T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GÜVENLİK KAMERALARINDAKİ YÜZ GÖRÜNTÜLERİNİN SÜPER ÇÖZÜNÜRLÜKLE NETLEŞTİRİLMESİ

YÜKSEK LİSANS TEZİ

Ali Hüsameddin ATEŞ

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

EYLÜL 2024

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GÜVENLİK KAMERALARINDAKİ YÜZ GÖRÜNTÜLERİNİN SÜPER ÇÖZÜNÜRLÜKLE NETLEŞTİRİLMESİ

YÜKSEK LİSANS TEZİ

Ali Hüsameddin ATEŞ

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

Tez Danışmanı: Dr.Öğr.Üyesi Hüseyin ESKİ

EYLÜL 2024

Ali Hüsameddin ATEŞ tarafından hazırlanan "GÜVENLİK KAMERALARINDAKİ YÜZ GÖRÜNTÜLERİNİN SÜPER ÇÖZÜNÜRLÜKLE NETLEŞTİRİLMESİ" adlı tez çalışması 12.09.2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı'nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı:

Jüri Üyesi:

Jüri Üyesi:

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 "GÜVENLİK KAMERALARINDAKİ YÜZ GÖRÜNTÜLERİNİN SÜPER ÇÖZÜNÜRLÜKLE NETLEŞTİRİLMESİ" 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.

(12/09/2024).

Ali Hüsameddin ATEŞ

İÇİNDEKİLER

Sayfa

KISALTMALAR

ANN : Artificial Neural Network
CNN : Convolutional Neural Network
CPU : Central Processing Unit
DNN : Deep Neural Network
ESA : Evrişimli Sinir Ağı
FSR : Face Super Resolution
FSR : Face Super-Resolution
GAN : Generative Adversarial Network
GPU : Graphics Processing Unit
HR : High-Resolution
ILoss : Identity Loss
LPIPS : Learned Perceptual Image Patch Similarity
LR : Low-Resolution
MAE : Mean Absolute Error
MSE : Mean Squared Error
PSNR : Peak Signal-To-Noise Ratio
RGB : Red Green Blue
RNN : Recurrent Neural Network
SÇ : Süper Çözünürlük
SR : Super-Resolution
SSIM : Structural Similarity Index Measure
Val. : Validation
Valid. : Validation
ViT : Vision Transformer
YSA : Yapay Sinir Ağı

TABLO LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

GÜVENLİK KAMERALARINDAKİ YÜZ GÖRÜNTÜLERİNİN SÜPER ÇÖZÜNÜRLÜKLE NETLEŞTİRİLMESİ

ÖZET

Günümüzde güvenlik kameraları kamu ve askeri alanlardan endüstriyel tesislere, iş yerlerinden trafik denetimine kadar geniş bir kullanım yelpazesine sahiptir. Bu sistemler genel olarak güvenliğin ve asayişin sağlanması, suçların önlenmesi ve delillerin toplanması gibi kritik amaçların yanı sıra endüstriyel iş takibi ve üretim hatlarının kontrolü gibi çok çeşitli alanlarda kullanılmaktadır. Bununla birlikte mevcut birçok güvenlik kamerasının düşük çözünürlükte ve detayda kayıt yapması, özellikle uzak mesafelerde ve düşük ışık koşullarında görüntü kalitesini olumsuz yönde etkilemektedir. Mevcut güvenlik sistemlerinin yeni nesil yüksek çözünürlükte kayıt yapabilen sensörlere sahip kameralarla güncellenmesi ise hem maliyetli hem de zamana yayılmış bir süreçtir. Günümüzde görüntü işleme alanında, kamera görüntülerinin iyileştirilmesinde derin öğrenme tabanlı süper çözünürlük mimarilerinin kullanımı gittikçe yaygınlaşmaktadır. Düşük çözünürlük ve detaydaki görüntülerin, görüntü detayları ve keskinliği korunarak daha yüksek çözünürlüklere üst örneklenmesi, dijital görüntü işleme alanında süper çözünürlük adı altında incelenmektedir. Görüntü işleme alanında süper çözünürlük teknikleri genel olarak fotoğraf ve video görüntülerinde, medikal görüntülemede, güvenlik sistemlerinde, uzaktan tespit yapan cihazlarda ve uydudan görüntüleme gibi birçok alanda görüntünün restorasyonu ve çözünürlüğünün artırılması amaçlarıyla yaygın olarak kullanılmaktadır. Bu çalışma, güvenlik sistemlerinden alınan görüntülerdeki şahısların görünürlüğünü süper çözünürlük yöntemleri ile artırarak, normal şartlarda insan gözü veya yüz tanıma sistemleri ile tespitin mümkün olmadığı durumlarda yazılımsal bir iyileştirme sunmayı amaçlamaktadır. Tez kapsamında önerilen VNet mimarisi, düşük çözünürlükteki yüz görüntülerini süper çözünürlük ile yüksek çözünürlük ve detaydaki görüntülere dönüştürmek amacıyla derin öğrenme tabanlı bir konvolüsyon ağı ve şahısların yüz biyometrisini korumak için önceden eğitilmiş bir FaceNet modelinden oluşmaktadır. Geliştirilen ağda kodlayıcı-çözücü mimarisinin avantajları kullanılarak hem görüntü detayları korunurken hem de kayıp bölgeler restore edilmektedir. Yüz görüntülerinin süper çözünürlükle restore edilmesi veya üst örneklenmesi konusunda yapılan çalışmalarda, elde edilen görüntünün kalitesi, yüksek metrik başarımları ve görsel iyileştirmeler çoğu zaman ön planda tutulmakta, biyometrik açıdan aslına uygunluk ikinci planda kalmaktadır. Yapılan çalışmada önceden eğitilmiş FaceNet modelinin biyometrik hata beslemesi sayesinde üretilen çıktı görüntüsü, asıl görüntüye olan kalıtsal benzerliğinden koparılmadan, bir yandan görsel iyileştirmeler yapılırken diğer yandan da yüksek doğrulukta çıktılar alınabilmektedir. VNet mimarisi, çeşitli yüz görüntü veri setleri ve gerçek dünya senaryosuna uygun güvenlik kamera görüntüleri kullanılarak yapılan testlerde yüksek metrik başarımlar elde etmiş ve düşük çözünürlüklü yüz görüntülerinden, referans görüntüye uygun bir biçimde, yüksek çözünürlüklü görüntüler elde edilmesi konusunda etkili olduğu görülmüştür.

FACE ENHANCEMENT IN SURVEILLANCE SYSTEMS USING SUPER-RESOLUTION TECHNIQUES

SUMMARY

In today's world, security cameras have found extensive applications across a variety of domains, ranging from public and military settings to industrial facilities, workplaces, and traffic monitoring. These systems are primarily employed for critical purposes such as ensuring security and public order, preventing criminal activities, and collecting evidence, in addition to being utilized in diverse areas such as industrial work tracking and production line management. However, the widespread deployment of security systems does not necessarily equate to enhanced security, as many existing security cameras operate at low resolution and detail. Consequently, image quality significantly deteriorates, particularly at greater distances and under low-light conditions.

Updating existing security systems with next-generation cameras equipped with high-resolution recording sensors presents both a costly and time-consuming challenge. Instead, software-based image enhancement provides a more scalable solution in terms of both cost and time. Software enhancement techniques can be applied to existing low-resolution systems to achieve substantial improvements in image quality, thereby modernizing systems without necessitating expensive hardware upgrades.

Currently, deep learning-based super-resolution architectures, which represent a new trend in image processing, are increasingly being adopted for enhancing camera images. Such software-based approaches offer a more flexible and economical solution for renewing security systems, presenting an attractive alternative to hardware-based solutions.

Super-resolution (SR) methods, which focus on upsampling images while preserving their details and sharpness, are being explored within the field of digital image processing. Super-resolution techniques are employed in various applications, including enhancing image quality in photographs or videos, improving the resolution of medical imaging devices such as magnetic resonance imaging (MRI) and computed tomography (CT), and refining satellite images to achieve greater clarity and detail.

In the context of security cameras, low-resolution images captured under conditions laden with artifacts such as low light, noise, and blur complicate face recognition and detection, thereby diminishing the performance of algorithms designed for this purpose. In such scenarios, the development of super-resolution techniques aimed at enhancing the effectiveness of security cameras and clarifying suspicious faces in low-resolution images holds significant promise.

This study employs deep learning-based super-resolution techniques to restore low-resolution camera images and enhance their resolution for improved usability. Super-resolution methods facilitate the reconstruction of low-resolution images at higher resolutions and with greater detail. These techniques enable the upsampling of low-resolution images while maintaining their details and sharpness. Notably, images

captured by older low-resolution security cameras are restored and upsampled, thereby enhancing the visibility of individuals within the image through super-resolution methods. This research aims to contribute to the literature advocating for the enhancement of security systems through super-resolution techniques and to further advance super-resolution architectures and deep learning-based methodologies for camera image analysis within the field of forensic informatics.

In this research, deep learning architectures serve as the foundation for achieving super-resolution. While numerous image scaling techniques based on interpolation and reconstruction exist in image processing, deep learning-based super-resolution techniques demonstrate superior performance compared to traditional methods. Specifically, CNN, GAN, and ViT-based approaches have emerged as pioneering methods in this domain by effectively restoring details in low-resolution images. Consequently, this study focuses on deep learning-based SR techniques to restore and upsample lost details in low-resolution face images captured by security cameras, thereby facilitating the detection of suspicious individuals.

This study proposes a VNet architecture designed to transform low-resolution face images obtained from real-world camera footage into high-quality, detailed images. The VNet architecture utilizes a deep learning-based convolutional network to convert low-resolution face images into high-resolution, detailed representations. Through the employed encoder-decoder structure, the image is upsampled while preserving image details and restoring lost regions.

In studies concerning the super-resolution or upsampling of face images, the quality of the resultant image, high metric performances, and visual enhancements are often prioritized, while fidelity in terms of identity remains a secondary concern. To address this issue, the VNet architecture incorporates identity loss feedback from a pre-trained FaceNet model as the loss function, in conjunction with mean squared error; thus, a model is proposed that harmonizes high metric performances and visual improvements without compromising the output image's inherent similarity to the original.

In this thesis, the proposed VNet network was initially trained using the VGGFace2 dataset. The VGGFace2 dataset proved beneficial for enabling the network to learn fundamental features due to its broad coverage and inclusion of face images from individuals of various ethnic backgrounds, ages, and professions, captured under diverse poses and lighting conditions.

However, since this dataset is general-purpose and predominantly consists of frontal face images, the model obtained as a result of training was unable to achieve sufficient success in improving faces in real-world security camera images, which is the main purpose of the thesis. Therefore, this and similar datasets remain incompatible both for video images and from the perspective of security cameras.

Factors such as how purpose-oriented the dataset used in training is, the diversity of examples in the dataset, and the presence of a sufficient number and quality of examples are critical to the success of deep learning models. For this reason, the ChokePoint dataset was used as a second dataset in the study. This dataset consists of video frames extracted from videos recorded by cameras placed at various checkpoints. This dataset was preferred because people pass the cameras at a natural angle and is more suitable for real-world security camera scenarios.

A hybrid approach was adopted by using multiple datasets for the training of the developed model. First, the training with the VGGFace2 dataset ensured that the

network generally recognizes the general structure of face images, facial components, and textural details. At this stage, the obtained weights were used to continue the training of the network with the ChokePoint dataset. During this process, the ChokePoint dataset significantly improved the model's performance in improving face images from security cameras, and after this second training stage, the model became more suitable for camera systems.

In the super-resolution tests performed, the VNet architecture achieved high metric performances in various face image datasets such as CelebA, UTKFace, FEIFace, and MultiPie, and proved effective in obtaining realistic and high-quality images from low-resolution face images.

Looking at the visual test outputs, in tests based on quadrupling the image resolution, VNet showed a performance of sampling in accordance with the original image in restoring missing eyes, mouth, nose and other facial components in the distorted image; it succeeded in repairing and upsampling blurred, noisy and distorted textures.

This study aims to contribute to the literature on improving security systems with super-resolution techniques and to further develop super-resolution architectures and deep learning-based techniques for camera image analysis in the field of forensic informatics.

1. GİRİŞ

Günümüzde güvenlik kameraları kamu ve askeri alanlarda, endüstride, iş yerlerinde, trafikte, toplumsal ve ticari birçok alanda vazgeçilmez bir unsur haline gelmiştir. Giderek yaygınlaşan kamera sistemleri; asayişin ve güvenliğin sağlanması, suçların önlenmesi ve delillerin toplanması gibi kritik amaçlar için kullanılmaktadır [1].

Güvenlik sistemlerinin yaygın kullanımı ve önemine karşın mevcut güvenlik kamera sistemlerinin bir dizi teknik kısıtları bulunmaktadır. Bu sistemlerin görüntü kalitesi genellikle istenen düzeyde değildir ve özellikle de eski tip sensörlere sahip kameralar düşük çözünürlükte ve detayda kayıt yapmaktadır. Bu durum, özellikle uzak mesafelerde netliğin kaybolmasına, düşük ışık koşullarında görüntü kalitesinin düşmesine ve görüntü bozulmalarının ortaya çıkmasına neden olmaktadır. Adli bilişim alanında delil toplama ve şüpheli kimselerin tespitinde, uygun ortam koşulları ve yakın plandaki görüntülerde eski tip güvenlik kameraları nispeten yeterli olsa da daha uzak görüntülerde detaylar kaybolmakta, netlik düşmekte, görüntüde bozulmalar olmakta ve kamera görüntülerinden yüz tespiti oldukça zorlaşmaktadır [2].

Halihazırda kullanılan milyonlarca kameranın yüksek çözünürlükte ve detayda kayıt yapabilen sensörlere sahip yeni kameralarla değiştirilmesi hem çok yüksek maliyetli hem de uzun zamana yayılmış bir süreçtir. Zamanla güvenlik sistemleri ve çoğu kayıt cihazı yenilense dahi yine önemli bir kısmı eski tip olarak kalmaya devam edecektir.

Eski tip düşük çözünürlük ve detayda kayıt yapan cihazlardan elde edilen görüntüler üzerinde adli bilişim alanında delil toplayabilmek, şahısların tespitini yapabilmek ve bu alanda kullanılan görüntü işleme yazılımlarına yeni bir perspektif kazandırmak adına yapılan bu çalışma; kaydedilen görüntünün yeterli çözünürlükte ve uygun ışık koşullarında olmaması, kişilerin kameraya uzak olması gibi çeşitli zorlukların bulunduğu görüntülerden, yeterince tanınabilir görüntüler elde etmeyi amaçlamakta ve bu sayede yüksek donanım maaliyetleri olmadan güvenlik sistemlerine yazılımsal bir iyileştirme sağlamayı amaçlamaktadır.

Düşük çözünürlük ve detaydaki görüntülerin yüksek çözünürlük, detay ve keskinlikteki görüntülere dönüştürülmesi, dijital görüntü işleme alanında süper

çözünürlük (SR) başlığı altında incelenmekte, çeşitli algoritmalar ve yöntemler kullanılarak süper çözünürlük elde edilmeye çalışılmaktadır. Süper çözünürlük genel olarak fotoğraf ve videoların görüntü kalitesinin iyileştirilmesinde ve restorasyonunda, tıbbi görüntüleme cihazlarından alınan görüntülerin netleştirilmesinde, uydu görüntülerinde ve diğer güvenlik sistemlerinde etkin bir şekilde kullanılmaktadır.

Güvenlik sistemlerinde süper çözünürlük, özellikle düşük ışık, gürültü ve bulanıklık gibi zorlu koşullarda, güvenlik kameralarının etkinliğinin artırılması, düşük çözünürlüklü görüntülerin daha yüksek detay ve çözünürlüğe çıkarılarak görüntüdeki kişilerin tanımlanabilir hale getirilmesi amacıyla bir görüntü işleme tekniği olarak kullanılmaktadır. Görüntü işleme teknikleri arasında enterpolasyon tabanlı ve benzeri birçok süper çözünürlük tekniği bulunmakla birlikte yapılan çalışmalar, derin öğrenme tabanlı süper çözünürlük tekniklerinin geleneksel yöntemlere kıyasla daha başarılı olduğunu göstermektedir. Özellikle GAN ve CNN tabanlı yaklaşımlar, düşük çözünürlüklü görüntülerdeki detayları başarılı bir şekilde iyileştirerek güvenlik kameralarındaki yüz tanıma performansını artırmaktadır [3]. Tez kapsamında da kullanılan derin öğrenme tabanlı bu teknikler, düşük çözünürlüklü yüz görüntülerinde kaybolan ince detayları geri getirebilmekte ve böylece görüntüdeki kişilerin tespitinde önemli avantajlar sağlamaktadır [4].

1.1. Literatür Araştırması

Görüntü netleştirme amacıyla kullanılan süper çözünürlük teknikleri Dong ve arkadaşlarının evrişimsel sinir ağlarını süper çözünürlükte kullanması ile derin öğrenme tabanlı yeni bir perspektif kazanmıştır. Önerilen SRCNN ağı üç katmanlı bir yapıya sahiptir. Enterpolasyon yöntemleriyle üst örneklenen girdi görüntüsü, ilk katmanda özellik haritaları çıkarılarak daha yüksek boyutlu temsillere dönüştürülür. İkinci katmanda, temsillerdeki eksik detaylar tamamlanarak yüksek çözünürlükte ve netlikte temsiller elde edilir. Üçüncü katmanda ise, yüksek çözünürlüklü temsiller birleştirilerek yeni görüntü oluşturulur. Bu üç aşamanın sonunda detaylar daha net hale gelmekte ve görüntünün yüksek çözünürlüklü hali ortaya çıkmaktadır [5].

Yine Dong ve ark. tarafından SRCNN'in hesaplama karmaşıklığını azaltmak için dekonvolüsyon kullanarak üst örnekleme yapan FSRCNN önerilmiştir [6].

Kim ve ark. Dong tarafından önerilen ağları daha da derinleştirerek Very Deep Super Resolution – VDSR ağını önermiştir. ImageNet sınıflandırması için kullanılan

VGGNet'e dayanan ağın her katmanında 64 filtre (3x3) kullanarak özellik çıkarımı yapmıştır. İlk ve son katman arasında artık öğrenme bağlantısı kurarak ağın derinleştikçe düşen öğrenme oranını iyileştirmeyi amaçlamıştır [7].

Ağların gittikçe derinleşmesi sonucu aktivasyon fonksiyonlarının sıfıra yaklaşması ve ağın eğitilememesi (Vanishing Gradient) problemine çözüm olarak He ve ark. Tarafından önerilen Artık Ağlar (Residual Network – ResNet) [8], Ledig ve ark. Tarafından SRResNet (Super Resolution Residual Network) olarak süper çözünürlüğe uyarlanmış, derinleşen ağı geriden besleyen bağlantılar sayesinde önceki tekniklere göre daha iyi bir model önerilmiştir [9].

Lim ve ark. SRResNet ağındaki normalizasyon katmanlarını kaldırarak eğitim sırasında gereksiz bellek ve hesaplama maliyetlerini azaltmış, daha derin ve bir yapı kullanılarak modelin kapasitesi artırmış, daha çok artık blok kullanarak daha yüksek PSNR ve SSIM değerleri elde etmiştir [10].

Huang ve ark. Tarafından Yoğun Bağlı Blokların (DenseNet) [11] önerilmesiyle Tong ve ark. Yoğun bağlantıların bilgi akışını artırması, gradyanların daha etkili bir şekilde geri yayılması ve artık öğrenmenin verdiği avantajları kullanarak SRDenseNet ağını önermiş, detay ve kenarları koruyarak yüksek netlik elde etmiştir [12].

Goodfellow ve ark. Tarafından Çekişmeli Üretici Ağların (Generative Adversarial Networks – GAN) [13] önerilmesi üzerine Lim ve ark. SRResNet ile birlikte SRGAN (Super-Resolution using a Generative Adversarial Network) ağını önermiş, klasik doğrusal ağlara alternatif olarak üretici ağları kullanarak daha yüksek doğrulukta çıktılar elde etmiştir [9].

Wang ve ark. SRGAN ağını temel alıp ESRGAN ağını önermiş, ancak artık bloktaki normalizasyon katmanını kaldırıp Üçlü Artık Yoğun Bloklar kullanmış, bir görüntünün diğerine kıyasla daha gerçekçi olup olmadığını öğrenen Relativistic Average GAN ile daha yüksek görsel kalite sağlamıştır [14].

