri kazandırılmıştır. Cilt tonlarındaki geçişler daha akıcı hale gelmiş ve genel görünüm doğal bir form kazanmıştır. Ancak SR görüntüde bazı ince detayların, örneğin dudak kenarındaki gölge derinliklerinin, hedef GT görüntüye kıyasla tam anlamıyla yeniden oluşturulamadığı fark edilmektedir. Yaşanılan zorluk, modelin yerel detayların bir kısmını üretirken sınırlı bir performansa sahip olduğunu göstermektedir.

Şekil 5.10’daki görüntüde dudaklar doğal bir dolgunluğa, belirgin kenar hatlarına ve akıcı renk geçişlerine sahiptir. DRL-SRFR modelinin LR görüntüsünden SR çıktısını üretme sürecinde GT’ye görsel açıdan oldukça yakın bir sonuç elde ettiği görülmektedir. Sonuçlar, DRL-SRFR’nin yerel ve küresel iyileştirme kabiliyeti açısından güçlü bir yöntem olduğunu, ancak daha ileri geliştirmeler için potansiyel taşıdığını da göstermektedir.

Harita	SR	LR	GT	Epok
				3
				6
				60
				100
				120

Şekil 5.11. Farklı eğitim epoklarında görsel çıktı ve parça dağılımı (8x).

Şekil 5.11’de, DRL-SRFR modelinin iteratif olarak nasıl çalıştığı gösterilmektedir. Sol sütunda, modelin farklı adımlarda (epoklarda) seçtiği parçalar gösterilmektedir. Epok 3’te, SR görüntüsü hala bulanık ve detaylardan yoksunken, epok 6 ve 60 arasında belirgin bir gelişim gözlenmekte, yüz hatları ve göz ve burun çevresinde detaylar daha net hale gelmektedir. Epok 100 ve 120’de, modelin neredeyse GT’ye yakın sonuçlar ürettiği görülmektedir. Ayrıca SR sütunundaki görüntülerde özellikle göz, ağız ve yüz hatlarının belirginleştiği fark edilmektedir. Özellikle 8x ölçekleme faktörü, model için oldukça zor bir süper çözünürlük görevini ifade etmektedir.

Çünkü bu seviyedeki bir ölçeklemede kaybolan ince detayların yeniden üretilmesi çok daha zordur. Ancak model, iteratif parça seçimi stratejisi sayesinde hem yerel detayları iyileştirmekte hem de küresel yüz bütünlüğünü koruyabilmektedir.

Harita	SR	LR	GT	Epok
				3
				6
				60
				100
				120

Şekil 5.12. Farklı eğitim epoklarında görsel çıktı ve parça dağılımı (4x).

Şekil 5.12’de, önerilen modelin düşük çözünürlüklü gri tonlamalı yüz görüntülerini süper çözünürlüklü hale getirirken izlediği aşamaları ve eğitim sürecindeki gelişimi gösterilmektedir. Şekilde modelin farklı epoklarda nasıl iyileştirmeler yaptığı görselleştirilmiştir. Başlangıçta geniş ve düzensiz seçimler yapılırken ilerleyen epoklarda modelin seçimleri daha stratejik ve detay odaklı hale gelmektedir. İlk

epoklarda görüntüler bulanık ve düşük detaylıken eğitim ilerledikçe model yüz hatlarını daha iyi öğrenerek göz, burun ve ağız gibi önemli detayları daha keskin şekilde üretmektedir.

Genel Değerlendirme:

Önerilen DRL-SRFR modeli, iteratif bir süreçle yüzün kritik bölgelerindeki detayları iyileştirmektedir. Aynı zamanda genel yüz yapısındaki algısal bütünlük sağlanmaktadır. Görsellerde görüldüğü üzere giriş görüntüsünde bulanık ve silik halde bulunan detaylar, çıkış görüntüsünde yüksek çözünürlük ve netlik kazanmaktadır. Gözlemler yöntemin yerel detayları iyileştirme ve global yapıyı koruma konusundaki başarısını ortaya koymaktadır. Model özellikle göz, burun ve ağız gibi yüzün önemli bölgelerinde daha yüksek çözünürlük ve daha gerçekçi detaylar sunarak, hem görsel hem de nicel metriklerde üstün bir performans sergilemektedir.

5.3.3. Ablasyon çalışmaları

İlerleyen sayfalarda çeşitli ablasyon çalışmaları ile farklı önemli bileşenlerin etkileri değerlendirilmektedir.

- Parça Sayılarının Etkisi

İlk olarak sabit bir parça boyutunda farklı iterasyon adımlar (T) kullanmanın etkisi araştırıldı. Model toplam parametre sayısı sabit tutularak dört farklı ayarla ($T = [5, 15, 25, 35]$) eğitilmiştir. Tablo 5.6, model performansının T arttıkça önce iyileştiğini ancak belirli bir seviyeden sonra azaldığını göstermektedir. Çıkarılan parçalar görüntünün tamamını kapsayamadığından düşük iterasyon adımı sayısı düşük PSNR değerlerine neden olmaktadır. Deneylerden parçaların tüm görüntüyü kaplaması için iterasyon adım sayısının en az 10 olması gerektiği bulunmuştur. T arttıkça PSNR değerlerindeki artış sınırlanmaktadır. $T=15$ değeri PSNR, hız ve hesaplama yükü açısından en uygun değer olarak bulunmuştur.

Tablo 5.6. T 'nin farklı değerlerinin 4 ve 8 ölçekleme faktöründe karşılaştırılması.

	LFW 4 x	LFW 8 x
T=5	28.96	23.62
T=15	34.88	28.53
T=25	35.92	29.22
T=35	34.12	28.01

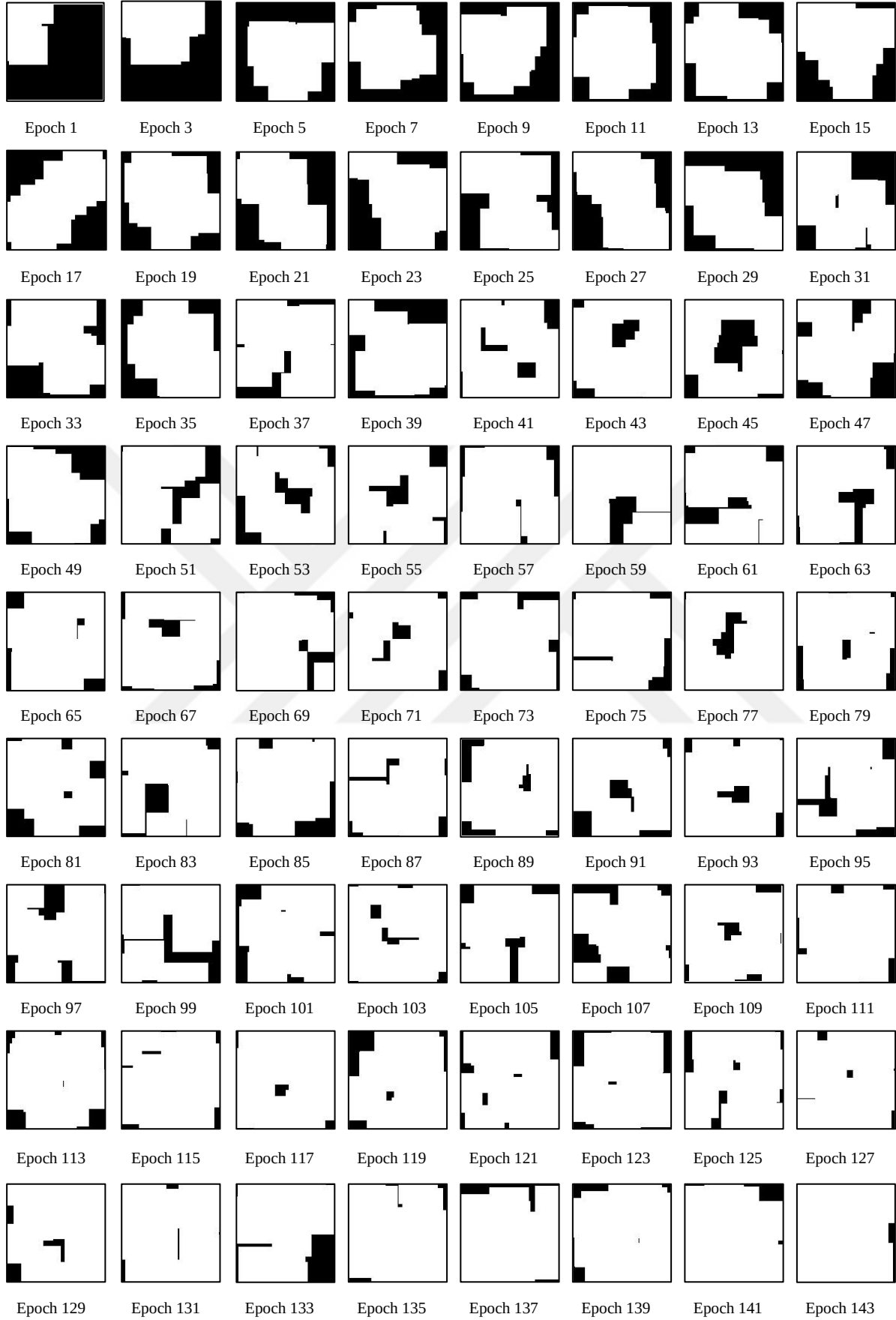
- Küresel Geliştirme Parametresinin (E_{pc}) Etkisi

Karar vericinin DRL-SRFR modeli ile çalışma prensibi, görüntü haritası üzerinde görselleştirilmektedir. Şekil 5.13'teki görseller modelin gelişim sürecini adım adım göstermektedir. DRL-SRFR modeli, öncelikle görüntünün belirli alanlarına odaklanarak yerel detayların iyileştirilmesini sağlamaktadır. Süreç boyunca model, karar verici yapısı sayesinde hangi alanların daha kritik olduğunu öğrenmekte ve iyileştirmeleri bu alanlardan başlatarak kademeli bir şekilde tüm görüntüye yaymaktadır.

Modelin aşamalı iyileştirme süreci görüntüdeki farklı alanların birbirinden bağımsız olarak işlenmesini sağlamaktadır. Örneğin görüntünün köşeleri genellikle düz bir arka plan içermekte ve bu bölgelerde ilgi çekici kişisel özellikler bulunmamaktadır. Dolayısıyla bu tür bölgelerin geliştirilmesi model için daha az karmaşıktır ve erken aşamalarda hızlı bir şekilde tamamlanabilmektedir. Bununla birlikte yüz bileşenleri (kulaklar, gözler, burun ve ağız gibi) daha karmaşık yapılar içerir ve bu bölgeler için daha ayrıntılı bir iyileştirme gereklidir. Yerel ve global süreç modelin bu bölgelerdeki özellikleri daha iyi tanımasını ve algısal iyileştirme sağlamasını mümkün kılmaktadır.

