T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DERİN ÖĞRENME İLE 2 BOYUTLU GÖRÜNTÜLERDE BELİRGİNLİK TESPİTİ

GÖNÜL SİNEM ÖZDOĞAN

YÜKSEK LİSANS

Bilgisayar Mühendisliği Anabilim Dalı

Nisan – 2025
KONYA

TEZ KABUL VE ONAYI

Gönül Sinem ÖZDOĞAN tarafından hazırlanan "Derin Öğrenme ile 2 Boyutlu Görüntülerde Belirginlik Tespiti" adlı tez çalışması ..14../..04../..2025…. tarihinde aşağıdaki jüri tarafından oy birliği / ~~oy çokluğu~~ ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Prof. Dr. Halife KODAZ (Konya Teknik Üniversitesi)	……
Danışman Doç. Dr. Nurdan BAYKAN (Konya Teknik Üniversitesi)	……
Üye Dr. Öğr. Üyesi Onur İNAN (Selçuk Üniversitesi)	……

Yukarıdaki sonucu onaylarım.

Prof. Dr. Mevlüt UYAN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

İmza

Gönül Sinem ÖZDOĞAN

Tarih: Nisan - 2025

ÖZET

YÜKSEK LİSANS TEZİ

DERİN ÖĞRENME İLE 2 BOYUTLU GÖRÜNTÜLERDE BELİRGİNLİK TESPİTİ

Gönül Sinem ÖZDOĞAN

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman:
Doç. Dr. Nurdan Baykan

2025, 67 Sayfa

Jüri
Prof. Dr. Halife KODAZ
Doç. Dr. Nurdan BAYKAN
Dr. Öğr. Üyesi Onur İNAN

Görüntü verilerinde nesnelerin tespit edilmesi, teknolojinin gelişmesiyle birlikte önemli bir konu haline gelmiştir. Belirgin nesne tespiti, görüntüdeki en dikkat çekici nesneyi belirlemeyi amaçlamaktadır. Karmaşık arka planlarda, nesnelerin iç içe olduğu durumlarda sınırlarının belirgin bir şekilde tespit edilmesi zorlu bir görevdir. Literatürde; nesnelerin dokuları, parlaklık düzeyleri, renkleri gibi özelliklerine dayanarak belirgin nesneleri tespit eden yöntemler denenmiştir ancak bu yöntemlerdeki özellikler nesneleri genel hatlarıyla sınıflandırsa da karmaşık arka planlı görüntülerde başarılı olamamıştır. Son zamanlarda farklı ön işleme metotları, farklı omurga ağları kullanan yapay sinir ağı mimarileri ile uygulanan yöntemler belirginlik tespitinde yaygın olarak kullanılmaktadır. Ancak nesnelerin genel hatları ve detay özellikleri aynı değildir. Bu sebeple yapay sinir ağı ve dikkat mekanizmalarını birlikte kullanan yöntemler geliştirilmiştir. Son yıllarda doğal dil işleme yöntemlerinden yola çıkarak transformer mimarileri görüntü verilerinde kullanılmaya başlanmıştır. Bu çalışmada, DUTS ve ECSSD veri setleri kullanılarak belirginlik tespiti yapılmıştır. Bunun için öncelikle Evrişimli Sinir Ağı kullanılmıştır. Daha sonra ise transformer kullanılarak belirginlik tespiti yapılmıştır. Transformer mimarilerinin başarımının artması için veri setlerinde öncelikle gürültü ekleme, rotasyon, blur gibi veri ön işleme yöntemleri ile verilerdeki çeşitlilik arttırılmıştır. Böylelikle çalışma kapsamında veri ön işleme metotlarının başarıya etkileri de incelenmiştir. Daha sonra belirginlik tespiti gerçekleştirilirken segmentasyon yöntemi kullanılmıştır. Mekansal dikkat, bir görüntüde hangi bölgelerin daha önemli olduğunu belirleyerek modelin bu alanlara odaklanmasını sağlar. Kanal tabanlı dikkat mekanizması ise her bir özellik haritasının ne kadar önemli olduğunu değerlendirerek modelin daha anlamlı kanallara ağırlık vermesine yardımcı olur. en başarılı sonuçlar DUTS veri kümesinde 0.019 Ortalama Mutlak Hata (OMH), 0.961 Geliştirilmiş Hizalama Ölçütü (GHÖ) ve 0.936 Yapısal Benzerlik Ölçütü (YBÖ) değerleri ile “Segment Anything Model UNet - 2.1” modeli ile elde edilmiştir.

Anahtar Kelimeler: Belirgin nesne tespiti, derin öğrenme, görüntü işleme

ABSTRACT

MS THESIS

SALIENCY DETECTION WITH DEEP LEARNING IN 2 DIMENSION IMAGES

Gönül Sinem ÖZDOĞAN

Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering

Advisor: Assoc. Prof. Dr. Nurdan BAYKAN

2025, 67 Pages

Jury
Prof. Dr. Halife KODAZ
Assoc. Prof. Dr. Nurdan BAYKAN
Assist. Prof. Dr. Onur İNAN

The detection of objects in image data has become an important topic with the advancement of technology. Salient object detection aims to identify the most attention-grabbing object in an image. In complex backgrounds or when objects are overlapping, clearly detecting object boundaries becomes a challenging task. In the literature, various methods have been proposed that rely on object features such as texture, brightness levels, and color to detect salient objects. However, while these features can generally classify objects, they have not been successful in images with complex backgrounds. Recently, methods that utilize different preprocessing techniques and neural network architectures with various backbone networks have been widely used in saliency detection. However, objects' overall structure and fine details differ. Therefore, approaches combining neural networks with attention mechanisms have been developed. Inspired by natural language processing, transformer architectures have recently begun to be applied to image data. In this study, saliency detection was performed using the DUTS and ECSSD datasets. First, a Convolutional Neural Network (CNN) was used, followed by a transformer-based saliency detection approach. To enhance the performance of transformer architectures, data augmentation techniques such as noise injection, rotation, and blurring were applied to increase dataset diversity. In this way, the study also investigated the impact of preprocessing methods on model performance. For saliency detection, a segmentation method was used. Spatial attention helps the model focus on important regions in an image by determining which areas are most significant. Channel-wise attention evaluates the importance of each feature map and helps the model focus on more meaningful channels. The best results were obtained on the DUTS dataset using the “Segment Anything Model UNet - 2.1,” achieving 0.019 Mean Absolute Error (MAE), 0.961 Enhanced Alignment Measure (E-measure), and 0.936 Structural Similarity Index (SSIM).

Keywords: Deep learning, image processing, saliency object detection

ÖNSÖZ

Bu tez çalışmasında “Derin Öğrenme ile 2 Boyutlu Görüntülerde Belirginlik Tespiti” çalışması gerçekleştirilmiştir.

Yüksek Lisans öğrenimim boyunca danışmanlığımı yapan, yol göstericiliği, tecrübesi, bilgisi, mentörlüğü ve değerli görüşleri ile yardım eden Doç. Dr. Nurdan Baykan’a;

Maddi ve manevi olarak her zaman yanımda olan sevgili eşim Ali Han ÖZDOĞAN’a

Eğitimim boyunca en büyük destekçilerim ve öğretmenlerim babam Oktay GÜMÜŞSOY’a, annem Rahşan GÜMÜŞSOY’a;

Çalışmalarım sırasında moral ve destekleriyle hep yanımda olan kardeşlerim Özge GÜMÜŞSOY KOLBAŞI’na ve Erol Murat GÜMÜŞSOY’a;

Ailemizin değerlisi kedimiz Mayıs’a teşekkür ederim.

Gönül Sinem ÖZDOĞAN
KONYA-2025

İÇİNDEKİLER

SİMGELER VE KISALTMALAR

Simgeler

L : Kayıp fonksiyonu (Loss function)
$\sum$: Toplam sembolü
E_{ϕ} : Toplam sembolü, Y vektörünün boyutu kadar olan elemanları toplamı
Yi : Gerçek değer vektörünün i-inci elemanı
Pi : Modelin tahmin ettiği değer vektörünün i-inci elemanı.
log : Doğal logaritma işlemi
αs : Bir ağırlık parametresi

Kısaltmalar

BNT : Belirgin Nesne Tespiti
BüK : Birlik Üzerinden Kesişim (IoU)
DVM : Destek Vektör Makineleri
ESA : Evrişimli Sinir Ağı
İÇE : İkili Çapraz Entropi
KAA : Küresel Anlamsal Ağ
KBD : Kanal Bazlı Dikkat
KSA : Küresel Semantik Ağ
KYBA : Küresel Yerel Birleşim Ağı
MA : Mekansal Dikkat
MD : Mesafe Dönüşümü
MÖ : Makine Öğrenmesi
ÖEA : Özellik Etkileşim Ağı
PÖD : Piramit Özellik Dikkati
RO : Rastgele Orman
TBSA : Tam Bağlantılı Sinir Ağı
TKSA : Tam konvolüsyonel sinir ağları
TSA : Tekrarlayan Sinir Ağı
YZ : Yapay Zeka
YBÖ : Yapısal Benzerlik Ölçütü (SSIM)
YİA : Yerel İyileştirme Ağı
YSA : Yapay Sinir Ağları

1. GİRİŞ

Yapay zekâ, bilgisayarlara insan gibi düşünme yeteneği kazandırmayı amaçlar ve genellikle insan benzeri yetenekler geliştirir. Makinelerin düşünme yeteneği kazanmasıyla birlikte günlük yaşamda insanların yaptığı işlerdeki kalite artar, efor azalır ve verimlilik artar.

Günümüzdeki yapay zekâ ve görüntü işleme alanındaki hızlı gelişmeler, nesnelerin belirginlik tespiti (saliency detection) konusunda önemli ilerlemelere yol açmıştır. Yapay zekâ, bilgisayar sistemlerine insan benzeri düşünme yetenekleri kazandırmayı hedeflerken; belirginlik tespiti, görsel olarak dikkat çekici nesneleri tanımlama amacını taşır. Yapılan çalışmalarda nesnelerin renk, doku gibi özellikleri tespit etmek amaçlanmıştır. Ancak karmaşık arka planlar ve nesnelerin düşük seviyeli özellikleri nesnelerin tespit edilmesini zorlaştırmaktadır (Zhao ve Wu, 2019). Bu bağlamda, belirgin nesne tespiti üzerine yapılan çalışmalarda derin öğrenme ağ mimarileri ve görüntü işleme modelleri son yıllarda ön plandadır.

Görüntü işleme; dijital görüntülerin analiz edilmesi, işlenmesi ve anlamlandırılmasını sağlar. Bu alanda; görüntüdeki nesnelerin, yüzeylerin, kenarların ve diğer özelliklerin tespiti için matematiksel ve algoritmik yöntemler kullanılmaktadır. Görüntüdeki desenlerin, renklerin ve şekillerin belirlenmesi, görüntü segmentasyonu, kenar tespiti ve obje tanıma gibi işlemler, görüntü işleme uygulamalarının temelini oluşturmaktadır. Bu sayede, makineler insan gözünün algılayamayacağı bilgileri işleyerek daha doğru ve verimli sonuçlar üretebilmektedir.

Görüntü segmentasyonu, bir görüntüyü anlamlı parçalara ayırmayı amaçlayan bir süreçtir. Segmentasyon sürecinin alt dallarından biri olan nesnelerin belirginlik tespiti, görüntü verisindeki görsel olarak dikkat çekici nesneleri veya bölgeleri tanımlama amacını taşır (Zhang ve ark., 2018). Ancak, nesnelerin renk ve kontrast farklılıkları, karmaşık arka planlar, ölçek ve perspektif değişiklikleri, gölgeler ve aydınlatma değişiklikleri, belirginlik tespitini karmaşık hale getiren etkenler arasındadır (J. J. Liu ve ark., 2023). Bu zorlukların üstesinden gelmek adına gelişmiş algoritmalar, derin öğrenme ve yapay zekâ yaklaşımları, nesnelerin belirginlik tespiti konusundaki araştırmalarda önemli bir role sahiptir.

Görüntü verilerindeki belirginlik tespiti, bir görüntü veya bir görüntü içindeki bir bölgenin diğerlerinden ne kadar farklı veya dikkat çekici olduğunu belirleme işlemidir (Liang ve Luo, 2024). Bu, genellikle görsel dikkatin odaklandığı önemli nesneleri veya

bölgeleri tanımlama amacını taşır. Belirgin nesne tespiti, görüntülerdeki en dikkat çekici bölgelerin veya nesnelerin tespit edilmesini sağlar. Bu yöntem; nesne segmentasyonu, görsel arama, robotik algılama ve tıbbi görüntü analizi gibi birçok uygulama alanında büyük avantajlar sunmaktadır. Örneğin, görsel arama sistemlerinde, dikkati çeken nesnelerin öne çıkarılması kullanıcı deneyimini iyileştirir. Robotik sistemlerde ise belirgin nesne tespiti, çevreyi anlamlandırma ve görevleri optimize etme açısından kritik bir rol oynar. Ayrıca, tıbbi görüntü analizinde belirginlik haritaları, radyologların tümör veya anormal bölgeleri daha hızlı ve doğru bir şekilde tespit etmelerine yardımcı olabilir. Bu avantajlar, belirgin nesne tespitinin geniş bir yelpazede fayda sağladığını göstermektedir.

Belirgin nesne tespitinde yüksek çözünürlüklü görüntülerde hassas piksel düzeyinde etiketleme yapmak zaman alıcıdır. Görüntülerin içerdiği, karmaşık arka planlar, düşük kontrastlı veya çoklu belirgin nesneler, tespit sürecini zorlaştıran koşullardandır. Bunun yanı sıra, kullanılan veri kümelerinin sınırlı çeşitliliği ve modelin genelleme kapasitesinin yetersiz olması durumu, performansın düşmesine neden olabilmektedir. Bu zorluklar, belirgin nesne tespiti yöntemlerinin daha fazla hassasiyet sağlayacak düzeyde geliştirilmesi çalışmalarını ortaya koymuştur.

Belirginlik tespitinde; renk tabanlı yöntemler, kenar algılama yöntemleri, parlaklık farkları gibi geleneksel yöntemlerle nesnelerin tespiti gerçekleştirilmiş ancak istenilen başarılara ulaşılamamıştır. Bu nedenle yapay sinir ağı tabanlı mimariler ile çalışmalar gündeme gelmiştir.

Tez çalışması kapsamında, literatürde belirgin nesne tespiti konusunda kullanılan veri kümeleri araştırılmıştır ve temin edilmiştir. Kaynak araştırması yapılarak belirgin nesne tespiti konusundaki çalışmalar incelenmiştir. Evrişimli Sinir Ağı (ESA) ve transformer tabanlı mimariler incelenmiş ve veri kümeleri kullanılarak performansları karşılaştırılmıştır. Yapılan deneyler sonucunda transformer tabanlı mimarilerin, gerekli verilerin sağlanması durumunda ESA tabanlı yöntemlerden daha iyi sonuçlar verdiği görülmüştür ve en iyi sonucu veren yöntem belirlenmiştir. Görüntü verilerinde belirgin nesne tespiti konusunun röntgen görüntülerinden tümör tanıma, otonom araç sistemlerinde etraftaki nesneleri tanıma, güvenlik sistemlerinde ısı haritaları ile nesnenin tespit edilmesini sağlama gibi gelişen teknolojik konularda kullanılması öngörülmektedir.

2. KAYNAK ARAŞTIRMASI

Belirgin Nesne Tespiti (BNT) (Saliency Object Detection - SOD), bir görüntü veya video karesi içinde görsel olarak belirgin, dikkat çekici nesneleri tanımlama amacını taşır. Bu amaçla literatürde pek çok çalışma yapılmıştır.

Histogram eşikleme, bir görüntüde belirgin nesneleri arka plandan ayırmak için yaygın olarak kullanılan bir tekniktir (Wayalun ve ark., 2012). Bu yöntemde, piksel yoğunluk değerlerinin dağılımı analiz edilerek, belirli bir eşik değeri belirlenir ve pikseller bu eşik değerine göre sınıflandırılır. Otsu yöntemi, görüntünün yoğunluk histogramını analiz ederek eşik değerini otomatik olarak belirleyen bir tekniktir ve özellikle iki sınıflı segmentasyon için etkilidir. Bu yöntem, sınıf içi varyansı minimize edecek şekilde eşik değerini hesaplar, böylece belirgin nesne ile arka plan ayrılır.

Belirgin nesne tespiti çalışmaları kapsamında 2009 yılında Itti-Koch modeli önerilmiştir (Sur ve ark., 2009). Önerilen model insan görme sisteminin dikkati nasıl yönlendirdiğini taklit ederek, görüntüdeki belirgin nesneleri tespit etmek için renk, yoğunluk (parlaklık) ve yönelim gibi farklı görsel ipuçlarını kullanmaktadır.

Makine Öğrenmesi (MÖ) (Machine Learning-ML), Yapay Zekânın (YZ) (Artificial Intelligence-AI) bir alt dalıdır. Makine öğrenmesi algoritmaları, büyük veri kümelerinde desenleri, benzerlikleri bulmak ve yapılan analize dayalı en iyi kararları veya tahminleri yapmak için kullanılan sistemlerdir. Rastgele Orman (RO) (Random Forest-RF), Destek Vektör Makineleri (DVM) (Support Vektör Machine-SVM) gibi makine öğrenmesi yöntemleri belirgin nesneleri sınıflandırmak için kullanılabilen temel yöntemlerdir.

Bölgesel tabanlı (Region-Based) belirginlik tespiti yöntemleri, bir görüntüde dikkat çekici bölgeleri belirlemek için bölgesel özellikleri analiz eden yaklaşımlardır. Bu yöntemler, görüntüyü anlamlı bölgelere ayırarak her bölgenin görsel dikkat açısından önemini değerlendirir. Özellikle kontrast tabanlı yaklaşımlar, bir bölgenin çevresiyle olan renk, parlaklık veya doku farklarını kullanarak belirgin nesneleri vurgular (Ren ve ark., 2003).

Belirgin nesne tespiti için DVM tabanlı yaklaşımlar ile hem yerel hem de küresel özellikleri kullanarak belirginlik haritaları oluşturulmuştur (Zhang ve ark., 2018). Literatürde DVM, genellikle sınıflandırma ve regresyon problemlerinde kullanılırken, belirgin nesne tespiti için optimize edilmiştir ve görüntülerin düşük ve yüksek seviyeli özelliklerini değerlendirerek belirginlik tahminini güçlendirmektedir.

Son yıllarda, belirginlik tespitinde yapay sinir ağ mimarileri ile ilgili çalışmalar önem kazanmıştır (Zhao ve Wu, 2019). Wei ve ark., bir etiket ayırma yöntemi kullanarak belirginlik haritasını vücut ve detay haritasına ayıran modellerini tanıtmışlardır (Wei ve ark., 2019). Wei ve ark., omurga ağı olarak ResNet50 modelini kullanmışlardır. Yaptıkları çalışmada, özellikler arasındaki etkileşimi kullanmak için ResNet50 modeline Özellik Etkileşim Ağı (ÖEA) (Feature Interaction Network - FIN) yapısını eklemişlerdir. ÖEA, etiket ayırma işlemine uyum sağlamak için tasarlanmış iki ağ yapısından oluşmuştur. Kullanılan yöntemde, bir pikselin tahmin zorluğunun konumuyla yakından ilişkili olduğunu belirtmişlerdir. Bu nedenle, pikselleri eşit bir şekilde işlemek yerine; orijinal, vücut, detay etiketi olarak üç bölüme ayırmışlardır. Orijinal etiketi ayırmak için geleneksel bir görüntü işleme algoritması olan Mesafe Dönüşümü (MD) (Distance Transformation - DT) yöntemini kullanmışlardır. MD, ikili bir görüntü her bir pikselin arka plana olan minimum mesafesine karşılık gelen bir değerle dönüştürebilir. Bu yöntem, merkeze yakın piksellerin daha yüksek tahmin doğruluğuna sahip olmasını kullanmaktadır (Wei ve ark., 2019). Çalışmada, tam bağlantılı katman çıkarılarak sadece konvolüsyon bloklarını tutmuşlardır. Kullanılan omurga ağı beş farklı özellik mekanizması üretmiştir. Düşük seviyeli özellikler kullanılarak vücut ve detay tahmin görevleri için ayrı ayrı uyarlanmış iki grup özellik elde etmişlerdir. Etiket ayırma kullanılarak da, belirginlik etiketi vücut ve detay haritasına dönüştürülmüştür.

Piao ve ark., belirginlik tespiti için önerdikleri Yinelemeli Sinir Ağı ile (Recurrent Neural Network - RNN) önemli bir başarı elde etmişlerdir. Önerilen ağın deneysel olarak gösterilen üç ana katkısı bulunmaktadır. İlk olarak, RGB (Red-Green-Blue) ve derinlik akışlarından çok seviyeli eşleşen tamamlayıcı ipuçlarını çıkarmak ve birleştirmek için artık bağlantıları kullanarak derinlik iyileştirme bloğu tasarlamışlardır. İkinci olarak, zengin mekânsal bilgi içeren derinlik ipuçlarını, çok ölçekli bağlam özellikleriyle birleştirerek, belirgin nesneleri doğru bir şekilde yerleştirmede kullanmışlardır. Üçüncü olarak, yeni bir tekrarlayan dikkat modülü geliştirmişler ve modelin performansını arttırmışlardır. Karmaşık senaryolar içeren büyük ölçekli "RGB-D" veri kümesi oluşturmuş, bu veri kümesini kullanarak seçkinlik modellerini kapsamlı bir şekilde değerlendirmişlerdir. Çalışma sonucunda, önerilen yöntemin belirgin nesneleri doğru bir şekilde tanımlayabildiğini belirlemişlerdir (Piao ve ark., 2019).

Tam Bağlantılı Sinir Ağları (TBSA) (Fully Connected Networks – FCNs), farklı çalışmalarda kullanılmış ve belirgin nesne tespiti görevinde avantajları gösterilmiştir

(Liu ve Han, 2016; P. Zhang ve ark., 2017; L. Zhang ve ark., 2018). Zhao ve ark., belirgin kenar bilgisi ile belirgin nesne bilgisi arasındaki tamamlayıcılığa odaklanarak üç adımlı bir kenar rehberlik ağı (EGNet) önermişlerdir (Zhao ve Wu, 2019). EGNet'in ilk adımında, belirgin nesne özellikleri ilerleyici bir birleştirme yöntemiyle çıkarılmıştır. İkinci adımda, yerel kenar bilgisi ve genel konum bilgisi birleştirilerek belirgin kenar özellikleri elde edilmiştir. Son olarak, bu tamamlayıcı özelliklerin yeterince kullanılabilmesi için aynı belirgin kenar özellikleri çeşitli çözünürlüklerde belirgin nesne özellikleri ile birleştirilmiştir. Zengin kenar ve konum bilgisi ile birleştirilmiş özellikler sayesinde, belirgin nesnelerin sınırları daha kesin bir şekilde tespit edilebilmektedir. Yaptıkları çalışmada piksel tabanlı belirgin nesne tespiti yöntemlerinin, bölge tabanlı yöntemlere kıyasla daha avantajlı olduğu sonucuna varmışlardır. Ancak bu yöntemler, görüntülerdeki mekânsal uyumun ihmaline sebep olabilmektedir, bu da belirgin nesne sınırlarının elde edilmesini zorlaştırabilmektedir. Zhao ve ark., seçkin kenar tespiti ile belirgin nesne tespiti arasındaki tamamlayıcılığa dikkat çekmişlerdir. Önerilen EGNet modeli, ana omurga ağından bağımsızdır ve VGG tabanlı modelleri temel almışlardır.

Zeng ve ark., yüksek çözünürlüklü belirginlik tespiti konusunda üç geleneksel yöntem bulunduğunu belirtmişlerdir (Zeng ve ark., 2019). İlk yöntem, sıralı bir dizi havuzlama işleminden sonra nispeten yüksek çözünürlük ve nesne detaylarını korumak için giriş boyutunu artırmaktır. Ancak, büyük giriş boyutu bellek kullanımında artışlara neden olabilmetedir ve çok yüksek donanım özellikleri gerektirir. İkinci yöntem, girişleri kısım kısım bölmek ve bölünen bölgelerde tahminler yapmaktır. Bu tür bir yöntem zaman alıcıdır ve arka plan gürültüsünden etkilenebilmektedir. Üçüncüsü, graf kesimleri gibi son işleme yöntemlerini içerir. Zeng ve ark., yaptıkları çalışmada yüksek çözünürlüklü görsellik tespiti için yeni bir veri kümesi olan HRSOD'u tanıtmışlardır (Zeng ve ark., 2019). Ayrıca, global anlamsal bilgiyi ve yerel yüksek çözünürlüklü detayları birleştiren Küresel Semantik Ağ (KSA) (Global Semantic Network - GSN), Yerel İyileştirme Ağı (YİA) (Local Refinement Network - LRN) ve Küresel ve Yerel Birleşim Ağı (KYBA) (Global-Local Fusion Network - GLFN) bileşenlerinden oluşan bir yaklaşım önermişlerdir. GSA, global bir bakış açısında semantik bilgi çıkarmayı amaçlamıştır. KSA'nın rehberliğiyle, YİA belirsiz alt bölgeleri rafine etmek üzere tasarlanmıştır. Son olarak, KYBA yüksek çözünürlüklü görüntüleri giriş olarak almakta ve KSA ve YİA'dan elde edilen birleştirilmiş tahminlerin mekânsal tutarlılığını arttırmıştır. Önerilen model, önceden eğitilmiş 16 katmanlı VGG ağı ile TBSA

mimarisine dayanmaktadır. Sonuçları birleştirmek için KSA ve YİA'nın sonuçlarını doğrudan birleştirmek yerine, bu sonuçları birbirine entegre etmek için yüksek çözünürlük bilgisini içeren bir KYBA eğitmeyi önermişlerdir. Yüksek çözünürlüklü RGB görüntüleri, KSA ve YİA'dan birleştirilmiş haritalar KYBA'nın girişleri olarak bir araya getirilmiştir. Bu mimariler sonucunda KYBA'nın son derece küçük bir model boyutuna sahip olduğunu belirtmişlerdir.