Süper çözünürlük ağları genel görüntüler üzerinde başarılı olurken yüz görüntüleri üzerinde aynı başarıyı gösterememektedir. Yüz bileşenlerinin ve cilt dokularının yüksek yapısal karmaşıklığından dolayı yüz görüntüleri diğer genel görüntülerden ayrı olarak değerlendirilmiş ve bu konuda özel çalışmalar yapılmıştır. Zhou ve ark. yüz görüntülerinin süper çözünürlüğü (Face Super-Resolution – FSR) için iki kanallı bir evrişimli sinir ağı (A Bi-Channel Convolutional Neural Network – BCCNN)

önermiştir. Süper çözünürlük elde etmek için kanallardan biri girdi görüntüsünden yüz bölgelerine ait özellik çıkarımları yaparken, diğeri de düşük çözünürlüklü girdi görüntüsü ile çıkarılan özellikleri uygun bir şekilde birleştirir. BCCNN, Gauss ve hareket bulanıklığına sahip yüz görüntülerinde klasik CNN mimarilerine kıyasla daha yüksek metrik başarımlar elde etmiştir [15].

Yu ve ark. yüz görüntülerinin restorasyonunda ilk defa GAN tabanlı bir ağ olan UR-DGN (Ultra-Resolution by Discriminative Generative Networks) ağını önermiştir. Burada ayrımcı ağ insan yüzündeki önemli bileşenleri öğrenirken, üretici ağ bu bileşenler ile girdi görüntüsünü birleştirmektedir. Üretici modelde yüz görüntülerini gerçek olanlara daha benzer hale getirmek için piksel bazlı bir hata fonksiyonu ile birlikte ayrımcı ağın hata beslemesi kullanılmıştır. Bu sayede çok küçük girdiler 8 katına kadar üst örneklenebilmiştir.

Klasik konvolüsyon ağlarında, girdi görüntüsü ağda ilerleyerek özellik çıkarımı yapılırken, bazı kanalların diğerlerine göre daha fazla özellik taşımasına rağmen her kanala aynı muamele yapılır. Zhang ve ark. ise her kanala farklı önem seviyeleri (ağırlıklar) atayarak kanal odaklı (Channel Attention) bir özellik çıkarımı önermiş, genel süper çözünürlük alanında verimli sonuçlar elde etmiştir [16].

Zhao ve ark. yüz görüntülerinde göz, burun ve ağız gibi önemli bölgeleri korumak, daha gerçekçi ve detaylı restorasyon yapabilmek için Semantik Kanal Odaklı (Semantic Channel Attention) özellik çıkarma yöntemini kullanmış, bu kanal odaklı mekanizma ile girdi görüntüsünün farklı bölgelerine farklı ağırlıklar vererek önemli yüz özelliklerini vurgulayan SAAN (Semantic Attention Adaptation Network for Face Super-Resolution) ağını önermiş, yüz görüntülerinin süper çözünürlüğünde verimli sonuçlar elde etmiştir [17].

Lu ve ark. SAAN (Face Hallucination via Split-Attention in Split-Attention Network) ağının önermiş, girdi kanallarını bölerek her biri için kanal odaklı (Split-Attention) özellik çıkarımını kullanmış, bu sayede ağın hem yüz dokularına hem de yapısal bölgelere odaklanmasını sağlayarak daha ayrıntılı yüzler elde etmiştir [18].

Dosovitskiy ve ark. doğal dil işlemedeki Transformer mimarisini Vision Transformers (ViT) adıyla görüntü işlemeye uyarlamasıyla görüntü işlemede kullanılan konvolüsyon ağlarına bir alternatif ortaya çıkmıştır [19]. Daha sonra Transformer mimarisini kullanan birçok süper çözünürlük ağı önerilmiştir.

Wang ve ark. yüz görüntülerinin daha etkin bir biçimde yeniden yapılandırılması için hem CNN'lerin hem de Transformer yapısının güçlü yönlerinden yararlanmış, CNN ve Vision Transformer (ViT) tabanlı TANet (A New Paradigm for Global Face Super-Resolution via Transformer-CNN Aggregation Network) mimarisini önermiştir. CNN kısmı yüzün ince detaylarının ve yerel özelliklerinin restorasyonunda kullanılırken, Transformer kısmı Global Attention mekanizması ile genel yüz yapısının tutarlılığını korumada kullanılmıştır. Bu sayede yeniden yapılandırılmış yüz görüntülerinde daha yüksek biyometrik doğruluk elde edilmiştir [20].

Gao ve ark. yine CNN-Transformer hibrit bir ağ olan CTCNet (A CNN-Transformer Cooperation Network for Face Image Super-Resolution) mimarisini önermiştir. U-Net benzeri bir yapısı olan CTCNet, kodlayıcı tarafında Facial Structure Attention Unit ve Transformer Blok birleşimi bir yapı kullanarak hem yerel yüz dokularını hem de genel yüz yapı bilgisini elde edebilmiştir. Ara geçiş tarafında Feature Refinement Modülü ile çıkarılmış özellikler arasında esas yüz yapısına odaklanılırken, kod çözücü tarafında Multi-Scale Feature Fusion Units modülü ile de çıkarılan lokal ve global özellikler birleştirilip üst örneklenerek yüksek çözünürlükte görüntüler elde edilmiştir. CTCNet'in karşılıklı bağlantılara sahip U-Net benzeri mimarisi ve gelişmiş modüler yapısı sayesinde çeşitli kıyaslamalarda birçok mimariye göre daha yüksek metrik başarımlar elde etmiştir [21].

Güvenlik kamera sistemlerinden elde edilen görüntülerin süper çözünürlükle iyileştirilmesi konusunda Alkanhal ve ark. Very-Deep Super-Resolution (VDSR) ağını CelebA veri seti [22] ile eğitmiş ve QMUL-SurvFace veri seti [23] üzerinde yapılan testlerde, bikübik enterpolasyona göre %7 daha yüksek PSNR, %3 daha yüksek SSIM ve %45 daha yüksek yüz tanıma iyileştirmeleri elde etmiştir [24].

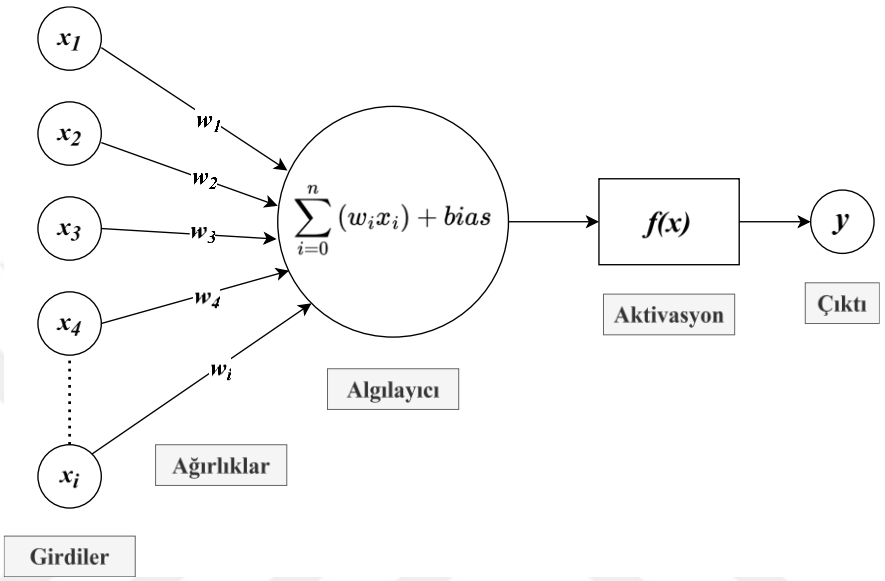
Aakerberg ve ark. eğitim için kullanılan veri setinden bozulmuş görüntüleri elde ederken, dış dünyaya daha uygun olması adına daha gerçekçi bir bulanıklık matrisi, gürültü dağılımı ve JPEG bozulma faktörleri kullanan yeni bir görüntü bozma yöntemi kullanmıştır. ESRGAN ağının kullanıldığı çalışmada, VGG-Loss yerine LPIPS-Loss kullanımı önerilmiştir. Eğitim için SiblingsDB [25], Radboud [26] ve FFHQ [27] veri setleri kullanılarak yüksek PSNR, SSIM ve LPIPS değerleri elde edilmiştir [28].

Yüz görüntülerinin süper çözünürlükle restore edilmesi veya üst örneklenmesi konusunda yapılan çalışmalarda elde edilen görüntünün kalitesi, yüksek metrik

başarımları ve görsel iyileştirmeler çoğu zaman ön planda tutulmakta, biyometrik açıdan aslına uygunluk (Fidelity) ikinci planda kalmaktadır. Güvenlik kameraları özelinde yapılan bu çalışmada, girdi görüntüsünden süper çözünürlükteki çıktı görüntüsü elde edilirken, bir yandan asıl görüntüye olan yapısal benzerliği korurken diğer yandan da yüksek metrik başarımlar ve görsel iyileştirmeler sağlayabilen bir model amaçlanmaktadır.

2. METODOLOJİ

2.1. Yapay Sinir Ağları



Şekil 2.1. Yapay Sini Ağı Yapısı.

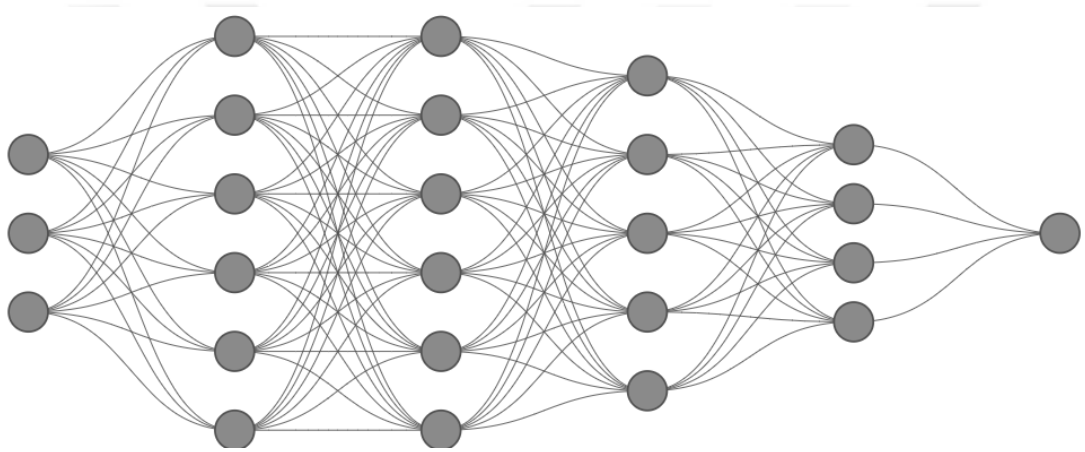
Yapay Sinir Ağları (Artificial Neural Networks – ANN), insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirebilmek amacı ile geliştirilen sistemlerdir. Beyindeki biyolojik sinir ağlarının yapısını, öğrenme ve hatırlama kabiliyetini taklit eden YSA'lar, insan beyni örnek alınarak öğrenme sürecinin matematiksel olarak modellenmesi sonucu ortaya çıkmıştır. Yapay sinir hücrelerinin yani nöronların birbirine bağlanmasıyla oluşan yapay sinir ağlarında öğrenme işlemi örnekler kullanılarak gerçekleştirilir, daha sonra görülmemiş girdiler için istenen sonuçların üretilmesi beklenir.

YSA'ların temel yapı taşı olan algılayıcıların (Single Layer Perceptron), giriş katmanı ve ağırlık değerleri, aktivasyon fonksiyonu ve çıkış katmanından oluşan basit bir yapısı vardır. Bu iki katman arasında herhangi bir gizli katman bulunmaz. Algılayıcı çalışma prensibinde, başlangıçta girdi verilerine rastgele bir ağırlık değeri atanır ve algılayıcı içerisinde her bir girdi ile ağırlıklar çarpılır. Daha sonra bir Bias ağırlığı eklenerek sayısal bir değer elde edilir.

Çıktı katmanında, olasılıksal bir değer elde etmek amacıyla, elde edilen değerler uygun bir aktivasyon fonksiyonundan geçirilir. Çıktı değeri ile gerçek değer uygun bir hata fonksiyonundan geçirilerek bir hata değeri elde edilir. Elde edilen hata istenen seviyede değilse tekrar en başa dönülerek ağırlık değerleri güncellenir. Bu güncelleme ve girdi işlemleri istenen hata değerine ulaşılana kadar devam eder [29].

Tek bir algılayıcı kompleks problemleri çözmekte yetersizdir. Algılayıcıların kullanımı özellikle basit lineer problemleri çözmek için uygundur. Karmaşık görevler için birden çok algılayıcı birleştirilerek çok katmanlı yapay sinir ağları (Multilayer Perceptrons) geliştirilmiştir.

2.1.1. Derin öğrenme



Şekil 2.2. Örnek Bir Derin Öğrenme Ağı.

Büyük veri kümelerinin erişilebilir hale gelmesi ve hesaplama gücünün istenen seviyelere gelmesiyle derin sinir ağları (Deep Neural Networks – DNN) olarak bilinen çok katmanlı yapılar ortaya çıkmıştır. Derin sinir ağları çok katmanlı algılayıcıların yapısal olarak bir uzantısı olmakla beraber çok daha fazla sayıda gizli katman içermektedir. Derin sinir ağları bu gizli katmanlar içerisinde veriyi işler ve her katman, bir önceki katmana göre giderek daha karmaşık ve soyut hale gelen girdi verilerine ait temsilleri öğrenip bir sonraki katmana aktarmak amacıyla birbirine bağlanmıştır. Bir ağ yüzlerce katmandan ve binlerce nörondan oluşabilir. Ağın derin olması, girdi verilerinin karmaşıklığını ve çok boyutluluğunu daha iyi kavrayabilmesi açısından gereklidir.

Derin öğrenme ağları üç ana katmandan oluşur; giriş katmanı (Input Layer), gizli katmanlar (Hidden Layers) ve çıkış katmanı (Output Layer). Bilgiler ağa girdi katmanından iletilir ve ara katmanlarda işlenerek çıktı katmanına gönderilir. Çıktı katmanında girdiler bir aktivasyon fonksiyonu kullanılarak çıkarım yapmaya uygun, olasılıksal bir değere dönüştürülür. Daha sonra elde edilen değerler amaca uygun olarak ağın hata fonksiyonuna girdi olarak verilir. Oluşan hata değeri istenen düzeyin üzerinde ise ağdaki tüm ağırlıklar güncellenir ve tekrar ağa girdi alınır. Burada asıl maksat, ağa gelen girdilerin, ağa ait ağırlık değerleri kullanılarak istenen çıktılara dönüştürülmesidir.

Derin öğrenme ağlarının girdilere uygun çıktılar üretebilmesi için ağdaki ağırlıkların doğru değerlere sahip olması gerekmektedir. Bu ağırlıkların optimize edilmesinde, ağın eğitimi için kullanılan veri seti oldukça önemlidir. Yeterli miktarda ve dengeli bir veri seti ile eğitilen ağırlıklar, modelin farklı senaryoları ve karmaşık özellikleri öğrenmesine olanak tanırken, yetersiz veya dengesiz bir veri seti modelin genelleme kabiliyetini olumsuz yönde etkileyebilmektedir.

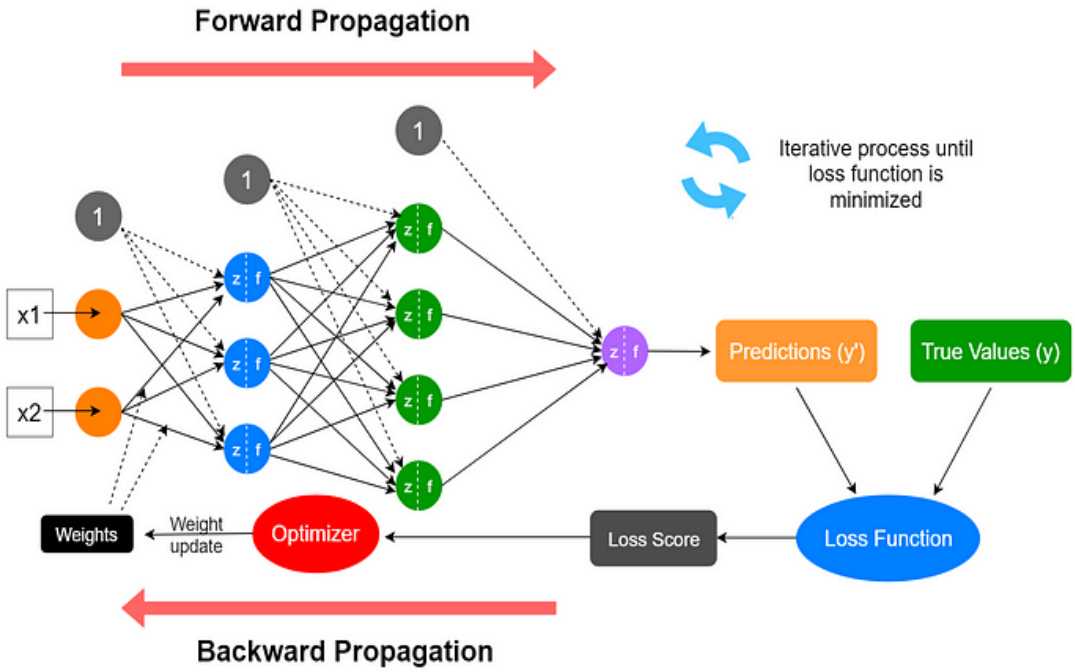
Derin öğrenme ağları görüntü işleme, doğal dil işleme ve otonom sistemler gibi alanlarda yaygın olarak kullanılmaktadır. Bu ağların en büyük avantajlarından biri, verilerden otomatik olarak özellik çıkarıp istenen çıktıyı üretebilmesidir. Bu ağları inşa ederken nöronların birbirleriyle bağlantı şekline ve nöronlarda yapılan işlemlere göre farklı ağ mimarileri geliştirilmiştir.

Yaygın olarak kullanılan derin öğrenme mimarilerine Evrişimli Sinir Ağları (Convolutional Neural Networks – CNN), Artık Ağlar (Recurrent Neural Networks – RNN) ve Çekişmeli Üretici Ağlar (Generative Adversarial Networks – GAN) örnek olarak gösterilebilir [29].

2.1.2. Ağırlıkların öğrenilmesi

Yapay sinir ağlarını eğitmek ve giriş verilerinden istenen çıktıları elde etmek için ağın eğitilmesi gerekir. Bir ağı eğitmekten kasıt, nöronları birbirine bağlayan mantıksal bağlantılara ait her bir ağırlığın doğruya en yakın değere sahip olması demektir. Çünkü giriş verileri sinir ağlarında ilerlerken bu ağırlıkların etkisine maruz kalarak manipüle olmakta ve ağırlıkların etkisiyle istenen çıktılar oluşmaktadır. Bu sebeple ağırlıklar ne kadar doğru ve dengeli değerlere sahip olursa çıkış değerlerinin doğruluğu da o derece yüksek olmaktadır [30].

2.1.3. İleri-Geri yayılım (Forward-Backward propagation)



Şekil 2.3. Forward-Backward Propagation [31].

İleri Yayılım (Forward Propagation), bir sinir ağında giriş verilerinin, girişten çıkışa doğru hareketidir. Girdiler her katmanda ağırlıklarıyla çarpılıp katmandan katmana aktarılarak ağ üzerinde ilerler ve çıkış katmanına ulaşır. Daha sonra bu girdiler çıkış katmanındaki aktivasyon fonksiyonundan geçirilerek ağın tahmin değeri hesaplanır. Tahmin değeri ile gerçek değer bir hata fonksiyonunda karşılaştırılır ve ne kadar sapma veya hata olduğu (loss) belirlenir. Bu işlemler İleri Yayılım olarak adlandırılır. Geri Yayılım (Backward Propagation) işleminde ise belirlenen hata değerini minimize etmek amacıyla nöronların ağırlıkları bir optimizasyon fonksiyonu ile güncellenir. Bu güncelleme işlemine de Geri Yayılım denir. Bu ileri ve geri yayılım döngüsü hata değerini belirlenen minimum değere düşürene kadar devam eder. Her bir döngü bir epok olarak adlandırılır ve döngü sayısı artırılarak ağın eğitimi gerçekleştirilir [32].