Modelin geliştirme sürecinde kullandığı parametrelerden biri olan küresel geliştirme parametresi (E_{pc}), yerel ve global piksellerin oranını belirleyerek modelin öğrenme sürecine yön vermektedir. E_{pc} değeri (0,1) aralığında bir parametre olup, harita üzerindeki kapsama alanını genişletmek veya daraltmak için kullanılmaktadır. Şekil 5.13'te gösterildiği gibi E_{pc} 'nin düşük olduğu erken aşamalarda, model öncelikle görüntünün küçük bir kısmına odaklanır ve bu bölgeleri iyileştirmektedir. Ancak E_{pc} arttıkça model geniş bir alanı kapsayarak küresel düzeyde daha kapsamlı bir iyileştirme sağlamaktadır.

Süreç, modelin yerel detaylardan global düzeye geçişini doğal bir şekilde düzenlemektedir. Erken aşamalarda odaklanan küçük alanlar belirli bir bölgenin özelliklerini detaylıca öğrenmeyi sağlarken, ilerleyen aşamalarda tüm görüntünün iyileştirilmesi hedeflenmektedir.



Şekil 5.13. İyileştirme aşamasındaki epoklara bağlı olarak seçilen parça haritası.

E_{pc} değeriindeki artış, haritanın daha geniş bir kapsama alanına sahip olduğunu göstermektedir. E_{pc} artışı, yalnızca yerel detaylardan elde edilen ödülleri optimize etmekle kalmaz aynı zamanda global iyileştirme sürecine de katkıda bulunmaktadır. E_{pc} 'nin yüksek olduğu durumlarda döngü sonunda elde edilen ödüllerin global iyileştirme için kullanılması tüm görüntünün kalitesini artırmaktadır. Böylelikle yerel detayların dikkatli bir şekilde öğrenilmesi ve bu detayların küresel iyileştirme ile birleştirilmesi yüz görüntülerinin hem algısal bütünlüğünü hem de kişisel özelliklerin doğru bir şekilde korunmasını sağlamaktadır.

- Parça Seçiminin Etkisi

Politika ağı bir dizi parçayı uyarlamalı olarak seçebilir. Parça seçim stratejilerinin etkisini değerlendirmek için Şekil 5.5'te farklı yöntemler karşılaştırılmaktadır. Karşılaştırma, standart çözünürlük iyileştirme yöntemleri ile parça seçimi arasında önemli farklar olduğunu göstermektedir. DRL ile daha iyi parça seçimi performansının iki nedeni vardır. İlk olarak, standart yöntemlerde deforme olabilen yüz parçaları (örneğin, Şekil 5.5 (c) ve (d)'de gösterilen burun veya ağız) gözlemlenebilir çünkü tüm görüntü veya kullanılan parçalar bozuk yüz parçaları (örneğin, çarpık burun, yanlış hizalanmış kaşlar, ayrık gözler) içermektedir. İkinci olarak, DRL kullanılarak seçilen parçalar birbirine yakın olan parçalar arasındaki yüzey ilişkilerini keşfetme yeteneğine sahiptir. Sağlam yüzey parçalarının seçilmesine yardımcı olur ve yerel iyileştirme ağını geliştirerek kimlik bilgilerinin kurtarılmasını sağlamaktadır.

- Parça Boyutunun Etkisi

Parça seçimi SR yöntemlerinin temel stratejilerinden biridir. Parçanın çözünürlük artışı üzerindeki etkisi parça boyutuna ve özelliklerine bağlıdır. Küçük parça boyutu ayrıntıların korunmasına olanak tanımaktadır. Ancak küçük parça boyutu daha fazla hesaplama gerektirir ve işlem süresini olumsuz etkileyebilmektedir. Öte yandan büyük bir parça boyutu işlem süresini azaltır. Ancak ayrıntıları korumayı zorlaştırır. Çünkü büyük bir parça boyutu görüntünün genel özelliklerine odaklanır ve görüntü hakkında genel bir fikir vermektedir. Parça boyutu seçimi ayrıntıların korunması ve işlem süresi arasında bir dengedir. İdeal parça boyutu hedeflenen çözünürlük, mevcut kaynaklar ve uygulama gereksinimleri göz önünde bulundurularak belirlenmelidir.

Ölçekleme faktörü 4 için parça boyutunun bir fonksiyonu olarak çözünürlükteki iyileşme Tablo 5.7’de verilmektedir.

Tablo 5.7. Parça boyutu karşılaştırması.

Parça boyutu	PSNR	SSIM
32 x 32	31.33	0.9025
52 x 52	34.88	0.9691
64 x 64	32.71	0.9187

- RRDB Blok Sayısının Etkisi

Önerilen çalışmada RRDB yapısındaki blok sayısı değiştirilerek farklı modeller oluşturulmuştur. Başlangıçta 8 bloklu temel bir RRDM modeli vardır. Temel blok uyarlanarak biri 12 bloklu diğeri 4 bloklu modeller türetilmiştir. Tüm modeller performans ölçütleri kullanılarak karşılaştırılmış ve sonuçlar Tablo 5.8’de sunulmaktadır. RRDB bloklarının sayısının artırılması belirli değerden sonra SR performansını önemli bir şekilde etkilememektedir. 12 bloklu yapının 4 bloklu yapıya kıyasla önemli PSNR ve SSIM iyileşmesi sağladığı ancak 8 bloklu yapıya kıyasla ilgili metriklerde marjinal iyileşme sağladığı görülmektedir. RRDB yapısındaki blok sayısı arttıkça belirli bir eşikten sonra özellik çıkarımındaki azalma nedeniyle performanstaki iyileşmenin sınırlı olduğu sonucuna varılmıştır.

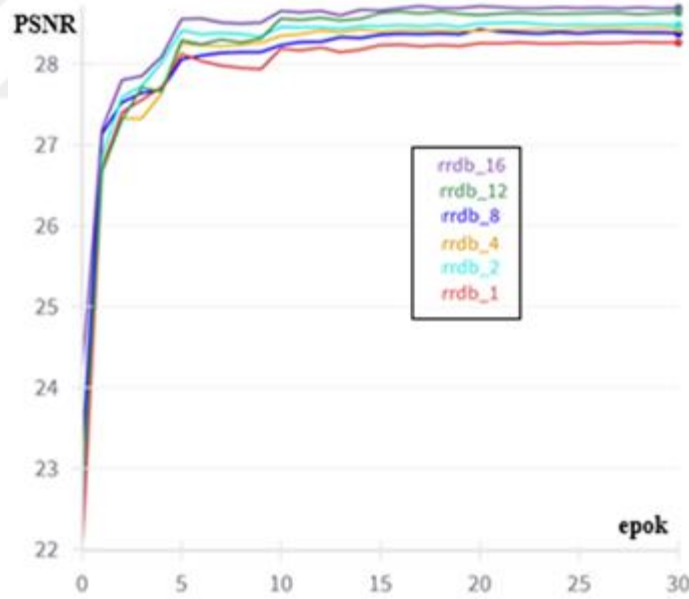
Tablo 5.8. Farklı sayıda RRDB katmanının karşılaştırılması(Ölçekleme 4x).

Blok sayısı	PSNR	SSIM	zaman(msn)
4	31.09	0.8820	0,114
8	34.45	0.9688	0,158
12	34.89	0.9710	0,213

Bulgular, SR uygulamalarında RRDB yapısının etkinliği konusunda önemli bir bakış açısı sağlamaktadır. Daha fazla blok eklemek başlangıçta modelin özellik temsilini artırmaktadır. Ancak temel özellikler önceki katmanlar tarafından zaten yakalandığı

için belirli bir noktadan sonra kazanımlar ihmal edilebilir hale gelmektedir. Gözlem, azalan getiri ilkesiyle tutarlıdır [104]. Sonuçlardan model yeterince derin olduğunda ek katmanların marjinal iyileştirmeler sağladığı açıktır. Ayrıca daha fazla blok kullanmak parametre sayısını artıracak, bu da modelin eğitim verilerine aşırı uyum sağlamasına ve dolayısıyla yeni verilere genelleme yapamamasına neden olabilmektedir. ESRGAN 8x ölçekleme faktörü için uygulandığında aşırı uyum sorunu gözlemlenmektedir. Bir diğer sorun ise ağ derinliği arttıkça geri yayılım sırasında gradyanların giderek aşırı büyümesidir. Bahsedilen sorun hem ESRGAN yönteminde hem de önerilen yöntemde yaşanmaktadır. Sorunu aşmak için gradyan kırpması kullanılsa bile belirli bir blok sayısından sonra kazanç çok küçük olmaktadır.

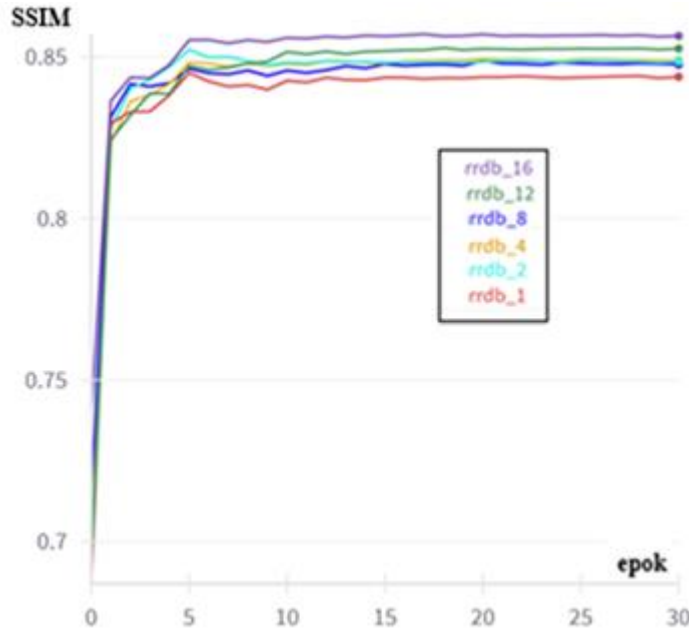
Şekil 5.14 ve Şekil 5.15, farklı RRDB blok sayıları ile oluşturulmuş modellerin eğitim sürecindeki ortalama PSNR ve SSIM değişimlerini göstermektedir. Grafikler, RRDB bloklarının sayısının performansa olan etkisini hem erken aşamalarda hem de modelin oturmuş durumunda incelemektedir. İncelemede, RRDB katman sayısı harici parametreler sabit tutularak sadece RRDB değişimi etkisi analiz edilmektedir.



Şekil 5.14. Eğitim anında RRDB katman sayısına bağlı ortalama PSNR değerleri.