Qin ve ark., Sınır Farkındalıklı Belirgin (Boundary-Aware Salient) nesne tespiti için BASNet adlı bir tahmin-rafine modeli önermişlerdir (Qin ve ark., 2019). Tahmin modülü, U-Net benzeri bir kodlayıcı-kod çözücü (Encoder-Decoder) ağıdır ve belirginlik haritasını giriş görüntülerinden tahmin etmek için yoğun bir şekilde denetlenmektedir. Bu tahminin ardından, çok ölçekli Artıklar Rafinasyon Modülü, tahmin modülünün ürettiği belirginlik haritasını gerçek değer ile karşılaştırarak rafine etmektedir. Hibrit bir kayıp fonksiyonu, ağın girdi görüntüsü ile gerçek değer arasındaki öğrenmeyi sağlamakta ve bu kayıp; piksel, yama ve harita seviyelerinde İkili Çapraz Entropi (İÇE) (Binary Cross Entropy - BCE), Yapısal Benzerlik Ölçütü (YBÖ) (Structural Similarity - SSIM) ve Birlik Üzerinden Kesişim (BüK) (Intersection-over-Union - IoU) kayıplarını birleştirmektedir. İÇE, YBÖ ve BüK kayıplarının etkileri görselleştirilmiş ve bu kayıpların birlikte kullanılmasının öğrenme sürecini optimize etmeye yardımcı olduğu gösterilmiştir.

Belirgin nesne tespiti, insan görsel sistemi ve bilişsel mekanizmalarını taklit ederek belirgin nesneleri tanımlama ve segmente etmeyi hedeflemektedir (Wang ve ark., 2023). Belirgin nesne tespiti konusunda birden fazla nesnenin ve karmaşık arka planın bulunduğu sahnelerde doğruluk ve sağlamlık daha fazla geliştirilmelidir. Wang ve ark. ilk olarak, ağın piksel düzeyi, bölge düzeyi ve nesne düzeyi özelliklerini öğrenmesini yönlendirmek için çok düzeyli hibrit bir kayıp fonksiyonu önermişlerdir. Ek olarak çok ölçekli özellik iyileştirme modülü (ME-Modülü) geliştirmişlerdir ve giriş özellik dizisinin boyut sırasını değiştirerek küresel ve ayrıntılı özelliklerin kademeli olarak birleştirilmesini ve iyileştirilmesini sağlamışlardır (Wang ve ark., 2023).

Yüksek çözünürlüklü görüntüler ve bu görüntülere ait piksel düzeyinde yapılan etiketlemeler, düşük çözünürlüklü görüntülere kıyasla çok daha zaman alıcı ve iş gücü gerektirmektedir (Kim ve ark., 2022). Kim ve ark., InSPyReNet adlı bir çerçeve önermişlerdir. Önerilen çerçeve, yüksek çözünürlüklü görüntülerin doğrudan kullanılmadan belirgin nesne tespiti yapmasına olanak sağlamaktadır. InSPyReNet ile, belirginlik haritalarının sıkı bir görüntü piramidi yapısına sahip olmasını sağlamışlardır.

Bu yapı, piramit tabanlı görüntü tanıma yöntemiyle birden fazla sonucun birleştirilmesine olanak tanımaktadır. Aynı görüntüden alınan düşük çözünürlüklü ve yüksek çözünürlüklü iki farklı görüntü piramidini sentezleyerek, etkili algılama alanı farkını aşmayı amaçlamışlardır.

PoolNet, belirgin nesne tespiti için tasarlanmış bir derin öğrenme modelidir (J. J. Liu ve ark., 2023). Bu model, görüntülerdeki önemli nesneleri tespit etmek için yenilikçi bir havuzlama yöntemi kullanmaktadır. PoolNet'in temel amacı, görsel sahnelerde insan gözünün doğal olarak odaklandığı belirgin nesneleri otomatik olarak belirlemektir. Model, farklı havuzlama tekniklerini bir araya getirerek, çok ölçekli özelliklerin daha etkili bir şekilde çıkarılmasını sağlamaktadır. PoolNet, farklı ölçeklerdeki özellik haritalarını birleştirerek, hem ince detayları hem de geniş bağlamsal bilgiyi yakalamaktadır. Bu sayede model, küçük ve büyük nesneleri tespit etmede daha başarılı olmuştur (J. J. Liu ve ark., 2023).

Liang ve Luo, çalışmalarında belirgin nesne tespitinin, insan dikkatini çeken nesnelerin segmentasyonunu hedefleyen bir problem olduğunu belirtmişlerdir. Tam Bağlantılı Sinir Ağlarının piksel seviyesinde segmentasyon başarısı sayesinde bu alanda ilerlemeler kaydedilmiş olsa da, belirgin nesne tespiti konusunda esas odak konusunun başarıyı arttırmak olduğunu ve bellek ile hesaplama maliyetlerinin göz ardı edildiğini göstermişlerdir (Liang ve Luo, 2024). Bu durum, kaynak sınırlı uygulamalarda belirtilen yöntemlerin kullanımını zorlaştırmaktadır. Yaptıkları çalışmada, bu sorunları ele almak amacıyla MEANet adında parametre sayısı az olan RSI-SOD ağını önermişlerdir. Önerilen yöntemde kenar bilgilerini mekânsal dikkat haritalarına entegre etmişlerdir ve böylece belirgin nesnelerin daha iyi tespit edilmesini sağlamışlardır. U-Net kodlayıcı – kod çözücü ağ ile segmentasyon performansını artırmışlardır.

Asheghi ve ark., belirgin nesne tespitindeki mevcut zorlukları dört grupta sınıflandırmışlardır: karmaşık arka planlı görüntüler, düşük kontrastlı görüntüler, şeffaf nesneler ve örtüşen nesneler (Asheghi ve ark., 2024). Bu zorlukların üstesinden gelmek için, Detay Farkındalıklı Belirgin Nesne Tespiti (DASOD) yöntemi önerilmiştir. Yazarlar DASOD'u, maske oluşturma ve iyileştirme olarak iki ana aşamadan oluşturmuşlardır. İlk aşamada; sözde maske, vücut etiketi ve süper çözünürlük tekniği kullanılarak oluşturulurken; ikinci aşamada, sözde maske iyileştirme modülü, sözde kenar oluşturucu tarafından üretilen detay haritasını kullanarak maske sınırlarını netleştirmektedir (Asheghi ve ark., 2024).

3. MATERYAL VE YÖNTEM

3.1. Yapay Zekâ

Yapay zekâ 4000 yıl önce insanların cansız nesneleri harekete geçirme amacıyla ortaya çıkmıştır. Yapay sinir ağları alanındaki gelişmeler son elli yılda gerçekleşmiştir. Modern yapay zekânın kökenleri, klasik filozofların insan düşünce sistemini sembolik bir sistem olarak tanımlamalarında görülebilir (Ozturk ve Şahin, 2018).

Yapay zekânın tarihi milattan önceki tarihlere dayanmaktadır. Antik Yunan döneminde, çeşitli insansı robot fikirlerinin uygulandığına dair kanıtlar bulunmaktadır. Modern yapay zekâ, filozofların insan düşüncesini sistematize etme amacıyla tarih boyunca belirgin hale gelmiştir. 1884'te Charles Babbage, zeki davranış sergileyecek mekanik bir makine üzerinde çalışmıştır. Ancak, bu çalışmaların sonucunda, insan gibi zeki davranışlar sergileyecek bir makine üretemeyeceğine karar vermiş ve çalışmasını askıya almıştır (Mijwel, 2015). 1950'de, Claude Shannon bilgisayarların satranç oynayabileceği fikrini ortaya atmıştır. Yapay zekâ üzerindeki çalışmalar, 1960'ların başına kadar yavaş bir şekilde devam etmiştir. 1956'da Dartmouth College'da bir yapay zekâ oturumu düzenlenmiştir. Bu dönemde ilk yapay zekâ uygulamaları tanıtılmıştır. Bu dönemde geliştirilen programlar, zekâ testlerinde kullanılan geometrik formlardan ayrılmış; bu da bilgisayarların zeki bir şekilde yaratılabileceği fikrini güçlendirmiştir (Mijwel, 2015).

3.2. Makine Öğrenmesi

Makine öğrenmesi, bilgisayar sistemlerinin algoritmalar aracılığıyla verilerden öğrenme yeteneğine sahip olmasını sağlayan yapay zekânın bir alt dalıdır. Bu yaklaşım, belirli bir görevi veya problemi çözmek için programlanmak yerine, veri kümelerini analiz ederek örüntüler ve ilişkiler bulma yeteneği üzerine odaklanır. Makine öğrenimi modelleri, genellikle deneme yanılma yoluyla geliştirilir ve veri kümelerindeki karmaşık ilişkileri anlamak için istatistiksel yöntemleri kullanır (Mohri, Rostamizadeh ve Talwalkar, 2018).

Makine öğrenimi, görüntü ve ses tanıma, dil işleme, oyun stratejileri, otomatik sürüş, sağlık sektörü analizi gibi birçok çeşitli uygulama alanlarında kullanılmaktadır. Temel olarak, bu yaklaşım, büyük veri kümelerinden öğrenilen modelleri kullanarak

gelecekteki olayları tahmin etmek veya önceki veri setlerinden öğrenilen örüntülere dayalı kararlar almak için kullanılır (Mohri, Rostamizadeh ve Talwalkar, 2018).

Makine Öğrenmesinin eğitme yapısına göre farklı türleri vardır:

Denetimli Öğrenme: Algoritma; eğitim veri kümesindeki giriş ve çıkışları kullanarak öğrenir. Amaç, algoritmanın girişleri çıkışlara eşleme yeteneğini geliştirmektir.

Denetimsiz Öğrenme: Etiketlenmemiş veri kümelerini analiz ederek veri içindeki doğal örüntüleri ve ilişkileri keşfetmeye odaklanır. Kümeleme ve boyut azaltma gibi teknikler bu kategoriye girer.

Pekiştirmeli Öğrenme: Bir ajanın belirli bir çevre içindeki durumları gözlemleyerek bu durumlara tepki vermesini ve çevresindeki ödüllere göre en iyi eylemleri seçmeyi öğrenmesini amaçlayan bir makine öğrenme türüdür. Temelde, ajan, çevresiyle etkileşimde bulunarak deneyimlerinden öğrenir ve bu süreçte aldığı ödüller veya cezalar aracılığıyla davranışlarını iyileştirir. Pekiştirmeli öğrenme, özellikle karmaşık ve dinamik ortamlarda optimal kararlar almak için kullanılır. Bu türdeki sistemler, belirli bir görevdeki başarılarını artırmak için deneyimleri üzerinden öğrenirler.

Yarı-Denetimli Öğrenme: Bu türler, denetimli ve pekiştirmeli öğrenmenin birleşimini içerir. Sınırlı sayıda etiketlenmiş veriyle çalışırken, daha fazla denetimsiz veya pekiştirmeli öğrenmeyi kullanarak genelleme yeteneklerini artırmayı amaçlarlar.

3.3. Derin Öğrenme

Derin öğrenme, yapay sinir ağları (YSA) gibi karmaşık matematiksel ve istatistiksel modeller kullanarak zorlu problemleri çözen bir makine öğrenme alt dalıdır. Derin öğrenme, büyük veri kümeleri üzerinde otomatik öğrenme yetenekleri ile bilinir ve genellikle çok katmanlı yapay sinir ağları ile uygulanır.

Yapay sinir ağları (YSA), biyolojik sinir sistemlerinden ilham alarak oluşturulan matematiksel modellerdir. YSA'lar, bilgisayar sistemlerinde karmaşık problemleri çözmek ve desenleri öğrenmek amacıyla kullanılır. Yapay sinir ağları, hücreler (veya sinir hücresi) adı verilen temel birimlerden ve bu hücreler arasındaki bağlantılardan oluşur.

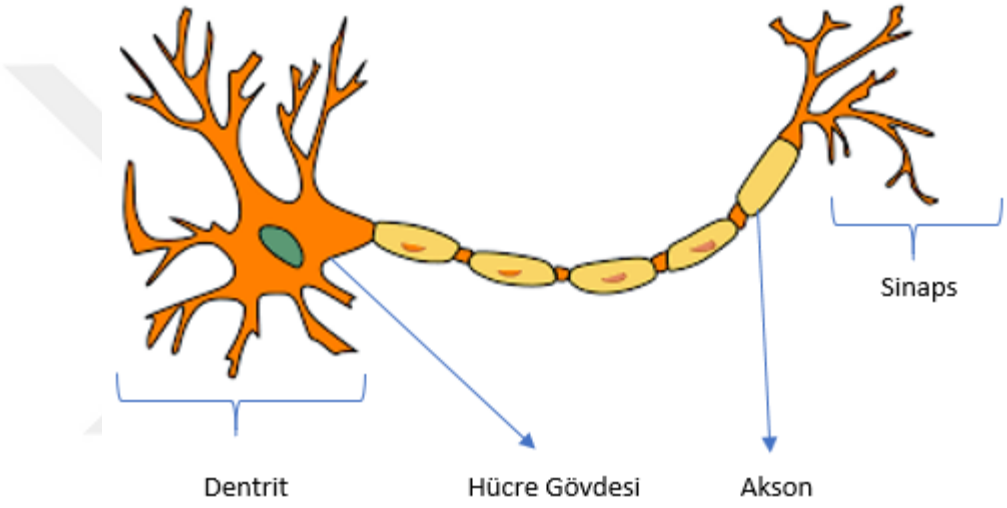
Hücre (Nöron):

Dendritler: Hücrenin giriş bölümüdür. Dendritler, diğer hücrelerden gelen sinirsel sinyalleri alır.

Hücre Gövdesi (Soma): Dendritlerden gelen sinyalleri entegre eder ve çıkış sinyali üretir.

Akson: Hücrenin çıkış bölümüdür. Hücreden çıkan sinirsel sinyalleri iletir.

Akson Ucu (Sinaps): Akson, diğer hücrelerle iletişim kurmak için sinapslar aracılığıyla bağlanır. Hücrenin temel bileşenleri Şekil 3.1'de gösterilmiştir.



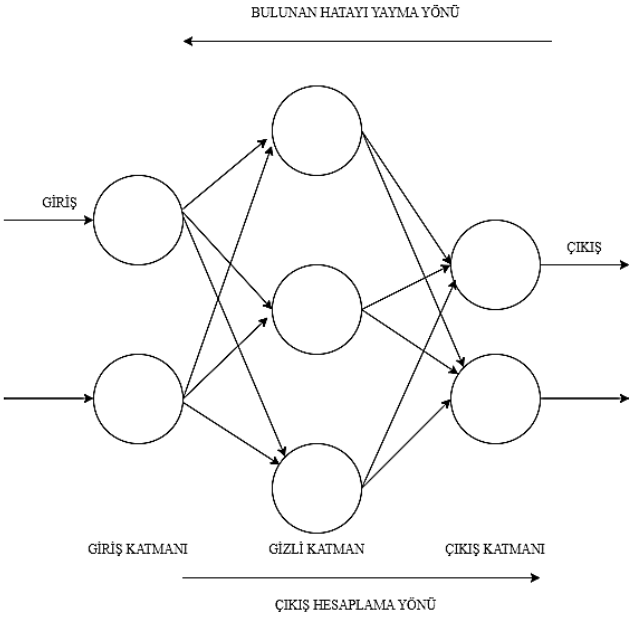
Şekil 3.1. Hücrenin temel bileşenleri (Çıtır, 2020)

Sinirsel Ağ Mimarisi:

Giriş Katmanı: Dış dünyadan alınan girdileri temsil eder.

Ara Katmanlar: Giriş ve çıkış katmanları arasında bulunan ve genellikle çeşitli özellikleri temsil eden katmanlardır.

Çıkış Katmanı: Sonuçları üreten katmandır. Genellikle bir sınıflandırma veya regresyon görevinde kullanılır. Sinirsel Ağ mimarisinin yapısı Şekil 3.2'de gösterilmiştir.



Şekil 3.2. Sinirsel Ağ mimarisinin yapısı

Aktivasyon Fonksiyonları:

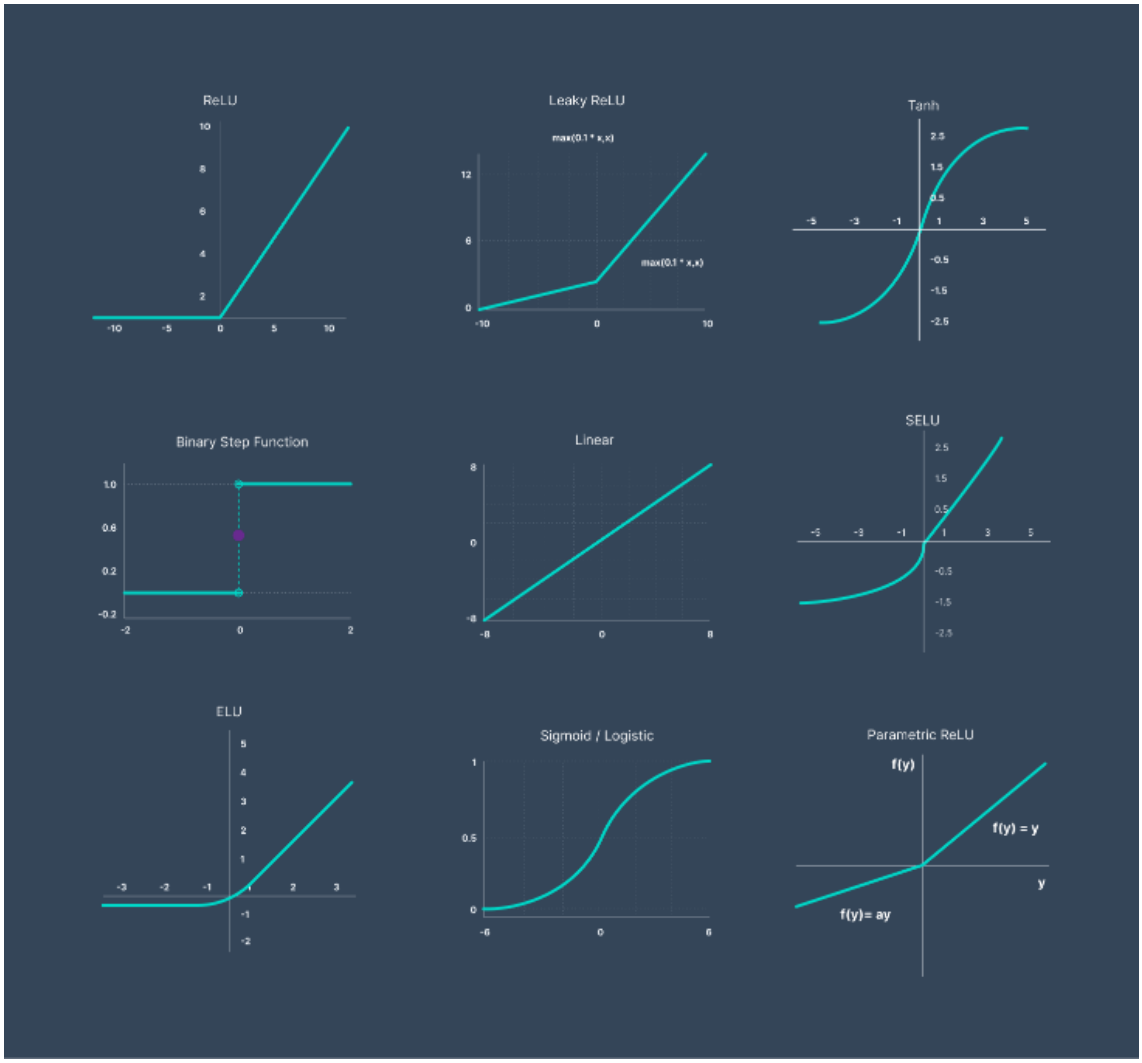
Aktivasyon fonksiyonları, yapay sinir ağlarında kullanılan matematiksel işlevlerdir. Bu fonksiyonlar, bir hücrenin çıkışını belirler ve yapay sinir ağlarının öğrenme yeteneklerini artırır. Aktivasyon fonksiyonlarının temel amaçları şunlardır:

Nonlinearite (Doğrusallık Dışı): Aktivasyon fonksiyonları, yapay sinir ağlarına doğrusallık dışı bir özellik kazandırır. Her katmandaki aktivasyon fonksiyonlarının doğrusal olması durumunda, tüm katmanlar birleştirildiğinde ağ doğrusal bir modeli temsil ederdi. Nonlinearite, ağın karmaşık yapıları ve görevlere uyum sağlayabilme yeteneğini artırır.

- Modelin Karmaşıklığını Artırma: Aktivasyon fonksiyonları, ağın karmaşıklığını artırarak daha geniş bir problem yelpazesine adapte olmasını sağlar. Nonlinearite sayesinde, ağ daha karmaşık ve hiyerarşik yapıları öğrenme yeteneğine sahip olur.
- Gradyanların Hesaplanmasını Sağlama: Aktivasyon fonksiyonlarının türevi, geriye yayılım (backpropagation) algoritması sırasında ağırlıkların güncellenmesi için kullanılır. Türev alınabilir olma özelliği, optimizasyon algoritmalarının ağırlıkları doğru bir şekilde ayarlamasını sağlar.

Sigmoid, Tanjant, ReLU, Leaky ReLU, Hiperbolik vb. farklı aktivasyon fonksiyonları bulunmaktadır (Şekil 3.3). Sigmoid fonksiyonu, çıkış değerlerini [0, 1] aralığına sıkıştırır. Özellikle çıkışın olasılık olarak yorumlandığı durumlarda kullanılır.

ReLU (Rectified Linear Unit), negatif girişleri sıfıra eşitler ve pozitif girişleri doğrudan ileterek hesaplama açısından daha hızlıdır.



Şekil 3.3. Aktivasyon fonksiyonları (Baheti, 2021)

Bağlantılar (Ağırlıklar):

Hücreler arasındaki bağlantılar, sinir ağlarının öğrenme yeteneğini sağlayan önemli unsurlardır. Her bağlantı bir ağırlığa sahiptir, bu ağırlıklar öğrenme sürecinde güncellenir.

Öğrenme Algoritmaları:

Dereceli Azalma (Gradient Descent): Ağın hatasını minimize etmek için kullanılır. Gerçek çıkış ile beklenen çıkış arasındaki farkı minimize eden ağırlık güncellemelerini gerçekleştirir.

Geriye Yayılım (Backpropagation): Hata, çıkış katmanından başlayarak giriş katmanına doğru geri yayılır ve ağırlıkların güncellenmesi için kullanılır.

Model Eğitim Aşamaları:

Yapay zekâ modelleri genellikle bir dizi hiperparametre içerir. Hiperparametreler, modelin eğitimi ve performansını kontrol etmek için kullanılan önceden belirlenmiş değerlerdir. Model eğitiminde kullanılan yapay zekâ hiperparametreleri şunlardır:

Öğrenme Oranı (Learning Rate): Modelin her iterasyonda ne kadar öğrenmesi gerektiğini belirleyen bir hiperparametredir. Yüksek bir öğrenme oranı, eğitim sırasında aşırı öğrenme riskini artırabilir, düşük bir öğrenme oranı ise koveryansın yavaşlamasına neden olabilir.

Bölüm (Çağ) Sayısı (Number of Epochs): Bölüm, tüm eğitim verisinin model tarafından bir kez görülmesini ifade eder. Eğitim süreci genellikle birkaç bölüm boyunca tekrarlanır. Ancak, çok fazla bölüm kullanmak aşırı öğrenmeye neden olabilir.

Küçük Grup Boyutu (Mini-Batch Size): Eğitim verisini küçük gruplara (mini-batch) bölmek, bellek kullanımını azaltabilir ve eğitimi hızlandırabilir.

Aktivasyon Fonksiyonları: Yapay sinir ağlarında kullanılan aktivasyon fonksiyonları, modelin öğrenme yeteneğini etkiler.

Dropout Oranı: Dropout, eğitim sırasında rastgele bir kısmı (örneğin, %20) nöronları devre dışı bırakarak aşırı öğrenmeyi azaltabilir.

Optimizasyon Algoritması: Optimizasyon algoritmaları, modelin ağırlıklarını güncelleme stratejilerini belirler.

Katman Sayısı ve Nöron Sayısı: Yapay sinir ağlarının mimarisi, katman sayısı ve katmanlardaki nöron sayısı gibi hiperparametreleri içerir.

Bu hiperparametreler, modelin eğitimini ve performansını etkileyen önemli faktörlerdir. Belirli bir görev veya veri kümesi için en iyi hiperparametre değerlerini bulmak modelin doğru çalışması için önemlidir.

3.4. Görüntü İşleme

Görüntü işleme, dijital olarak elde edilen veya bilgisayar tarafından üretilen görüntüler üzerinde çeşitli işlemlerin gerçekleştirilmesini ifade eden bir disiplindir.

Görüntü işleme, görüntüler üzerinde analiz, iyileştirme, nesne çıkarma veya bilgi çıkarma gibi çeşitli operasyonları içerir (Burger ve Burge, 2022).

Görüntü işleme, genellikle filtreleme, segmentasyon, desen tanıma, öznitelik çıkarma gibi temel teknikleri içerir. Bu teknikler, bir görüntüdeki bilgileri çıkarmak, anlamak veya görüntüyü iyileştirmek için kullanılır. Görüntü işleme algoritmaları matematiksel ve istatistiksel yöntemlere dayanır.

Görüntü formatları:

RGB (Kırmızı, Yeşil, Mavi), Gri renkli (Gray) ve Siyah-beyaz (Binary) (0,1) renk uzayları, bilgisayar grafikleri ve görüntü işleme alanlarında renkleri temsil etmek için kullanılan üç farklı renk modelidir (Garcia-Lamont ve ark., 2018).

RGB renk uzayı, renkleri üç ana renk bileşeni olan kırmızı, yeşil ve mavi bileşenlerin kombinasyonu ile ifade eder. Her piksel, kırmızı, yeşil ve mavi bileşenlerin belirli bir kombinasyonuyla temsil edilir. Bu üç bileşen 0 ile 255 arasında değerler alır.

Gri renk uzayında, her piksel sadece bir gri tonunu temsil eder ve her pikselde sadece bir sayısal değer bulunur, 0 (siyah) ile 255 (beyaz) arasında değerler alır. Gri renk uzayı, renk bilgisine ihtiyaç duyulmayan veya sadece parlaklık bilgisinin önemli olduğu uygulamalarda kullanılır. Örneğin, kenar tespiti veya görüntü analizi gibi çalışmalarda sıkça kullanılır.