2.1.4. Hata (Loss) fonksiyonları

Hata fonksiyonları, yapay sinir ağları ve diğer makine öğrenimi modellerinde, ağın tahmin ettiği değerlerin gerçek değerlerden ne kadar farklı olduğunu gösteren fonksiyonlardır. Bu fonksiyonlar, ağın performansını değerlendirmek ve öğrenme sürecindeki hataları azaltarak ağı optimize etmek için kullanılır.

Hata fonksiyonu seçimi öğrenme görevinin türüne ve ağın çıktı tipine bağlıdır. Kullanılan bazı hata fonksiyonlarına; Ortalama Mutlak Hata (Mean Absolute Error – MAE), Ortalama Karesel Hata (Mean Squared Error – MSE) ve İçerik Kaybı (Content Loss) fonksiyonları örnek verilebilir.

2.1.4.1. Ortalama mutlak hata (Mean absolute error – MAE)

Ortalama Mutlak Hata, regresyon problemlerinde tahmin ve gerçek değerlerin farkını hesaplarken, görüntü işlemede çıktı görüntüsü ile referans görüntü piksellerinin farkını hesaplarken yaygın olarak kullanılır. MAE değeri sıfıra ne kadar yakınsa hata o kadar az kabul edilir.

$$MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2.1}$$

Denklem 2.1'deki y gerçek değerleri, $\hat{y}$ tahmin değerlerini, n ise veri sayısını göstermektedir.

2.1.4.2. Ortalama karesel hata (Mean squared error – MSE)

Ortalama Karesel Hata, Ortalama Mutlak Hata ile benzer şekilde regresyon ve tahmin problemlerinde kullanılır. MSE, tahmin ile gerçek değerin farkının karesini aldığı için büyük farkların daha fazla ağırlığı vardır. Bu durum büyük hataların daha belirgin hale gelmesini sağlarken küçük hatalarda MSE daha yetersiz kalmaktadır.

$$MSE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.2}$$

Denklem 2.2'deki y gerçek değerleri, $\hat{y}$ tahmin değerlerini, n ise veri sayısını göstermektedir.

2.1.4.3. İçerik kaybı (Content loss)

İçerik kaybı, stil transferi bağlamında bir ağın, bir görüntünün içeriğini başka bir görüntüye aktarırken ne kadar başarılı olduğunu ölçer. Diğer yandan görüntü restorasyonunda iki görüntünün birbirinden ne kadar farklı olduğunu belirlemek için de kullanılır. İçerik kaybını elde etmek için önceden eğitilmiş bir sinir ağı kullanılarak her iki görüntü için özellik haritası (Feature Map) elde edilir ve bu iki özellik haritasının ne kadar farklı olduğuna bakılır. Özellik çıkarma işlemi için genellikle VGG veya ResNet gibi gelişmiş bir sinir ağı kullanılır [33].

$$L_{\text{content}} = \frac{\sum_{l=1}^{n} \left(\phi_l(y) - \phi_j(\hat{y})\right)^2}{C_l H_l W_l} \tag{2.3}$$

Denklem 2.3'teki y gerçek değerleri, $\hat{y}$ tahmin değerlerini, n katman sayısını, ϕ_l özellik haritası çıkarmak için kullanılan ağın ilgili katmanını, C_l, H_l, W_l sırasıyla kanal sayısını, düşeyde ve yatayda piksel sayısını göstermektedir.

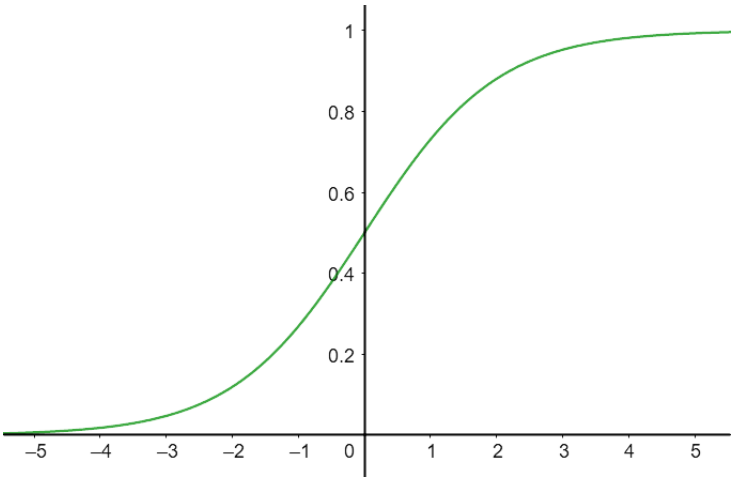
2.1.5. Aktivasyon fonksiyonları

Aktivasyon Fonksiyonları (Activation Functions), yapay sinir ağlarında bir birimin veya nöronun çıkışını belirlemek için kullanılan matematiksel fonksiyonlardır. Bu fonksiyonlar, sinir ağlarının öğrenme yeteneklerini artırmak ve daha karmaşık ilişkileri modellemek için kritik bir rol oynar.

Aktivasyon fonksiyonları sinir hücrelerinin çıkışını belirlerken çıkışa doğrusal olmayan bir yapı (Non-Linearity) katarak sinir ağının daha karmaşık öğrenme görevlerini başarıyla gerçekleştirebilmesini sağlarlar. Basit doğrusal ağların öğrenme ve daha karmaşık problemleri çözebilme kapasitesi sınırlıyken aktivasyon fonksiyonları sayesinde sinir ağları daha karmaşık veri yapılarını ve örüntülerini ele alabilmektedir.

Yapay sinir ağlarının geniş bir uygulama yelpazesi içinde başarılı olmalarının temel nedenlerinden biri de bu fonksiyonlardır. Birçok aktivasyon fonksiyonu bulunmakla beraber Sigmoid, Softmax, ReLU, LeakyReLU ve Tanh fonksiyonları genel olarak kullanılan aktivasyon fonksiyonlardır [34].

2.1.5.1. Sigmoid fonksiyonu



Şekil 2.4. Sigmoid Fonksiyonu.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.4}$$

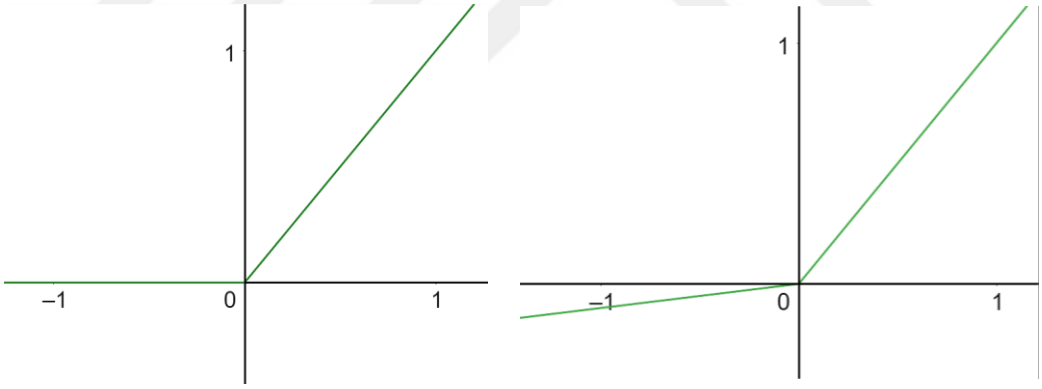
Sigmoid aktivasyon fonksiyonu (Denklem 2.4), giriş değeri ne olursa olsun çıkış değerini 0 ile 1 arasında bir değere dönüştürür. Sigmoid özellikle ikili sınıflandırma problemlerinde çıktıları olasılıksal değerlere dönüştürerek yorumlamak için kullanılır.

2.1.5.2. Softmax fonksiyonu

$$f(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_i}} \ (i = 1.,,,.K).\ (z = z_1,,,z_K) \tag{2.5}$$

Softmax aktivasyon fonksiyonu (Denklem 2.5), Sigmoid fonksiyonuna benzer şekilde giriş değerini 0 ile 1 arası olasılıksal bir değere dönüştürür. Sigmoid fonksiyonu tek giriş alıp ikili sınıflandırma yaparken, Softmax fonksiyonu bir girdi matrisi alır ve bu matrisin her bir değerinin olasılıksal karşılığını üreterek çoklu sınıflandırma gerektiren ve sınıfların sayısal olarak olasılığının belirlendiği durumlarda kullanılır.

2.1.5.3. ReLU ve LeakyReLU fonksiyonları



Şekil 2.5. ReLU ve LeakyReLU Fonksiyonları.

$$f(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \tag{2.6a}$$

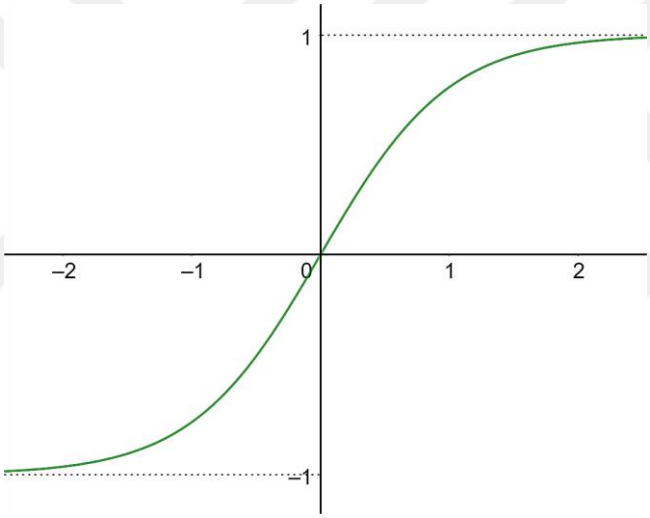
$$f(x) = \begin{cases} \alpha x & x < 0 \\ x & x \geq 0 \end{cases} \tag{2.6b}$$

ReLU (Rectified Linear Unit) aktivasyon fonksiyonu (Denklem 2.6a), pozitif girişler için lineer bir çıkış üretirken negatif girişler için sıfır çıkış üretir. Basit yapısı ve düşük hesaplama gereksinimi nedeniyle ReLU fonksiyonu görüntü işlemeye dayalı ağlarda yaygın olarak kullanılır.

LeakyReLU aktivasyon fonksiyonu (Denklem 2.6b) ise ReLU kullanıldığında bazı nöronların tamamen sıfır çıkış üretmesi sonucu oluşan "ölü nöron" problemini önlemek için kullanılmaktadır. Fonksiyonda kullanılan α, sabit bir sayı olup genellikle 0,01 gibi küçük bir sayıdır. Bu katsayı sayesinde negatif giriş değerleri için bir miktar tolerans tanınmaktadır.

Negatif eğimlerin optimizasyonu konusunda, LeakyReLU fonksiyonunda kullanılan α katsayısını sabit tutmak yerine, PReLU (Parametrik ReLU) gibi dinamik olarak değiştiren yöntemler de vardır. Bu durumda eğimin değeri, öğrenilebilen bir parametre olarak modelin bir parçasıdır ve her bir nöronun negatif eğimi eğitilebilmekte ve dinamik olarak optimize edilebilmektedir.

2.1.5.4. Tanh fonksiyonu

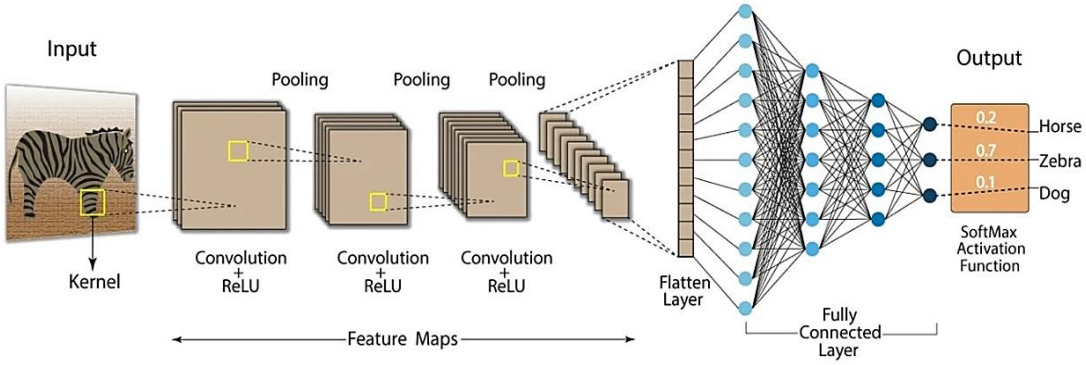


Şekil 2.6. Tanh Fonksiyonu.

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.7}$$

Tanh (Hiperbolik Tanjant) aktivasyon fonksiyonu (Denklem 2.7), Sigmoid fonksiyonuna benzer olmakla beraber Sigmoid fonksiyonunun sıfıra yaklaştığı uçlarda daha geniş bir çıkış aralığına sahiptir. Bu fonksiyon, girdi değerini [-1, 1] aralığına dönüştürerek negatif girdiler için negatif, pozitif girdiler için pozitif sonuçlar üretmek ve daha geniş bir çıktı aralığı elde etmek için kullanılır.

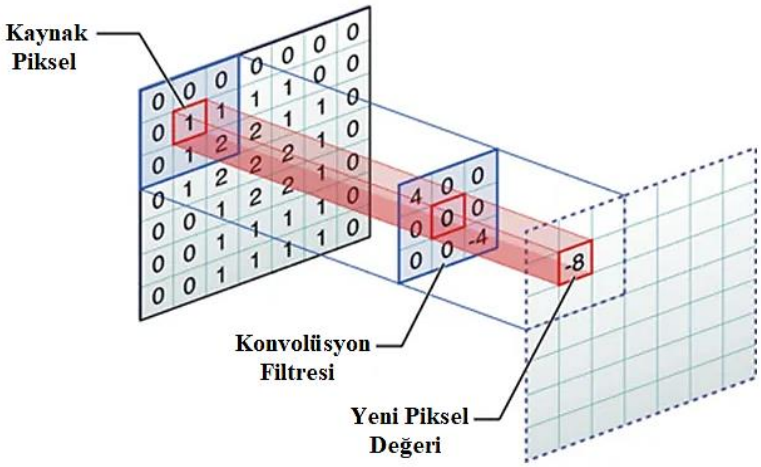
2.2. Evrişimli Sinir Ağları



Şekil 2.7. ESA Kullanarak Temel Bir Sınıflandırma Ağı Mimarisi [35].

Evrişimli Sinir Ağları (Convolutional Neural Networks – CNN), YSA'ların Konvolüsyon kullanılarak gerçekleştirilmiş halidir. Bu ağlar özellikle görsel verilerle çalışan derin öğrenme modelleridir, giriş verilerinin görsel yapısını kavramak ve bu yapılardan özellikler çıkarmak için kullanılır. ESA'lar özellikle görüntü tanıma, nesne tespiti, yüz tanıma ve diğer görsel işlemlerde büyük başarı elde etmiştir. Bu mimari veri içerisinde hiyerarşik özellikleri anlamak ve genelleştirmek için oldukça etkilidir. En bilinen ESA mimarileri; LeNet, AlexNet, VGG, ResNet, ImageNet ağlarıdır.

2.2.1. Konvolüsyon işlemi

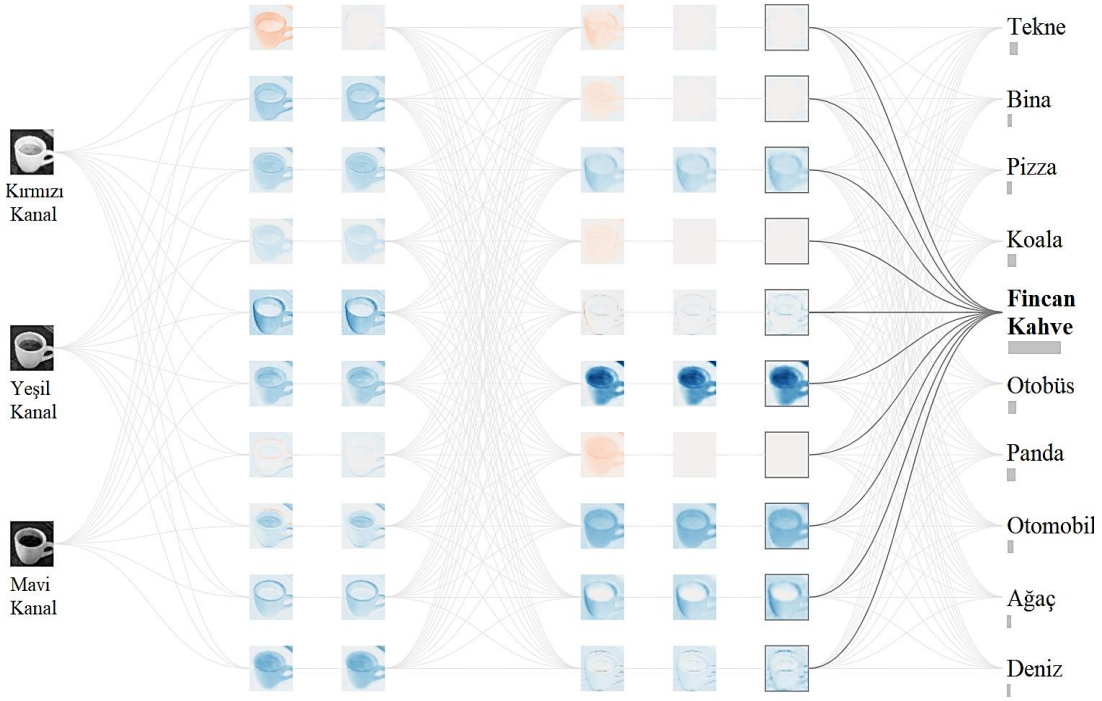


Şekil 2.8. Görüntü Matrisine Konvolüsyon Uygulanması [36].

ESA'ların en önemli özelliği olan Evrişim (Convolution), boyutu önceden belirlenen iki boyutlu bir konvolüsyon matrisinin (filtrenin) görüntü üzerinde gezdirilerek, filtre matrisindeki elemanlar ile ona karşılık gelen görüntü matrisindeki değerlerin çarpılıp toplanması ve bulunan değerin merkez piksele yazılmasıdır. Evrişim işlemi ağa girdi olarak verilen görüntünün her bir renk kanalına ayrı ayrı uygulanır.

Her bir görüntü pikseli için filtrelerin görüntü üzerinde kaydırılması suretiyle yapılan bu konvolüsyon işleminde amaç, görüntüyü tanımlayıcı ve ayırt edici bir özellik haritası oluşturmaktır. Bu filtre matrisleri ile giriş görüntüsünde, o görüntüyü tanımlayan kenarlar, köşeler, dokular ve desenler yakalanmakta, bu sayede görüntü öğrenilebilmektedir [37].

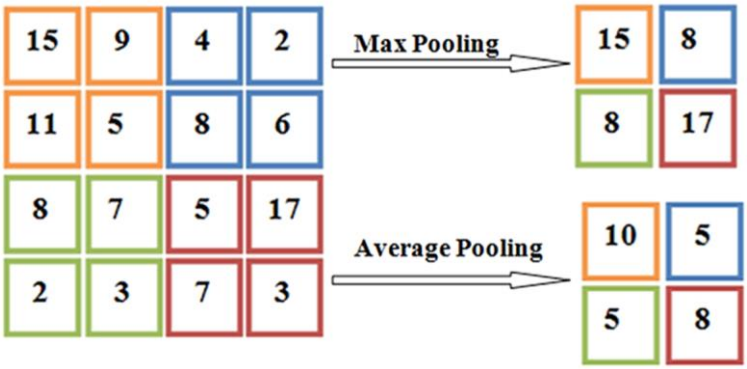
2.2.2. Evrişimsel katman



Şekil 2.9. Evrişim ile Özellik Çıkarımı ve Sınıflandırma [38].