Eğitim sürecinin ilk aşamalarında tüm modellerde hızlı bir performans artışı gözlemlenmektedir. PSNR açısından Şekil 5.14 incelendiğinde, RRDB sayısı 1 olan model erken aşamalarda diğer modellere kıyasla en düşük performansı sergilemektedir. PSNR değeri sınırlı bir artış göstermiş ve genel olarak daha düşük seviyelerde kalmıştır. Öte yandan, RRDB sayısı 8 ve 4 olan modeller, başlangıç

aşamasında hızlı bir şekilde yüksek PSNR değerlerine ulaşmaktadır. RRDB 2, 12, ve 16 modelleri ise bu ikisini takip etmiştir. Gözlemlenen durum, başlangıçta daha az sayıda RRDB bloklu modellerin daha derin modellere kıyasla daha hızlı öğrenme eğrisi sergilediğini göstermektedir.



Şekil 5.15. Eğitim anında RRDB katman sayısına bağlı ortalama SSIM değerleri.

SSIM açısından Şekil 5.15 incelendiğinde, RRDB 1 modeli yine en düşük SSIM değerini göstermektedir. RRDB 8 ve 4 modelleri, hızlı bir şekilde yüksek SSIM değerlerine ulaşmaktadır. RRDB 12, 16, ve 2 modelleri gözlemlenen davranışı bir miktar geriden takip etmiştir. Durum, daha az RRDB bloklu yapıların erken öğrenme aşamalarında yapısal detayları daha hızlı kavradığını göstermektedir.

Eğitim sürecinin orta aşamalarında tüm modellerde PSNR ve SSIM değerlerinde kararlılık sağlanmaya başlansa da RRDB blok sayısı performans açısından belirleyici olmaktadır. PSNR açısından incelendiğinde RRDB sayısı 8 ve 12 olan modeller, süreçte daha iyi performans sergileyerek diğer modellerden ayrılmıştır. RRDB 16 modeli, PSNR açısından hafif bir iyileşme sunmuş olsa da 8 ve 12 modellerine kıyasla kazanç istenilen seviyede artış gösterememektedir. RRDB 4 ve 2 modelleri, performans olarak daha dengeli bir yapıya sahip olmuş ve RRDB 1 modeli en düşük seviyede kalmıştır. SSIM açısından incelendiğinde RRDB 12 ve 16 modelleri, yapısal benzerlik açısından en yüksek değerleri üretmektedir. Özellikle RRDB 8 ve 4

modelleri bu iki yapılandırmayı takip etmiş, ancak RRDB 1 ve 2 modelleri SSIM açısından daha düşük değerlere ulaşmaktadır.

Eğitim sürecinin oturmuş durumunda modellerin performansları sabit bir noktada kararlı hale gelmektedir. PSNR açısından Şekil 5.14 incelendiğinde, RRDB 12 ve 16 modelleri en yüksek PSNR değerlerine ulaşmıştır. RRDB 8 modeli, bu iki yapılandırmayı çok az bir farkla takip ederek dengeli bir çözüm sunmaktadır. RRDB 4 ve 2 modelleri ise daha düşük PSNR değerleriyle sırayı almış, RRDB 1 modeli ise en düşük performansta kalmıştır. Tüm gözlemler düşünüldüğünde RRDB blok sayısındaki artışın belirli bir noktadan sonra azalan getiriler sunduğunu göstermektedir. Çünkü her RRDB katmanı ciddi hesaplama yükü doğurmaktadır. İncelemelerde özellikle 12 bloklu model hem yüksek PSNR değeri hem de hesaplama verimliliği açısından en optimal çözümü sunmaktadır.

SSIM açısından Şekil 5.15 incelendiğinde RRDB 12 ve 16 modelleri en iyi SSIM performansını sergilemektedir. Özellikle 12 bloklu yapı, hem oturmuş durumda hem de diğer aşamalarda daha tutarlı bir performans göstermektedir. RRDB 8, 4, ve 2 modelleri bu yapılandırmaları takip ederken, RRDB 1 modeli yapısal detayların korunması açısından en düşük performansı sergilemektedir.

Elde edilen sonuçlar, RRDB blok sayısındaki artışın hem PSNR hem de SSIM performansını artırdığını ancak bu artışın 12 blok seviyesinden itibaren sınırlı hale geldiğini göstermektedir. RRDB 12 modeli, PSNR ve SSIM açısından en dengeli ve yüksek performanslı yapı olarak öne çıkmaktadır. RRDB 16 modeline kıyasla hesaplama açısından daha verimli bir çözüm sunmuştur. RRDB 8 modeli ise hem erken hem de oturmuş durumda rekabetçi bir alternatif olmaktadır. RRDB 1 modeli, hem piksel seviyesinde (PSNR) hem de yapısal doğruluk (SSIM) açısından diğer RRDB sayısındaki modellere kıyasla belirgin şekilde düşük performans göstermektedir.

5.4. Optimum Hiperparametrelerin Ayarlanması

DRL-SRFR yönteminde bir görüntünün iyileştirilmesi sürecindeki parça sayısı T , RRDB sayısı, parça boyutu ve pekiştirmeli öğrenme yöntemindeki standart sapma olmak üzere dört hiperparametre bulunmaktadır. Başlık altında PSNR değerini maksimize etmek için T , RRDB sayısı ve σ standart sapmasının belirlenmesi tartışılmaktadır. Parametrelerin en iyi değerleri Ağaç Yapılı Parzen Tahmin Edici

(TPE) yöntemi kullanılarak belirlenmiştir. TPE, Bayes tabanlı bir optimizasyon yöntemidir. Tasarımdaki istek, verilen bir parametre kombinasyonunun PSNR'yi maksimize etmesidir. TPE'de olasılıklar ilk olarak rastgele başlangıç değerleriyle belirlenmektedir. Başlangıç değerleri modelin performansı hakkında bir ön fikir sağlamaktadır. Ardından her yeni denemeden sonra elde edilen sonuçlar kanıt olarak kullanılarak, belirli bir hiperparametre kombinasyonunun gözlenen PSNR'ye göre olasılığı hesaplanmaktadır. Son olarak güncellenen olabilirlik doğrultusunda yeni parametre kombinasyonları denenmektedir. Çalışmada en iyi %10 PSNR'ye karşılık gelen eşik değerleri belirlenerek olasılık dağılımları oluşturulmaktadır.

Ağaç Yapılı Parzen Tahmin Edici (TPE) yöntemi hiperparametre optimizasyonunda kullanılan etkili ve popüler bir yaklaşımdır. Özellikle karmaşık ve yüksek boyutlu optimizasyon problemlerinde başarıyla uygulanmaktadır. TPE yöntemi, olasılıksal bir modelleme yaparak hiperparametre arama uzayını verimli bir şekilde keşfetmeyi hedeflemektedir. TPE, belirli bir hedef fonksiyonun minimum ya da maksimum değerini bulmak için uygun hiperparametre kombinasyonlarını seçmektedir.

TPE, Bayes optimizasyonunun bir türüdür ve Gaussian Süreçler (GP) yerine, hiperparametrelerin olasılık dağılımlarını modellemek için Parzen tahmincilerini kullanmaktadır. Yöntem, aday hiperparametrelerin olasılıklarını şu şekilde bölümlendirmektedir:

- Hedef Fonksiyonu ve Likelihood Modelleri: Hedef fonksiyonun değerleri, bir eşik değerine (Y) göre iki gruba ayrılır:
 - $l(x)$: Performansı Y değerinden daha iyi olan hiperparametrelerin yoğunluk fonksiyonu.
 - $g(x)$: Performansı Y değerinden daha kötü olan hiperparametrelerin yoğunluk fonksiyonu.
- Bayesian Seçim Stratejisi: TPE, yeni hiperparametre örneklerini üretmek için, $P(y|x)$ olasılığını maksimize ederken, $l(x)$ ve $g(x)$ modellerini kullanır. Burada y , hedef fonksiyonun çıktısını temsil etmektedir.
- Ağaç Yapılı Yapı: Parzen tahmin edicilerinin ağaç tabanlı yapılara entegrasyonu, karmaşık uzaylarda daha detaylı bir optimizasyon sağlamaktadır.

TPE yönteminin teknik detayları matematiksel olarak şu şekilde ifade edilebilir:

Bayes Kuralı ve Yoğunluk Fonksiyonları

TPE, Bayes Kuralı'na dayanarak hiperparametrelerin seçim olasılığını modellemeye çalışmaktadır. İlgili formül denklem 5.3'te verilmiştir.

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (5.3)$$

Burada:

- $P(x|y)$: Hiperparametre x 'in, hedef fonksiyon çıktısı y 'ye göre olasılığı.
- $P(y|x)$: Hedef fonksiyon y 'nin x 'e bağlı olasılığı.
- $P(x)$: Hiperparametrelerin öncelik dağılımı (prior).

İki Yoğunluk Fonksiyonu

Performans eşik değeri (Y) kullanılarak, $l(x)$ ve $g(x)$ yoğunluk fonksiyonları denklem 5.4 ve 5.5'te tanımlanmaktadır:

$$P(x|y < Y) \propto l(x) \quad (5.4)$$

$$P(x|y \geq Y) \propto g(x) \quad (5.5)$$

Burada:

- $l(x)$: İyi performans gösteren hiperparametrelerin yoğunluk fonksiyonu.
- $g(x)$: Kötü performans gösteren hiperparametrelerin yoğunluk fonksiyonu.

TPE, iyi performans gösteren hiperparametreleri ($l(x)$) seçmek için bu yoğunluk fonksiyonlarını kullanmaktadır.

Kayıp Fonksiyonunun Maksimizasyonu

TPE, hiperparametrelerin seçiminde bir edinim fonksiyonunu maksimize etmeyi hedefler. Bu fonksiyon denklem 5.6'daki gibi tanımlanabilmektedir:

$$x^* = \operatorname{argmax} \frac{l(x)}{g(x)} \quad (5.6)$$

Oran, iyi hiperparametre değerlerini seçme olasılığını maksimize etmektedir.

Parzen Tahmin Edici

Parzen tahmin edicisi, yoğunluk fonksiyonlarını tahmin etmek için kullanılan bir yöntemdir. Olasılık yoğunluğu denklem 5.7'deki gibi hesaplanmaktadır.