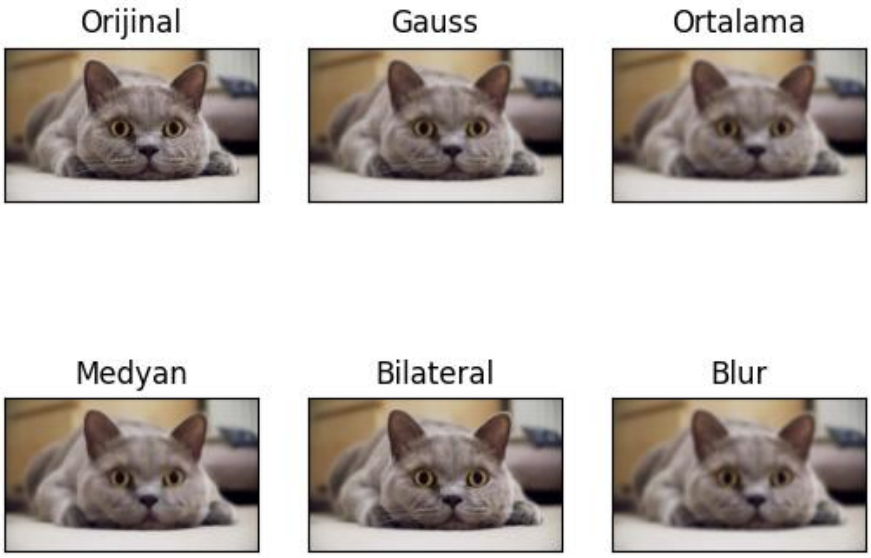
Siyah-beyaz görüntü, her pikseli sadece iki renk tonuyla (siyah ve beyaz) temsil eder. Her piksel sadece 0 (siyah) veya 1 (beyaz) değerlerinden birine sahiptir. Siyah-beyaz görüntü, basit nesne tespiti, kenar tespiti uygulamalarında kullanılır.

Bulanıklaştırma Filtreleri (Blurring Filters):

Bulanıklaştırma filtreleri, bir görüntü üzerindeki detayları azaltarak veya yumuşatarak görüntüye bir tür bulanıklık efekti ekleyen filtrelerdir. Bu filtreler, görüntü işleme ve grafik tasarımı gibi uygulamalarda kullanılır. Bulanıklaştırma, görüntüdeki gürültüyü azaltmak, keskin kenarları yumuşatmak veya belirli efektler elde etmeyi amaçlar. Literatürde yaygın kullanılan bulanıklaştırma filtreleri Gauss, Ortalama, Medyan, Bilateral filtrelerdir (Zhang, 2009, Chandel ve Gupta, 2013, Gupta ve Roy, 2019).

Bulanıklaştırma filtreleri, özellikle görüntüdeki detayları azaltma veya belirli efektleri elde etme ihtiyacı duyulan durumlarda kullanılır. Her bir filtre, farklı görsel sonuçlar üretir ve kullanım amacına bağlı olarak tercih edilir.

Gauss, ortalama, medyan, bilateral bulanıklaştırma filtrelerinin aynı görüntü üzerinde gösterimi Şekil 3.4'te gösterilmiştir.

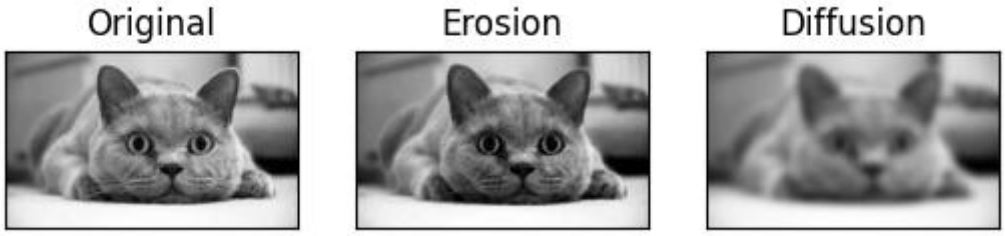


Şekil 3.4. Bulanıklaştırma filtreleri

Erozyon ve Difüzyon:

Erozyon ve difüzyon, morfolojik işlemler olarak bilinen görüntü işleme operasyonlarından ikisidir. Bu operasyonlar görüntü işleme ve görüntü analizi uygulamalarında kullanılır. Her biri görüntüdeki şekil ve yapıları değiştirmeye yönelik farklı matematiksel işlemler içerir. Bu iki işlem, görüntü işleme alanında yapısal ve kontrast iyileştirmeleri yapmak için kullanılır. Erozyon, nesnelerin kenarlarını vurgulamak veya bağlı bileşenleri genişletmek için kullanılır. Difüzyon ise, görüntüdeki detayları azaltmak ve daha düzgün bir görünüm elde etmek amacıyla kullanılır. Difüzyon, bir pikselin değerini, çevresindeki piksellerin değerleriyle birleştirerek belirler. (Kleefeld, Vorderwülbecke ve Burgeth, 2018).

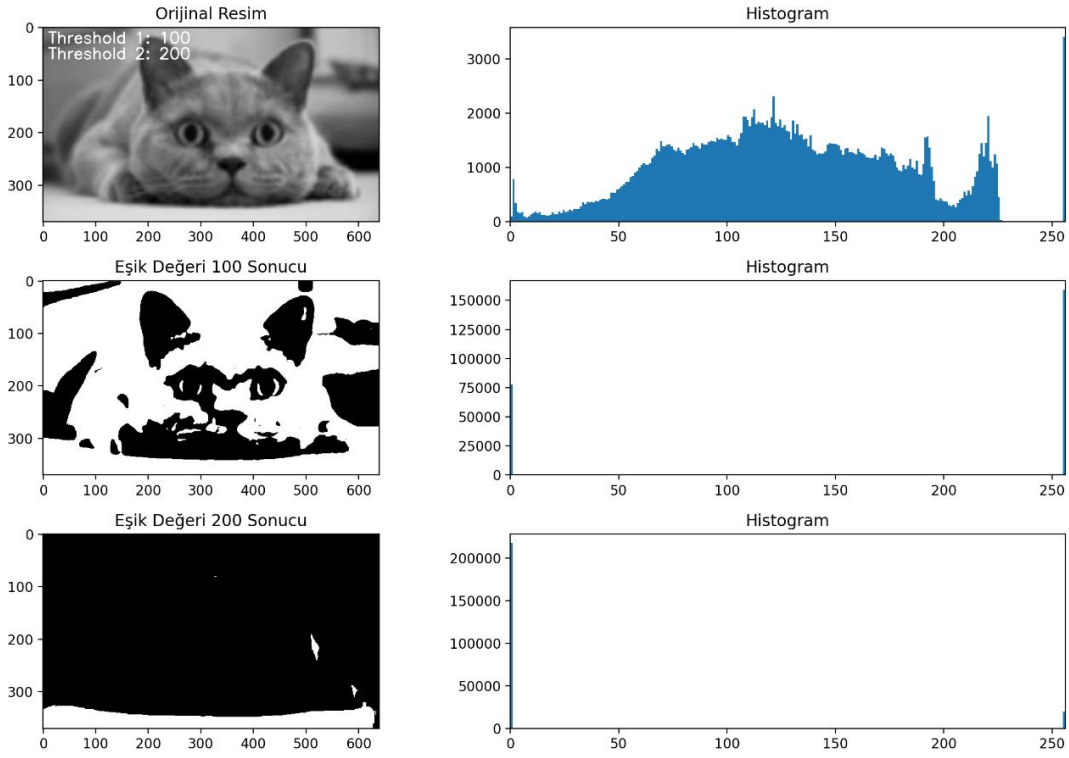
Erozyon ve Difüzyon işlemlerinin aynı görüntü üzerinde gösterimi Şekil 3.5'te gösterilmiştir.



Şekil 3.5. Görüntü üzerinde erozyon ve difüzyon işlemleri

Eşik değeri (Thresholding):

Eşikleme, görüntü işleme alanında kullanılan bir tekniktir. Bu teknik, bir görüntünün piksellerinin belirli bir eşik değeri ile karşılaştırıldığı ve belirlenen bir eşik değeri üzerindeki piksellerin bir duruma, diğerleri üzerindeki piksellerin ise başka bir duruma atanması işlemidir. Bu işlem, görüntüdeki kontrastı artırmak, nesneleri vurgulamak veya gürültüyü azaltmak amacıyla kullanılır. Histogram eşiklemede eşik değeri seçimi, uygulamanın amaçlarına ve görüntü özelliklerine bağlı olarak dikkatlice yapılmalıdır, çünkü eşik değeri seçimi sonuçları önemli ölçüde etkiler (Sezgin ve Sankur, 2004). Eşiklenmiş görüntü örnekleri Şekil 3.6.'da gösterilmiştir.



Şekil 3.6. Farklı eşik değerlerinde eşiklenmiş görüntüler

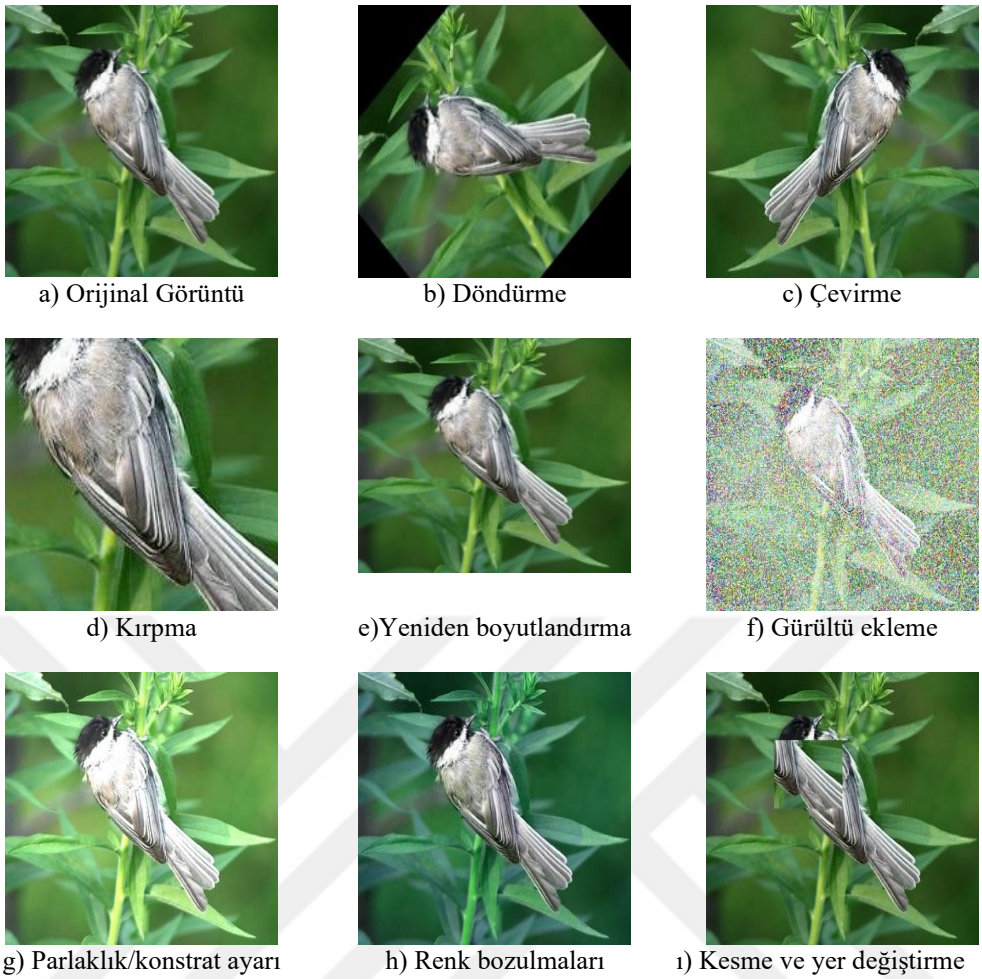
3.5. Veri Artırma

Veri arttırma (Data Augmentation), makine öğrenmesi ve özellikle derin öğrenme modellerinde kullanılan bir tekniktir. Bu yöntem, mevcut veri kümesini yapay olarak çoğaltmak amacıyla, var olan veriler üzerinde çeşitli dönüşümler uygulanarak yeni ve çeşitli örnekler üretmeyi hedefler. Veri arttırmanın temel amacı, modelin eğitim sürecinde aşırı ezberleme (overfitting) yapmasını engellemek ve modelin farklı durumlara karşı daha dayanıklı hale gelmesini sağlamaktır.

Gerçek dünyada veri toplamak hem zaman alıcı hem de maliyetlidir. Ancak makine öğrenmesi algoritmaları genellikle çalışılan problemlerin tüm çeşitliliğini öğrenebilmek için fazla sayıda veriye ihtiyaç duymaktadır. Bu yüzden, mevcut veriler üzerinde yapılan çeşitli işlemlerle daha fazla örnek elde edilmesi, modeli daha iyi eğitmenin ekonomik ve etkili bir yoludur.

Temel veri arttırma yöntemleri, kullanılan veri türüne göre değişiklik gösterse de özellikle görüntü verileri için en yaygın yöntemler (Şekil 3.7) şunlardır:

- Döndürme: Görüntü belirli açılarla döndürülerek yeni bir görünüm elde edilir.
- Çevirme: Görüntü yatay veya dikey olarak ters çevirilir.
- Kırpma: Görüntünün belli bir kısmı kesilerek yeni bir örnek oluşturulur.
- Yeniden boyutlandırma: Görüntünün boyutları değiştirilir.
- Gürültü ekleme: Görüntüye rastgele gürültüler eklenerek modelin farklılıklarla başa çıkması sağlanır.
- Parlaklık/Kontrast ayarı: Görüntünün parlaklığı ya da kontrastı değiştirilerek farklı ışık koşulları simüle edilir.
- Renk bozulmaları: Renk tonlarıyla oynanarak varyasyon yaratılır.
- Kesme ve yer değiştirme: Görüntülerin belli kısımları silinir, karıştırılır ya da başka görüntülerle birleştirilerek yeni örnekler oluşturulur.



Şekil 3.7. DUTS veri kümesinden veri arttırma teknikleri için örnek görüntüler

3.6. Evrişimli Sinir Ağları

Evrişimli Sinir Ağları (ESA), yapay sinir ağlarının bir alt türü olarak özellikle görüntü işleme alanında yaygın şekilde kullanılan derin öğrenme modelleridir. Temel olarak görsel verilerle çalışmak için tasarlanmış derin öğrenme algoritmalarıdır. ESA'lar, görüntülerdeki yerel özellikleri otomatik olarak öğrenebilme yeteneğine sahiptir ve bu sayede klasik sinir ağlarına göre daha verimli ve etkili sonuçlar üretir (Bharati ve ark., 2020). Bu ağlar, görüntülerdeki kenar, köşe, doku gibi temel desenleri tanıyarak bu bilgileri sınıflandırma ya da tespit işlemlerinde kullanabilir (Ji, Zhang, ve Wu, 2018). ESA'ların ortaya çıkış amacı, görüntü işleme gibi yüksek boyutlu verilerdeki özelliklerin manuel olarak çıkarılmasının zorluğunu ortadan kaldırmak ve bu süreci otomatik hale getirmektir. Özellikle görüntü sınıflandırma, yüz tanıma, nesne tespiti, tıbbi görüntü analizi ve otonom araç teknolojileri gibi alanlarda yaygın olarak kullanılmaktadırlar. Veriyi anlamlandırmak için ESA'nın temel yapısı birkaç ana

ardışık katmandan oluşur ve her katman, giriş verisinden daha soyut ve anlamlı özellikler öğrenir (Gidaris ve Komodakis, 2015; Maninis ve ark., 2018; Han, Li, ve Dong, 2019). İlk olarak girdi katmanına bir görüntü verilir. Ardından evrişim katmanları filtreler kullanarak görüntü üzerindeki önemli özellikleri çıkarır. Bu katmanlardan sonra, ReLU gibi aktivasyon fonksiyonları ile doğrusal olmayanlık sağlanır. Havuzlama katmanları ise bu özellik haritalarının boyutlarını küçülterek hesaplama yükünü azaltır ve genelleme yeteneğini artırır. Son olarak tam bağlantılı katmanlar ile öğrenilen özellikler kullanılarak sınıflandırma işlemi yapılır ve çıkış katmanında nihai sonuç üretilir. ESA'lar, bu katmanların ardışık bir şekilde çalışmasıyla veriden anlamlı sonuçlar çıkarmayı başarırlar ve derin öğrenmenin en güçlü araçlarından biri olarak kabul edilirler.

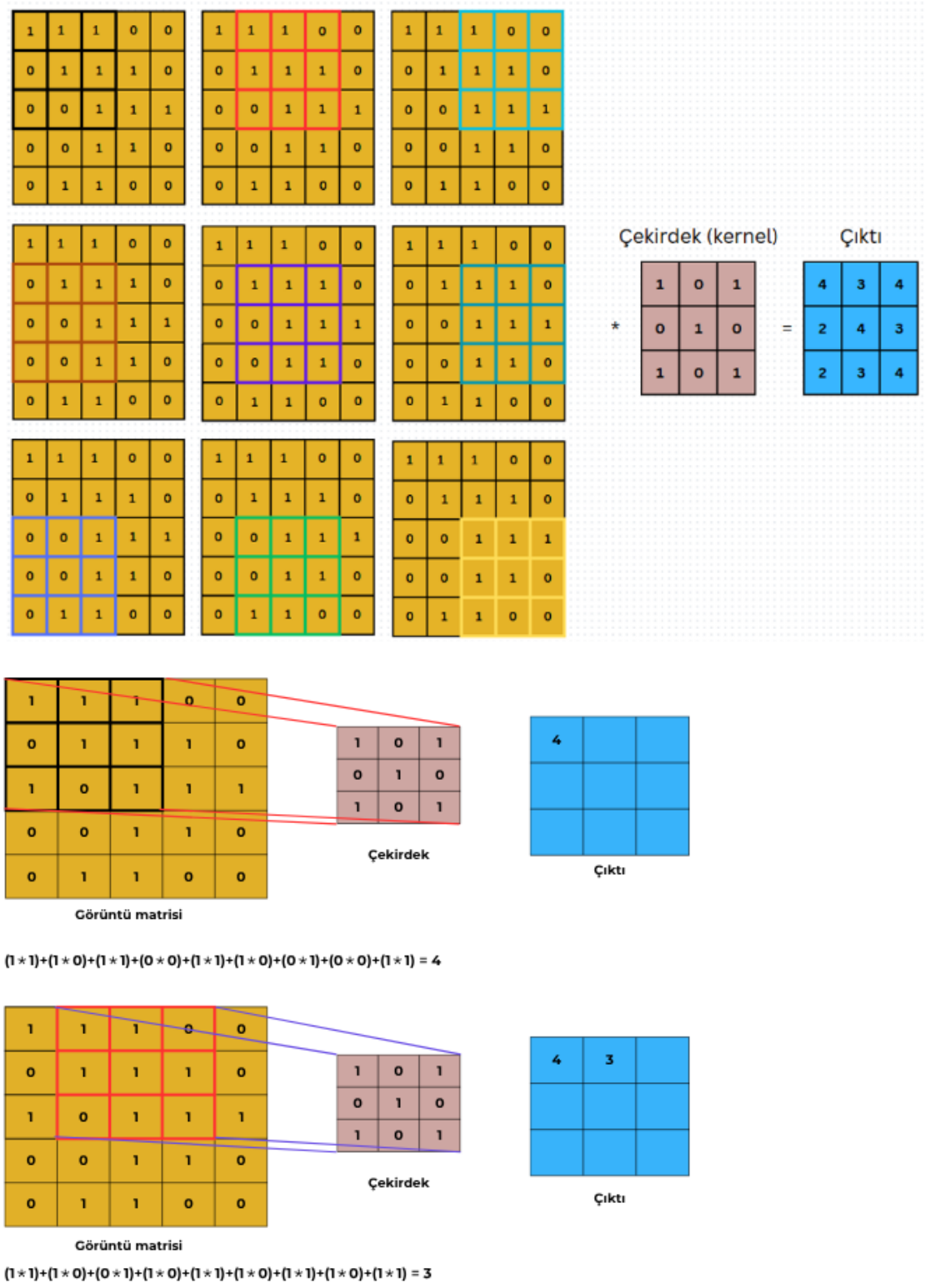
Belirginlik tespiti yapılırken Tam Bağlı Ağlar (Fully Connected Network – FCN) güncel yöntemlerde yaygın olarak kullanılmaktadır. Evrişimli Katman, ESA'daki ilk katmandır. Nesnelerin genişliklerinin ve yüksekliklerinin eşit kalmasını sağlamak için havuzlama katmanları kullanılmıştır. Nöronların toplam fonksiyonunda üretilen çıktıların geçmesi gereken değişimleri belirlemek amacıyla aktivasyon fonksiyonları kullanılmaktadır.

Evrişim katmanı (Convolutional Layer), bir ESA'nın temel yapı taşlarından biridir ve görüntülerdeki yerel özellikleri (kenar, doku, renk geçişi gibi) öğrenmek için kullanılır (Şekil 3.8) (O'Shea ve Nash, 2015). Bu katman, "filtre" veya "kernel" olarak adlandırılan matrislerle çalışır (Cui ve ark., 2017). Bu filtreler, giriş görüntüsünün üzerinde kaydırılarak nokta çarpımı işlemi uygulanır ve böylece öznitelik haritaları oluşturulur. Bu haritalar, görüntüde belirli desenlerin nerede bulunduğunu gösterir. Evrişim katmanı sayesinde ESA, model girdilerdeki önemli bilgileri otomatik olarak çıkarabilir hale gelir (Li ve ark., 2021). Şekil 3.8'de gösterilen ESA temel yapısında, 3x3 boyutunda bir çekirdek (kernel) ve 5X5 boyutunda bir giriş matrisi kullanılarak 3x3 boyutunda bir çıktı matrisi elde edilmektedir. Bu işlem, "evrişim" adı verilen temel bir hesaplama ile gerçekleştirilir. Her bir çıktı değeri, çekirdeğin ilgili konumda giriş matrisi üzerinde kaydırılmasıyla (sliding window) elde edilen 3x3 bölgedeki elemanların, çekirdekle çarpılıp toplanması sonucu elde edilir.

$(1*1)+(1*0)+(1*1)+(0*0)+(1*1)+(1*0)+(0*1)+(0*0)+(1*1) = 4$

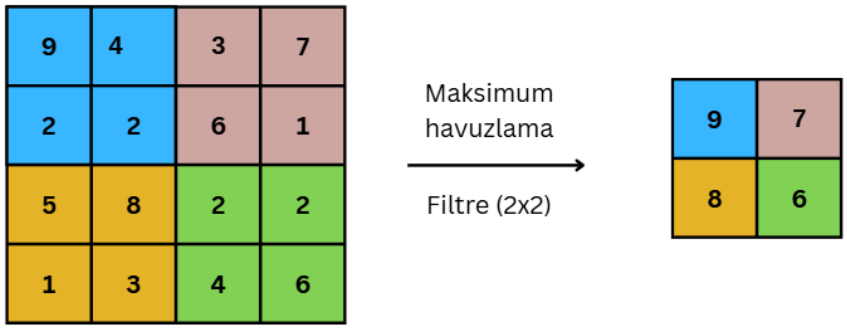
Bu işlem, çekirdeğin giriş matrisi üzerinde sağa ve aşağıya doğru kaydırılmasıyla tekrarlanarak tüm 3x3 çıktı matrisi elde edilir. Böylece, ESA modülünde dikkat haritası

oluşturmak için gerekli uzamsal ilişkiler, bu şekilde çekirdek ve giriş verisi arasında yapılan evrişimsel işlemlerle hesaplanır.



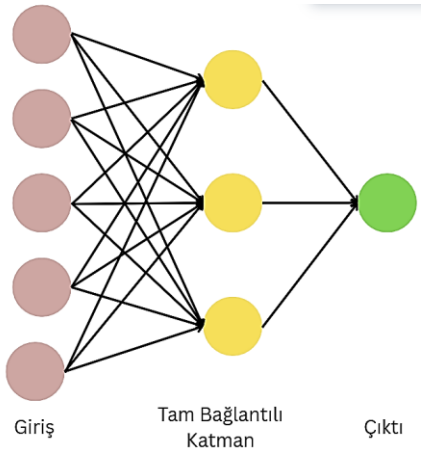
Şekil 3.8. ESA temel yapısı

Havuzlama katmanı (Pooling), evrişimle elde edilen öznitelik haritalarının boyutunu küçültmek ve modelin genel performansını iyileştirmek için kullanılır (Jie ve Wanda, 2020). En yaygın kullanılan türü "maksimum havuzlamadır" (Max pooling) ve belirli bir bölgedeki en yüksek değeri seçer (Şekil 3.9) (Christlein ve ark., 2019). Böylece modelin pozisyona duyarlılığı azalır ve genelleme gücü artar. Bu işlem, parametre sayısını ve hesaplama yükünü de azaltır. Ortalama havuzlama gibi diğer yöntemlerde ise ortalama değer alınır (ve ark., 2018). Bu katmanlar, modelin daha soyut ve özet temsil öğrenmesini sağlar.



Şekil 3.9. Maksimum havuzlama temel yapısı

Tam bağlantılı katmanlar (Fully Connected Layer), ESA'nın son katmanlarından biridir (Şekil 3.10). Önceki katmanlardan gelen çok boyutlu veriler düzleştirilir (flattening) ve bu katmana giriş olarak verilir. Bu düzleştirilmiş vektör, sinir ağının tüm sınıflar arasında karar vermesini sağlar. Buradaki her nöron, önceki katmanın tüm nöronlarına bağlıdır (Ma ve Lu, 2017).