Evrişim katmanı (Convolutional Layer), derin öğrenme modellerinde özellik çıkarma (Feature Extraction) işlemini yürüten, girdilerin bağlandığı ağın ilk katmanıdır. Evrişim katmanları girdi görüntüsünden belirli özellikleri çıkarmak için evrişim işlemini kullanmakta ve bu işlemi yapabilmek için çeşitli filtre matrislerine ihtiyaç duymaktadır. Bu filtreler belirli desenleri, kenarları, köşeleri, dokuları vb. algılamak için kullanılan matrislerdir. Her bir filtre girdi üzerinde kaydırılarak giriş verilerinden özellik haritaları (Feature Map) üretir ve bu haritalar sayesinde girdi verisi tanımlanıp ayırt edilebilir hale gelmektedir. Girdi görüntülerinden karmaşık yapıları ve özellikleri öğrenebilmek amacıyla derin öğrenme ağlarında evrişim katmanları sıkça kullanılmaktadır [37].

2.2.3. Ortaklama (Pooling) katmanı



Şekil 2.10. Max Pooling ve Average Pooling

Ortaklama katmanı (Pooling Layer), ESA'larda yaygın olarak kullanılan bir katmandır. Bu katmanların temel görevi özellik haritalarının boyutunu azaltarak hesaplama ve işlem yükünü azaltmak, her bir bölgenin en önemli özelliklerini koruyarak özellik haritalarını özetlemek, yüksek seviyeli özelliklerin öğrenimini ve daha etkili bir özellik çıkarımı yapılmasını sağlamaktır.

Ortaklama yapabilmek için öncelikle gezdirilecek çerçeve boyutu belirlenir. Genel olarak Max Pooling ve Average Pooling olarak iki türlü ortaklama vardır. Max Pooling yapılırken seçilen çerçeve, görüntü üzerinde gezdirilir ve kapsadığı hücrelerden sayısal olarak en büyüğü seçilir. Daha sonra tüm hücreler birleştirilerek bu maksimum değer içerisine yazılır. Average Pooling'te ise maksimum değer yerine tüm değerlerin ortalaması alınarak hücre içine yazılır [37].

2.2.4. Tam bağlantılı katman (Fully connected layer – Dense layer)

Tam bağlantılı katmanlar, önceki katmandaki her bir nöronun sonraki katmandaki tüm nöronlara bağlı olduğu yapılardır. Ağın eğitimi esnasında her bağlantı, belirli bir ağırlık değeri ile çarpılır ve bu ağırlıklar öğrenme sürecinde güncellenir. Tam Bağlantılı Katman genellikle sinir ağının son aşamalarında kullanılır ve ağın son kararını verirken önemli bir rol oynar.

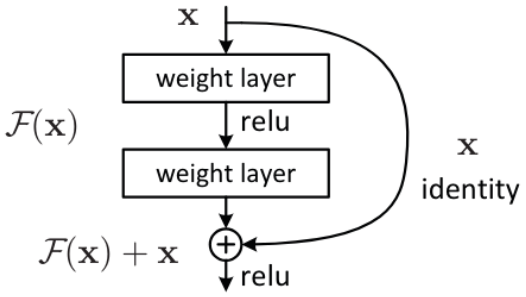
2.2.5. Çıktı katmanı (Output layer)

Çıktı katmanı bir yapay sinir ağının en son katmanıdır. Ağın gerçek dünya problemlerine yönelik tahminlerini ve çıktılarını sağladığı için bir yapay sinir ağının en önemli bileşenlerinden biridir. Ağın önceki katmanlarında yapılan işlemlerin sonucunu temsil eder ve nihai çıktıyı sağlar.

Bu katmanlar genellikle ağın amacına uygun bir görevi (Örneğin sınıflandırma veya regresyon gibi) gerçekleştirmek üzere özel bir aktivasyon ve hata fonksiyonu içerir. Problemin doğası, veri tipi ve ağın amacına bağlı olarak çıktı katmanı farklı yapılandırmalara ve işlevlere sahip olabilir. Sınıflandırma problemleri için genellikle çıktı katmanında Softmax aktivasyon fonksiyonu ve Çapraz Entropi (Cross-Entropy) hata fonksiyonu kullanılırken, görsel işlemler için ReLU aktivasyon fonksiyonu ve Ortalama Karesel Hata (Mean Squared Error) fonksiyonu kullanılmaktadır.

2.3. Artık Ağlar (Residual Networks)

Katman sayısını artırarak ağları derinleştirmek her zaman daha iyi sonuç vermemekte ve belirli bir derinliğin ardından ağın doğruluk oranı düşmektedir. Artık ağlar derin ağların eğitimini kolaylaştırmak ve ağ derinleştikçe performans düşüşü yaşamasını önlemek için tasarlanmıştır.



Şekil 2.11. Artık Ağ Bloğu.

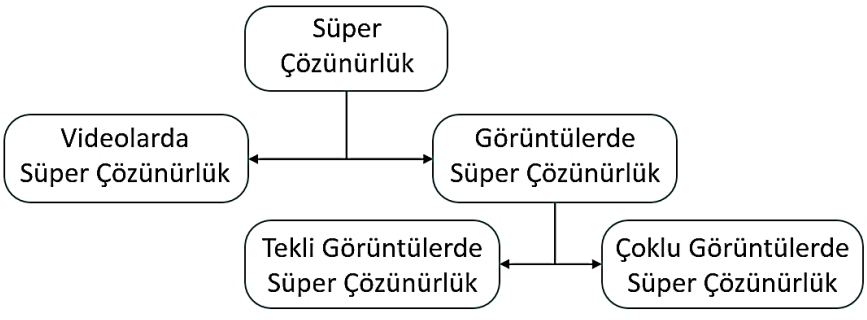
Artık ağların en küçük yapısal bloğu (Şekil 2.11) klasik ağlardan farklı olarak ekstra bağlantılar içermektedir. Katmanlar arası yapılan bağlantının yanı sıra iki veya daha fazla katman öncesi ile bağlantı kurularak derinleşen ağın önemli özellikleri kaybetmemesi amaçlanmaktadır.

Artık Ağlar derin öğrenme alanında önemli bir yere sahiptir. Karmaşık öğrenme görevlerinde yüksek doğruluk sağlayan bu mimari, görüntü sınıflandırma ve segmentasyonunda, nesne tespitinde, doğal dil işlemede ve birçok alanda yaygın olarak kullanılmaktadır [40].

3. SÜPER ÇÖZÜNÜRLÜK

Süper çözünürlük, düşük çözünürlüklü bir görüntüyü (Low-Resolution – LR), görüntü detaylarını ve netliğini koruyarak daha yüksek çözünürlüklü bir görüntüye (High-Resolution – HR) dönüştürme işlemidir. Süper çözünürlük, özellikle görüntü kalitesini artırma, görsel netliğin iyileştirilmesi ve çözünürlük artarken detayların geri kazanımı amaçlarıyla; görüntü işleme, video ve görüntü restorasyonu, tıbbi görüntüleme, uydudan izleme ve güvenlik sistemleri gibi çeşitli alanlarda kullanılmaktadır.

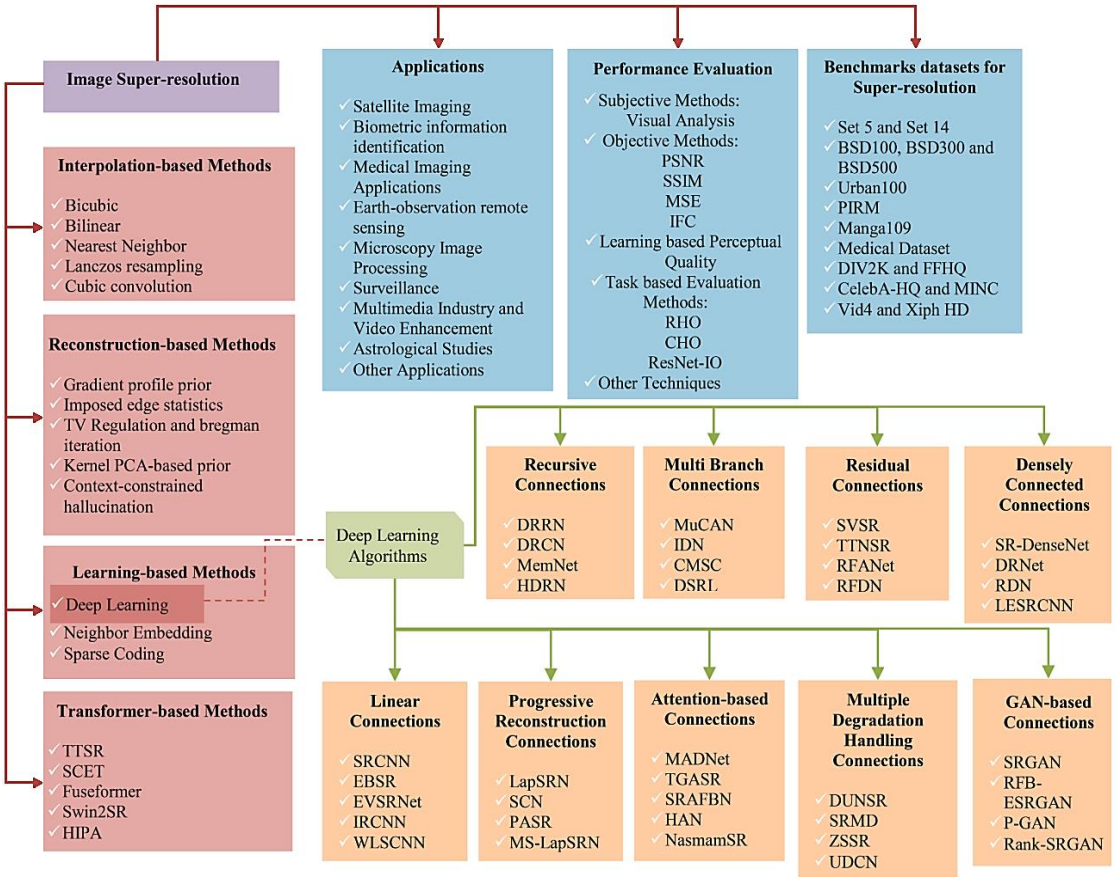
Süper çözünürlük teknikleri, geleneksel yöntemler ve öğrenme tabanlı yöntemler olmak üzere iki ana yaklaşım içermektedir. Geleneksel yöntemler enterpolasyon tekniklerine dayanmakta olup eksik kalan pikselleri matematiksel işlemlerle tahmin ederek üst örneklenmiş görüntüdeki boşlukları doldurulmaya çalışır. Ancak bu yöntemlerde sınırlı sayıda bilgiye sahip olunması nedeniyle çözünürlük artırıldığında detaylar ve netlik istenen düzeyde artmamaktadır. Son yıllarda yapay sinir ağları ve derin öğrenme ağlarının gelişmesiyle birlikte önerilen öğrenme tabanlı süper çözünürlük teknikleri, düşük çözünürlüklü görüntülerdeki karmaşık örüntüleri ve dokuları öğrenerek yüksek çözünürlüklü tahminler üretebilmektedir. Bu yöntemler, genellikle büyük veri kümeleri üzerinde eğitilerek pikseller arasındaki ilişkiyi öğrenir ve eksik kalan detayları doldurma konusunda oldukça etkilidir.



Şekil 3.1. Süper Çözünürlük Genel Uygulama Biçimleri.

Öğrenme tabanlı süper çözünürlük teknikleri hem videolar hem de görüntüler üzerine uygulanabilmektedir. Görüntülere uygulanırken hem tek bir görüntünün hem de birden çok görüntünün birlikte ele alındığı yöntemler vardır. Bu çalışmada ise tekli görüntüler üzerine uygulanan öğrenme tabanlı süper çözünürlük yöntemleri ele alınmaktadır.

Aşağıdaki şekilde (Şekil 3.2) detaylı olarak tekli görüntülerde süper çözünürlük taksonomisi gösterilmiştir.



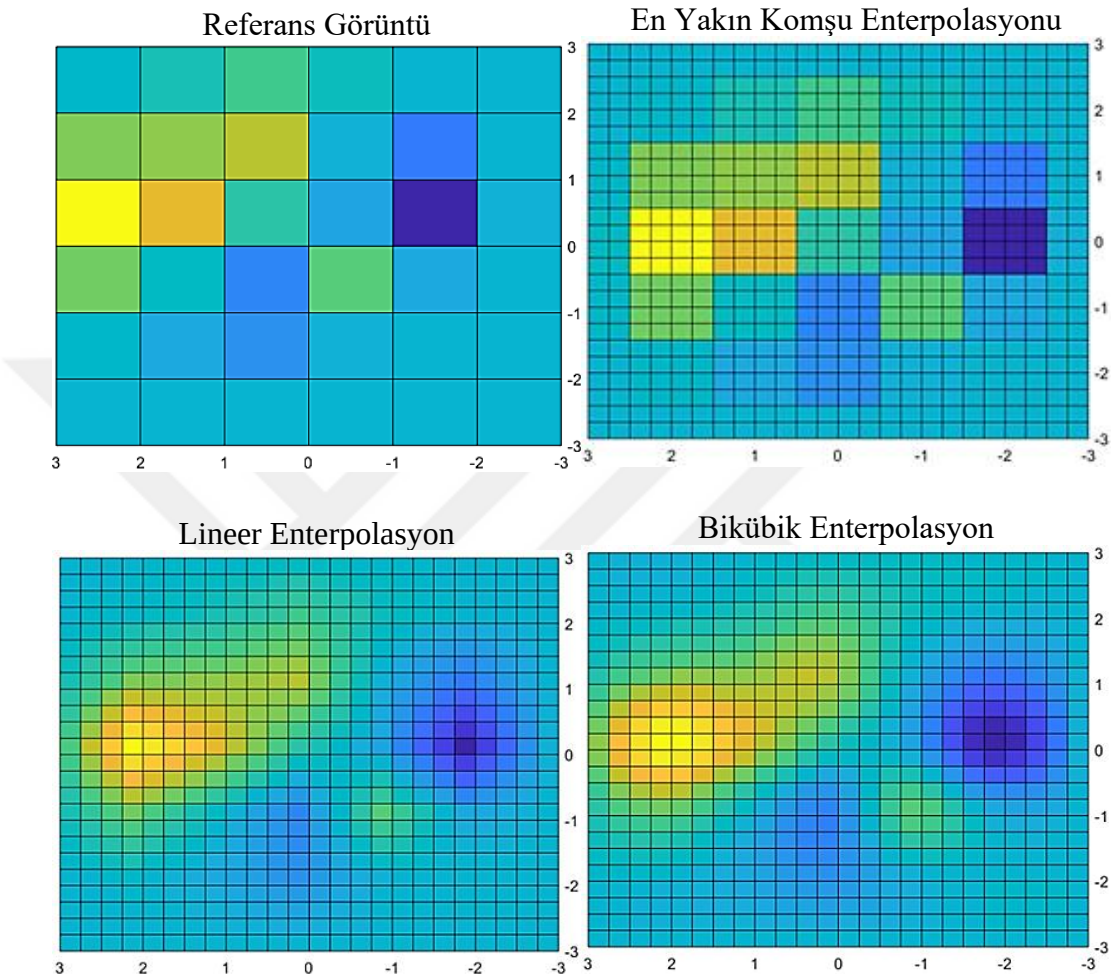
Şekil 3.2. Tekli Görüntü Süper Çözünürlük Taksonomisi [41].

3.1. Enterpolasyon Tabanlı Yöntemler

Görüntülerin çözünürlüğünü artırmak için kullanılan enterpolasyon tabanlı yöntemler, görüntüyü oluşturan anlamsal hiçbir özelliği dikkate almaksızın sadece çeşitli matematiksel işlemlerle piksellerin çoğaltılması mantığına dayanmaktadır. En Yakın Komşu enterpolasyonu, Lineer ve Bikübik enterpolasyon en yaygın kullanılan enterpolasyon tabanlı yöntemlerdendir [42].

En yakın komşu enterpolasyonu, bir noktanın değerini belirlemek için o noktaya en yakın noktaların değerlerini kullanarak tahmin yapa ve yeni görüntüdeki her piksel, orijinal görüntüdeki en yakın pikselin değerini alır. Lineer enterpolasyonda ise, bulunmak istenen yeni noktaya en yakın dört pikselin ortalama ağırlığı kullanılarak yeni piksel değeri belirlenir. En yakın komşu enterpolasyonuna göre daha yumuşak geçişler oluşturur.

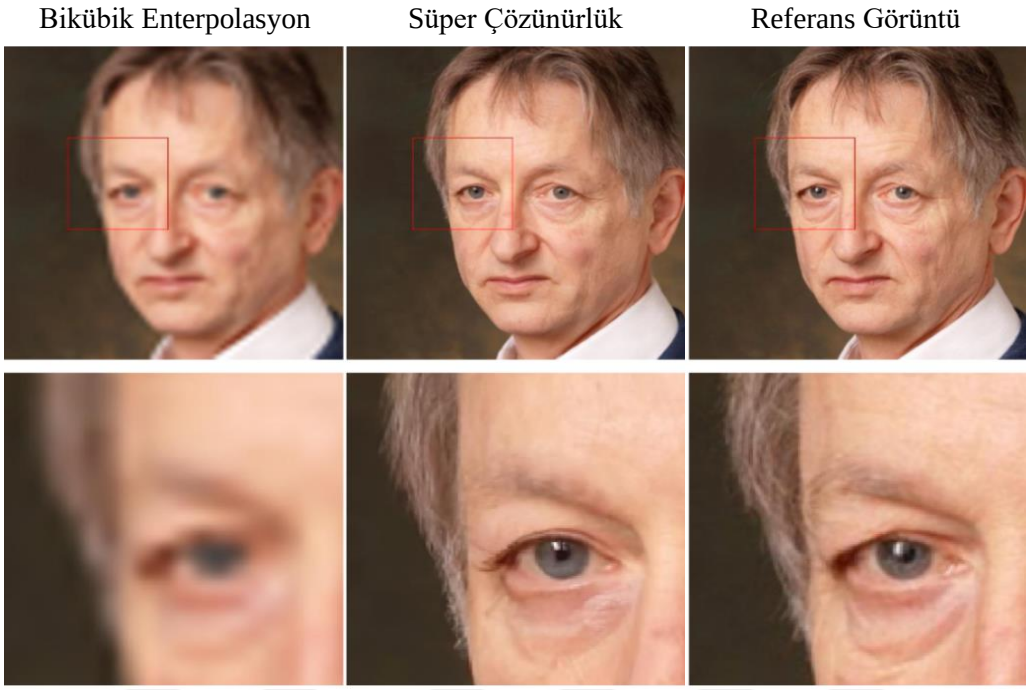
Bikübik enterpolasyon da benzer şekilde değeri bulunmak istenen noktaya en yakın on altı noktanın ağırlıklı ortalaması kullanılarak yeni piksel değeri hesaplanır. Bikübik enterpolasyonda polinom derecesi Lineer enterpolasyona göre daha yüksek olması sebebiyle daha pürüzsüz geçişler elde edilebilmektedir [43].



Şekil 3.3. Enterpolasyon Tabanlı Yöntemler [44].

3.2. Öğrenme Tabanlı Yöntemler

Yapay sinir ağlarının derinleşmesi ve evrişimsel ağların geniş veri setleri ile eğitilmeye başlamasıyla, süper çözünürlük alanında klasik enterpolasyon yöntemlerine göre büyük ilerlemeler kaydedilmiştir. Öğrenme tabanlı ağlar sayesinde, görüntüdeki karmaşık ilişkiler modellenerek geleneksel tekniklerin sınırları aşılmış, daha doğal ve gerçekçi sonuçlar üretilebilmiştir. Özellikle CNN, GAN ve ViT gibi mimarilerin kullanılmasıyla, düşük çözünürlüklü görüntülerdeki karmaşık ilişkileri öğrenen ve görüntüde olmayan ayrıntıları tahmin ederek daha doğal ve ayrıntılı görüntüler üretebilen teknikler geliştirilmiştir.



Şekil 3.4. Enterpolasyon ve Süper Çözünürlük Karşılaştırması [45].

Süper çözünürlükte öğrenmeye dayalı yaklaşımlar, düşük çözünürlükteki giriş görüntülerini kullanarak yüksek çözünürlüklü bir çıktı oluşturmayı amaçlayan, görüntüdeki karmaşık yapıları öğrenmek ve üst örneklemek için çok katmanlı sinir ağları kullanan yapılardır. Bu ağlara giriş olarak verilen düşük çözünürlüklü görüntü ağ boyunca ilerler ve ağdaki katmanlarda ağırlık değerleri oluşturur. Ağdan tamamen geçtikten sonra aynı görüntünün yüksek çözünürlükteki eşi ile karşılaştırılır ve iki görüntü arasındaki hata hesaplanır. Bu hata minimize edilene kadar ağdaki ağırlıklar revize edilir ve tekrardan düşük çözünürlüklü görüntü ağdan geçirilir [46].