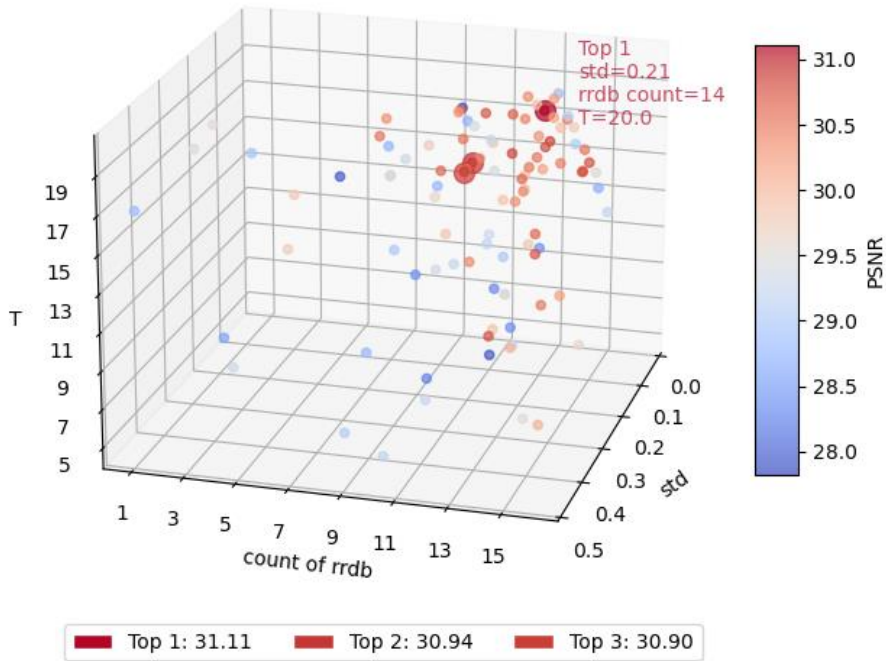
$$l(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (5.7)$$

Burada:

- K : Çekirdek (kernel) fonksiyonu.
- h : Bant genişliği (bandwidth).
- x_i : Gözlem verileri.

Ağaç Yapısı

Ağaç yapısı, hiperparametre uzayının daha iyi bir şekilde keşfedilmesine olanak tanımaktadır. Hiperparametrelerin farklı kombinasyonlarının performansı, hiyerarşik olarak modellenir ve ağaç yapısında optimize edilmektedir. TPE, hiperparametre optimizasyonunda hem verimlilik hem de esneklik sunmaktadır. Yöntem, hedef fonksiyonun olasılık dağılımını modelleyerek iyi performans gösteren hiperparametreleri seçmek için bir yol sağlamaktadır.



Şekil 5.16. Hiperparametre değerlerine göre PSNR performans grafiği.

Önerilen çalışmada TPE'nin uygulanmasında parametreleri belirlemek için LFW ve Celeba veri kümelerinden rastgele 3000 eğitim ve 600 test verisi seçilmiştir. En iyi sonuç $T=20$, RRDB sayısı 14 ve $\sigma=0.21$ değerleri için elde edilmektedir.

Şekil 5.16, modelin 100 kez çalıştırılmasından elde edilen sonuçları göstermektedir. Her yürütme 100 epoktan oluşur ve parametrelerin tahmin edilmesini içermektedir. Parametre tahmin aralıkları σ için (0.1, 0.5), RRDB sayısı için (1,16) ve T için (5,20)'dir. Grafikten de görülebileceği gibi, en iyi sonuçlar yeterli sayıda parçanın optimize edilmesi ve RRDB bloğunun kullanılması durumunda elde edilmektedir. σ değeri, daha önce hiç denenmemiş bir karar vermek ile önceki deneyimlerden bulunan en iyi eylemi yapmak arasındaki dengeyi belirlemektedir.

En iyi sonuç $T = 20$, RRDB sayısı 14 ve $\sigma = 0,21$ ile elde edilmiştir. Ancak, $\sigma = 0.37$, $T = 20$ ve $rrdb = 10$ değerleri ile de tatmin edici sonuçlar elde edilmektedir. Önerilen yöntemde RRDB ve T 'nin etkisi göz ardı edilemez. T ve RRDB'nin küçük değerleri düşük PSNR değerleri üretmektedir. T (14,20) aralığında ve RRDB (6,12) aralığında olduğunda PSNR değeri en iyi sonuca çok yakındır. Şekil 5.16'da bahsi geçen aralıklardaki bölgelerde yoğunlaşma görülmektedir.

5.5. Yöntemlerin Karşılaştırılması

Önerilen DRL-SRFR modeli, literatürde sıkça kullanılan yöntemler ile kapsamlı bir şekilde karşılaştırarak modelin performansını değerlendirmek için derinlemesine bir analiz yapılmaktadır. Karşılaştırmalarda hem genel süper çözünürlük yöntemleri (SRCNN, VDSR gibi) hem de yüz halüsinasyonu için özelleştirilmiş yöntemler (SRGAN, ESRGAN, SwinIR, IPFH, Attention-FH gibi) dikkate alınmaktadır. Yapılan analizler yöntemin hem görsel kalite hem de matematiksel doğruluk açısından literatüre yaptığı katkıyı net bir şekilde ortaya koymayı amaçlamaktadır.

5.5.1. Genel değerlendirme

DRL-SRFR modeli çeşitli veri kümelerinde (BioID, CelebA, LFW, PubFig) test edilmiş ve tüm yöntemler ortak performans ölçütleri olan PSNR ve SSIM ile değerlendirilmiştir. Modelin üstün performansı görsel kalite ve matematiksel başarı yönlerinden öne çıkmaktadır.

Görsel Kalite: DRL-SRFR, insan görsel algısına uygun şekilde daha doğal ve detaylı yüz görüntüleri üretmektedir. Özellikle göz, burun ve dudak gibi kritik detayların korunmasında üstün başarı sağlamaktadır.

Matematiksel Başarı: PSNR ve SSIM değerlerinde özellikle 8x ölçeklendirme faktörlerinde rakip yöntemlere göre belirgin bir üstünlük elde edilmektedir.

5.5.2. Yöntemler arası farklılıklar

Genel Süper Çözünürlük Yöntemleri:

SRCNN ve VDSR, genel görüntü süper çözünürlüğü için tasarlanmış olmasına rağmen yüz halüsinasyonu görevlerinde DRL-SRFR'ye kıyasla daha düşük performans sergilemiştir. Dolayısıyla sonuçlar yöntemlerin yüz özelliklerine özgü detayları korumakta yetersiz kaldığını göstermektedir.

GAN Tabanlı Yöntemler:

SRGAN ve ESRGAN, dokusal detayların iyileştirilmesinde etkili olmuş ancak küresel yüz yapısını korumada DRL-SRFR'nin gerisinde kalmıştır. Özellikle bu yöntemlerin PSNR ve SSIM gibi metriklerde daha düşük performans gösterdiği gözlemlenmektedir.

Dönüştürücü Tabanlı ve Difüzyon Yöntemleri:

SR3 modeli:

SR3 [79], difüzyon tabanlı bir yöntem olarak görsel algıyı geliştirmiş ancak işlem süresinin uzunluğu ve hesaplama maliyetinin yüksekliği nedeniyle pratik kullanımı sınırlı kalmaktadır.

SR3 (Super-Resolution via Repeated Refinement), görüntü süper çözünürlüğü için geliştirilmiş difüzyon gürültü giderme olasılıksal modeller “denoising diffusion probabilistic models (DDPM)” temel alan bir derin öğrenme modelidir. SR3, iteratif bir süreçle düşük çözünürlüklü görüntülerin kalitesini artırmaktadır. Model, gürültü ekleme (difüzyon) ve gürültü kaldırma (ters difüzyon) süreçlerini kullanarak düşük çözünürlüklü bir görüntüyü adım adım yüksek çözünürlüklü hale getirmektedir. Eğitim sürecinde düşük çözünürlüklü görüntü bir koşul olarak kullanılır ve model, bu bilgiye dayanarak detaylı bir yüksek çözünürlüklü görüntü üretmektedir.

SR3'ün temel yeniliği yüksek kaliteli görüntüler üretmek için bir difüzyon gürültü giderme olasılıksal modelin gücünü kullanmasıdır. SR3 yaklaşımı geleneksel SR

yöntemlerine kıyasla hem doğruluk hem de görsel kalite açısından üstün sonuçlar sunmaktadır.

SR3 modeli Difüzyon Süreci, Ters Difüzyon Süreci ve Koşullu Üretim olmak üzere üç ana bileşene dayanmaktadır.

1. Difüzyon süreci

Difüzyon süreci, giriş olarak verilen düşük çözünürlüklü bir görüntüye t zaman adımları boyunca Gauss gürültüsü eklenmesiyle gerçekleştirilmektedir. Süreç, veri dağılımını gürültü dağılımına dönüştürür ve denklem 4.8'deki şekli ile tanımlanmaktadır.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)I) \quad (4.8)$$

Burada:

- x_t , t zaman adımındaki görüntüyü,
- α_t , difüzyon sürecinde kullanılan bir sabit azaltma faktörünü
- $\mathcal{N}$, Gauss dağılımını ifade eder.

Süreç sonunda tamamen gürültülü bir görüntü x_T elde edilmektedir.

2. Ters Difüzyon Süreci

Ters difüzyon süreci, tamamen gürültülü bir görüntüden başlayarak iteratif olarak gürültüyü kaldırmayı ve yüksek çözünürlüklü bir görüntü elde etmeyi amaçlamaktadır. Sinir ağı p_θ tarafından modellenen bu süreç denklem 5.9'daki şekli ile ifade edilmektedir.

$$p_\theta(x_{t-1}|x_t, y) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t, y), \sum_\theta(x_t, t)) \quad (5.9)$$

Burada:

- y , düşük çözünürlüklü görüntüyü temsil eder ve koşullu bilgi olarak kullanılır.
- μ_θ , sinir ağı tarafından tahmin edilen beklenen değer,
- $\sum_\theta(x_t, t)$, sinir ağı tarafından tahmin edilen varyanstır.

Ters difüzyon süreci, gürültünün her adımda tahmin edilerek kaldırılmasıyla yüksek çözünürlüklü bir görüntü x_0 elde edilene kadar devam etmektedir.

3. Koşullu Üretim

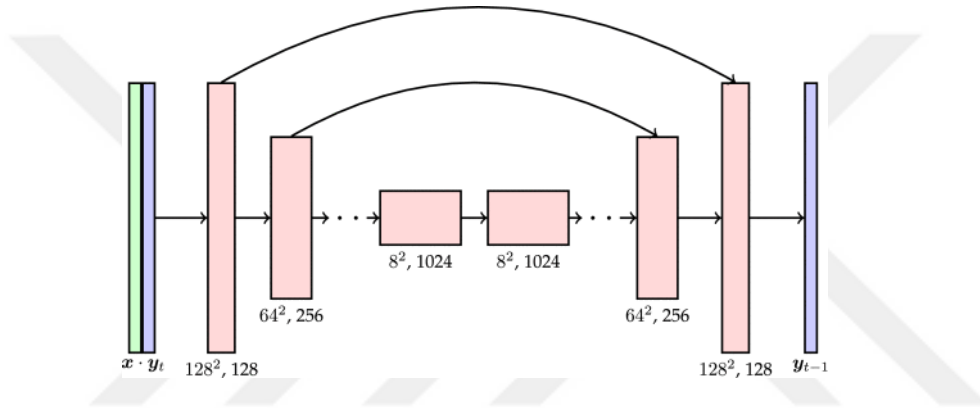
SR3 modeli, düşük çözünürlüklü görüntüyü koşullu bilgi olarak kullanarak yüksek çözünürlüklü görüntüyü üretmektedir. Süreç, denklem 5.10'daki kayıp fonksiyonunu optimize etmektedir.

$$L_{simple} = \mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(x_t, t, y)\|_2^2] \quad (5.10)$$

Burada:

- ϵ , difüzyon sürecinde eklenen gerçek gürültüyü,
- ϵ_θ , sinir ağı tarafından tahmin edilen gürültüyü ifade eder.