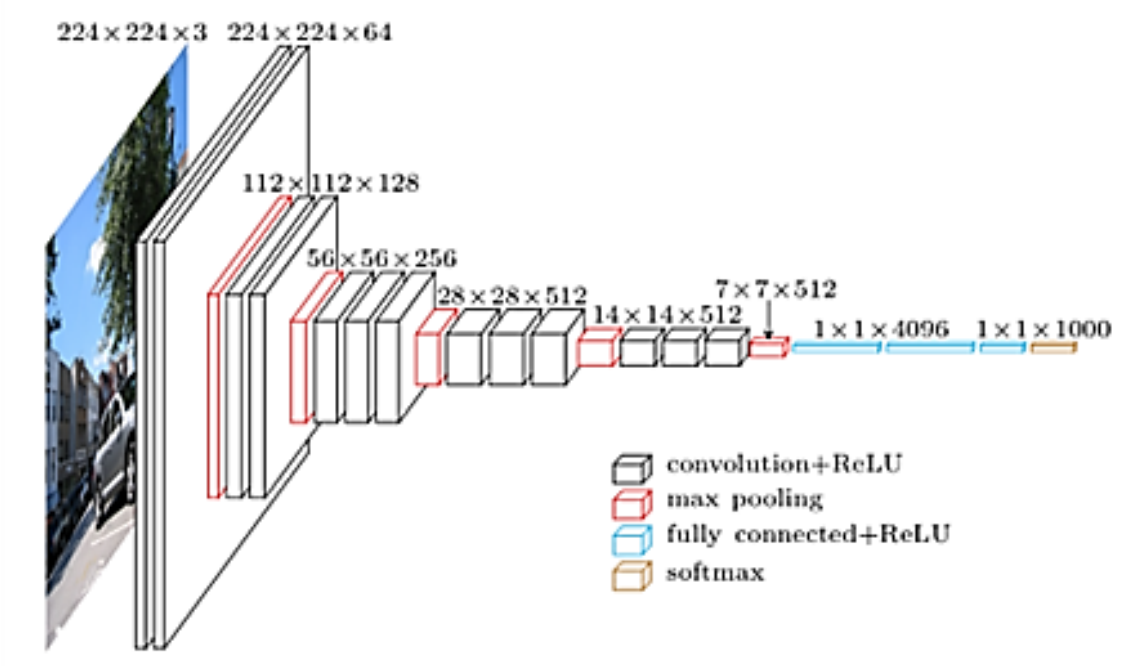
Şekil 3.10. Tam bağlantılı katman

Aktivasyon katmanı (Activation Layer), ağın doğrusal olmayan özellikler öğrenmesini sağlayan bir bileşendir. Genellikle her evrişim katmanından sonra yer alır ve öğrenme sürecinde doğrusal olmayanlık ekler (Sonoda ve Murata, 2017). Aktivasyon fonksiyonları, modelin daha karmaşık ve soyut ilişkileri öğrenmesine olanak tanır, çünkü evrişim katmanları yalnızca doğrusal işlemler yapar. Şekil 3.3'de verilen ReLU, Sigmoid, Softmax gibi aktivasyon fonksiyonları bulunur (Rasamoelina, Adjailia, ve Sinčák, 2020).

Zhao ve Wu (2019), PFA (Pyramid Feature Attention - Piramitsel Özellik Dikkat Mekanizması) adlı teknikle, çok ölçekli, çok alanlı ve yüksek seviyeli özellikler elde etmeyi hedeflemişlerdir. Bağlam Farkındalıklı Piramit Özellik Çıkarma Yöntemi ile elde edilen özellikler, dikkat mekanizmaları yardımıyla filtrelenmekte ve böylece daha değerli bilgilere odaklanılmaktadır. Araştırmacılar, farklı özelliklerin farklı anlamsal değerlere sahip olduğunu gözlemleyerek, tüm özellikleri ayrım yapmadan birleştirmek yerine, daha fazla bilgi içeren ve modelin başarısına katkı sağlayan özelliklerin ön plana çıkarılmasına önem vermişlerdir. Kullanılan dikkat mekanizmaları sayesinde bilgi tekrarını önlemeyi ve yanlış bilgilerin modelin performansını olumsuz etkilemesini engellemeyi amaçlamışlardır.

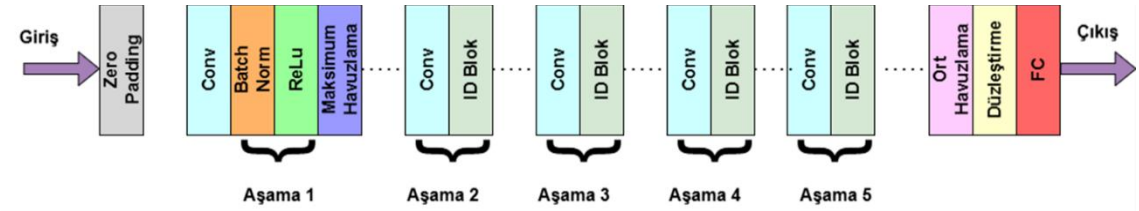
3.6.1. Derin Öğrenme Modelleri

VGG, 2014 yılında geliştirilmiş olan bir ESA mimarisidir. En çok bilinen versiyonu olan VGG-16, adını 16 katmandan oluşmasından alır. VGG mimarisi, özellikle basit ve tekrarlayan yapısı sayesinde birçok görüntü sınıflandırma ve nesne tespiti görevinde yüksek performans göstermiştir (Üzen ve Fırat, 2024). VGG-16, VGG-19 gibi versiyonları bulunmaktadır (Şekil 3.11) (Simonyan ve Zisserman, 2014)



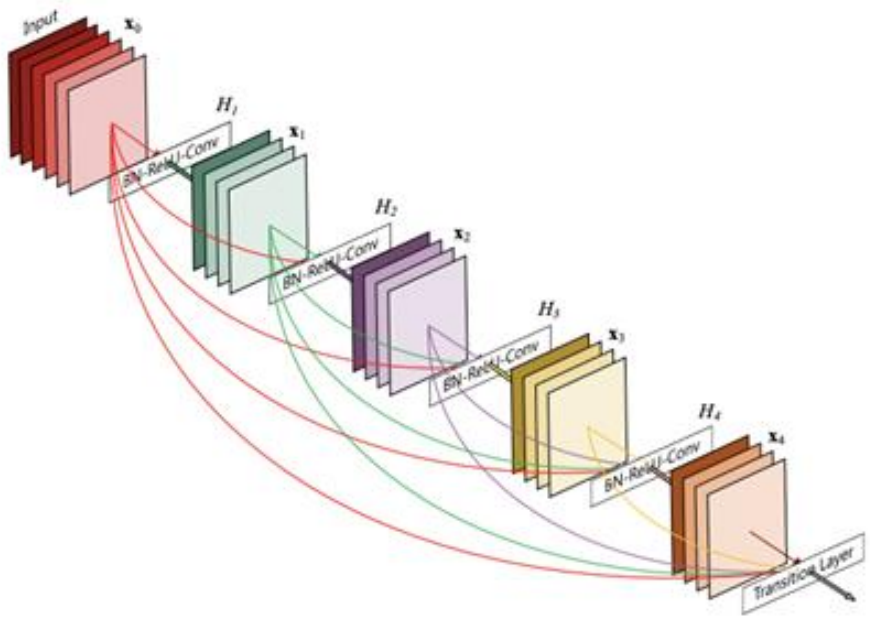
Şekil 3.11. VGG16 mimarisi, (Simonyan ve Zisserman, 2014)

ResNet, 2015 yılında Microsoft Research tarafından ImageNet yarışmasında tanıtılmıştır. ResNet'in temel yeniliği, "residual bloklar" adı verilen özel yapılardır. Bu bloklar, giriş sinyalini doğrudan çıkışa ekleyerek (skip connection) gradyan akışını iyileştirir. Bu sayede, ResNet mimarisi yüzlerce katman derinliğindeki ağların bile etkin bir şekilde eğitilmesine olanak tanır. ResNet18, ResNet34 ve ResNet50 gibi çeşitli versiyonları bulunan bu mimari, görüntü sınıflandırma, nesne tespiti ve segmentasyon gibi birçok bilgisayarla görme görevinde yaygın olarak kullanılmaktadır (Şekil 3.12) (He ve ark., 2016).



Şekil 3.12. ResNet50 model mimarisi

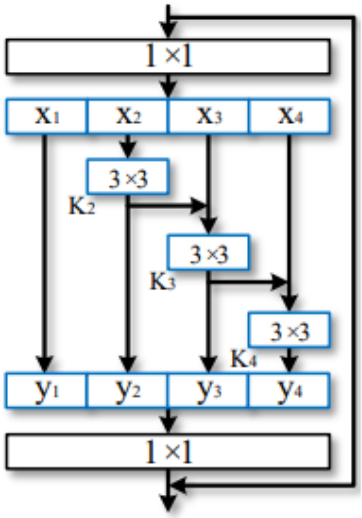
DenseNet, 2017 yılında Gao Huang ve arkadaşları tarafından geliştirilmiş, derin öğrenme mimarilerinde bilgi akışını optimize etmek için yoğun bağlantıları kullanan yenilikçi bir modeldir (Huang ve ark., 2017) (Şekil 3.13).



Şekil 3.13. DenseNet model mimarisi (Huang ve ark., 2017)

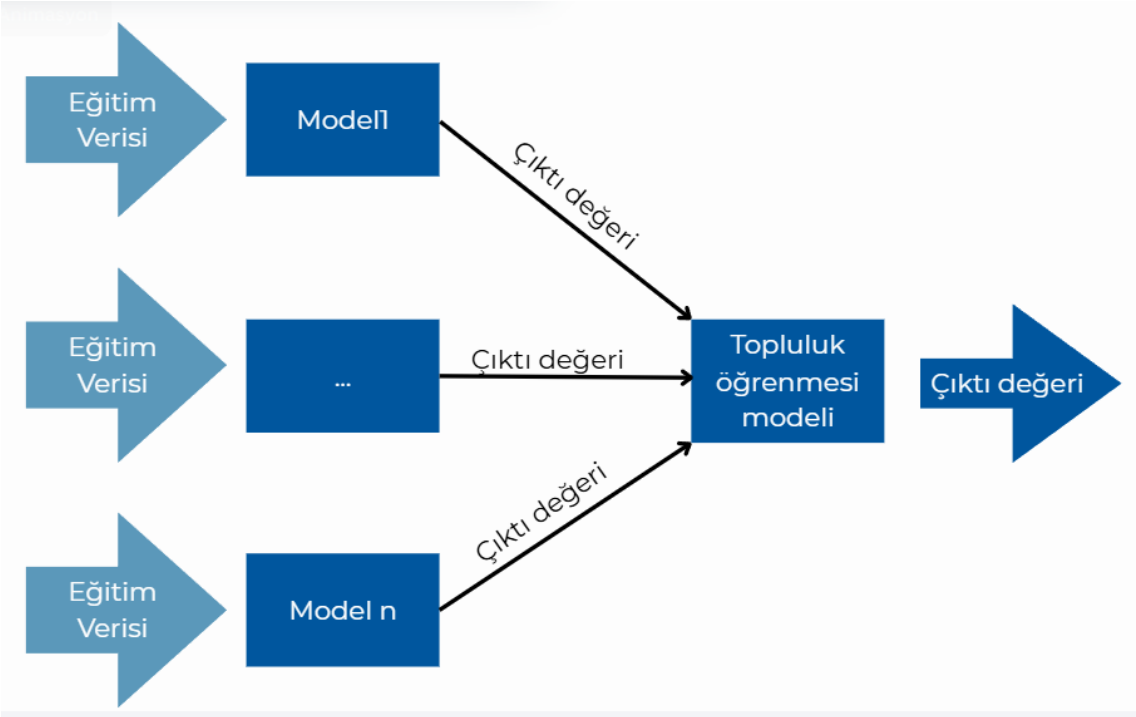
DenseNet'in temel özelliği, her katmanın, önceki tüm katmanlardan gelen özellik haritalarını giriş olarak alması ve kendi çıktılarını tüm sonraki katmanlara iletmesidir (Huang ve ark., 2017).

Res2Net50 (Gao ve ark., 2019), geleneksel ResNet mimarisine bir geliştirme olarak sunulan ve çoklu ölçekli özellik çıkarımını daha etkili hale getiren bir sinir ağı mimarisidir (Şekil 3.14). Bu model, her bir konvolüsyon bloğu içinde çoklu ölçekli (multi-scale) özellikleri aynı anda işleyebilen bir yapı içeren stratejiyi kullanır. Özellikle, her bloğun içinde bulunan kanallar daha küçük alt gruplara bölünerek farklı alan boyutlarıyla işlenir ve daha sonra bu alt gruplar birleştirilir. Bu sayede model, hem düşük hem de yüksek düzeyli detayları aynı anda öğrenebilir.



Şekil 3.14. Res2Net50 mimarisi (Gao ve ark., 2019)

Topluluk Öğrenmesi (TÖ) (Ensemble Learning), birden fazla makine öğrenmesi modelinin bir araya getirilerek, tek bir modelin başarabileceğinden daha yüksek doğruluk ve genelleme yeteneği elde etmeyi amaçlayan bir yöntemdir (Şekil 3.15.). Bu yaklaşım, her bir modelin farklı veri örüntülerini öğrenme yeteneğinden yararlanarak hata oranını azaltır ve daha kararlı sonuçlar üretir. Topluluk öğrenmesi yöntemleri arasında en yaygın olanlar bagging, boosting ve stacking teknikleridir. Bagging, modelleri farklı alt veri kümeleri üzerinde eğitirken; boosting, zayıf modelleri art arda eğiterek hataları azaltmayı hedefler. Stacking ise farklı türdeki modellerin çıktısını bir meta-model aracılığıyla birleştirir. Bu yöntem, özellikle sınıflandırma, regresyon ve görüntü işleme gibi alanlarda doğruluğu artırmak ve aşırı öğrenmeyi (overfitting) azaltmak için tercih edilmektedir.



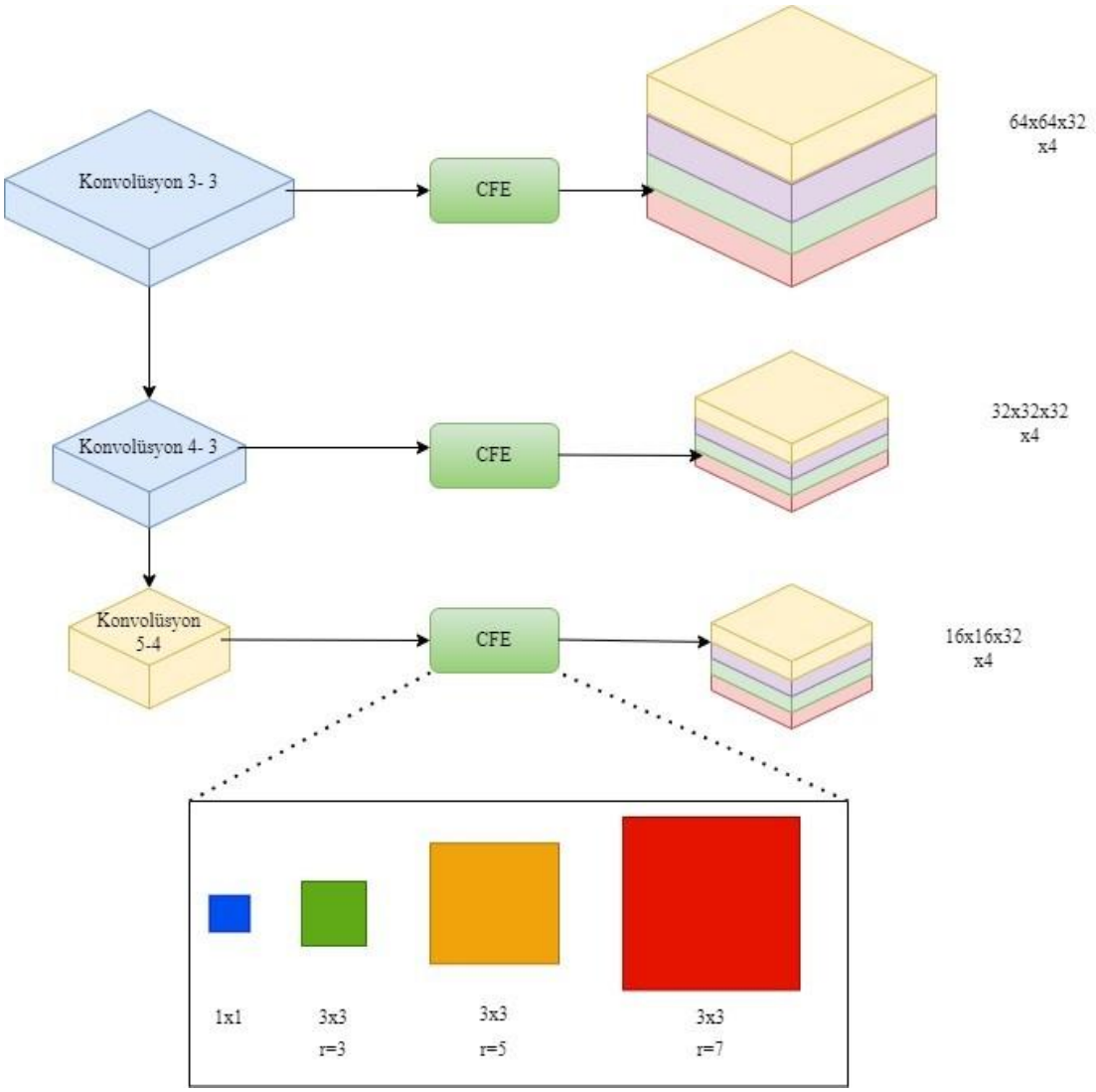
Şekil 3.15. Topluluk öğrenmesi şeması

3.6.2. Bağlam Farkındalıklı Piramit Özellik Çıkarma Yöntemi (Context-aware pyramid feature extraction)

Bağlam Farkındalıklı Piramit Özellik Çıkarma Yöntemi, farklı ölçeklerde bağlamsal bilgiler elde etmek amacıyla yapay sinir ağının belirli katmanlarından (VGG 16 için conv3-3, conv4-3 ve conv5-3 gibi) çıkarılan yüksek seviyeli özellik haritalarına odaklanır. Özellik haritaları, farklı genişleme oranlarına sahip "atrous" konvolüsyonlar ile işlenerek çoklu ölçekli ve farklı alıcı alanlarına sahip hale getirilir. Böylece model, nesnelerin ölçek ve şekil farklılıklarına karşı daha dirençli hale gelir ve belirgin nesnelerin daha tutarlı bir şekilde tespit edilmesi sağlanır (Zhao ve Wu, 2019).

Zhao ve Wu, VGG-16'daki konvolüsyon (convolutional); 3-3 (üçüncü bloktaki üçüncü evrişimli katman), 4-3 (dördüncü bloktaki üçüncü evrişimli katman) ve 5-3 katmanlarını (beşinci ve son bloktaki üçüncü evrişimli katman) temel yüksek seviyeli özellikler olarak kullanmışlardır. Son çıkarılan yüksek seviyeli özelliklerin ölçek ve şekil invariyanslarını içermesini sağlamak için 3, 5 ve 7 olarak farklı genişleme oranlarına sahip evrişimler (atrous evrişim) kullanılmıştır. Atrous evrişimde, geleneksel evrişimdeki gibi bir filtre (kernel) kullanılır, ancak bu filtredeki pikseller arasındaki

mesafe artırılır. Yani, filtredeki pikseller arasında boşluklar bırakılarak örneklem noktaları genişletilir. Bu, daha geniş bir örneklem alanına sahip olmayı sağlar ve özellik haritasındaki bağlamı artırarak daha geniş bir bilgi yelpazesi elde etmeye yardımcı olur. Zhao ve Wu, farklı atrous evrişim katmanlarından gelen özellik haritalarını ve 1×1 boyut indirme özelliğini kanallar arası birleşme (cross-channel concatenation) ile birleştirmişlerdir (Şekil 3.16) (Zhao ve Wu, 2019).



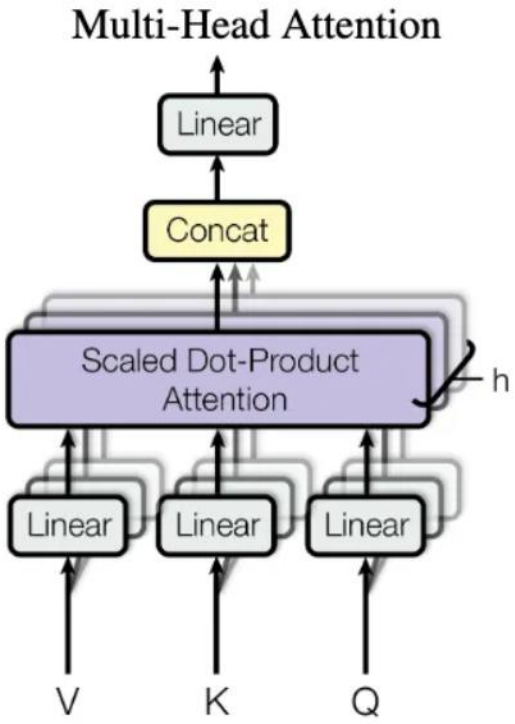
Şekil 3.16. Piramit özellik çıkarma modülü

3.6.3. Dikkat Mekanizması

Dikkat Mekanizması (Attention Mechanicsm), özellikle Doğal Dil İşleme (DDİ) (Natural Language Processing – NLP) ve derin öğrenme alanlarında çığır açıcı bir yenilik olarak ortaya çıkmıştır. Bu mekanizma ilk olarak 2014 yılında, makine çevirisi

gibi sıralı veri işlemede karşılaşılan uzun bağlam sorunlarını aşmak amacıyla geliştirilmiştir. Temel amacı, modelin, giriş verisinin tümüne eşit ağırlık vermek yerine, ilgili bölgelere odaklanmasını sağlamaktır. Dikkat mekanizması, genellikle "Kuyruk (Query)", "Anahtar (Key)" ve "Değer (Value)" adında üç temel yapısal bileşene dayanır. Bu yapıların bir araya gelmesiyle, model hangi bilginin daha önemli olduğunu dinamik olarak belirler (Vaswani ve ark., 2017).

En dış katmanda, genellikle çok başlı dikkat (multi-head attention) mekanizması yer alır; bu yapı sayesinde model farklı bağlamlardan gelen bilgileri paralel olarak işleyebilir (Saha ve ark., 2024). Alt katmanlarda ise doğrusal dönüşümler (linear transformations), katman normalizasyonu (layer normalization) ve artık bağlantılar (residual connections) gibi yapılar bulunur ve bunlar modelin öğrenme sürecini derinleştirip, daha kararlı hâle getirir (Li ve ark., 2025). Bu katmanların her biri, modelin girdiler arasındaki ilişkileri daha iyi anlamasını ve daha doğru çıktılar üretmesini sağlamada kritik rol oynar (Şekil 3.16). Bu bileşenler modelin öğrenme sürecini derinleştirip daha kararlı hâle getirir.



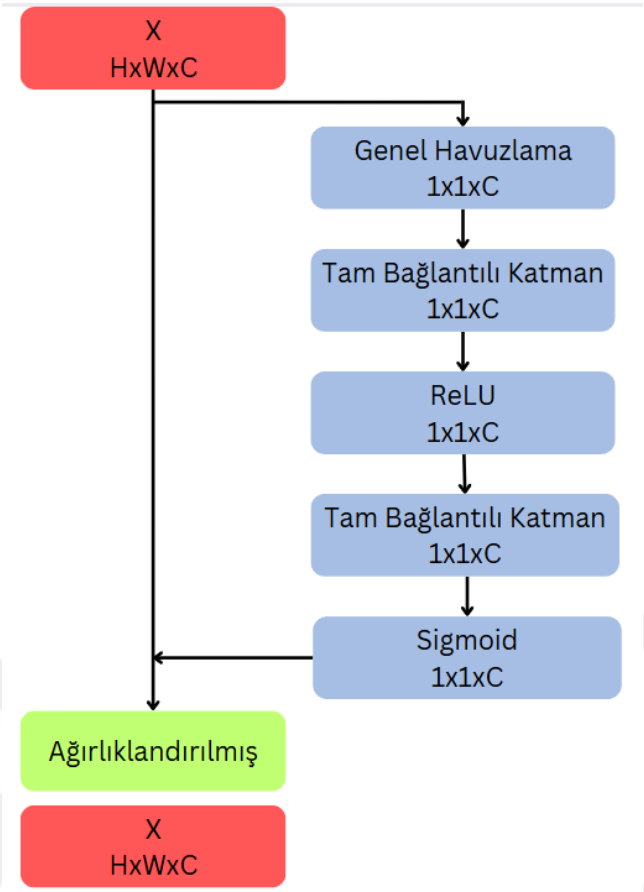
Şekil 3.17. Çok başlı dikkat çizimi (Vaswani ve ark., 2017)

PFA ağında kullanılan dikkat mekanizmalarını iki şekilde ele almışlardır. Farklı seviye özelliklerinin karakteristiklerine uygun olarak, yüksek seviyeli özellikler için

kanal tabanlı (Channel-wise attention) dikkat (Şekil 3.18) ve düşük seviyeli özellikler için mekânsal dikkat (Spacial attention) (Şekil 3.20) kullanmışlardır.

Yüksek seviyeli özellikler, daha anlamlı ve soyut bilgiler taşıdığı için, bu katmanlarda uzamsal bilgi korunmuştur. Buna karşılık, düşük seviyeli özellikler arasında anlam açısından büyük farklar bulunmadığından, bu katmanlarda kanal tabanlı dikkat mekanizması kullanılmasına gerek duyulmamıştır. Bu sayede, daha etkili ve özgün özelliklere daha çok önem verilmiş olacaktır. Şekil 3.18 ve Şekil 3.20'da gösterilen şemada H yükseklik (height), W genişlik (width) ve C kanal (channel) sayısını ifade eder. Her bir piksel konumunda C boyutlu bir özellik vektörü bulunur. Eğer boyutlandırma 1x1xC şeklindeyse, "Global Average Pooling" adı verilen katman sonucunda elde edilen özet bir kanal gösterir. Bazı durumlarda ifade edilen HxWxC / 2, kanal sayısının yarıya indirildiğini veya mekânsal boyutların (H ve W) azaltıldığını belirtir. W harfi modelin öğrenilebilir ağırlık parametrelerini (weights) temsil eder. Dikkat mekanizmasının uygulandığı durumlarda, orijinal girdi olan X, ağırlıklandırılarak X' haline getirilir; bu da dikkat ile zenginleştirilmiş, bağlama duyarlı bir özellik haritası anlamına gelir.