Düşük çözünürlüklü görüntü ağa verilmeden önce çözünürlüğünü artırmak için genel olarak dört farklı giriş yöntemi kullanılır. Bunlardan ön örnekleme (Pre-upsampling), başlangıçta herhangi bir enterpolasyon yöntemiyle girdi görüntüsünü üst örnekleyip ağa girdi olarak verilmesi, son örnekleme (Post-upsampling) çıktı katmanında ağın görüntüyü üst örneklemesi, progresif örnekleme (Progressive upsampling) görüntünün aşamalı olarak farklı ölçeklerde üst örneklenmesi, iteratif örnekleme (Iterative up-and-down sampling) üst ve alt örnekleme işlemlerinin ardışık olarak uygulandığı örnekleme yöntemidir.

Derin öğrenme tabanlı bir süper çözünürlük ağı tasarlarken görüntü bağlamına uygun ve ön işlemden geçirilmiş büyük miktarda eğitim verisi, doğru mimari seçimi ve uygun eğitim parametreleri gibi faktörlerin dikkatlice yönetilmesini gerekmektedir.

Eğitim parametreleri arasında tasarlanacak ağın yapısı, katman sayısı ve türü, konvolüsyon kullanımı durumunda filtre sayısı ve boyutu, kullanılacak aktivasyon ve hata fonksiyonları, kullanılan optimizasyon fonksiyonu vb. parametrelerin uygun bir şekilde belirlenmesi üretilecek modelin başarımını doğrudan etkilemektedir.

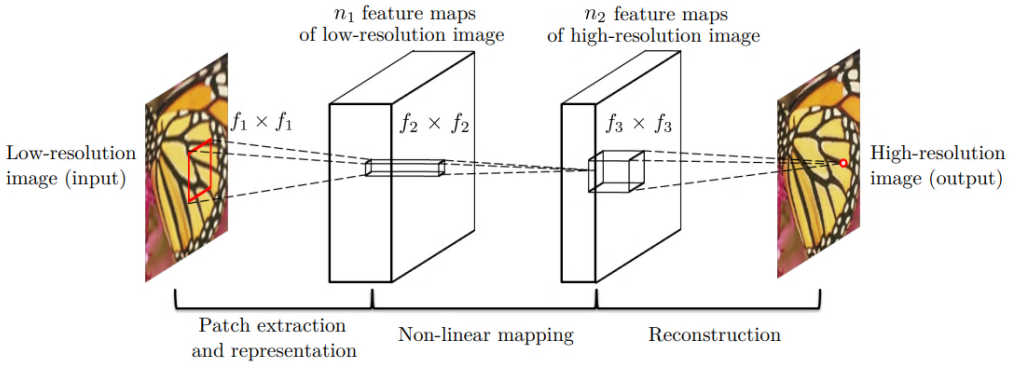
3.3. Tekli Görüntülerde Süper Çözünürlük

Tekli görüntülerde süper çözünürlük, aynı nesneye ait tek bir görüntünün bulunduğu durumlarda kullanılır. Genellikle bir nesnenin birden fazla görüntüsünün bulunmaması nedeniyle geliştirilen yöntemlerin çoğu tekli görüntüleri baz almaktadır.

Tekli görüntülerin süper çözünürlüğünde enterpolasyon tabanlı (Interpolation-based), yeniden yapılandırma tabanlı (Reconstruction-based), öğrenme tabanlı (Learning-based) ve dönüştürücü tabanlı (Transformer-based) yöntemler kullanılmaktadır. Bu çalışmada öğrenme tabanlı yöntemlerden derin öğrenme tabanlı süper çözünürlük yöntemleri ele alınmaktadır.

Derin öğrenme tabanlı süper çözünürlük yöntemleri arasında doğrusal bağlantılı ağlardan SRCNN, GAN tabanlı ağlardan SRGAN ve artık bağlantılı ağlardan EDSR ağları en temel süper çözünürlük yöntemleri arasındadır.

3.3.1. SRCNN



Şekil 3.5. SRCNN Mimarisi.

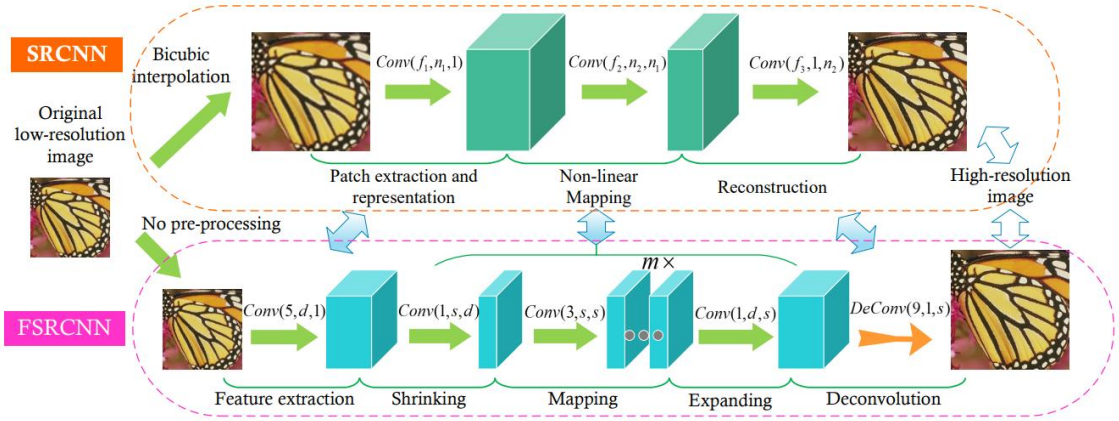
Derin öğrenme tabanlı tekli görüntü süper çözünürlük ilk defa SRCNN (Super-Resolution Convolutional Neural Network) ağları ile önerilmiştir.

Süper Çözünürlük Evrişimli Sinir Ağı, düşük çözünürlüklü (LR) bir giriş görüntüsünü daha yüksek çözünürlükteki (HR) bir çıkış görüntüsüne dönüştürmek için doğrusal bir derin öğrenme modeli kullanır.

SRCNN temelde üç aşamalı bir evrişimli sinir ağıdır. İlk aşamada düşük çözünürlüklü giriş görüntüsünden alt görüntüler oluşturulur. Daha sonra bu alt görüntüler konvolüsyon ve aktivasyon işlemlerinden geçirilerek özellik haritalarına dönüştürülür. Yeniden yapılandırma katmanında ise evrişim katmanlarının çıktıları birleştirilip HR çıktı görüntüsü oluşturularak girdi görüntüsü üst örneklenmektedir.

SRCNN ağının başarısı, diğer süper çözünürlük yöntemlerine kıyasla daha iyi sonuçlar elde etmesine ve daha az işlem zamanı gerektirmesine dayanmaktadır. Derin öğrenme tabanlı bu yaklaşım, tekli görüntülerde süper çözünürlük alanında bir dönüm noktası olarak kabul edilmiş ve bu alandaki diğer çalışmalara zemin oluşturmuştur [5].

3.3.2. FSRCNN



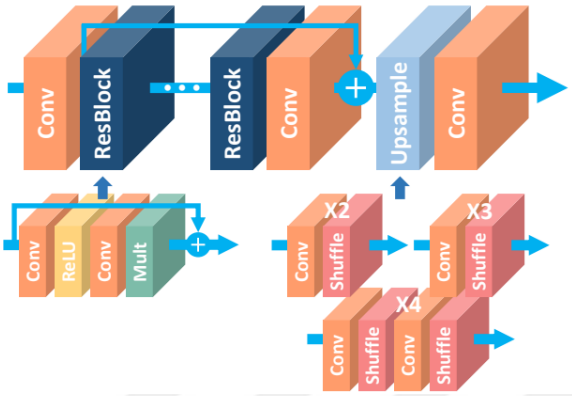
Şekil 3.6. FSRCNN Mimarisi.

FSRCNN (Fast Super-Resolution Convolutional Neural Network), diğer süper çözünürlük yöntemlerine kıyasla daha hızlı ve düşük maliyetli hesaplama gereksinimleri ile öne çıkmıştır. FSRCNN mimarisi, giriş görüntüsünü ön işleme tabi tutmadan direkt olarak özellik çıkarımı yapan konvolüsyon katmanı ile başlar. Bu katman, düşük çözünürlüklü giriş görüntülerinden özellik çıkarımı yapmak için bir dizi evrişim ve aktivasyon işlemi gerçekleştirir. Elde edilen özellikler ardışık katmanlarda işlenerek yüksek çözünürlüklü (HR) çıkış görüntüsü üretilir. Bu katmanlarda, düşük çözünürlüklü giriş görüntülerine öğrenilen filtreler uygulanarak yüksek çözünürlüklü çıkış görüntüleri elde edilir.

Mimarinin son kısmında ters evrişim (Deconvolution) katmanı ile görüntüler üst örneklenip çıkış katmanına verilerek, yüksek çözünürlüklü görüntü elde edilmektedir.

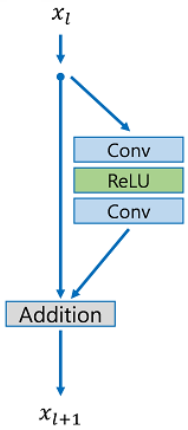
FSRCNN, kalite ve hız arasındaki dengeyi sağlamada önemli bir adım olarak görülmüş ve görüntülerin hızlı bir şekilde iyileştirilmesine ihtiyaç duyan sistemler için pratik bir seçenek sunmuştur [47].

3.3.3. EDSR



Şekil 3.7. EDSR Mimarisi.

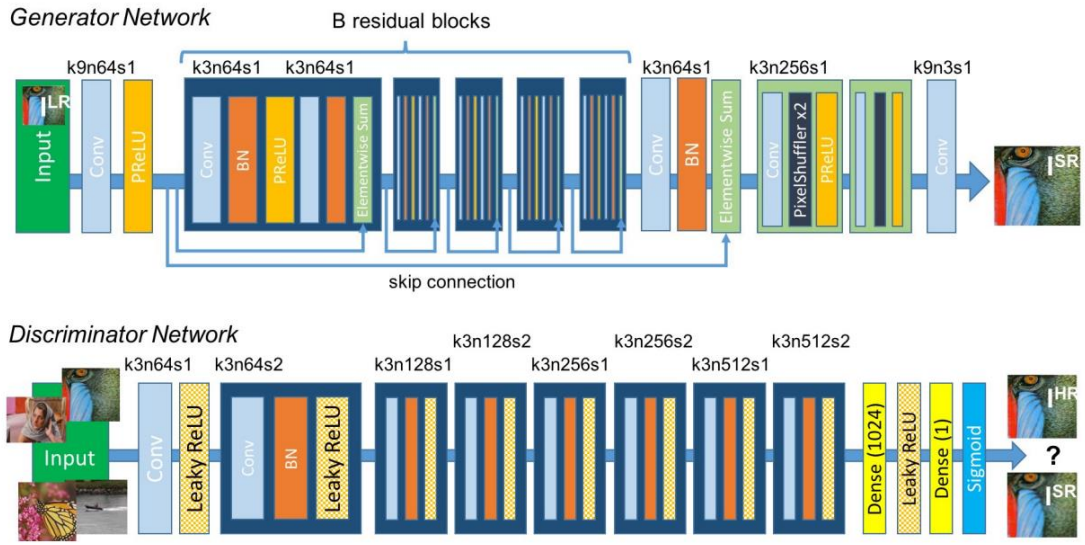
EDSR (Enhanced Deep Super-Resolution), evrişimli sinir ağlarının Artık (Residual) bloklar ile birlikte kullanıldığı bir derin öğrenme mimarisidir. Artık bloklar, bir katmanın çıkışına önceki katmanın girişini ekleyerek ağ derinleştikçe unutulmaması gereken temel özelliklerin korunmasını ve daha verimli bir eğitim yapılmasını sağlar.



Şekil 3.8. EDSR Mimarisinde Kullanılan Artık Blok

EDSR'de artık blokların kullanımıyla geleneksel derin öğrenme tabanlı süper çözünürlük modellerine yeni bir yaklaşım getirilmiştir. Bu teknik, derin öğrenme tabanlı süper çözünürlük modellerinin performansını artırırken eğitim sürecini de daha kararlı hale getirerek süper çözünürlük alanında önemli bir ilerleme sağlamıştır [10].

3.3.4. SRGAN



Şekil 3.9. SRGAN Mimarisi.

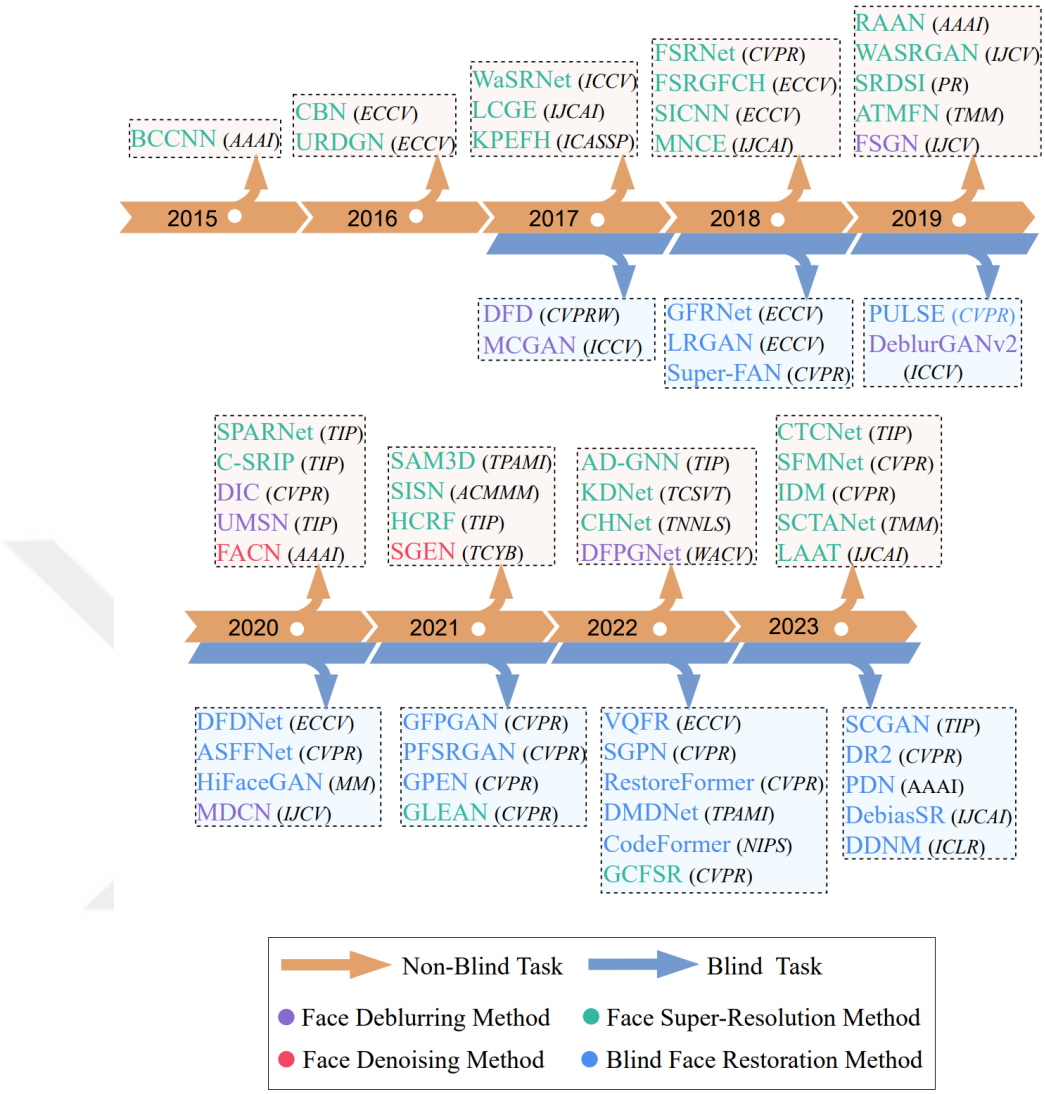
SRGAN (Super-Resolution Generative Adversarial Network), süper çözünürlük alanında derin öğrenme ve özellikle de GAN tabanlı bir yaklaşım kullanarak yüksek çözünürlüklü (HR) görüntüler üretmek için tasarlanmış bir mimaridir. SRGAN temel olarak üç bileşenden oluşur.

Üretici ağ (Generator), düşük çözünürlüklü (LR) giriş görüntülerini alıp yüksek çözünürlüklü (HR) görüntülere dönüştüren bir derin öğrenme ağıdır. Üretici, geleneksel süper çözünürlük yöntemlerinden farklı olarak GAN yapısının avantajlarını kullanarak daha gerçekçi görüntüler üretir.

Ayrıştırıcı ağ (Discriminator) ise, üretilen yüksek çözünürlüklü görüntülerin gerçeklik derecesini değerlendirir. Ayrıştırıcı, gerçek yüksek çözünürlüklü görüntülerle (HR), üretici ağ tarafından üretilen görüntülerin karşılaştırmasını yapar ve bu görüntülerin ne kadar gerçekçi olduğunu belirler. Bu değerlendirme işlemi Adversary Loss ve Perceptual Loss hata fonksiyonları kullanılarak gerçekleştirilir.

SRGAN, yüksek çözünürlük ve netlikte görüntüler üretebilmek için GAN yapısının avantajlarından faydalanarak geleneksel süper çözünürlük yöntemlerine kıyasla daha gerçekçi ve detay yönünden zengin görüntüler elde etmeyi başarmış ve süper çözünürlük alanında önemli bir ilerleme olarak kabul edilmiştir [9].

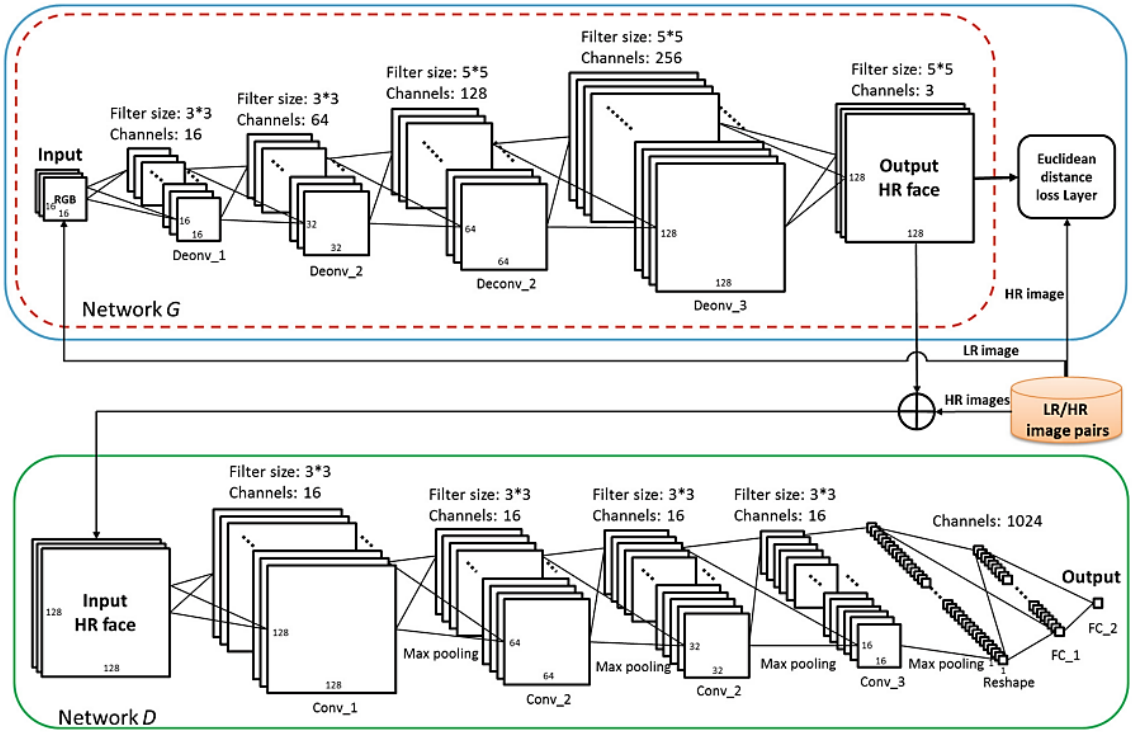
3.4. Yüz Görüntülerinde Süper Çözünürlük



Şekil 3.10. Yüz Görüntülerinde Süper Çözünürlük Ağları [48].