Amaç, gerçek gürültü ile tahmin edilen gürültü arasındaki farkı en aza indirmektir.



Şekil 5.17. SR3'ün U-Net mimarisi.

SR3, düşük çözünürlüklü görüntüyü giriş olarak alır ve iteratif bir süreçle yüksek çözünürlüklü bir görüntü üretir. Modelin yapısı, Şekil 5.17'de gösterilmektedir. Düşük çözünürlüklü giriş görüntüsü x , bikübik enterpolasyon kullanılarak hedef çözünürlüğe kadar örneklenir ve gürültülü yüksek çözünürlüklü çıkış görüntüsü y_t ile birleştirilmektedir. $16 \times 16 \rightarrow 128 \times 128$ süper çözünürlük modeli için aktivasyon boyutlarıdır. 16×16 özellik haritaları üzerinde öz dikkat gerçekleştirilmektedir. Bahsedilen yapı, modelin hem gürültü ekleme hem de gürültü kaldırma süreçlerini temsil eder ve SR3'ün iteratif doğasını vurgulamaktadır.

SwinIR modeli:

SwinIR [95], düşük kaliteli görüntülerden yüksek kaliteli görüntüler elde etmeyi amaçlayan bir derin öğrenme modelidir. Geleneksel olarak görüntü restorasyonu için kullanılan yöntemler genellikle konvolüsyonel sinir ağlarına (CNN) dayanırken SwinIR, Swin Transformer mimarisini temel alarak bu alanda yenilikçi bir yaklaşım sunmaktadır. Model, görüntü süper çözünürlüğü, görüntü gürültü giderme ve JPEG

sıkıştırma yapaylıklarının azaltılması gibi çeşitli görevlerde üstün performans sergilemektedir. SwinIR, büyük veri kümelerinde iyi sonuçlar üretmiş ancak küçük veri kümelerinde çalışma kapsamında önerilen model kadar başarılı olmamıştır.

SwinIR, üç ana bileşenden oluşur: sığ özellik çıkarımı, derin özellik çıkarımı ve yüksek kaliteli görüntü yeniden inşası yapılarıdır. Sığ özellik çıkarımı, giriş görüntüsünden temel özellikleri elde etmektedir. Derin özellik çıkarımı, özellikleri daha karmaşık ve yüksek seviyeli temsillere dönüştürmektedir. Son olarak yüksek kaliteli görüntü yeniden inşası, elde edilen derin özellikleri kullanarak nihai çıktıyı üretmektedir. Yapı, modelin hem yerel hem de küresel bağlam bilgilerini etkili bir şekilde işlemesini sağlamaktadır.

SwinIR modelinin metodoloji matematiği:

1. Sığ Özellik Çıkarımı

Giriş olarak verilen düşük kaliteli görüntü I_{LQ} , bir 3×3 konvolüsyon katmanı aracılığıyla sığ özellik haritası F_0 elde edilmektedir. F_0 formülü denklem 5.11’de verilmektedir.

$$F_0 = H_{SF}(I_{LQ}) \quad (5.11)$$

Burada H_{SF} , sığ özellik çıkarım fonksiyonunu temsil etmektedir.

2. Derin Özellik Çıkarımı

Sığ özellik haritası F_0 , K adet Artık Swin Transformer Bloğu (Residual Swin Transformer Block, RSTB) ve bir 3×3 konvolüsyon katmanından oluşan derin özellik çıkarım modülüne girmektedir. Yapı işleyişi denklem 5.12’de verilmektedir.

$$\begin{aligned} F_i &= H_{RSTB_i}(F_{i-1}), \quad i=1,2,\dots,K \\ F_{DF} &= H_{CONV}(F_K) \end{aligned} \quad (5.12)$$

Burada, H_{RSTB_i} i -inci RSTB’yi, H_{CONV} ise son konvolüsyon katmanını temsil etmektedir.

Her bir RSTB, birkaç Swin Transformer katmanı ve bir konvolüsyon katmanı içermektedir. Swin Transformer katmanları yerel dikkat mekanizması ve kaydırılmış pencere (shifted window) şeması kullanarak uzun menzilli bağımlılıkları ve yerel bağlamı etkili bir şekilde modellemektedirler.

3. Yüksek Kaliteli Görüntü Yeniden İnşası

Sıg ve derin özellikler birleştirilerek yüksek kaliteli görüntü I_{RHQ} yeniden inşa edilmekte ve formülü denklem 5.13'teki gibidir.

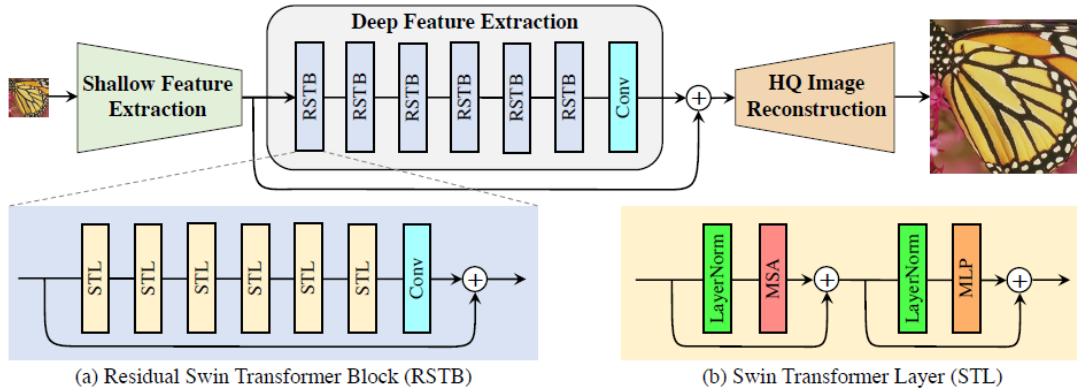
$$I_{RHQ} = H_{REC}(F_0 + F_{DF}) \quad (5.13)$$

Burada, H_{REC} , yeniden inşa fonksiyonunu temsil etmektedir.

Görüntü süper çözünürlüğü gibi görevler için yeniden inşa modülü, alt-piksel konvolüsyon katmanları kullanarak özellik haritalarını istenen çözünürlüğe yükseltmektedir. Görüntü gürültü giderme ve JPEG artefaktlarının azaltılması gibi görevler için ise tek bir konvolüsyon katmanı kullanılmaktadır. Ayrıca model, düşük kaliteli ve yüksek kaliteli görüntüler arasındaki farkı öğrenmek için denklem 5.14'ten de anlaşılabacağı üzere artık öğrenme stratejisini kullanmaktadır.

$$I_{RHQ} = H_{SwinIR}(I_{LQ}) + I_{LQ} \quad (5.14)$$

Burada, H_{SwinIR} , SwinIR modelinin fonksiyonunu temsil eder.



Şekil 5.18. Görüntü restorasyonu için önerilen SwinIR mimarisi.

Şekil 5.18'de SwinIR modelinin yapısı detaylı bir şekilde sunulmaktadır. Modelin temel bileşenleri olan Sıg Özellik Çıkarımı, Derin Özellik Çıkarımı, ve Yüksek Kaliteli Görüntü Yeniden İnşası, yukarıda açıklanan yöntem ile uyumlu olarak gösterilmektedir. İlk aşamada düşük kaliteli giriş görüntüsü, “Shallow Feature Extraction” kısmında temel özelliklere dönüştürülmektedir. Daha sonra “Deep Feature Extraction” aşamasında Artık Swin Transformer Blokları (RSTB) kullanılarak bu özellikler derinleştirilmektedir. Her bir RSTB, Swin Transformer Katmanları (STL) ve birleştirme işlemlerinden oluşurken katmanlar, yerel dikkat mekanizması ve kaydırılmış pencere yöntemi ile bağlamsal bilgiyi etkili bir şekilde işlemektedir. Son olarak, “HQ Image Reconstruction” aşamasında, sıg ve derin

özellikler birleştirilerek yüksek kaliteli görüntü yeniden inşa edilmektedir. Şekil 5.18, süreçlerin birbirine nasıl bağlandığını ve SwinIR modelinin yerel ve küresel bağlamı nasıl işlediğini net bir şekilde özetlemektedir.

Pekiştirmeli Öğrenme Yöntemleri:

IPFH ve Attention-FH, yerel detayların korunması açısından güçlü sonuçlar üretmiş ancak DRL-SRFR'nin global yapıyı geliştirme başarısına ulaşamamıştır. Özellikle, politika ağının etkinliği ve RRDB bloklarının performansı bu modellerden üstün bulunmuştur.

5.5.3. Önerilen modelin gücü

DRL-SRFR'nin başarısı, pekiştirmeli öğrenme tabanlı stratejik parça seçimi ile RRDB tabanlı süper çözünürlük mekanizmasını birleştirmesinden ilham almaktadır. Yapı hem yerel detayların korunmasını hem de global yüz yapısının iyileştirilmesini sağlamış, bu da rakip yöntemlere göre önemli bir avantaj sağlamaktadır.

Yöntemlerin karşılaştırılması, DRL-SRFR modelinin yüz süper çözünürlüğü alanında güçlü bir alternatif olduğunu açıkça ortaya koymaktadır. Modelin, literatürdeki mevcut yöntemlere kıyasla daha yüksek PSNR ve SSIM değerleri elde etmesi görsel kalitedeki başarısını desteklemektedir. Ayrıca modelin hem küçük hem de büyük veri kümelerinde genelleme yapabilmesi geniş bir uygulama alanı için uygun olduğunu göstermektedir. Sonuçlar, önerilen modelin yüz süper çözünürlüğü ve yüz halüsinasyonu alanına önemli bir katkı sağladığını ve literatürdeki diğer yöntemlere kıyasla rekabetçi bir performans sergilediğini ortaya koymaktadır.