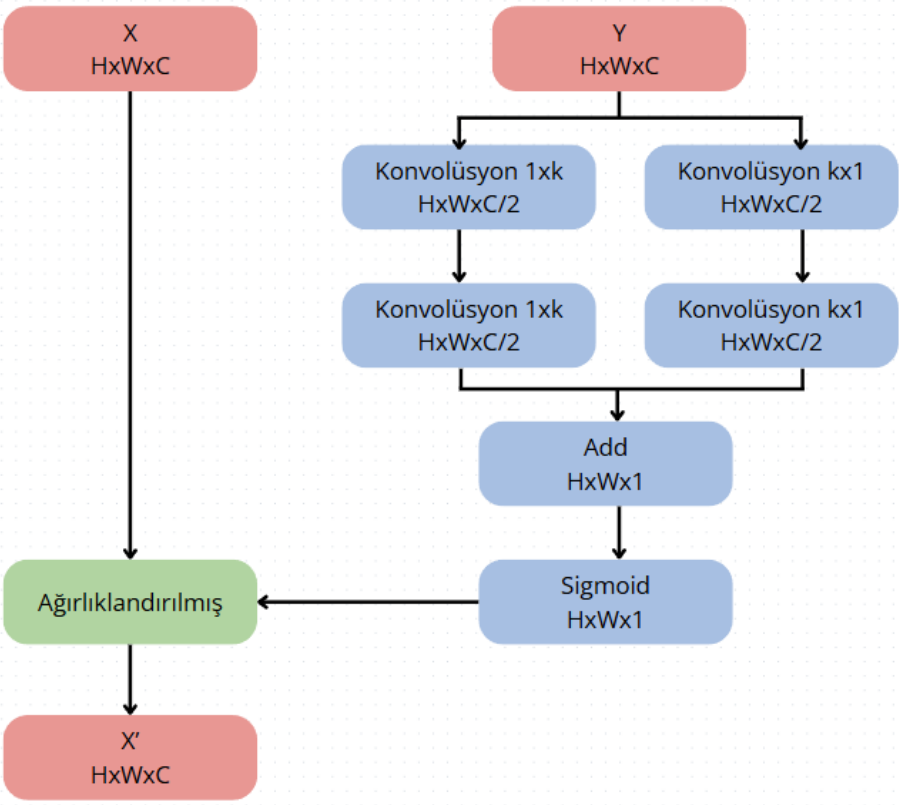
Kanal Tabanlı Dikkat (KTD) (Channel-wise Attention-CA): Bu mekanizma, yüksek seviyeli özellik haritalarında yer alan her kanalın ne kadar önemli olduğunu belirlemek için kullanılır. Her kanalın belirli anlamsal bilgilere duyarlılığı göz önüne alınarak, önemli kanallar daha fazla vurgulanır. Böylece model, belirgin nesneleri daha iyi ayırt edebilir (Şekil 3.18) (Şekil 3.19) (Zhao ve Wu, 2019).



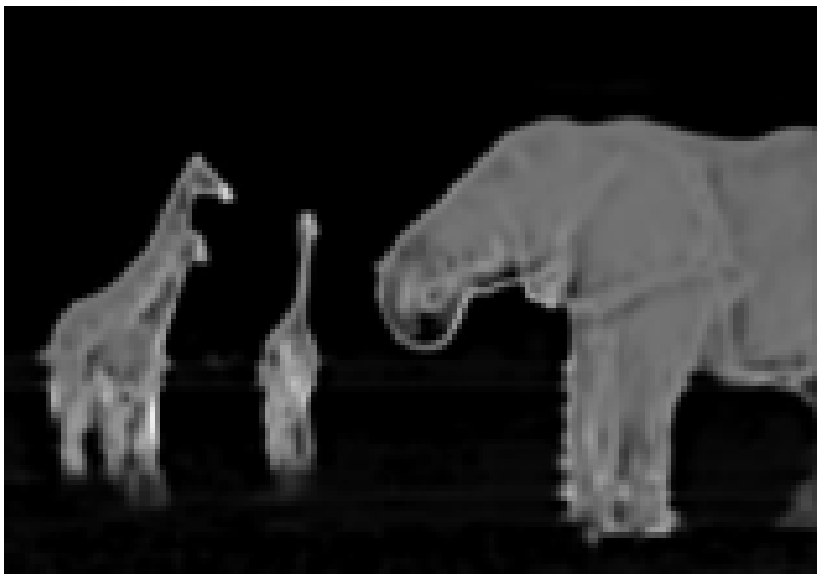
Şekil 3.18. Kanal tabanlı dikkat mekanizması

Şekil 3.19. Yüksek seviyeli özelliklere kanal tabanlı dikkat uygulanmasının örnek çıktısı (Zhao ve Wu, 2019)

Mekânsal (Uzamsal) Dikkat (MD) (Spatial Attention-SA): Düşük seviyeli özellik haritalarında özellikle ön plan bölgelerine odaklanan bu mekanizma, sınır detaylarının korunmasını sağlar. Arka plan gürültüsünü azaltarak, belirgin nesnelerin daha keskin bir şekilde tespit edilmesine katkıda bulunur (Şekil 3.20, Şekil 3.21).



Şekil 3.20. Mekânsal (Uzamsal) dikkat



Şekil 3.21. Düşük seviyeli özelliklere mekânsal dikkat uygulanmasının örnek çıktısı (Zhao ve Wu, 2019).

Belirginlik tespitinde amaç, dikkat dağıtıcı ayrıntılar olmadan, yalnızca öne çıkan nesneleri arka plandan net bir şekilde ayırarak, nesnelerin sınırlarını ve yerlerini

doğru şekilde tespit etmektir (Zhao ve Wu, 2019). Bu amaçla araştırmacılar, tüm mekânsal konumları eşit şekilde değerlendirmek yerine, mekânsal dikkat mekanizmasını özellikle ön plandaki bölgelere odaklamayı tercih etmişlerdir. Kullanılan dikkat mekanizmalarının görüntü üzerindeki örnek sonuçları Şekil 3.22'de gösterilmektedir. Şekil 3.22'de (a) ve (b) sırasıyla giriş görüntüsünü ve buna karşılık gelen Gerçek Değer (Ground Truth) haritasını temsil etmektedir. (c) ve (d), Mekânsal Dikkat (MD) (spatial attention) mekanizması olmadan ve MD ile elde edilen düşük seviye özellikleri göstermektedir. (e) ve (f) ise Kanal Tabanlı Dikkat (KTD) (channel-wise attention) mekanizması olmadan ve KTD ile elde edilen yüksek seviye özellikleri ifade etmektedir. (g) kullanılan dikkat mekanizmalarının görüntü üzerindeki örnek sonuçları ve (h) ise elde edilen görselinin oluşturulmuş sınır haritasını temsil etmektedir.

a) Orijinal görüntü | c) Düşük seviyeli özellikler | e) Yüksek seviyeli özellikler | g) Sonuç

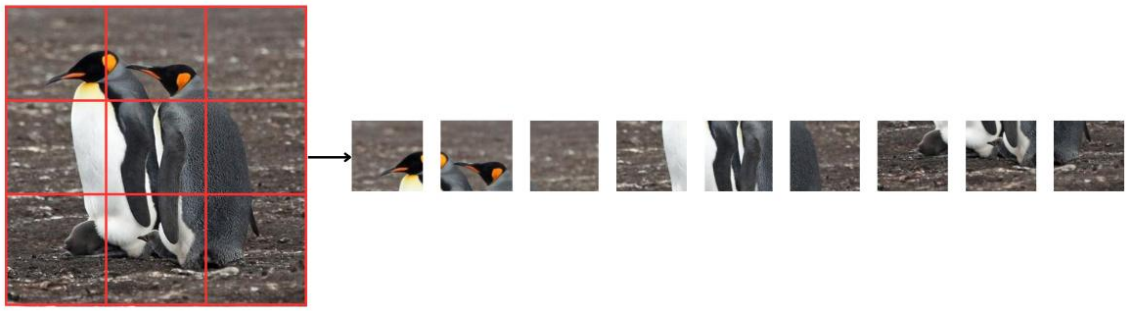
b) Doğru segmentasyon (Ground Truth) | d) Düşük seviyeli özellikler + MD | f) Yüksek seviyeli özellikler + KTD | h) Kenar sınırları

Şekil 3.22. Dikkat mekanizması sonuçları (Zhao ve Wu, 2019)

3.7. Dönüştürücü (Transformer) Yapısı

Dönüştürücü (Transformer) mimarisi, Doğal Dil İşlemede (DDİ) yaygın olarak kullanılan bir yöntemdir. Dönüştürücülerin DDİ alanındaki başarısından sonra görüntü sınıflandırma, segmentasyon gibi görevlerde kullanması araştırılmıştır (Chen ve ark., 2021).

Görüntü tabanlı transformer modellerinde, bir görüntü piksel düzeyinde işlenmez. Bunun yerine görüntü, “yama” (patch) adı verilen küçük bölgelere ayrılır. Her bir yama, transformer girişine uygun hale getirilmek üzere düzleştirilir, düzleştirici (flatten) katmanından geçirilir ve pozisyon bilgileriyle zenginleştirilir (Şekil 3.23).



Şekil 3.23. Görüntünün yamalara bölünmesi ve düzleştirilmesi (satır bazında düzleştirilme örneği)

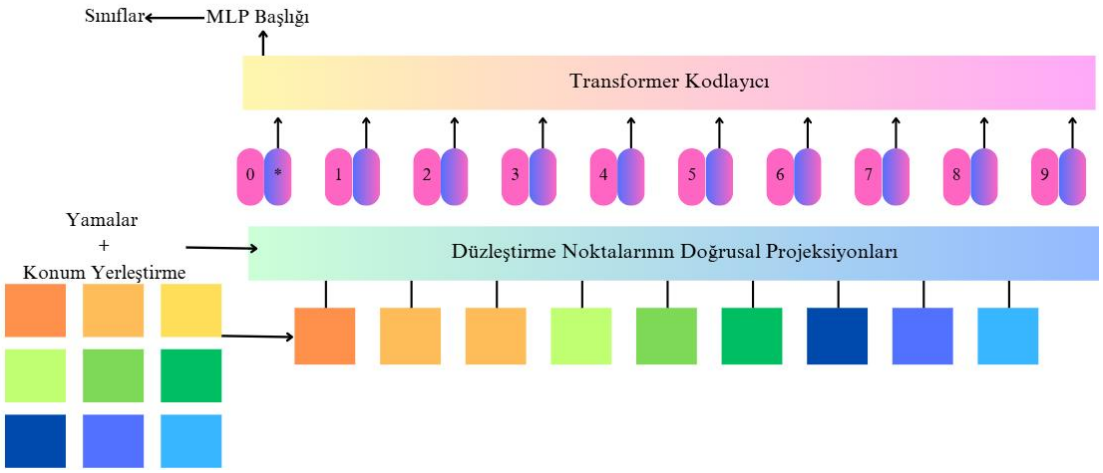
ESA tabanlı yöntemlerin yanında transformer tabanlı yöntemlerle de segmentasyon çalışmaları yapılmıştır. Transformer mimarilerinin ESA mimarilerine göre avantajı, uzun menzilli küresel bilgileri yakalama yeteneğidir. ESA, yerel kayar pencere mekanizmasıyla çalışarak sınırlı bir alanda bilgi çıkarabilir ve uzun mesafeli bağıntıları anlamakta zorlanır. Transformer ise, görüntüdeki her pikselin diğer tüm piksellerle olan ilişkisini öğrenerek küresel bağlamı etkili bir şekilde yakalayabilir. Bu yetenek, belirgin nesnelerin sınırlarının ve yerlerinin doğru bir şekilde belirlenmesinde önemli bir avantaj sağlar (Zheng ve ark., 2023).

Görüntü Dönüştürücü (GD) (Vision Transformers – ViT), bu alandaki geleneksel evrişimli sinir ağlarına alternatif olarak literatürde kullanılmaya başlanmıştır (Şekil 3.24). VvT, Veri kümesindeki verilerin sayısı arttıkça iyi sonuçlar vermektedir. ViT’lerin üç temel özelliği bulunur:

Yama (Patch): Görüntü verileri sabit boyutlu küçük parçalara ayrılır.

Transformer Mimarisi: Dikkat (attention) mekanizmaları kullanılır ve görüntülerdeki ayrılan parçaların birbiri ile ilişkisini öğrenir.

Verimlilik: GD’ler, büyük veri kümelerinde yüksek başarılar elde eder ancak küçük veri kümelerinde ESA mimarileri daha iyi sonuçlar verir.

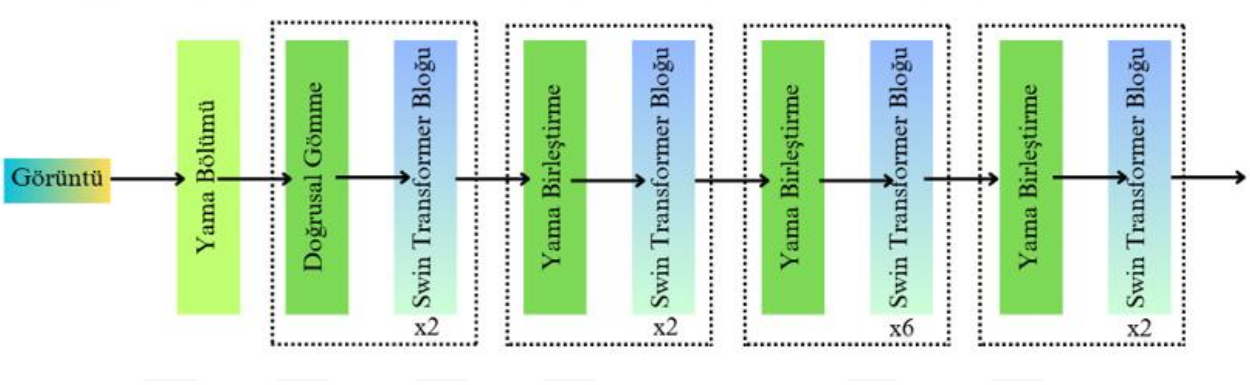


Şekil 3.24. Görüntü Dönüştürücü (ViT) mimarisi

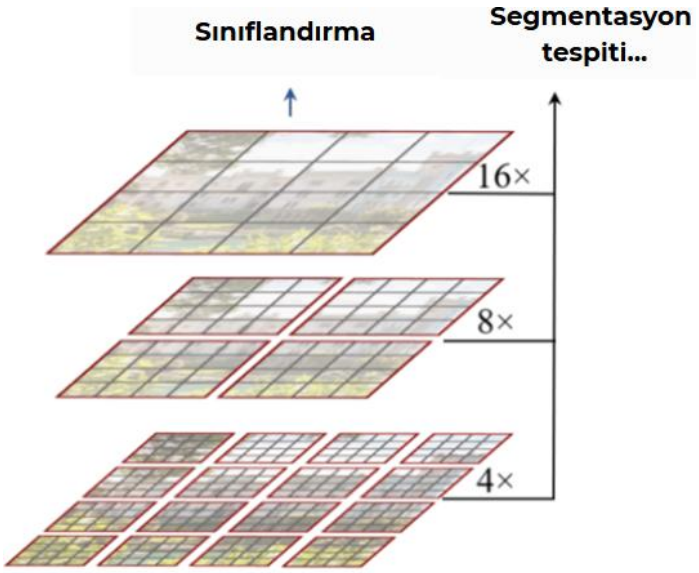
Swin Dönüştürücü (SD) (Shifted Window Transformer), görüntüler için kullanılan güncel bir transformer mimarisidir. Swin Transformer, görüntüyü yamalar halinde alır ve bu yamalar üzerinde işlemler yaparak bir hiyerarşik temsil oluşturur (Şekil 3.25). Görüntülerde oluşturulan yamalar üzerindeki özellikler, lineer bir katmanla projekte edilerek temsil edilir. Daha derin katmanlarda yama birleştirme (patch merging) mekanizması kullanılarak çözünürlük düşürülür ve kanal sayısı artırılır. Swin Transformer, dikkat mekanizmasını yerel pencerelerde uygular (Window-based Multi-Head Self-Attention, W-MSA). Ancak, pencere tabanlı dikkat mekanizması, pencereler arasında bağlantıyı kaybedebilir. Bu problemi çözmek için Shifted Window Attention (SW-MSA) yöntemi kullanılır. Ardışık katmanlarda, pencereler kaydırılarak komşu pencereler arasında bağlantı sağlanır (Liu ve ark., 2021a). ViT'ler düzlemsel özellik çıkarımı yaparken, Swin Transformer hiyerarşik dönüşüm gerçekleştirir.

Swin-B (Swin-Base) ve Swin-S (Swin-Small), Vision Transformer (ViT) mimarisine dayalı, görsel görevlerde güçlü performans sergileyen Swin Transformer mimarisinin iki farklı versiyonudur. Her ikisi de, görüntüleri sabit boyutlu pencerelere bölerek "Kayan pencereler" (Shifted Window) mekanizmasını kullanır (Şekil 3.26.). Bu yapı, hem hesaplama verimliliği sağlar hem de farklı ölçeklerde bilgi yakalamayı mümkün kılar. Kayan pencereler mekanizmasında, görüntü sabit boyutlu pencerelere bölünerek "self-attention" yalnızca bu pencereler içinde uygulanır, böylece hesaplama maliyeti düşürülür. Ancak sadece sabit pencereler kullanmak, pencere sınırlarındaki piksellerin birbirini görememesine neden olur. Bunu aşmak için her katmanda pencere konumu bir miktar kaydırılır (örneğin yarım pencere kadar), böylece bir önceki

katmanda aynı pencerede olmayan pikseller bir sonraki katmanda aynı pencere içine girer. Bu sayede farklı bölgeler arasında bilgi akışı sağlanır ve daha güçlü küresel bağlam öğrenimi gerçekleşir. Swin-S modeli, daha küçük bir yapıya sahip olup parametre sayısı ve hesaplama maliyeti açısından daha hafiftir; bu nedenle sınırlı kaynaklara sahip sistemlerde tercih edilir. Swin-B ise daha büyük ve derin bir model olup daha fazla parametreye sahiptir, bu da genellikle daha yüksek doğruluk sağlar ancak daha fazla hesaplama gücü gerektirir. Özetle, her iki model de benzer temel yapıya sahipken, Swin-B daha büyük ölçekli uygulamalar için optimize edilmiştir; Swin-S ise daha hızlı ve hafif çözümler sunar.



Şekil 3.25. Swin Dönüştürücü mimarisi

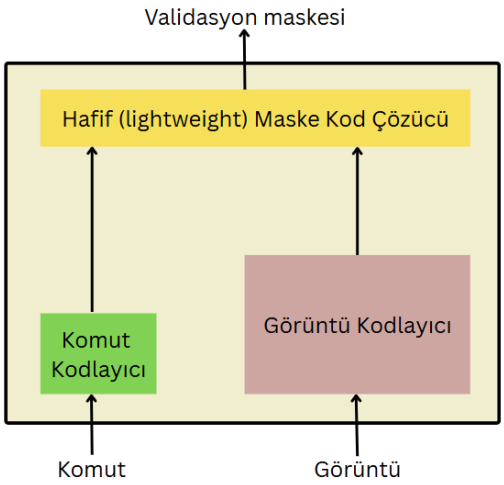


Şekil 3.26. Kayan pencereler mekanizması (Liu ve ark., 2021a).

3.8. SAM2-UNet Mimarisi

SAM (Segment Anything Model), farklı görüntü segmentasyon görevlerine genel bir çözüm sunmayı amaçlayan bir segmentasyon modelidir. SAM'in temelinde ViT yapısı yer alır. Bu yapı, görüntüleri yama seviyesinde işleyerek uzun menzilli bağımlılıkları öğrenmede etkilidir. ViT ve Swin Transformer, SAM gibi ileri seviye segmentasyon model transformer yapısını kullanırlar ve dikkat mekanizmalarına dayanırlar.

SAM; görüntü kodlayıcı, komut kodlayıcı ve maske çözücü olmak üzere 3 ana bileşenden oluşur (Şekil 3.27) (Kirillov ve ark., 2023).



Şekil 3.27. SAM bileşenleri

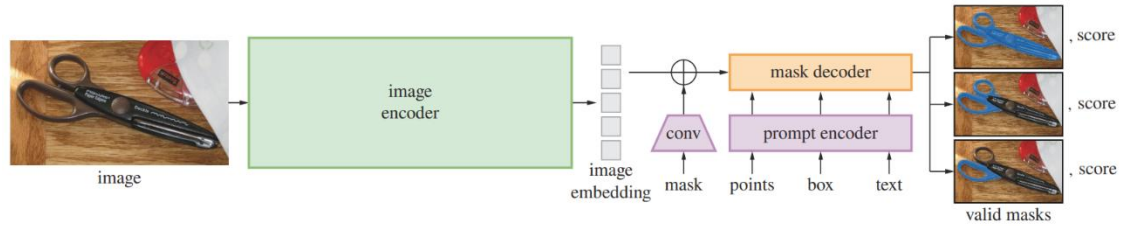
Görüntü Kodlayıcı (Image Encoder): Transformer tabanlı bir yapı kullanır. SAM'in görüntüleri yüksek çözünürlükte işlemeyi destekleyen Görüntü Dönüştürücü (ViT) mimarisi üzerine inşa edilmiştir.

Komut (Prompt) Kodlayıcı: Noktalar, kutular veya serbest metin gibi bilgiler seyrek ipuçları olarak kodlanır. Mevcut maskeler gibi yoğun ipuçları işlenir ve görüntü kodlamasıyla birleştirilir.

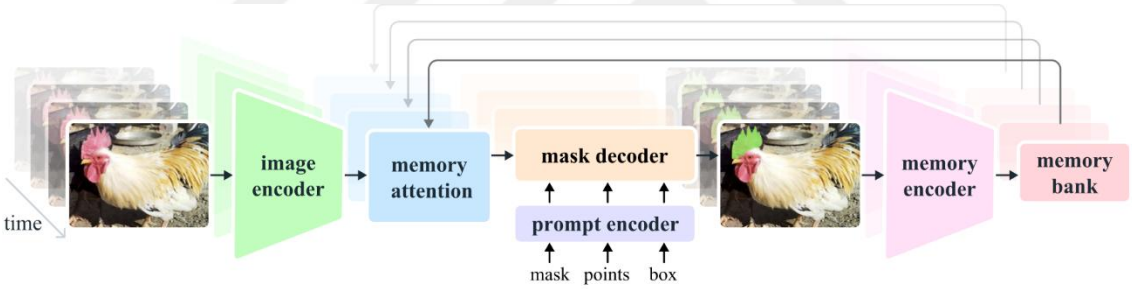
Maske Çözücü (Mask Decoder): Görüntü kodlaması ve komut kodlamasını birleştirerek segmentasyon maskesi oluşturur. Düşük gecikme ile çalışır ve belirsiz durumlarda birden fazla maske üretir (Kirillov ve ark., 2023).

Segment Anything Model-2 (SAM2), görüntü segmentasyonu alanında kullanılan bir modeldir ve Segment Anything Model (SAM)'in gelişmiş bir

versiyonudur (Şekil 3.28). SAM2 görüntü ve video segmentasyonu için optimize edilmiştir. SAM2, SAM'in güçlü özelliklerini temel alırken, özellikle video segmentasyonu gibi zaman boyutunu içeren görevlerde daha geniş bir uygulama alanı sunar. İstenilen görsel segmentasyon (Promptable Visual Segmentation - PVS) görevini destekler. Bir segmentasyon uygulamasında, bir video çerçevesine tıklama, kutu çizimi veya maske ipuçları gibi bir giriş verilerek nesne segmentasyonu başlatılır. Model, segmentasyonu zaman içinde (tüm video boyunca) takip eder ve gerektiğinde güncellenebilir.



a) SAM mimarisi (Kirillov ve ark., 2023)



b) SAM2 mimarisi (Xiong ve ark., 2024)

Şekil 3.28. SAM ve SAM2 mimarilerinin gösterimi

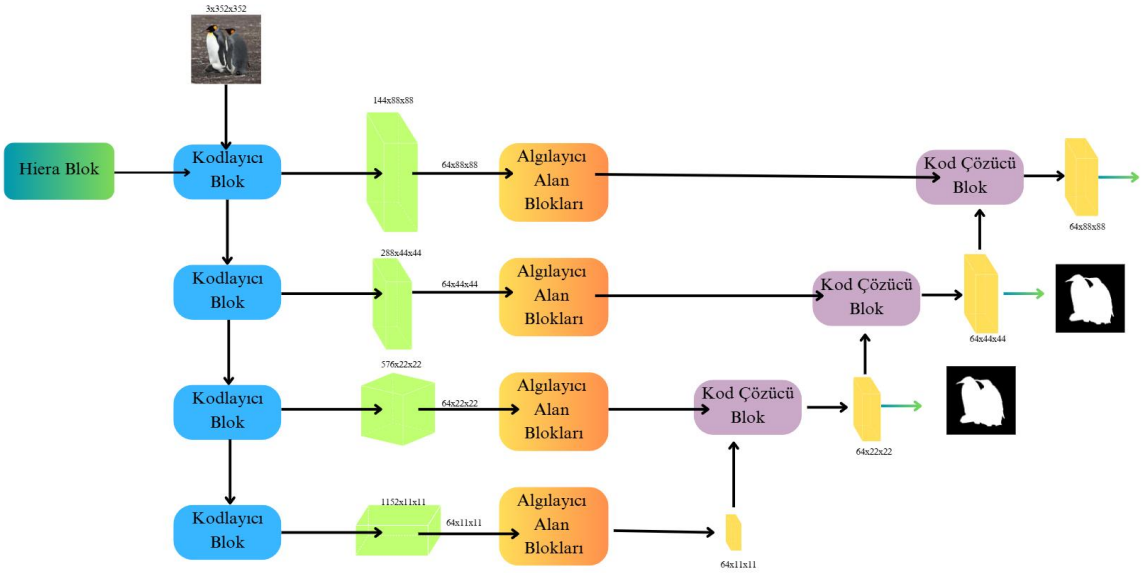
SAM2, "Hiera" adında hiyerarşik bir görüntü kodlayıcı kullanır. Hiera, birden fazla seviyede bilgi işlemeyi destekleyen hiyerarşik bir yapıya sahiptir (Xiong ve ark., 2024). Bu yapı, düşük seviyeli özelliklerden yüksek seviyeli özelliklere kadar kapsamlı bir bilgi temsili sağlar. Her seviyede, görüntü özellikleri farklı çözünürlüklerde işlenir. Bu, hem yerel detayların hem de global bağlamın yakalanmasını sağlar. Hiera blokları, görüntüyü küçük yamalara böler ve her yamayı bir diziye dönüştürür. Hiera, çoklu başlıklı dikkat mekanizmasını kullanır. Bu mekanizma, farklı başlıkların görüntüdeki farklı bölgelere dikkat etmesini sağlar. Bu, modelin görüntüdeki önemli detayları ve

ilişkileri öğrenmesini kolaylaştırır. Bu yamalar, ardından transformer katmanlarına beslenir ve özellik çıkarımı için kullanılır (Ryali vd., 2024).