Derin öğrenme tabanlı, yüz görüntülerinin süper çözünürlüğü ve restorasyonu alanında yapılan önemli çalışmalar yukarıdaki şekilde (Şekil 3.9.) gösterilmektedir. Süper çözünürlük, görüntü içeriği fark etmeksizin tüm görüntülere uygulanabilirken, yüz görüntülerinin yüksek frekans ve detaydaki yapısı nedeniyle klasik süper çözünürlük ağları yüz görüntülerinde tam bir başarı sağlayamamaktadır. Bu nedenle yüz görüntülerine özel ağlar geliştirilerek yine yüz görüntülerinden oluşan veri setleri ile eğitilmekte, yüksek çözünürlüklü yüz görüntüleri elde etme konusunda farklı mimariler önerilmeye devam etmektedir.

3.4.1. UR-DGN



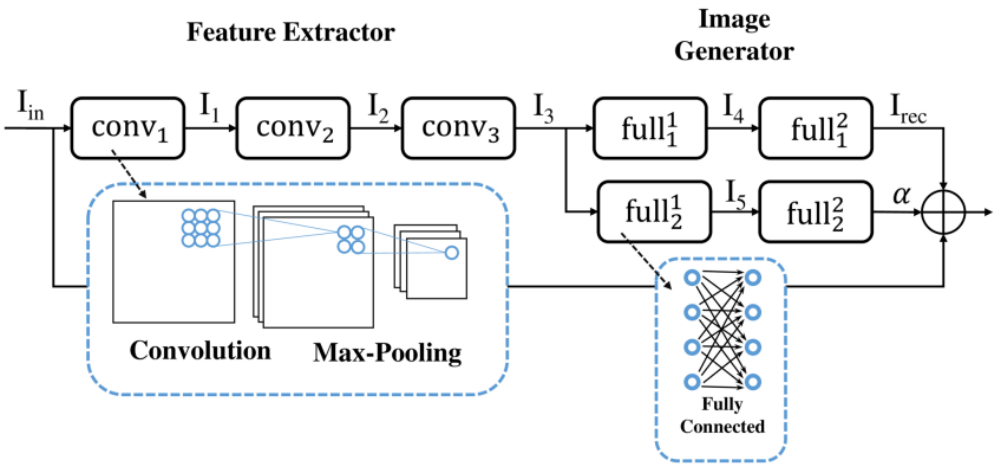
Şekil 3.11. UR-DGN Mimarisi.

UR-DGN (Ultra-Resolving Face Images by Discriminative Generative Networks), düşük çözünürlüklü (LR) yüz görüntülerini yüksek çözünürlüklü (HR) yüz görüntülerine dönüştürmek için GAN mimarisini kullanmıştır.

Üretici ağ (Network-G) LR görüntülerden gerçeğe yakın HR çıktılar üretmeyi öğrenirken, ayırıcı ağ (Network-D) LR ve HR görüntü çiftini alarak gerçek ve sahte görüntüler arasında ayrım yapmayı öğrenmektedir.

Eğitim aşamasında, bu iki ağ birlikte çalışır, ancak test aşamasında yalnızca üretici ağ (Şekil 3.11'de kırmızı çizgilerle gösterilen blok) kullanılır. Üst örneklenmiş yüz görüntülerinin referans görüntülere benzerliği, üretken modeldeki piksel bazında Öklid Uzaklık hata terimi ile kontrol edilmiştir. GAN mimarisinin avantajlarını özel ağ tasarımıyla destekleyen UR-DGN benzer ağlara kıyasla yüksek PSNR ve SSIM metrikleri elde etmiştir [49].

3.4.2. BCCNN



Şekil 3.12. BCCNN Mimarisi.

BCNN (Bi-Channel CNN) mimarisinde, yüz görüntülerinin süper çözünürlüğü ve restorasyonu için iki kanallı bir CNN mimarisi önerilmiştir. Bu mimaride giriş görüntüsü öncelikle klasik evrişim katmanlarından geçirilerek özellik çıkarma işlemi gerçekleştirilir ve görüntü tam bağlı katmanlara iletilir. Daha sonra tam bağlı katmanların çıkışı ile başlangıçtaki görüntü birleştirilerek yüksek çözünürlüklü çıkış görüntüsü elde edilir. BCNN girdi verilerini iki farklı kanal üzerinden işleyerek daha zengin bir öğrenme kapasitesi sunmasının yanı sıra farklı veri türleri için esnek bir yapı sunmuştur [50].

3.5. Değerlendirme Metrikleri

Düşük çözünürlüklü görüntülerden yüksek çözünürlüklü görüntüler elde etmek amacıyla geliştirilen algoritmaların, derin öğrenme ağlarının veya diğer metotların başarım ve performanslarının objektif bir şekilde değerlendirilebilmesi için bazı sayısal ölçütlere, metriklere ihtiyaç vardır. Bu metrikler süper çözünürlük algoritmalarının ürettiği yüksek çözünürlüklü görüntülerin gerçek referans görüntülere ne kadar yakın olduğunu ölçmek için kullanılmaktadır. PSNR, SSIM ve LPIPS metrikleri süper çözünürlükte yaygın olarak kullanılan değerlendirme metrikleri arasındadır [51].

3.5.1. PSNR

PSNR (Peak Signal-to-Noise Ratio), süper çözünürlük algoritmasının ürettiği görüntünün, gerçek referans görüntüye kıyasla ne kadar gürültülü olduğunu ölçer. Yüksek PSNR değeri ile görüntü çiftlerinin birbirlerine olan benzerliği genellikle doğru orantılıdır [41].

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \tag{3.1}$$

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3.2}$$

Denklem 3.1'deki MAX_I^2 görüntülerde bulunabilecek maksimum piksel değerini, denklem 3.2'deki y ve $\hat{y}$ girdi görüntü çiftlerini, n ise piksel sayısını göstermektedir.

3.5.2. SSIM

SSIM (Structural Similarity Index), iki görüntü arasındaki benzerliği ölçen 0 ile 1 arasında bir sayıdır. SSIM hesaplanırken parlaklık (luminance), karşıtlık (contrast) ve yapısal benzerlik (structure) olmak üzere üç bileşen çarpılır. Bu bileşenler, iki görüntünün benzerliğini farklı yönlerden analiz ettiği için SSIM, PSNR'a göre insan algısına daha yakın bir metriktir. Yüksek SSIM değeri, görüntü çiftlerinin benzerlik oranının da yüksek olduğunu gösterir [52].

$$\begin{aligned} &\text{SSIM}(y, \hat{y}) = l(y, \hat{y})^{\alpha} \cdot c(y, \hat{y})^{\beta} \cdot s(y, \hat{y})^{\gamma} \\ &l(y, \hat{y}) = \frac{2\mu_y \mu_{\hat{y}} + c_1}{\mu_y^2 + \mu_{\hat{y}}^2 + c_1} \\ &c(y, \hat{y}) = \frac{2\sigma_y \sigma_{\hat{y}} + c_2}{\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2} \\ &s(y, \hat{y}) = \frac{\sigma_{y\hat{y}} + c_3}{\sigma_y \sigma_{\hat{y}} + c_3} \end{aligned} \tag{3.3}$$

Denklem 3.3'teki $y \; ve \; \hat{y}$ girdi görüntü çiftleri, μ_y ve $\mu_{\hat{y}}$ görüntülerin piksel ortalamaları, $\sigma_y \; ve \; \sigma_{\hat{y}}$ görüntülerin varyansı, $\sigma_{y\hat{y}}$ görüntülerin kovaryansı, $c_1, c_2 \; ve \; c_3$ ise stabilizasyon sabitleridir. Çarpım durumundaki l, c, s fonksiyonları sırasıyla parlaklık, kontrast ve yapısal benzerliği ifade etmekte; $\alpha, \beta \; ve \; \gamma$ ise normal şartlarda 1 kabul edilmekle beraber bir fonksiyonun diğerine göre ne kadar öne çıkarıldığını göstermektedir.

3.5.3. LPIPS

LPIPS (Perceptual Image Patch Similarity), iki görüntü arasındaki algısal benzerliği ölçmek için kullanılan bir metriktir. Diğer metriklere göre insan algısına daha yakın olan LPIPS metriği hesaplanırken ilk olarak VGG veya AlexNet gibi önceden eğitilmiş başarılı bir derin öğrenme modeli seçilir ve görüntü çiftleri modele girdi olarak verilir. Model her iki görüntü için de çıkardığı özellik haritalarının farklarını toplayarak LPIPS değerini belirler. LPIPS değeri ne kadar düşükse, iki görüntü arasındaki algısal benzerlik o kadar fazladır [53].

$$d(x,y) = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(\hat{x}^l_{hw} - \hat{y}^l_{hw} \right) \right\|_2^2 \tag{3.4}$$

Denklem 3.4'teki $x\ ve\ y$ girdi görüntü çiftlerini, l özellik çıkarımı yapılan katmanı, $H_l\ ve\ W_l$ katmanın boyutlarını, $\hat{x}\ ve\ \hat{y}$ görüntülerin katman çıktılarını, w_l bazı özellikleri vurgulamak için kullanılan ağırlık matrisini göstermektedir.

3.6. Kullanılan Veri Setleri

Süper çözünürlük alanında veri seti seçimi modelin başarısı, genelleme yeteneği ve farklı uygulama alanlarında kullanılabilirliği açısından önemlidir. Veri seti seçimi, süper çözünürlük modelinin ayrıntı üretme kabiliyetini doğrudan etkiler. Örneğin insan yüzü gibi belirli nesneler üzerinde çalışan bir model için kullanılan veri setinin, yeterli miktarda ayrıntıya sahip çeşitli yüz görüntüleri içermesi, modelin hassasiyetini artırır ve daha gerçekçi sonuçlar üretmesini sağlar.

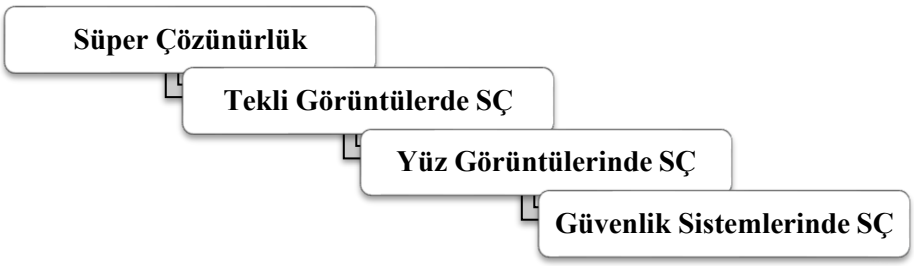
Süper çözünürlük algoritmalarının gerçek dünyada başarılı olabilmesi için, eğitim ve test verilerinin, kullanılan senaryoları doğru bir şekilde temsil etmesi gerekir. Örneğin yüz tanıma sistemleri için düşük çözünürlüklü yüz görüntüleri üzerinde çalışan bir modelin, gerçek dünya verilerini temsil eden veri setleriyle eğitilirken kamera açıları, ışık koşulları ve farklı çözünürlüklerdeki görüntüler bu veri setlerinde yer almalıdır.

Süper çözünürlük alanında kullanım alanına göre değişen, farklı niteliklere sahip birçok veri seti bulunmaktadır. Bu veri setlerinden bazıları genel özellikleriyle birlikte aşağıdaki tabloda (Tablo 3.1) gösterilmektedir.

Tablo 3.1. Süper Çözünürlük Veri Setleri [54].

Veri Seti Adı	Kullanım	Veri Sayısı	Alan	Açıklama
General-100	Train	100	SR	Net kenarlara sahip genel görüntüler.
T91	Train	91	SR	Genel görüntüler.
WED	Train	4744	SR	Genel görüntüler.
Flickr2K	Train	2650	SR	Genel görüntüler.
FFHQ	Train	70000	FSR	İnsan yüzlerinden oluşan yüksek çözünürlüklü görüntüler.
CelebA-HQ	Train/Val.	30000	FSR	GAN ile elde edilmiş sentetik insan yüzlerinden oluşan görüntüler.
CelebA	Train/Val.	202600	FSR	İnsan yüzlerinden oluşan yüksek çözünürlüklü görüntüler.
DRealSR	Train/Val.	31970	SR	Gerçek dünyadaki yüksek frekanslı genel görüntüler.
DIV2K	Train Val.	1000	SR	Yüksek çözünürlüğe sahip görüntüler (ECCV PIRM yarışması veri seti).
BSDS300	Train/Val.	300	SR	Genel görüntüler.
BSDS500	Train/Val.	500	SR	Genel görüntüler.
RealSR	Train/Val.	100	SR	Genel görüntü çiftleri.
OutdoorScene	Train/Val.	10624	SR	Manzara görüntüleri.
City100	Train/Test	100	SR	Genel görüntüler.
Flickr1024	Train/Test	100	SR	Stereo-SR için ikili görüntüler.
SR-RAW	Train/Test	3500	SR	Genel raw sensör veri seti.
PIPAL	Test	200	SR	Algısal görüntü kalitesi değerlendirme işlemi için genel görüntüler.
Set5	Test	5	SR	Genel görüntüler.
Set14	Test	14	SR	Genel görüntüler.
BSD100	Test	100	SR	Genel görüntüler.
Urban100	Test	100	SR	Gerçek dünya yapı görüntüleri.
Manga109	Test	109	SR	Japon Manga görüntüleri.
L20	Test	20	SR	Yüksek çözünürlüklü genel görüntüler.
PIRM	Test	200	SR	Genel görüntüler (ECCV PIRM yarışması veri seti).

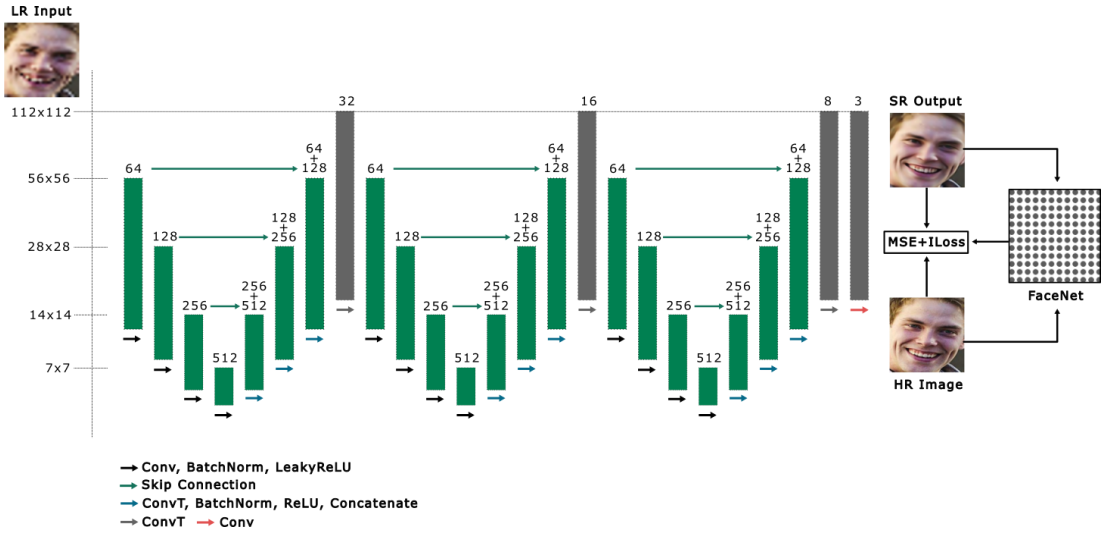
4. GÜVENLİK KAMERALARINDA SÜPER ÇÖZÜNÜRLÜK MODELİ



Şekil 4.1. Tez Kapsamı.

Süper çözünürlük temelde videolar ve görüntüler üzerine uygulanabilir yöntemler olmak üzere iki kısma ayrılmaktadır. Tez kapsamında güvenlik kameralarındaki kişilerin yüzlerini netleştirmek ve çözünürlüğünü artırmak amacıyla direkt olarak video görüntüleri üzerinden çalışmak yerine, ilgili yüz bölgesi video karesinden tespit edilip kesilerek tekli görüntü süper çözünürlük yöntemleri üzerine bir çalışma yürütülmüştür.

4.1. VNet Mimarisi



Şekil 4.2. VNet Mimarisi.

Bu çalışmada, düşük çözünürlükteki görüntülerden yüksek çözünürlük ve detayda görüntüler elde etmek amacıyla, kodlayıcı-çözücü (Encoder-Decoder) tabanlı bir evrişimli sinir ağı kullanan VNet mimarisi (Şekil 4.2) önerilmektedir. VNet mimarisinde U-Net benzeri "V" şeklindeki bloklar ile görüntü restore edilip üst örneklenirken kullanılan FaceNet modeli ile de kişinin biyometrik açıdan yüz özellikleri korunmaktadır.

Yüz görüntülerinin süper çözünürlük ile üst örneklenmesi amacıyla derin öğrenme mimarilerinden evrişimsel ağlar kullanılmaktadır. U-Net benzeri kodlayıcı-çözücü (Encoder-Decoder) mantığıyla tasarlanmış ağda, ResNet mimarisinde önerilen Artık Öğrenme (Residual Learning) kısayolları ile katmanlar karşılıklı bağlanmıştır. Artık Öğrenme sayesinde ağ derinleştikçe ve katman sayısı arttıkça ortaya çıkan "kaybolan gradyan" (Vanishing Gradient) probleminin önüne geçilmesi amaçlanmıştır.

VNet ağında, "V" şeklindeki yeşil bloklar ile görüntüden özellik çıkarımı yapılmakta ve gri ara katmanlar ile de çıkarılan özellikler, görüntünün giriş çözünürlüğü korunarak birleştirilmektedir. Toplamda 33 milyon eğitilebilir parametreye sahip ağa girdi olarak verilecek olan görüntüler, öncelikle veri ön işleme adımından geçirilmektedir. Ağ, orijinal ve bozulmuş (HR-LR) görüntü çiftleriyle beslenerek çalıştığı için, orijinal görüntü bir bozulma fonksiyonuna tabi tutulmakta ve bozuk görüntü (LR Image) elde edilmektedir. Bozulma fonksiyonu 112x112 piksel görüntüyü 4 kat küçültüp, 28x28 piksel boyutuna indirgemekte ve tekrar eski boyutuna çevirerek bozulmuş görüntü (LR Image) elde edilmektedir.

Mimaride herhangi bir öncül (Prior) kullanılmamıştır. Ağ, girdi görüntülerini bütüncül bir yaklaşımla ele alarak çeşitli katmanlar ve bağlantılar ile HR-LR görüntü çiftleri arasındaki bozulma fonksiyonunu öğrenmeye ve bu fonksiyonun tersini alarak bozulmuş görüntüden asıl görüntüyü tahmin etmeye çalışmaktadır.

Ağın genel yapısının gösterildiği şekilde (Şekil 4.2) soldaki çizelgede her bir katman sonrası görüntünün çıkış çözünürlüğü gösterilmiştir. Her katman sonrası sürekli yarılanan görüntü çözünürlüğü 32, 16 ve 8 filtreli ara konvolüsyonlar ile tekrar giriş boyutuna dönmekte, en sonunda ise 3 filtreli konvolüsyon ile süper çözülmüş RGB görüntü (SR Output) elde edilmektedir. VNet ağı sabit girdi boyutlu olup, 112x112 piksel boyutunda girdiler almaktadır. Girdi görüntüsü, (Şekil 4.2'de görüldüğü üzere) filtre sayıları üzerinde belirtilen katmanlar üzerinden soldan sağa doğru ilerlemektedir.

Giriş katmanı 64 tane konvolüsyon filtresinden oluşmakta olup kodlayıcı katmanlar boyunca filtre sayısı iki katına çıkarak en son 512 filtreye kadar ağ derinleşmektedir. Daha sonra çözücü katmanlarda ise filtre sayısı yarılanarak azaltılmakta ve ağ daha sığ bir hal almaktadır. Ağda kullanılan bütün konvolüsyon filtreleri 4x4 boyutundadır.