6. SONUÇLAR

Önerilen çalışmada, yüz halüsinasyonu problemini çözmek için DRL-SRFR adı verilen özgün bir yaklaşım önerilmektedir. Görüntüde parça seçimi için DRL tabanlı bir çerçevenin kullanılması, yüz bileşenleri arasındaki bağlantıların değerlendirilmesinin verimliliğini artırmaktadır. Önerilen model DRL ve RRDB'ye dayalı bir çözünürlük iyileştirme mekanizması içermektedir. Klasik ödül mekanizması yerine bağlantılı bir yapı kurularak daha özel bir çözüm elde edilmiştir. DRL-SRFR yapısı tekrarlayan bir politika ağı ve yüzden seçilen parça için yerel bir iyileştirme ağı içermektedir. Kapsamlı deneysel sonuçlar, bu yaklaşımın insan görsel algısı tarafından tanınabilen yüksek çözünürlüklü yüz görüntüleri üretebildiğini ve PSNR ve SSIM açısından son teknoloji yüz halüsinasyon yöntemlerinden daha iyi performans sergilediğini göstermektedir. Önerilen yöntem daha hafif ağ yapıları kullanarak eğitim verimliliğini potansiyel olarak artırabilmektedir. Strateji, gerekli hesaplama kaynaklarını en aza indirir ve modeli daha geniş bir uygulama yelpazesine uyacak şekilde ayarlayabilmeye olanak tanımaktadır. Ayrıca farklı dikkat mekanizmalarının entegre edilmesi, yüz bileşenleri arasındaki ilişkilerin daha ayrıntılı bir analizini sağlayarak model performansını önemli ölçüde artırabilmektedir. Yöntem yüz bileşenlerini birbirleriyle uyumlu hale getirerek potansiyel olarak daha doğal ve gerçekçi yüksek çözünürlüklü yüz görüntüleri oluşturabilmektedir.

Model yalnızca PSNR ve SSIM gibi sayısal metriklerde değil, aynı zamanda insan görsel algısına yakın sonuçlar üretmekte de başarılıdır. Yüz ifadelerindeki doğallık, dokuların korunması ve genel görünüm açısından elde edilen sonuç görseller, modelin çeşitli uygulamalarında kullanılabilirliğini göstermektedir. Deneysel sonuçlar önerilen yöntemin yalnızca sayısal olarak güçlü olmadığını, aynı zamanda kullanıcı algısını tatmin edici seviyede karşılayabildiğini göstermektedir.

Hafif bir ağ yapısı kullanan model rakip yöntemlere kıyasla daha az hesaplama kaynağına ihtiyaç duymakta ve bu da eğitim sürecini hızlandırmaktadır. FLOP sayısının düşük olması, modelin düşük donanım kapasitelerine sahip cihazlarda bile çalışmasını mümkün kılarken, bu özelliği sayesinde geniş bir kullanıcı kitlesine hitap

eder hale gelmektedir. Eğitim verimliliği modelin farklı veri kümelerinde hızlı bir şekilde uygulanabilir olmasını sağlarken küçük veri kümelerinde gösterdiği performans, genelleme kapasitesinin yüksek olduğunu ortaya koymaktadır. Ayrıca modelin esnek ve modüler yapısı, yeni dikkat mekanizmaları ve mimarilerin entegrasyonu için uygun bir altyapı sunmaktadır. DRL-SRFR'nin bu özelliği araştırmacılar ve geliştiriciler için daha fazla optimizasyon ve özelleştirme fırsatı yaratmakta, farklı uygulamalara kolayca uyarlanmasını mümkün kılmaktadır.

DRL-SRFR modeli, birçok açıdan güçlü avantajlar sunsa da bazı sınırlamaları ve iyileştirme alanlarını da beraberinde getirmektedir. Örneğin RRDB bloklarının sayısının artırılması modelin doğruluğunu artırırken beraberinde işlem süresini ve hesaplama yükünü önemli ölçüde artırmaktadır. Dolayısıyla oluşan bu durum özellikle büyük ölçekli veri kümelerinde eğitim sürecini zorlaştırabilmektedir. Gelecekte RRDB bloklarının daha hafif varyantlarının geliştirilmesi veya hesaplama maliyetini azaltan optimizasyon tekniklerinin kullanılması bahsedilen sınırlamayı aşmada etkili olabilir. Ayrıca güncel dönüştürücü ve difüzyon tabanlı modellerle karşılaştırıldığında DRL-SRFR'nin bazı noktalarda rekabet gücünün artırılması gerekmektedir. Özellikle SR3 ve SwinIR gibi yöntemler, büyük veri kümelerinde üstün sonuçlar elde edebilmekte ve bu modellerle rekabet edebilmek için DRL-SRFR'nin dikkat mekanizmaları veya dönüştürücü tabanlı yapılarla zenginleştirilmesi önerilmektedir.

Modelin genel uygulanabilirliğini artırmak için daha çeşitli ve zorlu veri kümelerinde test edilmesinin model eğitime katkılar sağlayabileceği ön görülmektedir. Özellikle farklı yaş gruplarına, etnik kökenlere ve ışıklandırma koşullarına sahip görsel veri kümelerinin kullanımı modelin pratikteki başarısını değerlendirmek açısından faydalı olacaktır.

DRL-SRFR'nin bir diğer geliştirme alanı pratik uygulamalar için optimize edilmesidir. Modelin daha hafif mimarilerle yeniden yapılandırılması pratik yüz süper çözünürlüğü gerektiren uygulamalarda kullanılabilirliğini artıracaktır. Bunun yanı sıra insan görsel algısını daha iyi yansıtan yeni kayıp fonksiyonlarının geliştirilmesi, modelin hem algısal hem de sayısal performansını daha ileriye taşıyabilir. Algısal kaliteyi artırmak için VGG tabanlı kayıp fonksiyonlarının yanı sıra, insan algısına uygun metriklerin entegrasyonu süreçte etkili olabilir.

Son olarak modelin hiperparametre ayarlarının otomatik olarak optimize edilmesi için yöntemlerin geliştirilmesi, kullanıcı dostu bir yapı sunarak araştırmacıların ve geliştiricilerin iş yükünü azaltacaktır. Yukarıda bahsedilen geliştirme önerilerinin, DRL-SRFR'nin hem teorik hem de pratik anlamda daha güçlü bir model haline gelmesine katkı sağlayacağı düşünülmektedir.





KAYNAKLAR

- [1] Jiang, J., Wang, C., Liu, X., & Ma, J. (2021). Deep Learning-based Face Super-resolution: A Survey. *ACM Comput. Surv.*, 55(1), 13:1-13:36. <https://doi.org/10.1145/3485132>
- [2] Kanakaraj, S., V K, G., & Kalady, S. (2017). Face Super Resolution: A Survey. *International Journal of Image, Graphics and Signal Processing*, 9, 54-67. <https://doi.org/10.5815/ijigsp.2017.05.06>
- [3] Sharma, A., Srivastava, B. P., & Shankar, P. N. (2023). Facial Image Super-Resolution with CNN, "A Review". 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 1-6. <https://doi.org/10.1109/SCEECS57921.2023.10063110>
- [4] Shah, A. J., & Gupta, S. B. (2012). Image super resolution-A survey. 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking, 1-6. <https://doi.org/10.1109/ET2ECN.2012.6470098>
- [5] Baker, S., & Kanade, T. (2002). Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9), 1167-1183. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2002.1033210>
- [6] Raschka, S. (2020). Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning (arXiv:1811.12808). *arXiv*. <https://doi.org/10.48550/arXiv.1811.12808>
- [7] Shi, Y., Li, G., Cao, Q., Wang, K., & Lin, L. (2020). Face Hallucination by Attentive Sequence Optimization with Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11), 2809-2824. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2019.2915301>
- [8] Howard, R. A. (1960). *Dynamic programming and Markov processes* (ss. viii, 136). John Wiley.
- [9] Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, 9(5), 1054-1054. *IEEE Transactions on Neural Networks*. <https://doi.org/10.1109/TNN.1998.712192>
- [10] Bellman, R. E. (1954). *The Theory of Dynamic Programming*. RAND Corporation. <https://www.rand.org/pubs/papers/P550.html>
- [11] Samuel, A. L. (1960). Programming Computers to Play Games. İçinde F. L. Alt (Ed.), *Advances in Computers* (C. 1, ss. 165-192). Elsevier. [https://doi.org/10.1016/S0065-2458\(08\)60608-7](https://doi.org/10.1016/S0065-2458(08)60608-7)

- [12] Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2), 267-279. <https://doi.org/10.1016/j.neuron.2013.11.005>
- [13] Seijen, H. van, Mahmood, A. R., Pilarski, P. M., Machado, M. C., & Sutton, R. S. (2016). True Online Temporal-Difference Learning (arXiv:1512.04087). arXiv. <https://doi.org/10.48550/arXiv.1512.04087>
- [14] Ahmed, M., Lim, C. P., & Nahavandi, S. (2021). A Deep Q-Network Reinforcement Learning-Based Model for Autonomous Driving. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 739-744. <https://doi.org/10.1109/SMC52423.2021.9658892>
- [15] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. ArXiv. <https://www.semanticscholar.org/paper/Proximal-Policy-Optimization-Algorithms-Schulman-Wolski/dce6f9d4017b1785979e7520fd0834ef8cf02f4b>
- [16] Sarkar, A. (2018). A Brandom-ian view of Reinforcement Learning towards strong-AI. ArXiv. <https://www.semanticscholar.org/paper/A-Brandom-ian-view-of-Reinforcement-Learning-Sarkar/501daf146f2be2126b4182e4555953d62961e7e3>
- [17] Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (1999). Policy Gradient Methods for Reinforcement Learning with Function Approximation. *Advances in Neural Information Processing Systems*, 12. https://papers.nips.cc/paper_files/paper/1999/hash/464d828b85b0bed98e80ade0a5c43b0f-Abstract.html
- [18] Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3), 229-256. <https://doi.org/10.1007/BF00992696>
- [19] Marhon, S. A., Cameron, C. J. F., & Kremer, S. C. (2013). Recurrent Neural Networks. In M. Bianchini, M. Maggini, & L. C. Jain (Eds.), *Handbook on Neural Information Processing* (ss. 29-65). Springer. https://doi.org/10.1007/978-3-642-36657-4_2
- [20] Schmidt, R. M. (2019). Recurrent Neural Networks (RNNs): A gentle Introduction and Overview. ArXiv. [https://www.semanticscholar.org/paper/Recurrent-Neural-Networks-\(RNNs\)%3A-A-gentle-and-Schmidt/9c8a6ab02a048de36aa11bd03ca80203e34de359](https://www.semanticscholar.org/paper/Recurrent-Neural-Networks-(RNNs)%3A-A-gentle-and-Schmidt/9c8a6ab02a048de36aa11bd03ca80203e34de359)
- [21] Ghogh, B., & Ghodsi, A. (2023). Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey (arXiv:2304.11461). arXiv. <https://doi.org/10.48550/arXiv.2304.11461>
- [22] Smagulova, K., & James, A. P. (2019). A survey on LSTM memristive neural network architectures and applications. *The European Physical Journal Special Topics*, 228(10), 2313-2324. <https://doi.org/10.1140/epjst/e2019-900046-x>
- [23] Hassanin, M., Anwar, S., Radwan, I., Khan, F. S., & Mian, A. (2024). Visual attention methods in deep learning: An in-depth survey. *Information Fusion*, 108, 102417. <https://doi.org/10.1016/j.inffus.2024.102417>