Hiera, çok ölçekli özellik çıkarımı sağlar, bu da özellikle yoğun segmentasyon görevlerinde etkilidir. SAM2, SAM modelindeki ViT tabanlı kodlayıcıdan daha hızlı ve daha az karmaşıktır. SAM2'nin maske çözücü yapısı SAM ile benzerdir ancak bellek ve yeni komut türlerini işlemek için iyileştirilmiştir. Ayrıca SAM modeline ek olarak bir görünmezlik tahmin başlığı (occlusion prediction head) eklenmiştir. Bu, nesnenin belirli bir karede görünür olup olmadığını tahmin eder (Ravi ve ark., 2024).

SAM2-UNet, SAM2 üzerine inşa edilmiş bir görüntü segmentasyonu mimarisidir. SAM2 – UNet mimarisi, kodlayıcı, kod çözücü, adaptörler ve alıcı alan bloklarından oluşur (Şekil 3.29).

SAM2.1, SAM2'nin ve SAM2.1-UNet de SAM2-UNet'in gelişmiş bir versiyonudur ve özellikle bellek yönetimi, zamansal dikkat ve maske tahminlerinin iyileştirilmesi konusunda çeşitli yenilikler sunmaktadır.



Şekil 3.29. SAM2-UNET mimarisi

Kodlayıcı (Encoder): SAM (Kirillov ve ark., 2023) mimarisindeki gibi Hiera omurga ağı kullanır ve çok ölçekli özelliklerin çıkarılmasını sağlayan hiyerarşik bir yapıya sahiptir (Ryali vd., 2024). Bu, ViT tabanlı düz yapılardan daha gelişmiştir ve segmentasyon görevlerinde daha etkilidir.

Alıcı Alan Blokları (Receptive Field Blocks - RFBs): Kodlayıcıdan elde edilen özellik haritalarını işleyerek kanalların boyutunu azaltır ve daha hafif özellikler oluşturur. Bu bloklar, hızlı ve etkili bir özellik çıkarımı sağlar.

Adaptörler: Parametre verimliliğini sağlar. Hiera'nın büyük parametre boyutu (örneğin, 214 milyon parametre) tam ince ayarlama (fine-tuning) yapmayı zorlaştırabilir. Bu nedenle, adaptörler kullanılarak Hiera parametreleri dondurulur ve yalnızca adaptörler üzerinden ince ayar yapılır.

Çözücü (Decoder): UNet'in klasik U-şekilli yapısını takip eder. Çıktılar, hedef segmentasyon maskesiyle karşılaştırılarak eğitilir.

3.9. Kayıp Fonksiyonu

Çapraz Entropi (ÇE) (Cross-Entropy) (Denklem 3.1) kaybı, her pikselin doğru sınıfa (örneğin, arka plan veya nesne) ne kadar doğru sınıflandığını ölçer. Çapraz entropi kaybına ek olarak, kenar bölgelerindeki detayların daha iyi öğrenilmesini sağlamak amacıyla Kenar Koruma Kaybı (KKK) (Edge Preservation Loss-EPL) uygulanır. Bu kayıp fonksiyonu, Laplace operatörü yardımıyla sınır bölgelerini belirleyerek, modelin bu bölgeleri daha doğru tahmin etmesini teşvik eder. Böylece, belirgin nesnelerin sınırları daha keskin hale gelir ve modelin genel performansı artar (Denklem 3.1).

$$L_S = -\sum_{i=0}^{size(Y)} (\alpha_s Y_i \, log(P_i) + (1 - \alpha_s)(1 - Y_i) \, log(1 - P_i)) \tag{3.1}$$

Toplam kayıp fonksiyonu, genel belirginlik haritasını oluşturmak için çapraz entropi kaybı (Denklem 3.1) ve sınırları vurgulamak için Laplace operatörü (Denklem 3.2) ile elde edilen sınırların çapraz entropi kaybının ağırlıklı toplamını içerir. Bu tez çalışmasında evrişimli sinir ağı deneylerinde ÇE ve KKK kullanılmıştır.

$$\Delta f = \frac{\{\partial^2 f\}}{\{\partial x^2\}} + \frac{\{\partial^2 f\}}{\{\partial y^2\}} \tag{3.2}$$

Birleşim Üzerine Kesişim (BüK) (Intersection Over Union-IoU), nesne algılama algoritmalarının performansını ve doğruluğunu değerlendirmek için bilgisayarlı görüşte kullanılan önemli bir ölçümdür. Jaccard İndeksi (Jaccard Index) olarak da bilinmektedir. Bir görüntüde sınırları tespit edilen bir nesnenin, gerçek nesne sınırları ile ne kadar iyi hizalandığını ölçer. Daha yüksek bir BüK puanı daha doğru bir tespit anlamına gelir.

Ağırlıklı BüK kaybı (IoU Loss), segmentasyon maskesindeki nesne ve arka plan bölgelerinin örtüşme oranını ölçer ve küçük nesneleri daha fazla önemseyerek modelin duyarlılığını artırır. İkili çapraz entropi kaybı (BCE Loss), her pikselin doğru sınıfa ait olup olmadığını değerlendirir. Bu iki kayıp fonksiyonunu birleştiren kayıp fonksiyonu kullanılabilir (Denklem 3.3). Bu iki kaybın bir arada kullanılması, piksel düzeyinde segmentasyon kalitesini artırır ve tıbbi görüntü analizi, uydu görüntülerinin işlenmesi ve otonom sistemlerdeki nesne tanıma gibi hassas segmentasyon gerektiren alanlarda tercih edilir. Transformer kullanılarak yapılan çalışmalarda görsel ögeler arasındaki uzun menzilli ilişkileri daha iyi modellemek ve segmentasyon doğruluğunu artırmak amacıyla kullanılmıştır.

$$L = L_{IoU}^{w} + L_{BCE}^{w} \tag{3.3}$$

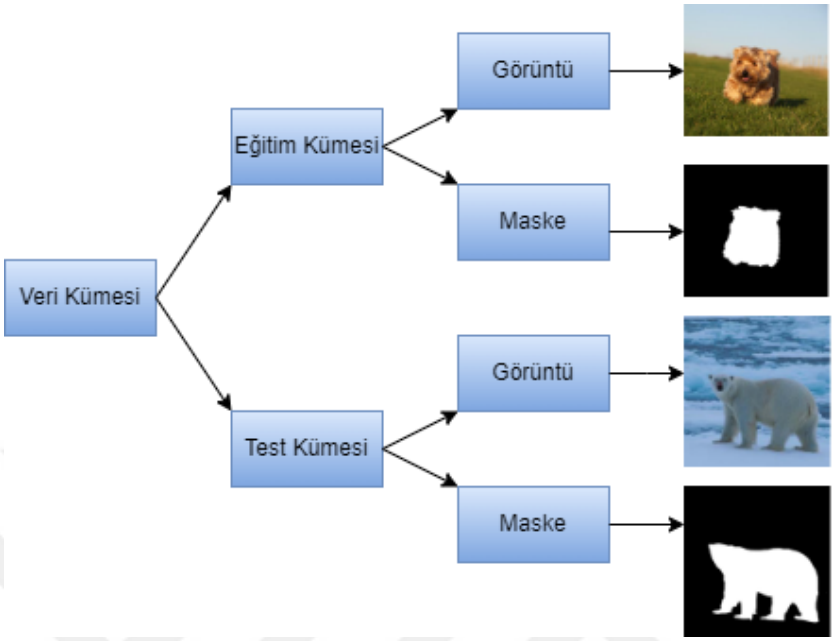
Bu tez çalışmasında transformer modellerinde segmentasyon çıktıları için derin denetim (deep supervision) uygulanmıştır. Bu, modelin ara katmanlarından alınan tüm çıktıları da eğitmeyi ve modelin üç farklı segmentasyon çıktısının her biri ile gerçek segmentasyon maskesi arasındaki kaybı hesaplar ve toplar (Denklem 3.4). Bu, modelin her seviyede daha iyi öğrenmesini sağlar (Xiong ve ark., 2024).

$$L_{toplam} = \sum_{i=1}^{3} L(G, S_i) \tag{3.4}$$

3.10. Veri Kümeleri

DUTS veri kümesi (Wang ve ark., 2017), 10553 eğitim ve 5019 test görüntüsü içeren bir veri kümesidir (Şekil 3.30) (Çizelge 3.1). Doğru piksel düzeyinde zemin gerçekleri, 50 kişi tarafından manuel olarak etiketlenmiştir. Veri kümesi yapay zekâ

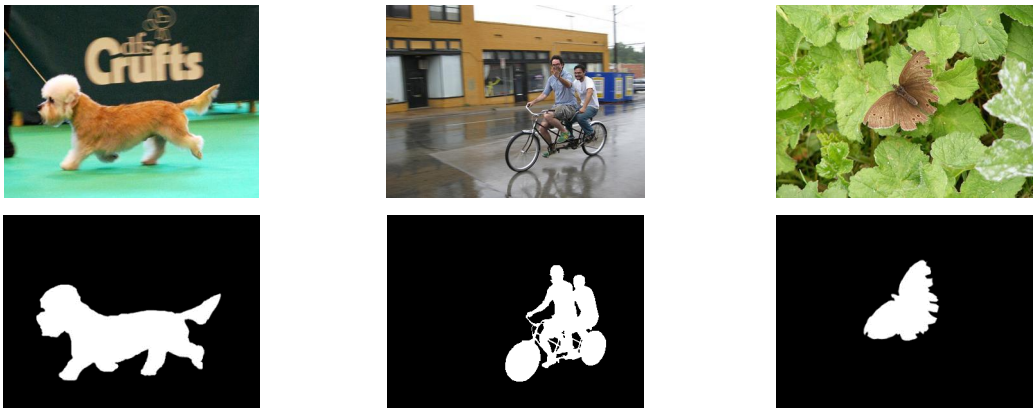
modellerinin eğitilmesi için tasarlanmıştır. Görüntü verileri RGB'dir ve ".jpg" uzantılıdır. Maske (ground truth) verileri ikili (binary) görüntülerdir ve ".png" formatındadır (Şekil 3.31).



Şekil 3.30. DUTS veri kümesi

ECSSD (Extended Complex Scene Saliency) veri kümesi, karmaşık sahnede öne çıkan görüntülerin segmentasyonu konusunda araştırmaları geliştirmek için oluşturulmuştur (Shi ve ark., 2016). Bu veri kümesinde, doğal ve birbirinden farklı içerikte görüntüler vardır (Şekil 3.32). Tüm görüntüler internet üzerindeki görüntülerden toplanmış ve 5 yardımcı kişinin katkısıyla etiketlenerek maskeleri üretilmiştir.

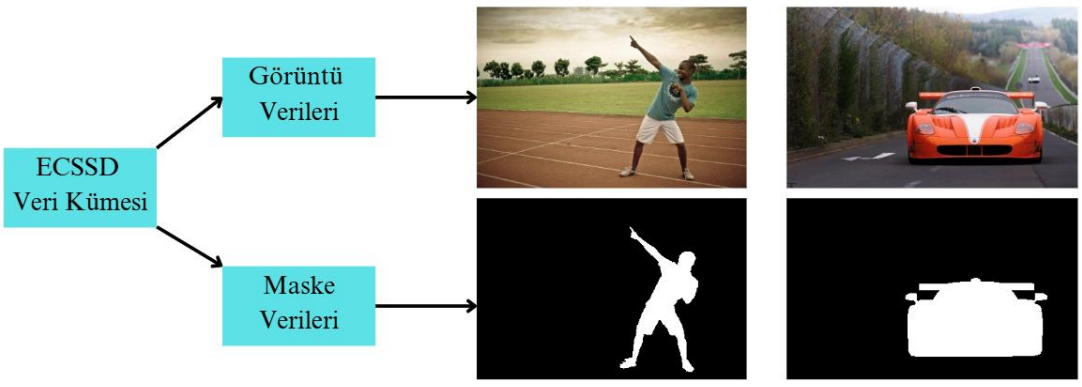
DUT-OMRON veri kümesi, belirgin nesne tespiti konusunda kullanılır ve karmaşık sahne görüntülerinden oluşur (Şekil 3.33). Görseller, öne çıkan nesnelerin belirginliğinin düşük olduğu, karmaşık arka planlar ve zengin içerikli sahneler içermesiyle dikkat çeker (Yang ve ark., 2013). Veri kümelerindeki eğitim ve test görüntülerine ait bilgiler Çizelge 3.1'de gösterilmiştir.



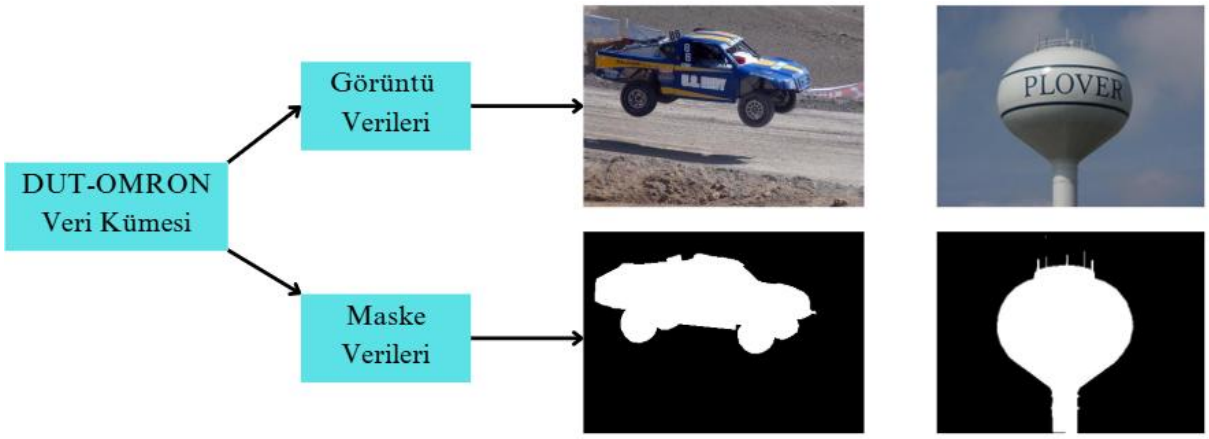
a) Eğitim veri kümesinden örnek görüntüler ve maske görüntüleri

b) Test veri kümesinden örnek görüntüler ve maske görüntüleri

Şekil 3.31 Eğitim ve test veri kümesinden örnek görüntüler



Şekil 3.32. ECSSD veri kümesinden örnek görüntüler



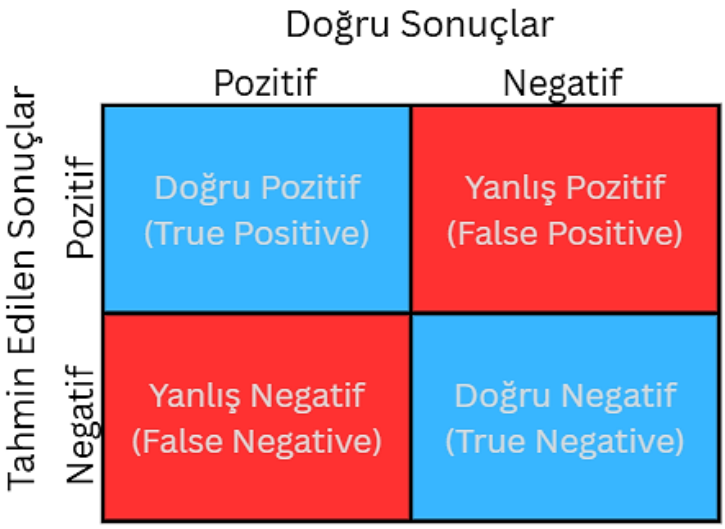
Şekil 3.33. DUT-OMRON veri kümesinden örnek görüntüler

Çizelge 3.1. DUTS, DUT-OMRON ve ECSSD veri kümeleri

Veri Kümesi	Eğitim Kümesi	Test Kümesi
DUT-OMRON (Yang ve ark., 2013).	X	5168
ECSSD (Shi ve ark., 2016).	X	1000
DUTS (Wang ve ark., 2017),	10553	5019
x: Değer bulunmuyor.		

3.11. Karşılaştırmada Kullanılan Metrikler

Konfüzyon matrisi (Confusion Matrix), yapay zekâ modelinin tahminleri ile gerçek sınıflar arasındaki ilişkiyi tablo şeklinde gösteren bir performans ölçüm aracıdır. Bu matris, modelin hangi çıktılarda doğru tahmin yaptığını veya hata yaptığını ayrıntılı olarak görmeyi sağlar. Genellikle dört temel bileşenden oluşur: Doğru Pozitif (True Positive-TP), Yanlış Pozitif (False Positive-FP), Doğru Negatif (True Negative-TN) ve Yanlış Negatif (False Negative-FN). Bu bileşenler sayesinde modelin başarımı yalnızca genel doğrulukla değil, aynı zamanda duyarlılık (recall), özgüllük (specificity), kesinlik (precision) ve F1 skoru gibi daha kapsamlı metriklerle de analiz edilebilir. Özellikle sınıf dengesizliği bulunan veri kümelerinde, konfüzyon matrisi modelin hangi sınıflarda daha başarılı veya başarısız olduğunu anlamak açısından kritik öneme sahiptir (Şekil 3.34).



Şekil 3.34. Konfüzyon matrisi

Duyarlılık (Recall), modelin gerçek pozitifleri ne kadar iyi tespit ettiğini gösterir (Denklem 3.5). Yani, modelin pozitif sınıfları doğru bir şekilde tanıyabilme kabiliyetidir. Yanlış negatifler (gerçekte pozitif olan ancak model tarafından negatif olarak sınıflandırılan) üzerinden hesaplanır.

$$Duyarlılık = \frac{Doğru\ Pozitif}{Doğru\ Pozitif + Yanlış\ Negatif} \tag{3.5}$$

Kesinlik (Precision), modelin pozitif olarak sınıflandırdığı örneklerin ne kadarının gerçekten pozitif olduğunu gösterir (Denklem 3.6). Yanlış pozitifler (gerçekte negatif olan ancak model tarafından pozitif olarak sınıflandırılan) üzerinden hesaplanır.

$$Kesinlik = \frac{Doğru\ Pozitif}{Doğru\ Pozitif + Yanlış\ Pozitif} \tag{3.6}$$

OMH (Ortalama Mutlak Hata), sürekli değerlerin tahmin edildiği regresyon problemlerinde değerlendirme metriği olarak kullanılır (Nečasová vd., 2022) (Denklem 3.7). Gerçek değerler ile tahmin edilen değerler arasındaki hataların mutlak değerlerinin ortalamasını hesaplamaktadır.

$$OMH = \frac{1}{N}\sum_{i=1}^{N}|P_i - G_i| \tag{3.7}$$

F1 Skor, sınıflandırma için popüler bir performans ölçüsüdür ve örneğin, verilerin dengesiz olduğu, örneğin bir sınıfa ait örneklerin miktarının diğer sınıfta bulunanlardan önemli ölçüde fazla olduğu durumlarda doğruluktan daha çok tercih edilir. F1 Skor Kesinlik (Precision) ve Duyarlılık (Recall) skorlarının harmonik ortalamasıdır (Denklem 3.8).

$$F1\ Skor = \frac{2\ x\ Doğruluk\ x\ Kesinlik}{Doğruluk + Kesinlik} \tag{3.8}$$

F_β skoru, β parametresi ile ayarlanabilir bir F1 Skor değeridir (Denklem 3.9). F1 skoru, β = 1 olduğunda elde edilir ve hem kesinlik (precision) hem de duyarlılık (recall) eşit olarak dikkate alır. β'nin değeri arttıkça, kesinliğe daha fazla ağırlık verilir; β'nin değeri azaldıkça ise duyarlılığa daha fazla ağırlık verilir (Xiong vd., 2024).

$$F_\beta = \frac{(1\ +\ \beta^2)\ x\ Doğruluk\ x\ Kesinlik}{\beta^2\ x\ Doğruluk + Kesinlik} \tag{3.9}$$

Sα (Yapısal Benzerlik Ölçütü), görüntü segmentasyonunda tahmin edilen segmentlerin yapısal benzerliğini ölçmek için kullanılır ve, segmentasyon sonuçlarının nesnenin yapısal bilgilerini ne kadar iyi koruduğunu değerlendiren bir metriktir (Xiong vd., 2024).

Eϕ (Geliştirilmiş Hizalama Ölçütü), segmentasyon sonuçlarının görsel algılanabilirliğini ve tahmin edilen maskelerin gerçek maskelere hizalanma derecesini ölçer ve hem genel doğruluğu hem de yerel ayrıntıları değerlendirir (Xiong vd., 2024) (Denklem 3.10).

$$E_\phi = \frac{1}{N}\sum_{i=1}^{N}\frac{2xP_\phi(i)xG_\phi(i)}{P_\phi(i) + G_\phi(i)} \tag{3.10}$$

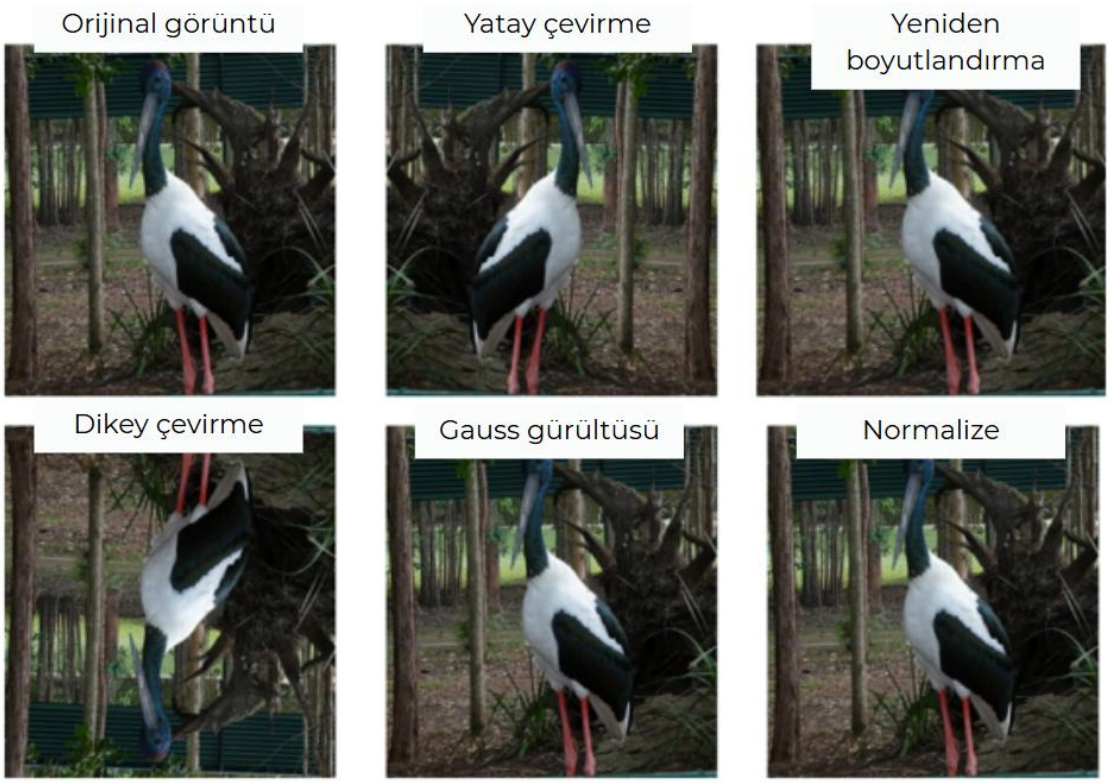
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Yapay zekâ günümüzde tıp, endüstri ve eğitim gibi birçok alanda kullanılmaktadır. İnsan gibi düşünebilen makineler sayesinde, hayatın içerisinde kalite-kontrol işlemleri gibi alanlarda insanlardan daha hızlı, daha başarılı sonuçlar elde edilebilmektedirler. Yapay zekâ tabanlı projelerde görüntü işleme teknikleri sıklıkla kullanılmaktadır. Nesneleri sınıflandırma, bölütleme, tespit etme, izleme gibi alt başlıklarda doğru sonuçlar elde etmek için projeye özel geliştirilecek görüntü işleme yöntemleri oldukça önemlidir. Görüntü verilerinde Derin Öğrenme (DÖ) ile belirginlik tespiti konusunda güncel literatürde Evrişimsel Sinir Ağları (ESA) ve transformer tabanlı yöntemler kullanılmaktadır.

Bu tez çalışmasında "Derin öğrenme ile 2 Boyutlu Görüntülerde Belirginlik Tespiti" çalışması yapılmıştır. Belirgin Nesne Tespiti (BNT) gerçekleştirilirken segmentasyon yöntemi kullanılmıştır. Tez kapsamında bu amaçla ESA ve transformer mimarileri ile deneyler gerçekleştirilmiştir.

BNT, karmaşık arka planlarda, iç içe geçmiş nesnelerde, farklı kontrast ve parlaklık koşullarında zorlu bir problemdir. Tez kapsamında çalışmalar gerçekleştirilirken, literatürdeki BNT konusundaki veri kümeleri ile çalışılmıştır.

Tez kapsamında BNT, Python programlama dili kullanılarak gerçekleştirilmiştir. Kullanılan görüntülere, literatürde kullanılan veri arttırma yöntemlerinden yatay ve dikey çevirme, yeniden boyutlandırma, gürültü ekleme ve normalizasyon yöntemleri uygulanmıştır (Şekil 4.1). Sonuçların değerlendirilmesinde, literatürde kullanılan değerlendirme metrikleri araştırılmıştır ve eğitilen yapay zekâ modellerinin test veri kümelerinde S_α, E_ϕ ve OMH değerleri karşılaştırılmıştır. Sonuçlar, SAM2.1-UNet mimarisinin önceden eğitilmiş (pre-trained: sam2.1 hiera large) modeli ile elde edilmiştir.