Kodlayıcı katmanlardaki konvolüsyon işlemleri 2 adımlıdır (stride) ve eşit dolgu (same-padding) kullanılmıştır. Bu sayede her katman sonrası görüntü çözünürlüğü düzgün bir şekilde yarılanmaktadır. Filtre ağırlıkları rasgele bir başlatıcı ile 0-0,02 arası bir başlangıç değeri ile başlatılmaktadır. Konvolüsyon sonrası optimizasyon amacıyla Batch normalizasyonu kullanılmakta ve aktivasyon fonksiyonu olarak da LeakyReLU kullanılmaktadır.

Çözücü katmanlardaki ters konvolüsyonlar yine 2 adımlı, eşit dolgulu ve 0-0,02 arası rasgele bir başlangıç değerleri ile başlatılmaktadır. Optimizasyon için Batch normalizasyonu, aktivasyon fonksiyonu olarak da ReLU kullanılmıştır. Kullanılan ters konvolüsyon ve eşit dolgu sayesinde yarılanmış olan görüntü çözünürlüğü düzgün bir şekilde katmandan katmana iki katına çıkmaktadır. Ağın en derin kısmında kullanılan 512 filtre, çözücü katmanlar boyunca yarılanırken, her katmanın filtre ağırlıkları, katmanda bulunan görüntü ile aynı çözünürlüğe sahip karşı katmanın filtre ağırlıkları ile birleştirilmektedir. Birleştirme işlemi (Şekil 4.2'de yeşil oklarla gösterildiği üzere) ağ derinleştikçe asıl görüntüden uzaklaşan özellik haritalarını güncelleyerek filtrelerin daha verimli bir şekilde özellik çıkarımı yapmasını sağlamaktadır.

Çözücü katmanlardan sonraki 32-16-8 filtreli gri katmanlarda, ters konvolüsyon ile görüntü giriş çözünürlüğüne döndürülürken, "V" şeklindeki yeşil bloklar arasında da özellik haritalarının aktarımı yapılmaktadır. Ters konvolüsyon işleminde 2 adımlı ve eşit dolgulu konvolüsyon ve Tanh aktivasyon fonksiyonu kullanılmıştır.

Son katmanda giriş görüntüsünün ağda ilerlemesi sonucu elde edilen özellik haritaları, görüntünün RGB kanallarını temsil eden 3 filtreli konvolüsyon katmanında birleştirilip Tanh aktivasyon fonksiyonundan geçirilerek yüksek çözünürlük ve detaydaki görüntü elde edilmektedir.

VNet mimarisindeki kodlayıcı-çözücü katmanların temel görevi, girdi görüntüsündeki önemli bilgileri sıkıştırarak gereksiz ayrıntıları filtrelemek ve temel özellikleri ortaya çıkarmak, daha sonra filtrelenmiş temsili görüntüyü hedef çözünürlüğe sahip yeni görüntüye dönüştürmektir.

Girdi boyutunun sürekli küçültülmesi bir yandan bellek kullanımı ve hesap karmaşıklığını azaltırken, diğer yandan modele dokular ve diğer yüz bölgeleri gibi hem düşük hem de yüksek seviyeli detayları öğrenme yeteneği kazandırmaktadır. Örneğin mimarideki 512 konvolüsyona sahip katmanlara ait filtreler 7x7 piksel boyutundaki ince detaylar üzerinde dururken, 128 konvolüsyona sahip katmanlardaki filtreler 28x28 piksel boyutundaki bölgelere odaklanmaktadır. Diğer yandan katmanlar arasında kullanılan artık (Residual) bağlantılar, ağ derinleştikçe azalan öğrenme oranını tekrar artırarak ağın eğitimini daha verimli hale getirmektedir.

4.2. Hata Fonksiyonu

Ağın hata fonksiyonu olarak, Khazaie ve ark. önerdiği gibi, Ortalama Karesel Hata (Mean Squared Error – MSE) ve Biyometrik Hata (Identity Loss – ILoss) fonksiyonları bir arada kullanılmaktadır. Bu iki hata fonksiyonu, toplam hata fonksiyonu adı altında birleştirilerek ağın genel hata değerini belirtmektedir [19].

Biyometrik hata hesaplanırken ağın çıktısından elde edilen süper çözülmüş görüntü ile referans görüntü çifti, FaceNet [55] yüz tanıma ağı ile eğitilmiş bir modele girdi olarak verilmektedir. FaceNet ön eğitilmiş (pre-trained) modeli, VGGFace2 veri seti kullanılarak eğitilmiş olup insan yüzüne ait filtrelenmiş verileri (Embeddings) elde etme konusunda başarılı bir modeldir [56]. Bu modelin, tahmin edilen görüntü ile referans görüntü için ürettiği çıktılar karşılaştırılmakta ve bu iki görüntünün benzerlik oranı, ağa geri besleme olarak verilmektedir. Bu sayede görsel olarak kaliteli ve detaylı görüntüler elde edilirken, çıktı görüntüsünün asıl kişiye olan benzerliği de korunmaktadır.

MSE hata fonksiyonu ise orijinal görüntü ile tahmin edilen görüntü arasında piksel bazındaki sayısal farka bakmakta ve bu görüntülerin piksel değerlerinin farkını ağa geri besleme olarak vererek, sayısal benzerliğin artırılmasında kullanılmaktadır.

4.3. Değerlendirme Metrikleri

Görüntüler arasındaki nicelik ve nitelik bakımından benzerlik oranını ölçebilmek için birçok değerlendirme metriği kullanılmaktadır [51]. Değerlendirme metrikleri, birbirinden farklı yol ve yöntemler kullanarak, görüntü çiftleri arasında bir benzerlik bağıntısı elde eder ve bu benzerliği sayısal bir veriye dönüştüren niceliksel metrikler yaygın olarak kullanılır.

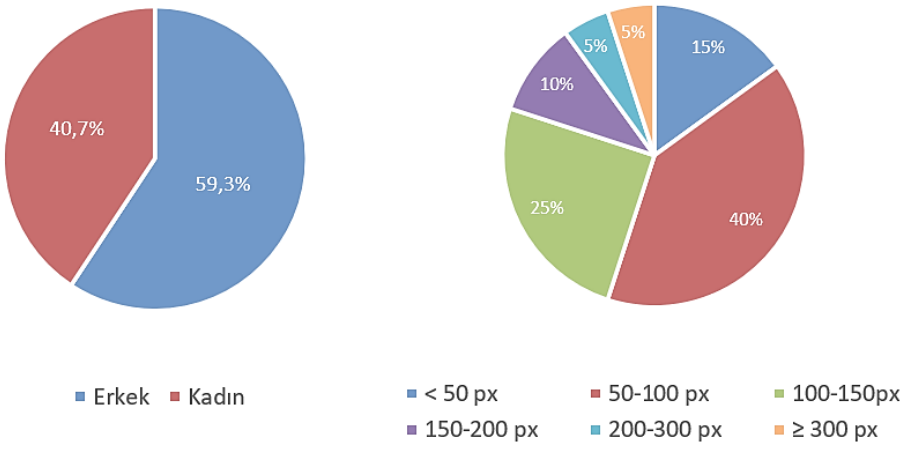
Bu çalışmada yaygın olarak kullanılan niceliksel metriklerden; MAE (Mean Absolute Error), MSE (Mean Squared Error), PSNR (Peak Signal-To-Noise Ratio), SSIM (Structural Similarity Index Measure) ve LPIPS (Learned Perceptual Image Patch Similarity) metrikleri kullanılmıştır.

MAE ve MSE iki görüntünün pikselleri arasındaki sayısal farka odaklanır. Görüntüler birbirine benzeyip aralarındaki sayısal fark azaldıkça, bu hata değerleri de azalmaktadır. PSNR da yine MSE kullanarak iki görüntü arasındaki maksimum piksel değerine göre hata oranını logaritmik olarak ölçer. Yüksek PSNR değeri, iki görüntü arasındaki benzerliğin de yüksek olduğunu gösterir.

SSIM yapısal benzerlik indeksi, iki görüntü arasındaki parlaklık, kontrast ve yapısal benzerlikleri karşılaştırır. Benzerlik arttıkça 0-1 arası değişen SSIM değeri de artmaktadır. LPIPS, öğrenilmiş algısal görüntü yama benzerliği, görüntülerin algısal benzerliklerini ölçmek için görüntü çiftleri arasındaki benzerlikleri değerlendiren, AlexNet ve VGGNet gibi önceden eğitilmiş bir modelden elde edilmiş özellikler kullanır. Görüntüler, modelin ara katmanlarından geçirilir ve elde edilen çıktıların farkı alınarak sayısal bir değer elde edilir.

LPIPS, diğer sayısal metriklerden farklı olarak insan algısına daha yakın bir benzerlik ölçütü sağlar. Düşük LPIPS değeri, algısal olarak benzerliğin yüksek olduğunu göstermektedir.

4.4. VGGFace2 Veri Seti



Şekil 4.3. VGGFace2 Veri Seti Cinsiyet ve Yüz Bölgesi Dağılımları.

Önerilen VNet ağının eğitimi için Cao ve ark. tarafından duyurulan VGGFace2 veri seti kullanılmıştır [57]. Bu veri setinde Google görsellerden elde edilmiş çok çeşitli poz, aydınlatma, etnik köken, yaş ve meslekteki (aktörler, sporcular, politikacılar vb.) 9131 kişiye ait 3,31 milyon yüz görüntüsü bulunmaktadır. Kişi başı ortalama 362 resmin bulunduğu sette, yüz görüntülerinin sınır noktaları da paylaşılmış olup çözünürlükleri 50px ile 300px arası değişmektedir.

Veri ön hazırlığı aşamasında eğitim sürecini optimize etmek adına, veri setinden her kişi için yaklaşık 100 görüntü olmak üzere toplamda 863.732 görüntü rastgele seçilmiştir. Daha sonra seçilen görüntülerdeki yüz bölgeleri kesilmiş ve 112x112 piksel boyutuna getirilmiştir.

4.5. ChokePoint Veri Seti

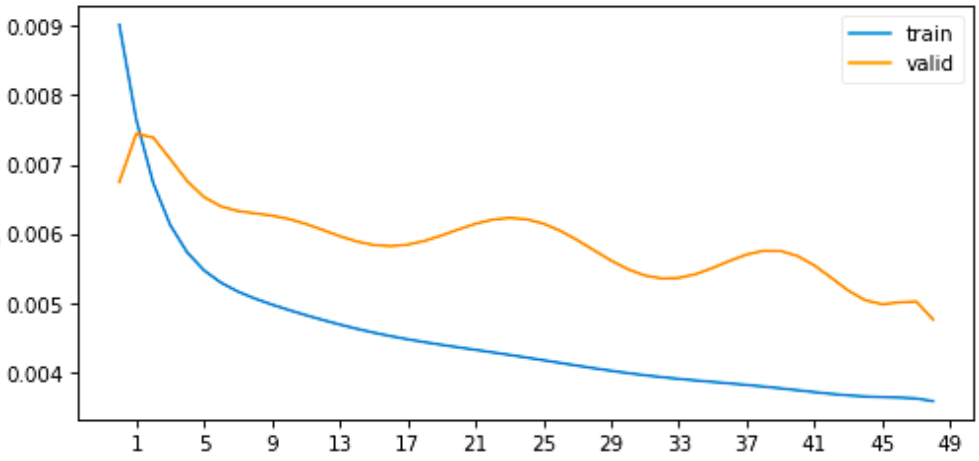
Şekil 4.4. ChokePoint Veri Seti Örnekleri.

ChokePoint veri seti [58], yüz tanıma algoritmalarının performansını test etmek amacıyla tasarlanmış bir veri setidir. Bu veri seti, kontrol noktalarından geçen insanların farklı açılarda ve ışık koşullarında çekilmiş videolarından elde edilen video karelerinden oluşmaktadır. Gerçek dünya uygulamalarına uygun olarak, insanların hareket halinde olduğu gerçekçi senaryoların canlandırılmaya çalışıldığı ChokePoint veri seti, diğer yüz görüntü veri setleriyle karşılaştırıldığında, güvenlik sistemlerinde yüz görüntülerinin süper çözünürlüğü için daha uygun bir veri setidir.

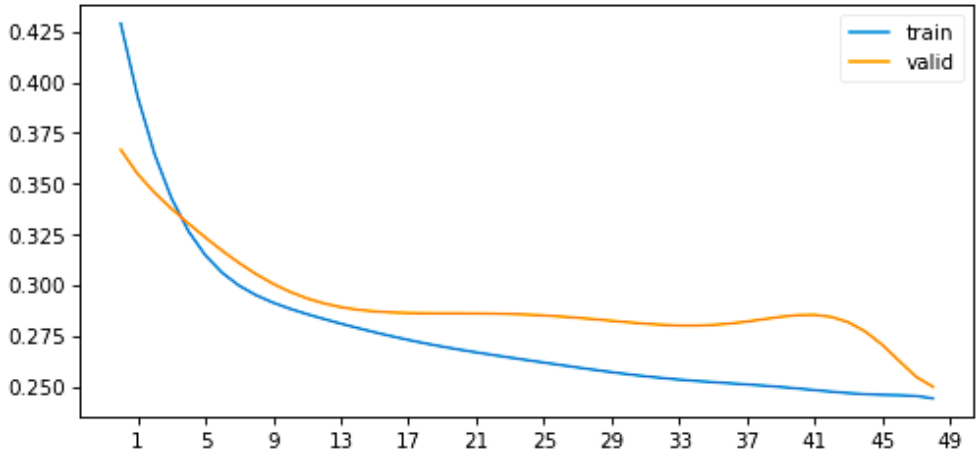
Veri setinde, kontrol noktalarında çekilmiş videolara ait 800x600 piksel boyutunda yaklaşık 180 bin video karesi bulunmaktadır. Bu kareler OpenCV kütüphanesi ile taranarak kare olacak şekilde yüz görüntüleri kesilmiştir. Düşük çözünürlüklü görüntülerden ziyade 100x100 piksel ve üzeri çözünürlükteki yüz görüntüleri kullanılarak 4 bin yüz görüntüsü elde edilmiştir. Daha sonra bu görüntüler yatayda aynalanarak 100x100 ile 286x286 piksel arası 8 bin yüz görüntüsünden oluşan bir veri seti oluşturulmuştur.

5. UYGULAMA VE SONUÇLAR

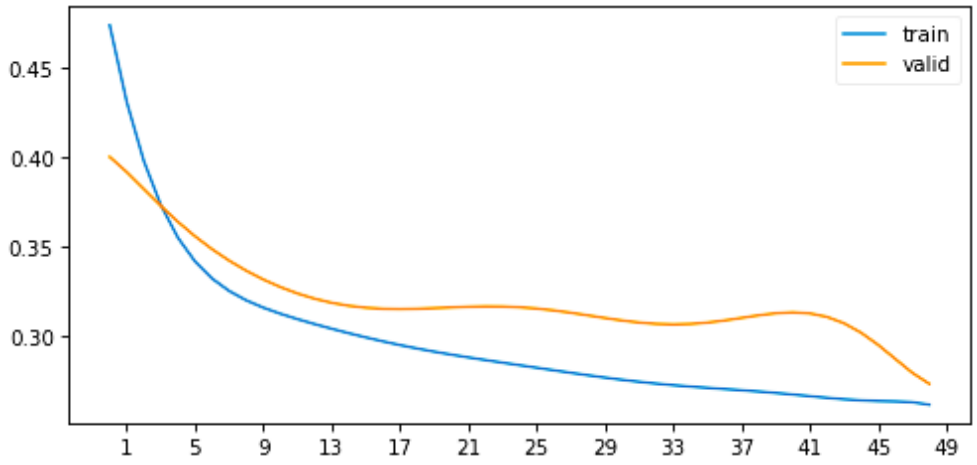
Eğitim için donanımsal olarak Intel Core i9-7960X CPU ve Nvidia RTX 4090 24gb GPU, yazılımsal olarak da Python 3.11.9 ve Tensorflow 2.15.0 kullanılmıştır. VNet ağı, VGGFace2 veri setinden seçilen 863.732 görüntü ile 49 epok boyunca eğitilmiştir. Ağın eğitiminde kullanılan Ortalama Karesel Hata (MSE), Biyometrik Hata (Identity Loss) ve toplam hata fonksiyonlarının eğitim ve doğrulama (Train-Validation) aşamalarındaki çıktıları aşağıdaki grafiklerde (Şekil 5.1, 5.2 ve 5.3) gösterilmektedir. Birbirine yakın seyreden eğitim ve doğrulama grafikleri, eğitim sürecinin ve ağın stabilitesini göstermektedir.



Şekil 5.1. Ortalama Karesel Hata (MSE).



Şekil 5.2. Biyometrik Hata (Identity Loss).



Şekil 5.3. Toplam Hata.

5.1. Karşılaştırmalı Sonuçlar

Eğitim işleminin tamamlanmasıyla, eğitilen modelin başarımını ölçmek için bazı test veri setleri ve MAE, MSE, PSNR, SSIM ve LPIPS değerlendirme metrikleri kullanılmıştır [51].

Test veri seti olarak, farklı yöntemler kullanılarak oluşturulmuş, yüz görüntülerinden oluşan; VGGFace2, CelebA, UTKFace, MultiPie ve FEIFace veri setlerinin her birinden rastgele seçilen 2.000 görüntü kullanılmıştır. Test aşamasında 28x28 piksel boyutlarına indirgenmiş görüntülerin 4 kat büyütülmesi sonucu elde edilen metrik başarımlar Tablo 5.1'de gösterilmektedir.

Tablo 5.1. Test Veri Setleri ×4 Skorları.

Test Seti	MAE↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
VGGFace2 [57]	.02812	.00169	28.730	0.8806	0.0937
CelebA [22]	.02757	.00172	28.271	0.8891	0.0835
UTKFace [59]	**.02650**	**.00163**	**28.982**	**0.8999**	**0.0626**
FEIFace [60]	.02717	.00165	28.769	0.8980	0.0747
MultiPie [61]	.03984	.00362	25.884	0.8800	0.1131

Elde edilen sonuçlara bakıldığında, UTKFace veri seti en yüksek metrik başarımlara sahipken, diğer test veri setleri de yakın değerler almıştır. Test veri setleri üzerinde elde edilen sonuçların birbirine yakın değerler olması dengeli ve genellenebilir bir model elde etme konusunda önemlidir.

Geliştirilen VNet ağını diğer süper çözünürlük ağları ile karşılaştırmak için VGGFace2 ve CelebA veri setlerinden rastgele seçilen, eğitim esnasında kullanılmayan 2.000 görüntü ile yapılan testlerin PSNR, SSIM ve LPIPS metrik çıktıları, karşılaştırmalı olarak Tablo 5.2 ve 5.3'te gösterilmiştir.

Tablo 5.2. VGGFace2 ×4 Test Skorları [62].

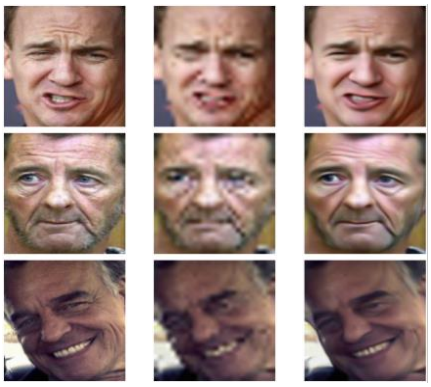
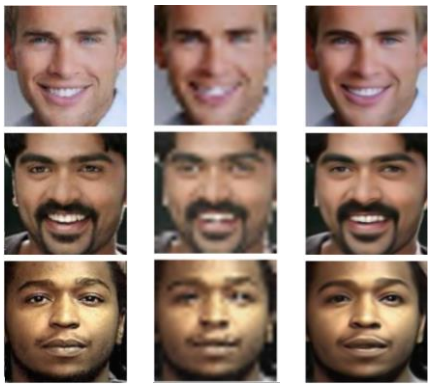
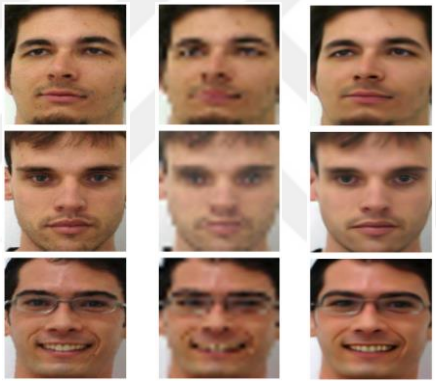
Metot	PSNR↑	SSIM↑	LPIPS↓
RCAN [63]	24.87	.889	.283
ESRGAN [64]	24.13	.876	.223
WaveletSR [65]	24.30	.878	.236
GFRNet [66]	27.13	.912	.132
DFDNet256 [62]	27.54	**.923**	.114
VNet	**28.73**	.880	**.093**

Tablo 5.3. CelebA ×4 Test Skorları [62].