- [24] Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. *Information*, 15(9), Article 9. <https://doi.org/10.3390/info15090517>
- [25] Datta, S. K., Shaikh, M. A., Srihari, S. N., & Gao, M. (2021). Soft Attention Improves Skin Cancer Classification Performance. In M. Reyes, P. Henriques Abreu, J. Cardoso, M. Hajij, G. Zamzmi, P. Rahul, & L. Thakur (Eds.), *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data* (ss. 13-23). Springer International Publishing. https://doi.org/10.1007/978-3-030-87444-5_2
- [26] Papadopoulos, A., Korus, P., & Memon, N. (2024). Hard-attention for scalable image classification. *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 14694-14707.
- [27] Wang, Z., Miao, Z., Jonathan Wu, Q. M., Wan, Y., & Tang, Z. (2014). Low-resolution face recognition: A review. *The Visual Computer*, 30(4), 359-386. <https://doi.org/10.1007/s00371-013-0861-x>
- [28] Zhang, L., Zhang, H., Shen, H., & Li, P. (2010). A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 90(3), 848-859. <https://doi.org/10.1016/j.sigpro.2009.09.002>
- [29] Isaac, J. S., & Kulkarni, R. (2015). Super resolution techniques for medical image processing. *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, 1-6. <https://doi.org/10.1109/ICTSD.2015.7095900>
- [30] Zha, Y., Huang, Y., Sun, Z., Wang, Y., & Yang, J. (2015). Bayesian Deconvolution for Angular Super-Resolution in Forward-Looking Scanning Radar. *Sensors*, 15(3), Article 3. <https://doi.org/10.3390/s150306924>
- [31] Lu, T., Hao, X., Zhang, Y., Liu, K., & Xiong, Z. (2019). Parallel Region-Based Deep Residual Networks for Face Hallucination. *IEEE Access*, 7, 81266-81278. <https://doi.org/10.1109/ACCESS.2019.2923023>
- [32] Keys, R. (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6), 1153-1160. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. <https://doi.org/10.1109/TASSP.1981.1163711>
- [33] Parker, J. A., Kenyon, R. V., & Troxel, D. E. (1983). Comparison of Interpolating Methods for Image Resampling. *IEEE Transactions on Medical Imaging*, 2(1), 31-39. *IEEE Transactions on Medical Imaging*. <https://doi.org/10.1109/TMI.1983.4307610>
- [34] Catmull, E., & Smith, A. R. (1980). 3-D transformations of images in scanline order. *SIGGRAPH Comput. Graph.*, 14(3), 279-285. <https://doi.org/10.1145/965105.807505>

- [35] Liu, C., Shum, H.-Y., & Zhang, C.-S. (2001). A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1, I-I. <https://doi.org/10.1109/CVPR.2001.990475>
- [36] Wang, X., & Tang, X. (2005). Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35(3), 425-434. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. <https://doi.org/10.1109/TSMCC.2005.848171>
- [37] Roweis, S. T., & Saul, L. K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500), 2323-2326. <https://doi.org/10.1126/science.290.5500.2323>
- [38] Chang, H., Yeung, D.-Y., & Xiong, Y. (2004). Super-resolution through neighbor embedding. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 1, I-I. <https://doi.org/10.1109/CVPR.2004.1315043>
- [39] Zhuang, Y., Zhang, J., & Wu, F. (2007). Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation. *Pattern Recognition*, 40(11), 3178-3194. <https://doi.org/10.1016/j.patcog.2007.03.011>
- [40] Liu, W., Lin, D., & Tang, X. (2005). Neighbor combination and transformation for hallucinating faces. *2005 IEEE International Conference on Multimedia and Expo*, 4 pp.-. <https://doi.org/10.1109/ICME.2005.1521381>
- [41] Li, X., & Orchard, M. T. (2001). New edge-directed interpolation. *IEEE Transactions on Image Processing*, 10(10), 1521-1527. *IEEE Transactions on Image Processing*. <https://doi.org/10.1109/83.951537>
- [42] Sun, J., Xu, Z., & Shum, H.-Y. (2008). Image super-resolution using gradient profile prior. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1-8. <https://doi.org/10.1109/CVPR.2008.4587659>
- [43] Dai, S., Han, M., Xu, W., Wu, Y., Gong, Y., & Katsaggelos, A. K. (2009). SoftCuts: A Soft Edge Smoothness Prior for Color Image Super-Resolution. *IEEE Transactions on Image Processing*, 18(5), 969-981. *IEEE Transactions on Image Processing*. <https://doi.org/10.1109/TIP.2009.2012908>
- [44] Yan, Q., Xu, Y., Yang, X., & Nguyen, T. Q. (2015). Single Image Superresolution Based on Gradient Profile Sharpness. *IEEE Transactions on Image Processing*, 24(10), 3187-3202. *IEEE Transactions on Image Processing*. <https://doi.org/10.1109/TIP.2015.2414877>
- [45] Marquina, A., & Osher, S. (2008). Image Super-Resolution by TV-Regularization and Bregman Iteration. *Journal of Scientific Computing*, 37, 367-382. <https://doi.org/10.1007/s10915-008-9214-8>
- [46] Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2, 60-65 c. 2. <https://doi.org/10.1109/CVPR.2005.38>

- [47] Hardie, R. C., Barnard, K. J., & Armstrong, E. E. (1997). Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6(12), 1621-1633. [IEEE Transactions on Image Processing. https://doi.org/10.1109/83.650116](https://doi.org/10.1109/83.650116)
- [48] Farsiu, S., Robinson, M. D., Elad, M., & Milanfar, P. (2004). Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10), 1327-1344. [IEEE Transactions on Image Processing. https://doi.org/10.1109/TIP.2004.834669](https://doi.org/10.1109/TIP.2004.834669)
- [49] Irani, M., & Peleg, S. (1993). Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency. *Journal of Visual Communication and Image Representation*, 4(4), 324-335. <https://doi.org/10.1006/jvci.1993.1030>
- [50] Lin, Z., & Shum, H.-Y. (2004). Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1), 83-97. [IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2004.1261081](https://doi.org/10.1109/TPAMI.2004.1261081)
- [51] Wang, Z., Chen, J., & Hoi, S. C. H. (2021). Deep Learning for Image Super-Resolution: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10), 3365-3387. [IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2020.2982166](https://doi.org/10.1109/TPAMI.2020.2982166)
- [52] Huang, H., & He, H. (2011). Super-Resolution Method for Face Recognition Using Nonlinear Mappings on Coherent Features. *IEEE Transactions on Neural Networks*, 22(1), 121-130. [IEEE Transactions on Neural Networks. https://doi.org/10.1109/TNN.2010.2089470](https://doi.org/10.1109/TNN.2010.2089470)
- [53] Nguyen, K., Sridharan, S., Denman, S., & Fookes, C. (2012). Feature-domain super-resolution framework for Gabor-based face and iris recognition. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*, 2642-2649. <https://doi.org/10.1109/CVPR.2012.6247984>
- [54] Wu, L., & Wang, X. (2011). A fast algorithm for learning-based super-resolution reconstruction of face image. 2011 4th *International Congress on Image and Signal Processing*, 2, 1049-1053. <https://doi.org/10.1109/CISP.2011.6100303>
- [55] Zou, W. W. W., & Yuen, P. C. (2010). Learning the Relationship Between High and Low Resolution Images in Kernel Space for Face Super Resolution. 2010 20th *International Conference on Pattern Recognition*, 1152-1155. <https://doi.org/10.1109/ICPR.2010.288>
- [56] He, Y., Yap, K.-H., & Chau, L.-P. (2009). A learning approach for single-frame face super-resolution. 2009 *IEEE International Symposium on Circuits and Systems*, 770-773. <https://doi.org/10.1109/ISCAS.2009.5117862>
- [57] Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., & Liao, Q. (2019). Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Transactions on Multimedia*, 21(12), 3106-3121. [IEEE Transactions on Multimedia. https://doi.org/10.1109/TMM.2019.2919431](https://doi.org/10.1109/TMM.2019.2919431)

- [58] Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1132-1140. <https://doi.org/10.1109/CVPRW.2017.151>
- [59] Kong, Y., Zhang, S., & Cheng, P. (2013). Super-resolution reconstruction face recognition based on multi-level FFD registration. *Optik*, 124(24), 6926-6931. <https://doi.org/10.1016/j.ijleo.2013.05.175>
- [60] Lu, T., Wang, H., Xiong, Z., Jiang, J., Zhang, Y., Zhou, H., & Wang, Z. (2017). Face hallucination using region-based deep convolutional networks. 2017 IEEE International Conference on Image Processing (ICIP), 1657-1661. <https://doi.org/10.1109/ICIP.2017.8296563>
- [61] Lu, T., Wang, Y., Zhang, Y., Wang, Y., Wei, L., Wang, Z., & Jiang, J. (2021). Face Hallucination via Split-Attention in Split-Attention Network. Proceedings of the 29th ACM International Conference on Multimedia, 5501-5509. <https://doi.org/10.1145/3474085.3475682>
- [62] Cheng, X., Lu, J., Yuan, B., & Zhou, J. (2020). Identity-Preserving Face Hallucination via Deep Reinforcement Learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(12), 4796-4809. *IEEE Transactions on Circuits and Systems for Video Technology*. <https://doi.org/10.1109/TCSVT.2019.2961629>
- [63] Cao, Q., Lin, L., Shi, Y., Liang, X., & Li, G. (2017). Attention-Aware Face Hallucination via Deep Reinforcement Learning. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1656-1664. <https://doi.org/10.1109/CVPR.2017.180>
- [64] Pan, J., Sun, D., Zhang, J., Tang, J., Yang, J., Tai, Y.-W., & Yang, M.-H. (2022). Dual Convolutional Neural Networks for Low-Level Vision. *International Journal of Computer Vision*, 130(6), 1440-1458. <https://doi.org/10.1007/s11263-022-01583-y>
- [65] Dong, J., Pan, J., Ren, J. S., Lin, L., Tang, J., & Yang, M.-H. (2022). Learning Spatially Variant Linear Representation Models for Joint Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 8355-8370. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2021.3102575>
- [66] Dong, J., Roth, S., & Schiele, B. (2022). DWDN: Deep Wiener Deconvolution Network for Non-Blind Image Deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12), 9960-9976. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2021.3138787>
- [67] Zhou, E., Fan, H., Cao, Z., Jiang, Y., & Yin, Q. (2015). Learning Face Hallucination in the Wild. Proceedings of the AAAI Conference on Artificial Intelligence, 29(1). <https://doi.org/10.1609/aaai.v29i1.9795>
- [68] Huang, H., He, R., Sun, Z., & Tan, T. (2017). Wavelet-SRNet: A Wavelet-Based CNN for Multi-scale Face Super Resolution. 2017 IEEE International Conference on Computer Vision (ICCV), 1698-1706. <https://doi.org/10.1109/ICCV.2017.187>