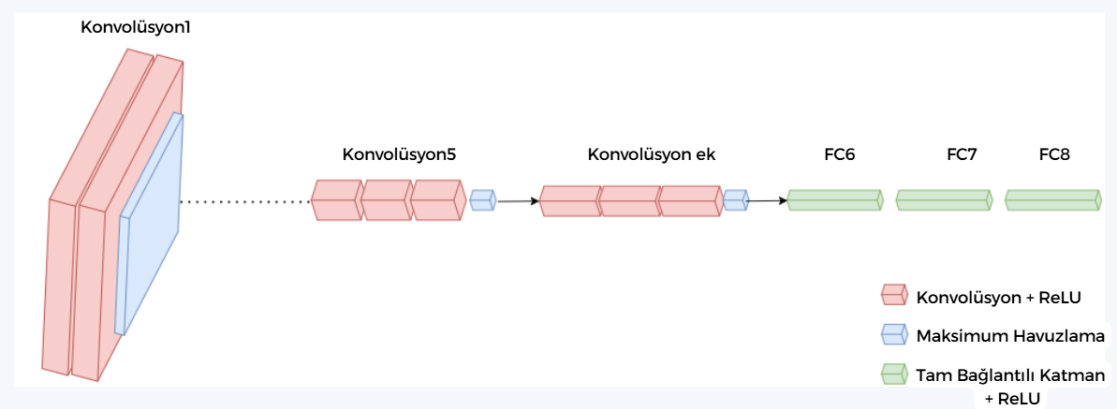
Şekil 4.1. Tez kapsamında kullanılan veri arttırma yöntemlerinden örnekler

4.1. ESA Tabanlı Yöntemler

Yapılan tez çalışmasında, DÖ ile görüntü verilerinde nesnelerin belirginlik tespiti gerçekleştirilmiştir. Belirginlik tespiti yapılırken tam bağlı ağlar (Fully Connected Network) güncel yöntemlerde yaygın olarak kullanılmaktadır. Evrişimli Katman, evrişimli sinir ağlarındaki ilk katmandır. Nesnelerin genişliklerinin ve yüksekliklerinin eşit kalmasını sağlamak için havuzlama katmanları kullanılmıştır. Nöronların toplam fonksiyonunda üretilen çıktıların geçmesi gereken değişimleri belirlemek amacıyla aktivasyon fonksiyonları kullanılmıştır.

Tez kapsamında ilk önce görüntülerdeki belirgin nesnelerin tespiti için VGG-16 modeli temel alınarak daha fazla semantik özellik dâhil etmek için ekstra bir konvolüsyon katmanı eklenmiş ve katman sayısındaki değişikliğin modelin doğruluğuna etkisi araştırılmıştır (Şekil 4.2). Bu eklenen katman, önceki konvolüsyon katmanlarından elde edilen yüksek seviyeli özellikleri geliştirmek ve özellikle ölçek, şekil ve konum değişkenliklerine karşı modeli daha hassas hale getirmek için tasarlanmıştır. Bu amaçla modelde Pyramid Feature Attention (PFA) kullanılmıştır. Geliştirilen nesne belirginlik tespit yönteminde, “bağlam farkındalıklı piramit özellik

çıkarma yöntemi" (context-aware pyramid feature extraction), "kanal bazlı dikkat modülü" (channel-wise attention module) ve düşük seviye özellik haritalarını iyileştirmek için "uzamsal dikkat modülü" (spatial attention module) kullanılmaktadır.

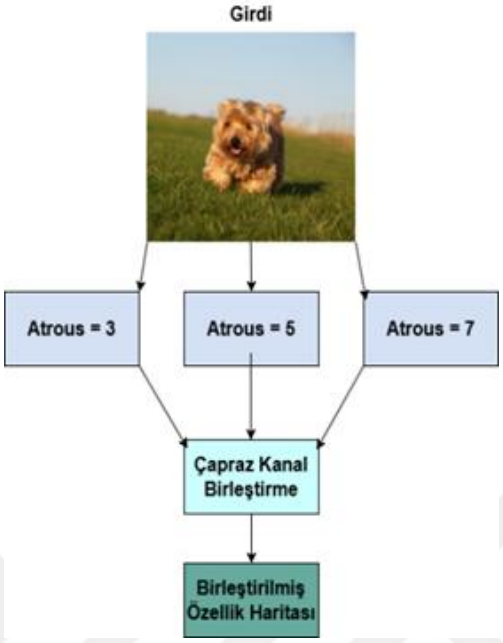


Şekil 4.2. Tez kapsamında oluşturulan VGG16 modeli

VGG16'daki konvolüsyon katmanlarından 3-3 (üçüncü bloktaki üçüncü evrişimli katman), 4-3 (dördüncü bloktaki üçüncü evrişimli katman) ve 5-3 (beşinci ve son bloktaki üçüncü evrişimli katman) temel yüksek seviyeli özellikler olarak kullanılmıştır. Bu katmanlar, daha derin seviyelerde yer aldığından, ağın daha soyut ve anlamlı özellikler çıkarmasına yardımcı olur. Son çıkarılan yüksek seviyeli özelliklerin ölçek (scale) ve şekil (shape) varyantlarını içermesini sağlamak için, farklı genişleme oranlarına sahip 3, 5 ve 7 değerlerinde atrous evrişimler kullanılmıştır (Şekil 4.3). Atrous evrişimde, geleneksel evrişimdeki gibi bir filtre (kernel) kullanılır, ancak bu filtredeki pikseller arasındaki mesafe artırılarak daha geniş bir algılama alanı elde edilir. Bu sayede, farklı ölçeklerde bağlamsal bilgi yakalanarak, özellik haritasındaki daha geniş bir bilgi yelpazesi elde edilir. Farklı genişleme oranlarına sahip evrişimlerden elde edilen özellik haritaları, çapraz kanal birleştirme yöntemiyle (cross-channel concatenation) birleştirilerek daha güçlü bir özellik gösterimi oluşturulmuştur.

Farklı atrous evrişim katmanlarından gelen özellik haritaları ve 1×1 boyut indirgeme ile elde edilen özellikler, kanallar arası birleşme (cross-channel concatenation) yöntemiyle birleştirilmiştir. Bu işlem, farklı genişleme oranlarına sahip evrişimlerin (3, 5 ve 7) yakaladığı çok ölçekli bağlamsal bilgileri ortak bir temsil halinde birleştirerek daha zengin bir özellik haritası oluşturmayı amaçlar. Bu çalışmada VGG-16 modeline ek olarak, ResNet50 ve DenseNet169 modelleri de ayrı omurga

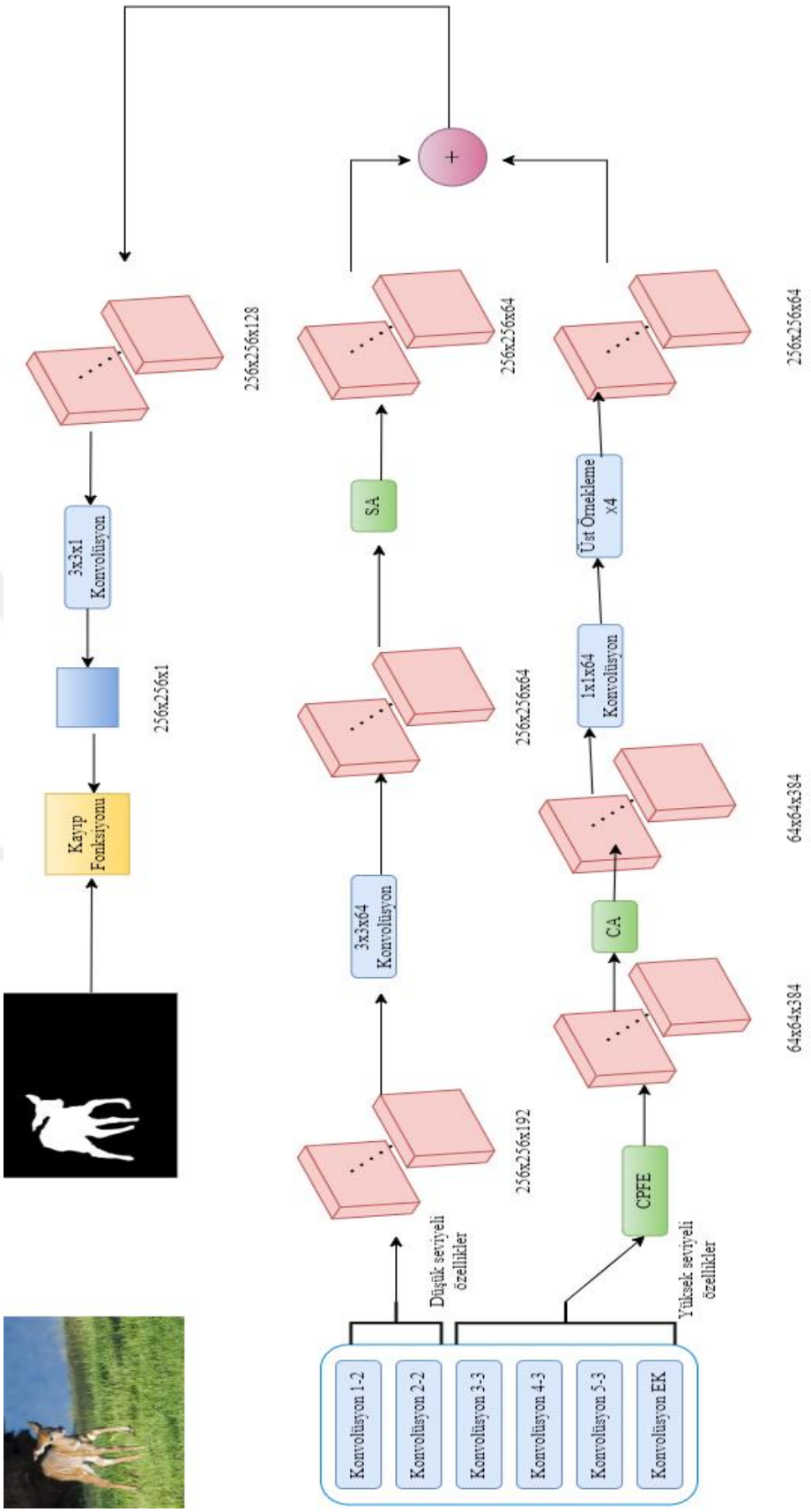
ağları (backbone) olarak kullanılmış ve her biri için aynı kanallar arası birleşme işlemi uygulanmıştır. Ancak, bu modeller birlikte değil, bağımsız olarak değerlendirilmiştir.



Şekil 4.3. Farklı genişleme oranlarında atrous evrişimler

Deneysel çalışmada kullanılan yöntemin şeması Şekil 4.4'te gösterilmiştir.

Bu tez çalışmasında Kanal Tabanlı Dikkat (KTD) mekanizması ve Mekânsal Dikkat (MD) mekanizması kullanılmıştır. Çalışmalarda model eğitim aşamasında DUTS (Wang ve ark., 2017) veri kümesi kullanılmıştır. Görüntüler alınırken veri yükleyicide parlaklık, rotasyon gibi çeşitli veri ön işleme yöntemleri kullanılmıştır. Veriler normalize edilmiştir. Normalizasyon, verilerin belirli bir aralığa (genellikle [0,1] veya [-1,1]) ölçeklenmesi işlemidir. Bu işlem, modelin daha hızlı ve dengeli öğrenmesini sağlar. Görüntü verileri normalize edilirken, her piksel değeri 255'e bölünerek [0,1] aralığına indirgenmiştir. Böylece, tüm görüntüler benzer ölçeklerde işlenmiş ve modelin eğitimi sırasında kararlılık sağlanmıştır.



Şekil 4.4. Tez kapsamında geliştirilen yöntemin şeması

Tez kapsamında geliştirilen modelin doğruluğunu artırmak için geleneksel çapraz-entropi (Denklem 3.1) kaybına ek olarak, kenar bölgelerindeki detayların daha iyi öğrenilmesini sağlamak amacıyla Kenar Koruma Kaybı (KKK) (edge preservation loss) (Denklem 3.2) uygulanmıştır. Bu fonksiyon, Laplace operatörü yardımıyla sınır bölgelerini belirleyerek, modelin bu bölgeleri daha doğru tahmin etmesini sağlamaktadır. Böylece, belirgin nesnelerin sınırları daha keskin hale gelmekte ve modelin genel performansı artmaktadır. Bu bileşenlerin birlikte kullanılması, modelin hem yüksek seviyeli bağlamsal bilgileri hem de düşük seviyeli mekânsal detayları etkili bir şekilde işleyerek belirgin nesne tespiti görevlerinde yüksek başarı elde etmesini sağlamıştır.

Bağlam farkındalıklı özellik çıkarım modülü, ağın ara katmanlarından alınan bir özellik haritasını giriş olarak kullanmıştır. Bu modül, farklı genişleme oranlarına sahip üç adet 3×3 konvolüsyon katmanı ile bir adet 1×1 konvolüsyon katmanından oluşmaktadır. Her bir konvolüsyon katmanında 32 çıkış kanalı bulunur ve bu yapı, farklı ölçeklerde bağlamsal bilgi yakalamayı amaçlar (Şekil 4.3).

Geliştirilen ESA tabanlı yöntem ile DUTS veri kümesinde yapılan segmentasyon çalışmalarında test verilerinde elde edilen sonuçlar Çizelge 4.1'de verilmiştir. Bu aşamada ESA yöntemleri kullanılarak Belirgin Nesne Tespiti (BNT gerçekleştirilmiştir. ESA mimarilerinde farklı omurga ağları kullanılmış ve temel derin öğrenme omurga ağlarına dikkat mekanizması eklenmesinin sonuca etkileri gözlemlenmiştir Modellere ait sonuçlar Şekil 4.5'te gösterilmiştir.

Çizelge 4.1. DUTS veri kümesinde ESA modelleri ile elde edilen sonuçlar

Veri Kümesi	**OMH**	**F1 Skor**
VGG16 [19]	0.0646	0.8136
ResNet50 [20]	0.0539	0.8355
DenseNet169 [21]	0.0563	**0.8405**
ResNet50 (dikkat mekanizması ile)	0.0495	0.8192
VGG16 (dikkat mekanizması ile)	**0.0473**	0.8164
DenseNet169 (dikkat mekanizması ile)	0.0579	0.7754


(a) orijinal görüntü


(c-1)


(d-1)


(e-1)


(b) gerçek maske

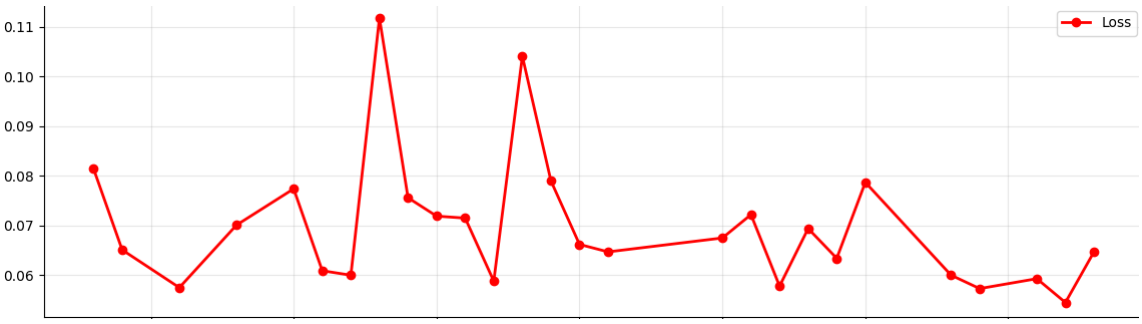

(c-2)


(d-2)

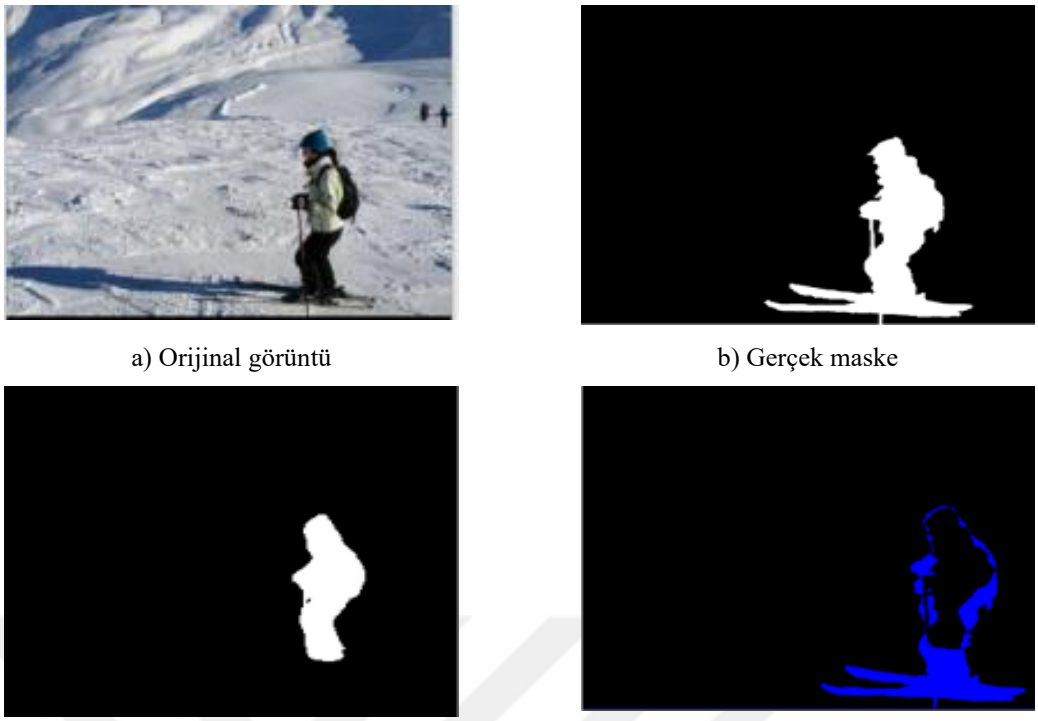

(e-2)

Şekil 4.5. Farklı derin öğrenme modellerine ait sonuçlar
(c-1) VGG16 sonucu (c-2) Dikkat mekanizması eklenen VGG-16 sonucu,
(d-1) DenseNet-169 sonucu (d-2) Dikkat mekanizması eklenen DenseNet-169 sonucu,
(e-1) ResNet-50 sonucu (e-2) Dikkat mekanizması eklenen ResNet-50 sonucu

Yapılan segmentasyon çalışmalarında en düşük OMH değeri, Çizelge 4.1'de görüldüğü gibi Dikkat mekanizması ile kullanılan VGG-16 modeli ile elde edilmiştir. Şekil 4.5'teki örnek çıktılar incelendiğinde ise en başarılı görsel sonucun da yine aynı yöntem ile elde edildiği görülmektedir. Şekil 4.6'da Dikkat mekanizması ile kullanılan VGG-16 yöntemine ait kayıp fonksiyonuna ait grafik ve Şekil 4.7'de ise bu modelin sonuç çıktısı ile gerçek maske (Ground Truth) arasındaki görsel fark verilmiştir.



Şekil 4.6. DUTS eğitim veri kümesinde Dikkat mekanizması eklenen VGG16 modelinin kayıp fonksiyonunun sonuçları



a) Orijinal görüntü b) Gerçek maske

c) Modelin segmentasyon sonucu (VGG16 dikkat mekanizması ile) d) Gerçek maske ve segmentasyon sonucun farkı

Şekil 4.7. DUTS-TE test kümesinde gerçek maske (b), segmentasyon sonucu (c), arasındaki fark (d)

4.2. Transformer Tabanlı Yöntemler

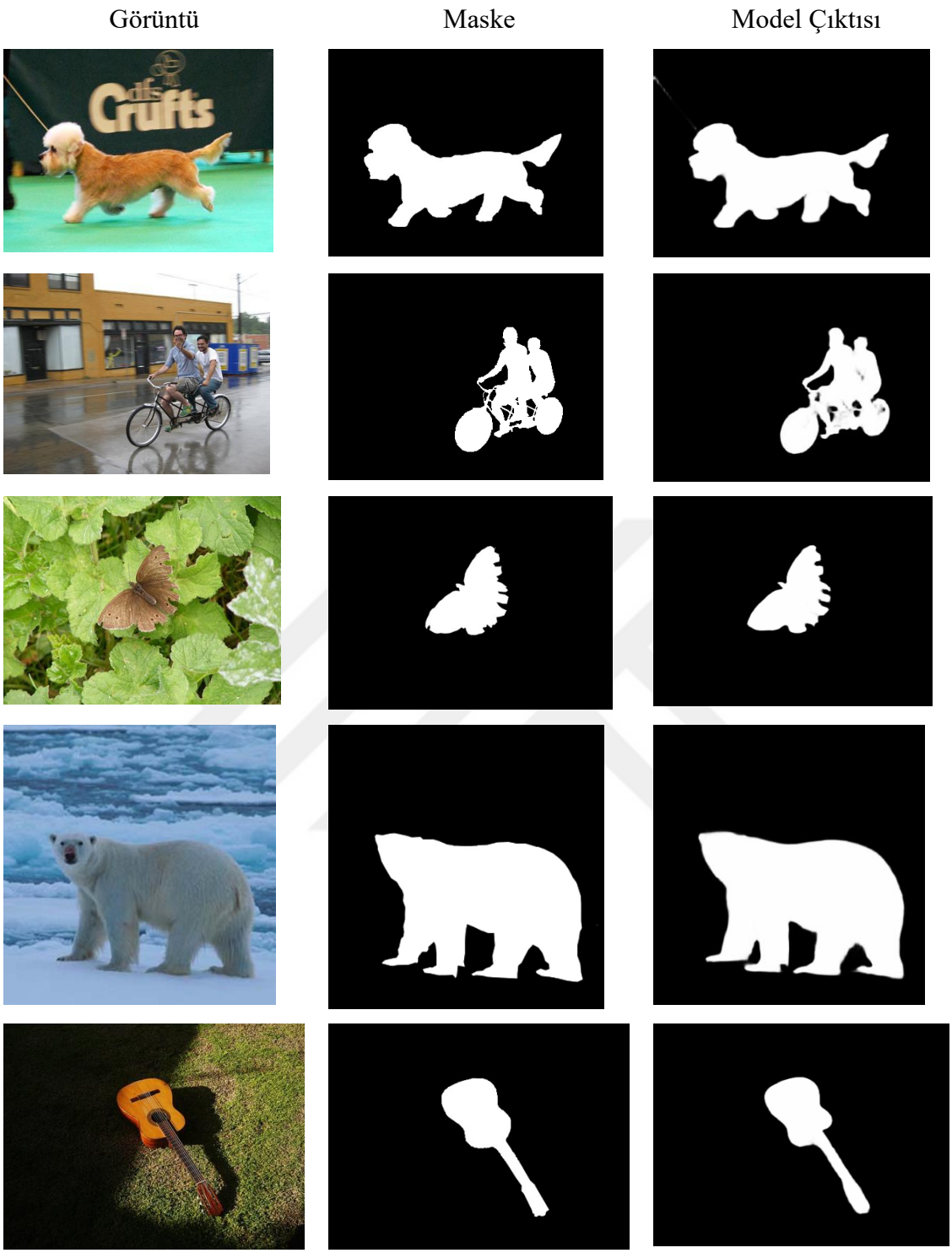
Tez çalışması kapsamında ESA modelleri sonrasında ViT ile de Belirgin Nesne Tespiti (BNT) çalışmaları yapılmıştır. Bu amaçla, Swin transformer mimarisinden "Swin-b" ve "Swin-s" önceden eğitilmiş modelleri kullanılmıştır. Çalışmalarda eğitim ve test aşamasında DUTS (Wang ve ark., 2017) veri kümesi kullanılmıştır. Topluluk öğrenmesi aşamasında Swin-b %70, Res2Net50 %30 olarak ağırlık ataması yapılmıştır. Yapılan deneylerin sonuçları Çizelge 4.2 de verilmiştir. Literatürde yapılan karşılaştırmalarda (Kim ve ark., 2022), Swin ve Res2Net50 modellerinin güçlü yönleri ayrı ayrı incelenmiştir. Bu tez çalışmasında da benzer şekilde, ilk aşamada Swin ve Res2Net50 modelleri birbirinden bağımsız olarak değerlendirilmiş ve topluluk (ensemble) öğrenmesine yer verilmemiştir. Bu tercih, her bir modelin veri üzerindeki performansını izole şekilde gözlemlemek ve güçlü-zayıf yönlerini daha net ortaya koymak amacıyla yapılmıştır. Yapay zekâ modelleri (Swin, Res2Net50) yapısı gereği topluluk öğrenmesiyle çalışmak zorunda değildir; ancak farklı mimarilerin bir araya getirilmesiyle elde edilen topluluk modelleri, çoğunlukla daha dengeli ve yüksek

doğrulukta sonuçlar verir. Bu doğrultuda, tez çalışmasında yapılan çalışmalarda Swin-b ve Res2Net50'nin bir araya getirildiği bir topluluk modeli oluşturulmuş ve bu modelin çıktıları, her bir alt modelin başarımına göre ağırlıklandırılmıştır. Buradaki "yüzdelik oranlar", her bir modelin tahmin sonucuna katkısını ifade etmektedir. Örneğin %70 Swin ve %30 Res2Net50 oranı, sonucun %70'inin Swin modelinden gelen tahmine dayandığını gösterir. Bu ağırlıklar, model başarımlarına göre sabit olarak belirlenmiştir. Swin mimarisi sadece Res2Net50 ile çalışmak zorunda değildir; farklı modellerle de entegre edilebilir. Tez kapsamında çalışmada Res2Net50 tercih edilmiştir çünkü çok ölçekli özellik çıkarımı ile Swin'in kümeleme temelli dikkat mekanizması tamamlayıcı niteliktedir. Bu uyum, literatürde daha önce açıkça vurgulanmamış olsa da, bu tez çalışmasında deneysel olarak değerlendirilmiştir (Liu ve ark., 2021b).