Metot	PSNR↑	SSIM↑	LPIPS↓
RCAN [63]	24.93	.892	.267
ESRGAN [64]	24.31	.878	.210
WaveletSR [65]	24.51	.884	.247
GFRNet [66]	27.32	.915	.124
DFDNet256 [62]	27.77	**.925**	.103
VNet	**28.27**	.889	**.083**

VGGFace2, CelebA, UTKFace, MultiPie ve FEIFace veri setlerinden alınan test görüntülerinin asılları (HR), 28x28 piksel boyutlarına düşürülerek bozulmuş versiyonları (LR) ve bozulan görüntülerin 4 kat büyütülerek süper çözülmesi sonucu elde edilen çıktılar (SR) Tablo 5.4'te sırasıyla gösterilmiştir.

Tablo 5.4. Test Veri Setleri ×4 Çıktıları.

VGGFace2			UTKFace		
					
HR	LR	SR	HR	LR	SR
CelebA			**FEIFace**		
					
HR	LR	SR	HR	LR	SR
MultiPie					
					
HR	LR	SR			

5.2. Gerçek Dünya Senaryosu

Tez çalışmasında önerilen VNet ağını eğitmek için öncelikle VGGFace2 veri seti kullanılmıştır. VGGFace2 veri seti, geniş kapsamlı ve oldukça çeşitli yüz görüntüleri içerdiği için ağın temel özellikleri öğrenmesi açısından faydalı olmuştur. Ancak bu veri seti, çoğunlukla önden çekilmiş yüzlerden oluştuğu için tezin asıl amacı olan gerçek dünya güvenlik kamera görüntülerini iyileştirme konusunda yeterli başarıyı sağlayamamıştır.

VNet ağı, VGGFace2 veri seti ile 49 epok boyunca eğitildikten sonra elde edilen modelin ağırlıkları kullanılarak, ChokePoint veri seti ile ağın eğitimine devam edilmiştir. ChokePoint veri seti, çeşitli noktalardan geçen insan görüntülerinden oluşması ve gerçek dünya güvenlik kamerası görüntülerine yapısal olarak daha benzer olması sebebiyle tercih edilmiştir. Bu süreçte ChokePoint veri seti, özellikle güvenlik kameralarından elde edilen yüz görüntüleri üzerinde modelin performansını optimize etme konusunda kritik bir rol oynamış, bu ikinci eğitim aşaması sonunda model, kamera sistemleri için daha uygun hale gelmiştir.

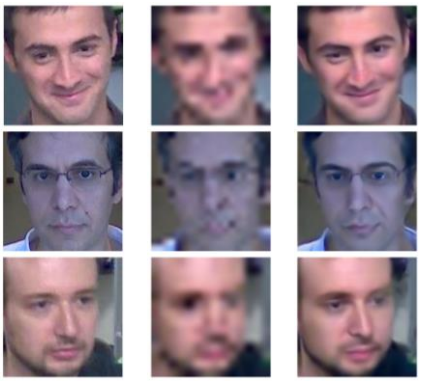
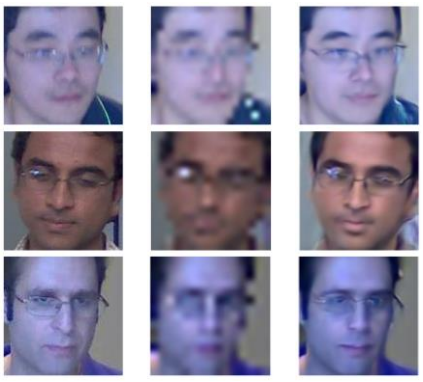
Ağ tasarımında ve eğitim parametrelerinde herhangi bir değişikliğe gidilmeden, VNet ağı ChokePoint veri seti ile öncelikle 100 epok boyunca, girdi görüntülerini 4 kat süper çözünürlüğe çıkartabilmek için eğitilmiş, daha sonra 200 epok boyunca da 8 kat süper çözünürlük için eğitilerek, 4 ve 8 kat süper çözünürlüğe uygun hale getirilmiştir. Elde edilen model, eğitim esnasında kullanılmayan 2.000 test görüntüsü ile hem ×4 ölçekte hem de ×8 ölçekte test edilmiş ve ilgili metrik değerleri Tablo 5.5'te gösterilmiştir.

Tablo 5.5. ChokePoint Veri Seti Test Skorları.

Ölçek	MAE↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
×4	.04648	.00439	24.575	0.7977	0.1359
×8	.05128	.00542	23.921	0.7722	0.1612

ChokePoint veri setinden alınan test görüntülerinin asılları (HR), 28x28 piksel boyutlarına düşürülerek bozulmuş versiyonları (LR) ve bozulan görüntülerin 8 kat büyütülerek süper çözülmesi sonucunda elde edilen çıktılar (SR) Tablo 5.6'da sırasıyla gösterilmiştir.

Tablo 5.6. ChokePoint ×8 Test Çıktıları.

ChokePoint					
					
HR	LR	SR	HR	LR	SR

6. TARTIŞMA

Elde edilen sayısal veriler ışığında, önerilen VNet mimarisinin yüz görüntülerinden oluşan VGGFace2 ve CelebA veri setlerinde, görüntü boyutunun 4 katına çıkarılmasına dayalı testlerde, literatürdeki diğer bazı ağlara göre daha yüksek PSNR ve daha düşük LPIPS değeri elde etmiştir. SSIM değerinin nispeten düşük çıkmasının sebebi ise, bu metriğin görüntüleri karşılaştırırken yapısal benzerliğin yanı sıra parlaklık ve kontrast benzerliğini de ele almasıdır. VNet modeli yapısal olarak yeterli iyileştirmeler sağlasa da parlaklık ve kontrast konusunda daha fazla eğitilmesi, optimize edilmesi gerekmektedir. Stabil ve genellenebilir bir mimari ortaya koymak adına, test için kullanılan diğer UTKFace, FEIFace ve MultiPie veri setlerinde de benzer PSNR, SSIM ve LPIPS değerleri elde edilmiştir. Tablo 5.1'de gösterilen test veri setleri üzerinde MAE ve MSE metriklerinden elde edilen sonuçlar da PSNR değerlerini destekler niteliktedir.

Görsel test çıktılarında da görüleceği üzere, görüntü boyutu 4 katına çıkarıldığında, bozulmuş görüntüdeki bulanık, gürültülü ve çarpık dokular; göz, ağız, burun ve diğer yüz bölgeleri onarılarak asıl görüntüye uygun bir biçimde restore edilmektedir.

Diğer yandan MultiPie veri seti üzerinde yapılan testlerde düşük metrik başarımlar alınmıştır. Bunun sebebi ise Tablo 5.1'deki diğer dört veri setindeki görüntüler çoğunlukla önden çekilmişken, MultiPie veri seti birçok açıdan ve farklı pozlamalar ile çekilmiş yüz görüntülerinden oluşmaktadır. Bu tür handikapların önüne geçmek adına, eğitim için kullanılan veri setlerinin açısal ve pozlama yönünden zengin olması ve örnek sayısının da olabildiğince fazla olması gerekmektedir.

Tez kapsamında öncelikle yüz görüntülerinin süper çözünürlüğünde kullanılan en kapsamlı veri setlerinden olan VGGFace2 veri seti ile eğitim yapılmıştır. Ancak bu ve benzeri veri setleri hem video görüntüleri için hem de güvenlik kameraları perspektifinden bakıldığında uyumsuz kalmaktadır. Bu bağlamda direkt olarak kameralarla çekilmiş videolardan elde edilen ChokePoint veri setinin kullanılmasıyla önerilen ağ gerçek dünya senaryolarına daha uygun hale gelmiştir.

KAYNAKLAR

[1] M. T. AĞDAŞ and S. GÜLSEÇEN, "Automatic Weapon and Knife Detection System on Security Cameras: Comparative YOLO Models," *European Journal of Science and Technology*, Sep. 2022, doi: 10.31590/ejosat.1163675.

[2] O. Elharrouss, N. Almaadeed, and S. Al-Maadeed, "A review of video surveillance systems," May 01, 2021, *Academic Press Inc.* doi: 10.1016/j.jvcir.2021.103116.

[3] J. Jiang, C. Wang, X. Liu, and J. Ma, "Deep Learning-based Face Super-Resolution: A Survey," *ACM Comput Surv*, vol. 55, no. 1, Jan. 2021, doi: 10.1145/3485132.

[4] N. Singh, S. S. Rathore, and S. Kumar, "Towards a super-resolution based approach for improved face recognition in low resolution environment," *Multimed Tools Appl*, vol. 81, no. 27, pp. 38887–38919, Nov. 2022, doi: 10.1007/S11042-022-13160-Z/FIGURES/16.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 38, no. 2, pp. 295–307, Dec. 2014, doi: 10.1109/TPAMI.2015.2439281.

[6] C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9906 LNCS, pp. 391–407, Aug. 2016, doi: 10.1007/978-3-319-46475-6_25.

[7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 1646–1654, Nov. 2015, doi: 10.1109/CVPR.2016.182.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.

[9] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 105–114, Sep. 2016, doi: 10.1109/CVPR.2017.19.

[10] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2017-July, pp. 1132–1140, Jul. 2017, doi: 10.1109/CVPRW.2017.151.

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, Aug. 2016, doi: 10.1109/CVPR.2017.243.

[12] T. Tong, G. Li, X. Liu, and Q. Gao, "Image Super-Resolution Using Dense Skip Connections".

[13] I. J. Goodfellow *et al.*, "Generative Adversarial Nets", Accessed: May 07, 2024. [Online]. Available: http://www.github.com/goodfeli/adversarial

[14] X. Wang *et al.*, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11133 LNCS, pp. 63–79, Sep. 2018, doi: 10.1007/978-3-030-11021-5_5.

[15] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, in AAAI'15. AAAI Press, 2015, pp. 3871–3877.

[16] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," 2018.

[17] T. Zhao and C. Zhang, "SAAN: Semantic Attention Adaptation Network for Face Super-Resolution," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*, 2020, pp. 1–6. doi: 10.1109/ICME46284.2020.9102926.

[18] T. Lu *et al.*, "Face Hallucination via Split-Attention in Split-Attention Network," in *Proceedings of the 29th ACM International Conference on Multimedia*, in MM '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 5501–5509. doi: 10.1145/3474085.3475682.

[19] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, Oct. 2020, Accessed: Jul. 10, 2024. [Online]. Available: https://arxiv.org/abs/2010.11929v2

[20] Y. Wang *et al.*, "TANet: A new Paradigm for Global Face Super-resolution via Transformer-CNN Aggregation Network," Sep. 2021, Accessed: Jul. 10, 2024. [Online]. Available: https://arxiv.org/abs/2109.08174v1

[21] G. Gao, Z. Xu, J. Li, J. Yang, T. Zeng, and G.-J. Qi, "CTCNet: A CNN-Transformer Cooperation Network for Face Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 32, pp. 1978–1991, Apr. 2022, doi: 10.1109/TIP.2023.3261747.

[22] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," *CoRR*, vol. abs/1411.7766, 2014, [Online]. Available: http://arxiv.org/abs/1411.7766

[23] Z. Cheng, X. Zhu, and S. Gong, "Surveillance Face Recognition Challenge," *arXiv preprint arXiv:1804.09691*, 2018.

[24] L. Alkanhal, D. Alotaibi, N. Albrahim, S. Alrayes, G. Alshemali, and O. Bchir, "Super-Resolution using Deep Learning to Support Person Identification in Surveillance Video," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, p. 2020, Accessed: Jul. 25, 2024. [Online]. Available: www.ijacsa.thesai.org

[25] T. F. Vieira, A. Bottino, A. Laurentini, and M. De Simone, "Detecting siblings in image pairs," *Vis Comput*, vol. 30, no. 12, pp. 1333–1345, 2014, doi: 10.1007/s00371-013-0884-3.

[26] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the Radboud Faces Database," *Cogn Emot*, vol. 24, no. 8, pp. 1377–1388, Dec. 2010, doi: 10.1080/02699930903485076.

[27] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019.

[28] A. Aakerberg, K. Nasrollahi, and T. B. Moeslund, "Real-world super-resolution of face-images from surveillance cameras," *IET Image Process*, vol. 16, no. 2, pp. 442–452, Feb. 2022, doi: 10.1049/IPR2.12359.

[29] Ö. İnik *et al.*, "Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri", Accessed: May 06, 2024. [Online]. Available: http://dergipark.gov.tr/gbad

[30] M. H. Sazlı, "A brief review of feed-forward neural networks," *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, vol. 50, no. 01, pp. 11–17, Jan. 2006, doi: 10.1501/COMMUA1-2_0000000026.

[31] "Overview of a Neural Network's Learning Process | by Rukshan Pramoditha | Data Science 365 | Medium." Accessed: May 07, 2024. [Online]. Available: https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa

[32] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature 2015 521:7553*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[33] L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," 2015.

[34] S. Kılıçarslan, K. Adem, and M. Çelik, "An overview of the activation functions used in deep learning algorithms," *Journal of New Results in Science*, vol. 10, no. 3, pp. 75–88, Dec. 2021, doi: 10.54187/JNRS.1011739.

[35] A. Editors, Z. Zhou, H. Shan, R. Feng, A. Naebi, and Z. Feng, "The Performance of a Lip-Sync Imagery Model, New Combinations of Signals, a Supplemental Bond Graph Classifier, and Deep Formula Detection as an Extraction and Root Classifier for Electroencephalograms and Brain–Computer Interfaces," *Applied Sciences 2023, Vol. 13, Page 11787*, vol. 13, no. 21, p. 11787, Oct. 2023, doi: 10.3390/APP132111787.

[36] "Convolutional Neural Network – What Is It and Why Does It Matter?" Accessed: May 07, 2024. [Online]. Available: https://www.nvidia.com/en-us/glossary/convolutional-neural-network/

[37] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” 2015, doi: 10.1016/j.neucom.2015.09.116.

[38] Z. J. Wang *et al.*, “CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization,” *IEEE Trans Vis Comput Graph*, vol. 27, no. 2, pp. 1396–1406, Apr. 2020, doi: 10.1109/tvcg.2020.3030418.

[39] R. D. Rakshit, D. R. Kisku, P. Gupta, and J. K. Sing, “Cross-resolution face identification using deep-convolutional neural network,” *Multimed Tools Appl*, vol. 80, no. 14, pp. 20733–20758, Jun. 2021, doi: 10.1007/S11042-021-10745-Y.

[40] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, Accessed: May 07, 2024. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[41] D. C. Lepcha, B. Goyal, A. Dogra, and V. Goyal, “Image super-resolution: A comprehensive review, recent trends, challenges and applications,” *Information Fusion*, vol. 91, pp. 230–260, Mar. 2023, doi: 10.1016/J.INFFUS.2022.10.007.

[42] Krishnamoorthy and S. Goswami, “Comparison of Image Interpolation Methods for Image Zooming,” *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*, pp. 1413–1417, 2022, doi: 10.1109/SMART55829.2022.10047094.

[43] T. M. Lehmann, C. Gönner, and K. Spitzer, “Survey: Interpolation methods in medical image processing,” *IEEE Trans Med Imaging*, vol. 18, no. 11, pp. 1049–1075, 1999, doi: 10.1109/42.816070.

[44] AHENK VURAL, “Görüntü işlemede derin öğrenme tabanlı süper çözünürlük uygulamaları,” YÜKSEK LİSANS TEZİ, FEN BİLİMLERİ ENSTİTÜSÜ, İSTANBUL TEKNİK ÜNİVERSİTESİ, 2021.

[45] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, “Image Super-Resolution via Iterative Refinement,” *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 4, pp. 4713–4726, Apr. 2021, doi: 10.1109/TPAMI.2022.3204461.

[46] Z. Wang, J. Chen, and S. C. H. Hoi, “Deep Learning for Image Super-Resolution: A Survey,” *IEEE Trans Pattern Anal Mach Intell*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021, doi: 10.1109/TPAMI.2020.2982166.

[47] C. Dong, C. C. Loy, and X. Tang, “Accelerating the Super-Resolution Convolutional Neural Network,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9906 LNCS, pp. 391–407, Aug. 2016, doi: 10.1007/978-3-319-46475-6_25.

[48] W. Li *et al.*, “Survey on Deep Face Restoration: From Non-blind to Blind and Beyond”.

[49] X. Yu and F. Porikli, “Ultra-resolving face images by discriminative generative networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9909 LNCS, pp. 318–333, 2016, doi: 10.1007/978-3-319-46454-1_20/TABLES/1.

[50] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, “Learning Face Hallucination in the Wild,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, pp. 3871–3877, Mar. 2015, doi: 10.1609/AAAI.V29I1.9795.

[51] T. Wang *et al.*, “A Survey of Deep Face Restoration: Denoise, Super-Resolution, Deblur, Artifact Removal,” Nov. 2022, Accessed: May 08, 2024. [Online]. Available: https://arxiv.org/abs/2211.02831v1

[52] J. Nilsson and T. Akenine-Möller NVIDIA, “Understanding SSIM”, Accessed: May 08, 2024. [Online]. Available: https://research.nvidia.com/publication/2020-07_Understanding-SSIM

[53] R. Kazmierczak, G. Franchi, N. Belkhir, A. Manzanera, and D. Filliat, “A STUDY OF DEEP PERCEPTUAL METRICS FOR IMAGE QUALITY ASSESSMENT”.

[54] LiJuncheng *et al.*, “A Systematic Survey of Deep Learning-based Single-Image Super-Resolution,” *ACM Comput Surv*, May 2023, doi: 10.1145/3659100.

[55] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 815–823, Mar. 2015, doi: 10.1109/cvpr.2015.7298682.

[56] “davidsandberg/facenet: Face recognition using Tensorflow.” Accessed: Jul. 15, 2024. [Online]. Available: https://github.com/davidsandberg/facenet?tab=MIT-1-ov-file#readme

[57] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” in *International Conference on Automatic Face and Gesture Recognition*, 2018.

[58] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, “Patch-based Probabilistic Image Quality Assessment for Face Selection and Improved Video-based Face Recognition,” in *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*, IEEE, Jun. 2011, pp. 81–88.

[59] S. Y. Zhang Zhifei and H. Qi, “Age Progression/Regression by Conditional Adversarial Autoencoder,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[60] C. E. Thomaz and G. A. Giraldi, “A new ranking method for principal components analysis and its application to face image analysis,” *Image Vis Comput*, vol. 28, no. 6, pp. 902–913, Jun. 2010, doi: 10.1016/J.IMAVIS.2009.11.005.

[61] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE,” *Image Vis Comput*, vol. 28, no. 5, pp. 807–813, May 2010, doi: 10.1016/J.IMAVIS.2009.08.002.

[62] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, “Blind Face Restoration via Deep Multi-scale Component Dictionaries,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12354 LNCS, pp. 399–415, Aug. 2020, doi: 10.1007/978-3-030-58545-7_23.

[63] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image Super-Resolution Using Very Deep Residual Channel Attention Networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, pp. 294–310, Jul. 2018, doi: 10.1007/978-3-030-01234-2_18.

[64] X. Wang *et al.*, “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11133 LNCS, pp. 63–79, Sep. 2018, doi: 10.1007/978-3-030-11021-5_5.

[65] H. Huang, R. He, Z. Sun, and T. Tan, “Wavelet-SRNet: A Wavelet-Based CNN for Multi-Scale Face Super Resolution,” 2017.

[66] X. Li, M. Liu, Y. Ye, W. Zuo, L. Lin, and R. Yang, “Learning Warped Guidance for Blind Face Restoration,” 2018.

ÖZGEÇMİŞ

Ad-Soyad : Ali Hüsameddin ATEŞ

ÖĞRENİM DURUMU:

- **Lisan :** 2019, Konya Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği.
- **Yüksek Lisans :** 2024, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği Programı.

TEZDEN TÜRETİLEN ESERLER:

- Face Super Resolution Based on Identity Preserving V-Network, Sakarya University Journal of Computer and Information Sciences.