- [69] Chen, Y., Tai, Y., Liu, X., Shen, C., & Yang, J. (2018). FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2492-2501. <https://doi.org/10.1109/CVPR.2018.00264>
- [70] Ma, C., Jiang, Z., Rao, Y., Lu, J., & Zhou, J. (2020). Deep Face Super-Resolution With Iterative Collaboration Between Attentive Recovery and Landmark Estimation. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5568-5577. <https://doi.org/10.1109/CVPR42600.2020.00561>
- [71] Wang, C., Jiang, J., Zhong, Z., & Liu, X. (2023). Spatial-Frequency Mutual Learning for Face Super-Resolution. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 22356-22366. <https://doi.org/10.1109/CVPR52729.2023.02141>
- [72] Dastmalchi, H., & Aghaeinia, H. (2022). Super-resolution of very low-resolution face images with a wavelet integrated, identity preserving, adversarial network. *Signal Processing: Image Communication*, 107, 116755. <https://doi.org/10.1016/j.image.2022.116755>
- [73] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 105-114. <https://doi.org/10.1109/CVPR.2017.19>
- [74] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual Dense Network for Image Super-Resolution. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2472-2481. <https://doi.org/10.1109/CVPR.2018.00262>
- [75] Xu, J., Chae, Y., Stenger, B., & Datta, A. (2018). Dense Bynet: Residual Dense Network for Image Super Resolution. 2018 25th IEEE International Conference on Image Processing (ICIP), 71-75. <https://doi.org/10.1109/ICIP.2018.8451696>
- [76] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., & Loy, C. C. (2019). ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In L. Leal-Taixé & S. Roth (Ed.), *Computer Vision – ECCV 2018 Workshops* (ss. 63-79). Springer International Publishing. https://doi.org/10.1007/978-3-030-11021-5_5
- [77] Li, H. (2021). Image Super-Resolution Algorithm Based on RRDB Model. *IEEE Access*, 9, 156260-156273. <https://doi.org/10.1109/ACCESS.2021.3118444>
- [78] Chen, Y.-Z., Liu, T.-J., & Liu, K.-H. (2022). Super-Resolution of Satellite Images by two-Dimensional RRDB and Edge-Enhancement Generative Adversarial Network. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1825-1829. <https://doi.org/10.1109/ICASSP43922.2022.9747063>

- [79] Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., & Norouzi, M. (2023). Image Super-Resolution via Iterative Refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4), 4713-4726. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2022.3204461>
- [80] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Commun. ACM*, 63(11), 139-144. <https://doi.org/10.1145/3422622>
- [81] Yu, J. J., Derpanis, K. G., & Brubaker, M. A. (2020). Wavelet Flow: Fast Training of High Resolution Normalizing Flows. *Advances in Neural Information Processing Systems*, 33, 6184-6196. https://proceedings.neurips.cc/paper_files/paper/2020/hash/4491777b1aa8b5b32c2e8666dbe1a495-Abstract.html
- [82] Kingma, D. P., & Dhariwal, P. (2018). Glow: Generative Flow with Invertible 1x1 Convolutions. *Advances in Neural Information Processing Systems*, 31. https://proceedings.neurips.cc/paper_files/paper/2018/hash/d139db6a236200b21cc7f752979132d0-Abstract.html
- [83] Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio (arXiv:1609.03499). *arXiv*. <https://doi.org/10.48550/arXiv.1609.03499>
- [84] Van den Oord, A., Kalchbrenner, N., Espeholt, L., Kavukcuoglu, K., Vinyals, O., & Graves, A. (2016). Conditional Image Generation with PixelCNN Decoders. *Advances in Neural Information Processing Systems*, 29. <https://proceedings.neurips.cc/paper/2016/hash/b1301141feffabac455e1f90a7de2054-Abstract.html>
- [85] Noroozi, M., Hadji, I., Martinez, B., Bulat, A., & Tzimiropoulos, G. (2024). You Only Need One Step: Fast Super-Resolution with Stable Diffusion via Scale Distillation (arXiv:2401.17258). *arXiv*. <https://doi.org/10.48550/arXiv.2401.17258>
- [86] Corvi, R., Cozzolino, D., Poggi, G., Nagano, K., & Verdoliva, L. (2023). Intriguing Properties of Synthetic Images: From Generative Adversarial Networks to Diffusion Models. 973-982. https://openaccess.thecvf.com/content/CVPR2023W/WMF/html/Corvi_Intriguing_Properties_of_Synthetic_Images_From_Generative_Adversarial_Networks_to_CVPRW_2023_paper.html
- [87] Benny, Y., Galanti, T., Benaim, S., & Wolf, L. (2021). Evaluation Metrics for Conditional Image Generation. *International Journal of Computer Vision*, 129(5), 1712-1731. <https://doi.org/10.1007/s11263-020-01424-w>
- [88] Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., & Verdoliva, L. (2023). On The Detection of Synthetic Images Generated by Diffusion Models. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1-5. <https://doi.org/10.1109/ICASSP49357.2023.10095167>

- [89] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (arXiv:2010.11929). arXiv. <https://doi.org/10.48550/arXiv.2010.11929>
- [90] Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., & Zeng, T. (2022). Transformer for Single Image Super-Resolution. 457-466. https://openaccess.thecvf.com/content/CVPR2022W/NTIRE/html/Lu_Transformer_for_Single_Image_Super-Resolution_CVPRW_2022_paper.html
- [91] You, W., Sun, S., & Iyyer, M. (2020). Hard-Coded Gaussian Attention for Neural Machine Translation (arXiv:2005.00742). arXiv. <https://doi.org/10.48550/arXiv.2005.00742>
- [92] Tai, Y., Yang, J., & Liu, X. (2017). Image Super-Resolution via Deep Recursive Residual Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2790-2798. <https://doi.org/10.1109/CVPR.2017.298>
- [93] Lai, W.-S., Huang, J.-B., Ahuja, N., & Yang, M.-H. (2017). Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution (arXiv:1704.03915). arXiv. <https://doi.org/10.48550/arXiv.1704.03915>
- [94] Chen, C., Gong, D., Wang, H., Li, Z., & Wong, K.-Y. K. (2021). Learning Spatial Attention for Face Super-Resolution. IEEE Transactions on Image Processing, 30, 1219-1231. <https://doi.org/10.1109/TIP.2020.3043093>
- [95] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image Restoration Using Swin Transformer. 1833-1844. https://openaccess.thecvf.com/content/ICCV2021W/AIM/html/Liang_SwinIR_Image_Restoration_Using_Swin_Transformer_ICCVW_2021_paper.html
- [96] Thanh, T. N. (t.y.). Deep Learning Based Approach Implemented to Image Super-Resolution. Journal of Advances in Information Technology. Geliş tarihi 10 Şubat 2024, gönderen https://www.academia.edu/72768607/Deep_Learning_Based_Approach_Implemented_to_Image_Super_Resolution
- [97] Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the Super-Resolution Convolutional Neural Network (arXiv:1608.00367). arXiv. <https://doi.org/10.48550/arXiv.1608.00367>
- [98] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307. IEEE Transactions on Pattern Analysis and Machine Intelligence. <https://doi.org/10.1109/TPAMI.2015.2439281>
- [99] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1646-1654. <https://doi.org/10.1109/CVPR.2016.182>

- [100] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep Learning Face Attributes in the Wild. 3730-3738. https://openaccess.thecvf.com/content_iccv_2015/html/Liu_Deep_Learning_Face_ICCV_2015_paper.html
- [101] Jesorsky, O., Kirchberg, K. J., & Frischholz, R. W. (2001). Robust Face Detection Using the Hausdorff Distance. İçinde J. Bigun & F. Smeraldi (Ed.), Audio- and Video-Based Biometric Person Authentication (ss. 90-95). Springer. https://doi.org/10.1007/3-540-45344-X_14
- [102] Huang, G. B., Jain, V., & Learned-Miller, E. (2007). Unsupervised Joint Alignment of Complex Images. 2007 IEEE 11th International Conference on Computer Vision, 1-8. <https://doi.org/10.1109/ICCV.2007.4408858>
- [103] Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009). Attribute and simile classifiers for face verification. 2009 IEEE 12th International Conference on Computer Vision, 365-372. <https://doi.org/10.1109/ICCV.2009.5459250>
- [104] Guha, E., & Lakshman, V. (2024). On the Diminishing Returns of Width for Continual Learning (arXiv:2403.06398). arXiv. <https://doi.org/10.48550/arXiv.2403.06398>

ÖZGEÇMİŞ

Ad-Soyad : Emre ALTINKAYA

ÖĞRENİM DURUMU:

- **Lisans** : 2016, Sakarya Üniversitesi, Mühendislik Fakültesi, Elektrik & Elektronik Mühendisliği Bölümü
- **Yükseklisans** : 2018, Sakarya Üniversitesi, Elektrik & Elektronik Mühendisliği Anabilim Dalı, Elektrik Programı

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2019-devam ediyor. Bilecik Şeyh Edebali Üniversitesinde Öğretim Görevlisi olarak çalışmaya devam etmekte.

TEZDEN TÜRETİLEN ESERLER:

Makale

- E. Altinkaya ve B. Barakli, “Enhancing Face Image Quality: Strategic Patch Selection With Deep Reinforcement Learning and Super-Resolution Boost via RRDB,” in IEEE Access, vol. 12, pp. 120142-120164, 2024, doi: 10.1109/ACCESS.2024.3450571.

Hakemli Kongre / Sempozyum Bildiri

- E. Altinkaya ve B. Barakli, “Yüz Görüntülerinde Derin Evrişimli Ağ Kullanan Süper Çözünürlük Modeli,” 11. Uluslararası Akademik Araştırmalar Kongresi (ICAR), Turkey, 25 - 26 July 2023.

DİĞER ESERLER:

- E. Altinkaya ve B. Barakli, “Comparison Of Reversible Image Watermarking Methods Based On Prediction-Errors,” Sakarya University Journal of Science, 23(4), 493-500, DOI: 10.16984/aufenbilder.466656
- E. Altinkaya, K. Polat, ve B. BARAKLI, (2019), “Detection of Alzheimer’s Disease and Dementia States Based on Deep Learning from MRI Images: A Comprehensive Review,” Journal of the Institute of Electronics and Computer, 1, 39-53. <https://doi.org/10.33969/JIEC.2019.11005>.