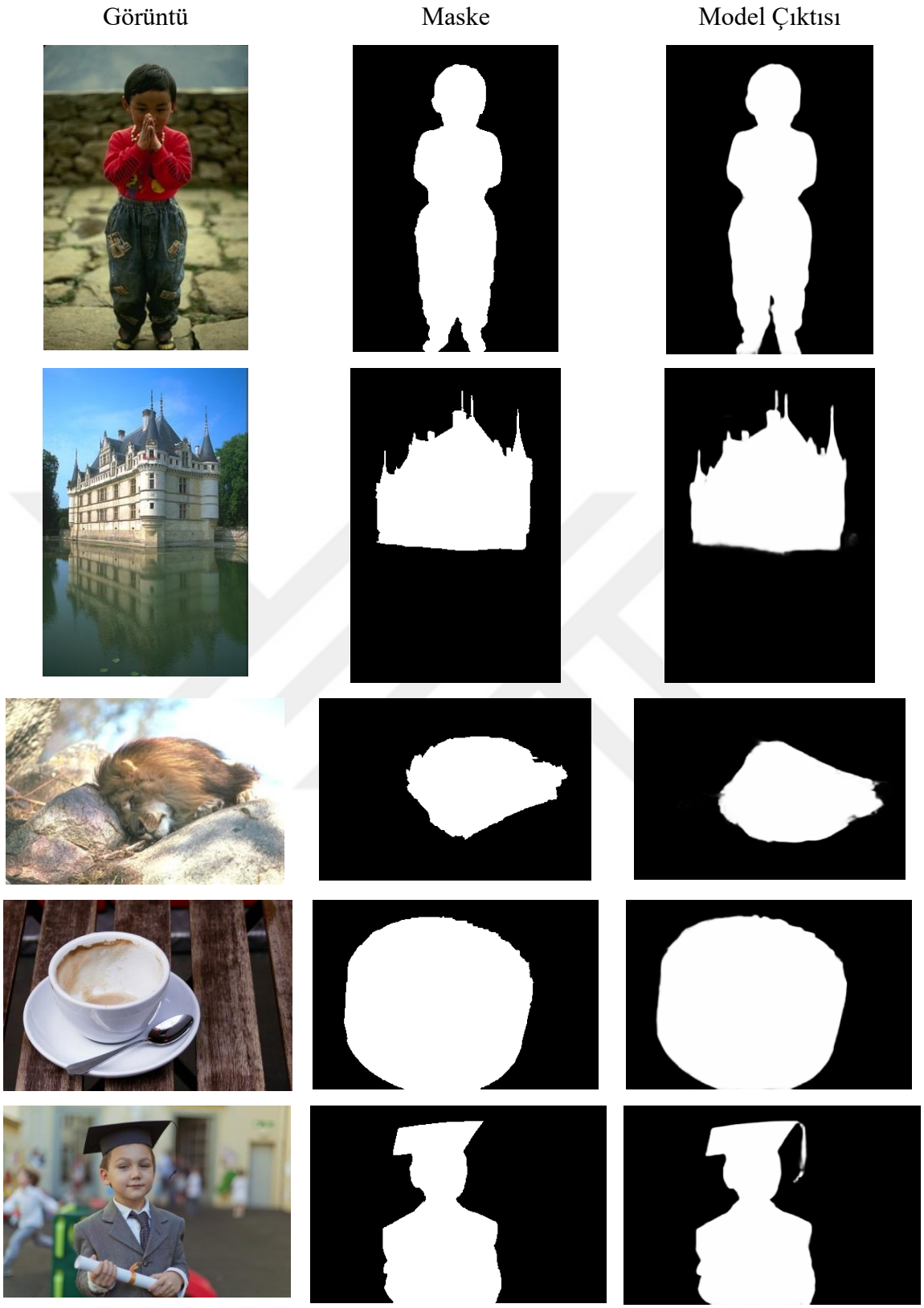
Çizelge 4.2. Swin Trasnformer ve Res2Net50 modelinin sonuçları

Veri Kümesi	**OMH**	**F1 Skor**
Swin - s	0.0356	0.8911
Swin- b	**0.0239**	0.9253
Res2Net50	0.4260	0.8714
Swin – b & Res2Net50	0.0273	**0.9291**

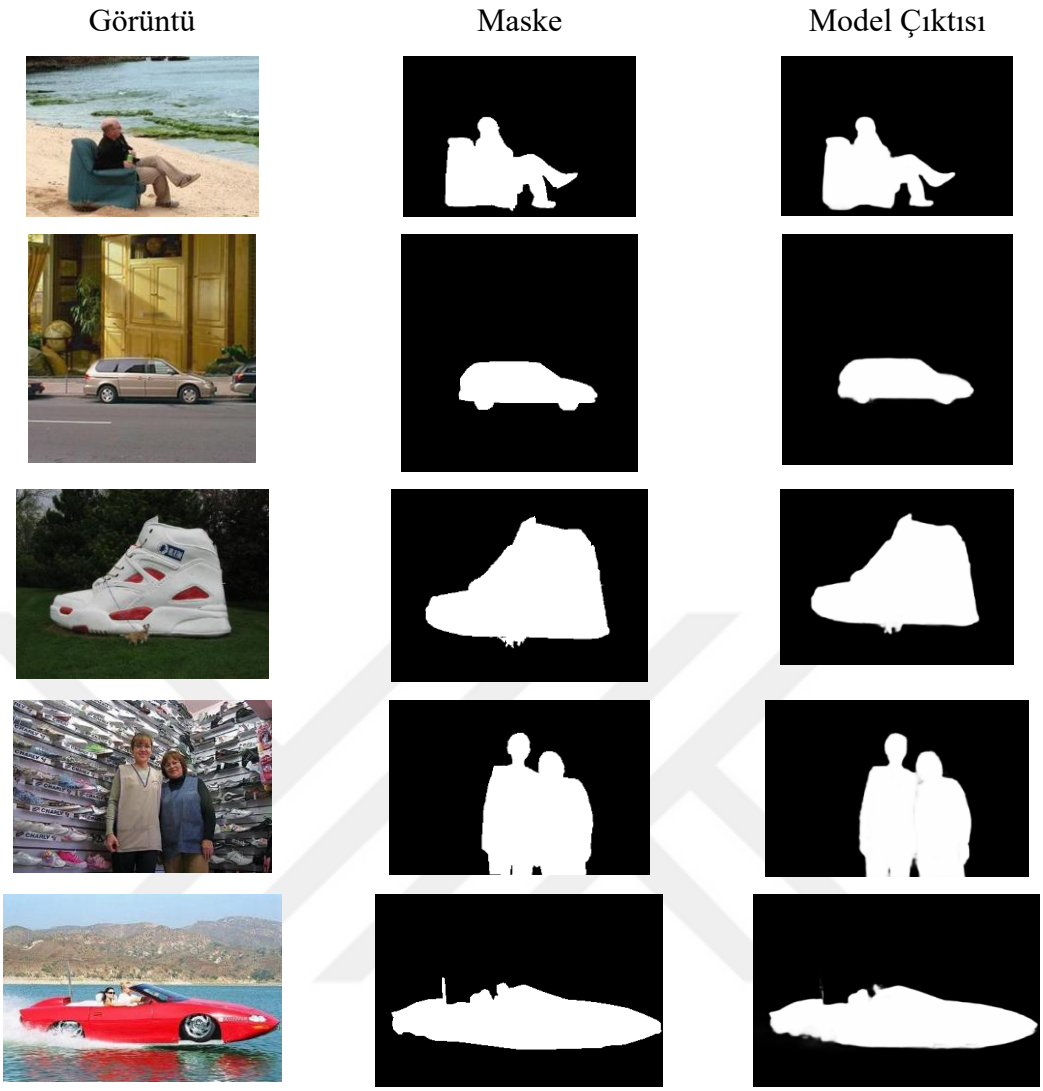
Transformer modellerinden SAM2.1-UNet kullanılarak elde edilen tahminleme sonuçları Şekil 4.8. – 4.10.'da gösterilmiştir. Eğitim kayıp fonksiyonun grafiği Şekil 4.11'dedir. Eğitim gerçekleştirilirken değişen bölüm (epoch) sayılarında test veri kümelerinde segmentasyon sonuçları ise Şekil 4.12'de verilmiştir.



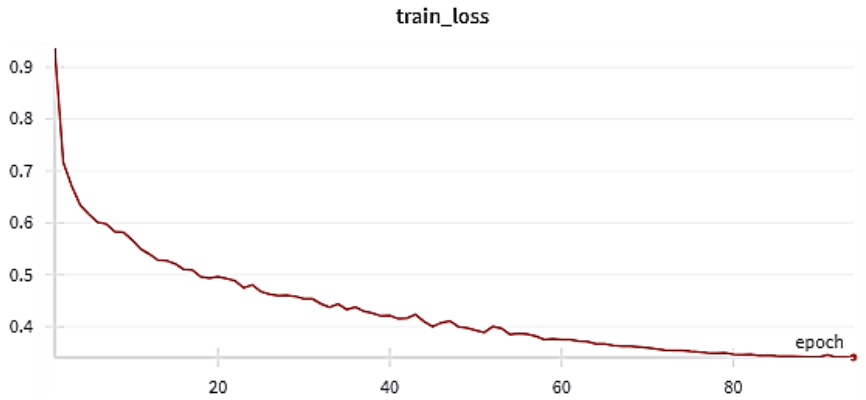
Şekil 4.8. DUT-TE test veri kümesinin SAM2.1-UNet sonuçları



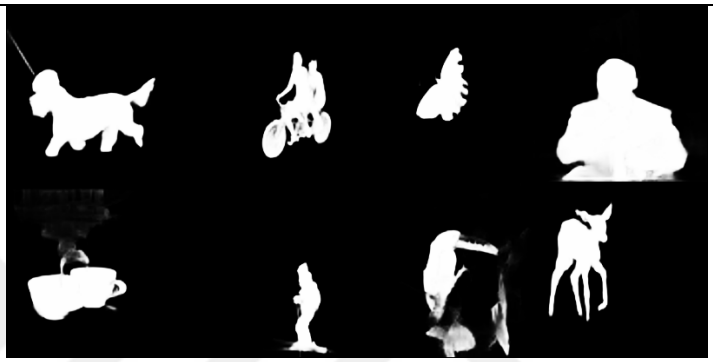
Şekil 4.9. ECSSD test veri kümesinin SAM2.1-UNet sonuçları



Şekil 4.10. DUT-OMRON test veri kümesinin SAM2.1-UNet sonuçları



Şekil 4.11. DUTS eğitim veri kümesinde SAM2.1-UNet modelinin kayıp fonksiyon grafiği

Bölüm sayısı (Epoch)	Segmentasyon sonuçları
20	
40	
60	
80	

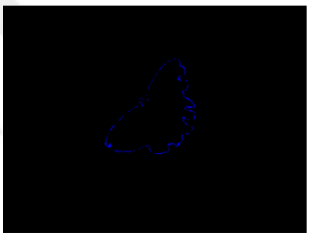
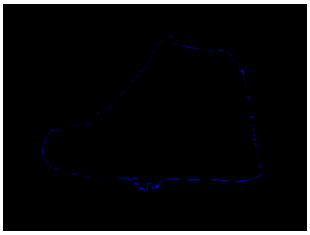
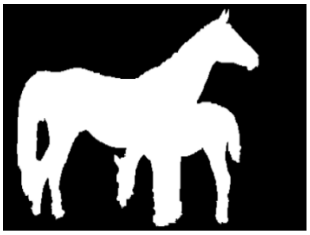
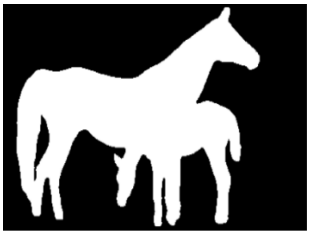
Şekil 4.12. DUTS eğitim veri kümesinde SAM2.1-UNet modelinin farklı bölüm sayılarındaki sonuçları

Deneyler gerçekleştirilirken DUTS veri kümesinde SAM2.1-UNet modelinin farklı bölümlerde eğitim ve doğrulama (validasyon) veri kümelerinde kayıp fonksiyonun değerleri Çizelge 4.3'te gösterilmiştir.

Çizelge 4.3. DUTS veri kümesinde eğitim ve doğrulama için kayıp (loss) değerleri

Bölüm sayısı (Epoch)	**Kayıp Değerleri**	
	Eğitim	**Doğrulama**
10	0.5817	0.2380
20	0.4961	0.2239
30	0.4537	0.2014

DUTS-TE, ECSSD ve DUT- Omron veri kümelerinde SAM2.1-UNet modelinin gerçek maskeler (ground truth) ve segmentasyon modellerinin farklarına ait örnekler Şekil 4.13'te gösterilmiştir.

Test veri kümesi	**Görüntülerin etiketleri**	**Segmentasyon çıktıları**	**Farkları**
DUTS-TE			
DUT-Omron			
ECSSD			

Şekil 4.13. Veri kümelerindeki gerçek maske ve segmentasyon çıktıları ile farkları

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Görüntü segmentasyonu, bir görüntüyü anlamlı bölgelere ayırma işlemidir ve her bir pikseli doğru olarak segmente etmek hedeflenir. Bu süreç, derin sinir ağları, özellikle ESA'lar kullanılarak oldukça başarılı bir şekilde gerçekleştirilmektedir. Gelişen teknoloji ile birlikte ESA mimarilerine ek olarak transformer mimarileri görüntü segmentasyonunda sıklıkla kullanılmaktadır. Bu mimariler, görüntülerdeki her bir nesnenin sınırlarını öğrenerek, segmente edilecek nesnelerin sınırlarını doğru bir şekilde belirleyebilmektedir.

Belirgin nesne tespiti, bir görüntüdeki dikkat çekici nesneleri tespit etmeyi amaçlar ve tıbbi görüntüleme, otonom araçlar, uydu görüntüleri gibi alanlarda sıklıkla kullanılmaktadır. Güncel literatürde belirgin nesne tespiti için temel segmentasyon yöntemleri kullanılmaktadır. Görüntü verilerindeki karmaşık arka planlar, tespit edilecek nesnelerin ince detayları belirgin nesne tespitini zorlaştıran koşullardandır.

Tez kapmasında ESA ve transformer tabanlı yöntemlerle çalışmalar gerçekleştirilmiştir. Yapılan deneylerde transformer tabanlı mimarilerin daha iyi sonuçlar verdiği gözlemlenmiştir. Nihai sonuca SAM2-UNet (Xiong vd., 2024) çalışması temel alınıp geliştirilerek varılmıştır. Deneyler yapılırken farklı veri ön işleme ve veri arttırma teknikleri uygulanmış, optimum parametreler bulunmuştur.

Bu tez çalışmasında belirgin nesne tespiti Python programlama dili kullanılarak gerçekleştirilmiştir. Eğitim sonuçlarının Sα, Eϕ ve OMH değerleri karşılaştırılmıştır. Yapılan deneylerde nihai sonuç SAM2.1-UNet mimarisinin önceden eğitilmiş (pre-trained: sam2.1 hiera large) modeli eğitim aşamasında kullanılmıştır. Literatür ile karşılaştırmada başarılı sonuçlar elde edilmiştir. Sonuçların karşılaştırılması Çizelge 5.1'de verilmiştir.

Çizelge 5.1. DUTS-TE, ECSSD, DUT OMRON test kümelerinde değerlendirme metriklerinin sonuçları

YÖNTEM		VERİ KÜMESİ											
		DUTS-TE				ECSSD				DUT OMRON			
		Sα	**Eϕ**	**OMH**	**F_β**	**Sα**	**Eϕ**	**OMH**	**F_β**	**Sα**	**Eϕ**	**OMH**	**F_β**
ESA tabanlı	Zhao ve Wu, 2019	X	X	0.0405	X	X	X	**0.0328**	X	X	X	**0.0414**	X
	Wei ve ark., 2020	X	0.9100	**0.0340**	**0.8970**	X	0.9300	0.0340	**0.9300**	X	**0.8730**	0.0510	**0.8730**
	Pang ve ark., 2020	**0.8840**	**0.9170**	0.0370	0.8250	**0.9250**	**0.9530**	**0.0330**	0.9110	**0.8330**	**0.8730**	0.0550	0.7380
	Yapılan Çalışma	0.8453	0.8401	0.0473	0.8164	0.8973	0.9308	0.0452	0.9059	0.8111	0.8128	0.0570	0.7683
Transformer tabanlı	Kim ve ark., 2022	0.9310	X	0.0240	**0.9270**	X	X	X	X	0.8750	X	0.0450	**0.8320**
	Liu ve ark., 2023	0.8900	X	0.0390	0.8940	0.9250	X	0.0440	**0.9490**	0.8170	X	0.0590	0.8170
	Liang ve Luo, 2024	0.8687	0.9097	0.0433	0.8392	0.9194	0.9482	0.0344	0.9266	0.8336	0.8728	0.0588	0.7749
	Asheghi ve ark., 2024	0.8930	0.9320	0.0330	0.8480	0.9310	0.9600	0.0270	0.9260	0.8490	0.8930	0.0520	0.7670
	Xiong ve ark., 2024	0.9340	0.9590	0. 0200	X	**0.9500**	**0.9700**	**0.0200**	X	**0.8840**	**0.9120**	**0.0390**	X
	Yapılan Çalışma	**0.9360**	**0.9610**	**0.0190**	0.9090	**0.9500**	0.9690	**0.0200**	**0.9440**	**0.8820**	**0.9100**	**0.0390**	**0.8220**

X: Değerler bulunmamaktadır.

5.2 Öneriler

Bu tez kapsamında, görüntü verilerinde derin öğrenme teknikleri kullanılarak segmentasyon yöntemleri uygulanmıştır. Literatürdeki çalışmalar incelendiğinde, birçok çalışmanın genellikle DUTS veri kümesi üzerinde eğitim yaptığı gözlemlenmiştir. Ancak, belirgin nesne tespiti konusunda daha geniş kapsamlı ve zengin etiketlenmiş veri kümelerinin oluşturulması gerektiği ortaya çıkmaktadır. Bu tür veri kümeleri, piksel bazlı etiketleme yöntemleriyle daha hassas sonuçlar elde edilmesini sağlayabilir, böylece modelin daha doğru ve genel sonuçlar üretmesi mümkün olabilir.

Eğitim aşamasında, özellikle transformer tabanlı mimarilerin kullanılması durumunda güçlü donanım gereksinimlerinin önemli bir sınırlayıcı faktör olduğu gözlemlenmiştir. Bu durum büyük veri kümelerinin ve karmaşık modellerin işlenmesi sırasında donanımın yüksek işlem gücüne ihtiyaç duymasından kaynaklanmaktadır. Bu sorunu aşmak için, daha hafif ve optimize edilmiş modeller tercih edilebilir. Hafif modeller, daha az hesaplama gücü gerektirecek şekilde tasarlandığından, daha geniş veri kümeleriyle eğitim yapılmasına olanak tanımaktadr. Bu sayede, daha fazla veri kullanarak modelin daha fazla görüntüyle eğitilmesi ve dolayısıyla daha kapsamlı bir öğrenme süreci gerçekleştirilmesi mümkün olabilir. Bu strateji, eğitim verisinin çeşitliliğini artırarak modelin daha genel bir başarıya ulaşmasını sağlayacaktır.

KAYNAKLAR

Asheghi, B., Salehpour, P., Khiavi, A. M., Hashemzadeh, M., & Monajemi, A. (2024). DASOD: Detail-aware salient object detection. *Image and Vision Computing*, *148*. https://doi.org/10.1016/j.imavis.2024.105154

Baheti P. (2021). Neural Networks Activation Functions. V7 Labs Blog.

Bharati, P., & Pramanik, A. (2020). Deep learning techniques—R-CNN to mask R-CNN: a survey. *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, 657-668.

Chen, J., He, Y., Frey, E. C., Li, Y., & Du, Y. (2021). Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. *arXiv preprint arXiv:2104.06468*.

Cheng, M. M., Mitra, N. J., Huang, X., Torr, P. H., & Hu, S. M. (2014). Global contrast based salient region detection. IEEE transactions on pattern analysis and machine intelligence, 37(3), 569-582.

Christlein, V., Spranger, L., Seuret, M., Nicolaou, A., Král, P., & Maier, A. (2019, September). Deep generalized max pooling. In *2019 International conference on document analysis and recognition (ICDAR)* (pp. 1090-1096). IEEE.

Cui, Y., Zhou, F., Wang, J., Liu, X., Lin, Y., & Belongie, S. (2017). Kernel pooling for convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921-2930).

Çıtır E. (2020). Yapay Sinir Ağları. Ennurcitir Blog.

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., & Torr, P. (2019). *Res2Net: A New Multi-scale Backbone Architecture*. https://doi.org/10.1109/TPAMI.2019.2938758

Gidaris, S., & Komodakis, N. (2015). Object detection via a multi-region and semantic segmentation-aware cnn model. In *Proceedings of the IEEE international conference on computer vision* (pp. 1134-1142).

H.Üzen and H. Fırat, "Deep learning-based adaptive ensemble learning model for classification of monkeypox disease", Konya Journal of Engineering Sciences (Konjes), vol. 12, no. 4, pp. 822-837, 2024.

Han, L., Li, X., & Dong, Y. (2019). Convolutional edge constraint-based U-net for salient object detection. *IEEE Access*, *7*, 48890-48900.

HE, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

HUANG, Gao, et al. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4700-4708.

Ji, Y., Zhang, H., & Wu, Q. J. (2018). Salient object detection via multi-scale attention CNN. *Neurocomputing*, *322*, 130-140.

Jie, H. J., & Wanda, P. (2020). RunPool: A dynamic pooling layer for convolution neural network. *International Journal of Computational Intelligence Systems*, *13*(1), 66-76.

K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.Kim, T., Kim, K., Lee, J., Cha, D., Lee, J., & Kim, D. (2022). *Revisiting Image Pyramid Structure for High Resolution Salient Object Detection*. http://arxiv.org/abs/2209.09475

Kim, T., Kim, K., Lee, J., Cha, D., Lee, J., & Kim, D. (2022). Revisiting image pyramid structure for high resolution salient object detection. In *Proceedings of the Asian Conference on Computer Vision* (pp. 108-124).

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4015-4026).

Li, G., & Yu, Y. (2016). Deep contrast learning for salient object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 478–487).

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.

Li, X., Chen, C., Chen, Y., Yu, M. A., & Xiao, R. (2025). MSAR-Net: A multi-scale attention residual network for medical image segmentation. Biomedical Signal Processing and Control, 104, 107543.

Liang, B., & Luo, H. (2024). MEANet: An effective and lightweight solution for salient object detection in optical remote sensing images. *Expert Systems with Applications*, *238*, 121778.

Liu, N., & Han, J. (2016). Dhsnet: Deep hierarchical saliency network for salient object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 678–686).

Liu, J. J., Hou, Q., Liu, Z. A., & Cheng, M. M. (2023). PoolNet+: Exploring the Potential of Pooling for Salient Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 887-904. https://doi.org/10.1109/TPAMI.2021.3140168

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021a). *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. http://arxiv.org/abs/2103.14030

Liu, Z., Tan, Y., He, Q., & Xiao, Y. (2021b). SwinNet: Swin transformer drives edge-aware RGB-D and RGB-T salient object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(7), 4486-4497.

Ma, W., & Lu, J. (2017). An equivalence of fully connected layer and convolutional layer. *arXiv preprint arXiv:1712.01252*.

Maninis, K. K., Caelles, S., Pont-Tuset, J., & Van Gool, L. (2018). Deep extreme cut: From extreme points to object segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 616-625).

Nečasová, T., Burgos, N., & Svoboda, D. (2022). Validation and evaluation metrics for medical and biomedical image synthesis. In *Biomedical Image Synthesis and Simulation* (pp. 573-600). Academic Press.

O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Piao, Y., Ji, W., Li, J., Zhang, M., & Lu, H. (2019). Depth-induced multi-scale recurrent attention network for saliency detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 7254–7263).

Prasanth. (2021, March 31). *Activation functions in neural networks*. Medium. https://medium.com/@prasanNH/activation-functions-in-neural-networks-b79a2608a106

Rasamoelina, A. D., Adjailia, F., & Sinčák, P. (2020, January). A review of activation function for artificial neural network. In *2020 IEEE 18th world symposium on applied machine intelligence and informatics (SAMI)* (pp. 281-286). IEEE.

Ren, Z., Gao, S., Chia, L. T., & Tsang, I. W. H. (2013). Region-based saliency detection and its application in object recognition. IEEE Transactions on Circuits and Systems for Video Technology, 24(5), 769-779.

Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., & Jagersand, M. (t.y.). *BASNet: Boundary-Aware Salient Object Detection*. https://github.com/NathanUA/BASNet.

Saha, S., Kumar, A., & Nandi, D. (2024). AUNet-MHA: an attention U-Net based multi-head self attention for lung lesion segmentation from CT images. Procedia Computer Science, 235, 1860-1869.

Sezgin, M., & Sankur, B. L. (2004). Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic imaging, 13(1), 146-168.

Sonoda, S., & Murata, N. (2017). Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, *43*(2), 233-268.

Sur, A., Sagar, S. S., Pal, R., Mitra, P., & Mukhopadhyay, J. (2009, December). A new image watermarking scheme using saliency based visual attention model. In 2009 Annual IEEE India Conference (pp. 1-4). IEEE.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., & Ruan, X. (2017). Learning to detect salient objects with image-level supervision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 136–145).

Wang, T., Zhang, L., Wang, S., Lu, H., Yang, G., Ruan, X., & Borji, A. (2018). Detect globally, refine locally: A novel approach to saliency detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3127–3135).

Wang, Y., Wang, R., Fan, X., Wang, T., & He, X. (2023). *Pixels, Regions, and Objects: Multiple Enhancement for Salient Object Detection*. https://github.com/yiwangtz/MENet.

Wayalun, P., Chomphuwiset, P., Laopracha, N., & Wanchanthuek, P. (2012, August). Images enhancement of G-band chromosome using histogram equalization, OTSU thresholding, morphological dilation and flood fill techniques. In 2012 8th International Conference on Computing and Networking Technology (INC, ICCIS and ICMIC) (pp. 163-168). IEEE.

Wei, J., Wang, S., Wu, Z., Su, C., Huang, Q., & Tian, Q. (2020). Label decoupling framework for salient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13 025–13 034).

Xiong, X., Wu, Z., Tan, S., Li, W., Tang, F., Chen, Y., Li, S., Ma, J., & Li, G. (2024). *SAM2-UNet: Segment Anything 2 Makes Strong Encoder for Natural and Medical Image Segmentation*. http://arxiv.org/abs/2408.08870

Yang, C., Zhang, L., Lu, H., Ruan, X., & Yang, M. H. (2013). Saliency detection via graph-based manifold ranking. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3166-3173).

Yang, J., Xie, F., Fan, H., Jiang, Z., & Liu, J. (2018). Classification for dermoscopy images using convolutional neural networks based on region average pooling. *IEEE Access*, *6*, 65130-65138.

Y. Pang, X. Zhao, L. Zhang, and H. Lu, "Multi-scale interactive network for salient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 9413–9422.

Zeng, Y., Zhang, P., Zhang, J., Lin, Z., & Lu, H. (2019). Towards high-resolution salient object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 7234–7243).

Zhang, L., Dai, J., Lu, H., He, Y., & Wang, G. (2018). *A Bi-directional Message Passing Model for Salient Object Detection*.

Zhang, M. (2009). Bilateral filter in image processing. Louisiana State University and Agricultural & Mechanical College.

Zhao, J.-X., Liu, J.-J., Fan, D.-P., Cao, Y., Yang, J., & Cheng, M.-M. (2019). Egnet: Edge guidance network for salient object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8779–8788).

T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 3085–3094.