Sinopfbd, 2026; 11(1): 1-13 https://doi.org/10.33484/sinopfbd.1670411

Sinop Üniversitesi Fen Bilimleri Dergisi
Sinop Uni J Nat Sci
E-ISSN: 2564-7873

https://dergipark.org.tr/tr/pub/sinopfbd

Derin Öğrenme ile 2D Görüntülerde Belirgin Nesne Tespiti

Gönül Sinem ÖZDOĞAN[1] ve Nurdan AKHAN BAYKAN[2]

How to cite: Özdoğan, G. S., & Akhan Baykan, N. (2026). Derin öğrenme ile 2D görüntülerde belirgin nesne tespiti. *Sinop Üniversitesi Fen Bilimleri Dergisi, 11*(1), 1-13. https://doi.org/10.33484/sinopfbd.1670411

Open Access

Araştırma Makalesi

Sorumlu Yazar
Gönül Sinem Özdoğan
gonulsinemgumussoy@gmail.com

Yazarlara ait ORCID
G.S.Ö: 0009-0000-1661-8591
N.A.B: 0000-0002-4289-8889

Geliş Tarihi: 05.04.2025
Kabul Tarihi: 03.12.2025

Özet

Belirginlik tespiti, bir görüntü veya bir görüntü içindeki bir bölgenin diğerlerinden ne kadar farklı veya dikkat çekici olduğunu belirleme işlemidir. Bu çalışmada, belirginlik tespiti için derin öğrenme yöntemleri ve görüntü işleme teknikleri kullanılmıştır. DUTS veri kümesi kullanılarak yapılan çalışmada; görüntülere yatay çevirme, rastgele açılarda döndürme gibi veri artırma işlemleri uygulanmıştır. Çalışmada dikkat mekanizmaları kullanılarak gerçekleştirilen Evrişimli Sinir Ağları ile belirginlik tespiti, yüksek ve düşük seviyeli özelliklere etkili bir şekilde odaklanmayı sağlamaktadır. Bu yaklaşım, belirgin nesnelerin tespiti konusunda başarılı sonuçlar elde edilmesini sağlamış ve farklı dikkat mekanizmalarının performansı değerlendirilmiştir. Çalışma kapsamında VGG-16, ResNet-50, DenseNet-169 gibi farklı derin öğrenme modellerinde belirgin nesne tespiti yapılmıştır.

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

Deep Learning-based Salient Object Detection in 2D Images

[1]Konya Teknik Üniversitesi, Bilgisayar Müh. Bölümü, Lisansüstü Eğitim Enstitüsü Türkiye

[2]Konya Teknik Üniversitesi, Bilgisayar Müh. Bölümü, Türkiye

This work is licensed under a Creative Commons Attribution 4.0 International License

Abstract

Saliency detection is the process of determining how different or attention-grabbing a region within an image is compared to others. In this study, deep learning methods and image processing techniques were used for salient object detection. Using the DUTS dataset, images were augmented with horizontal flipping and random rotations at arbitrary angles. Salient object detection with convolutional neural networks (CNNs) incorporating attention mechanisms effectively focuses on both high- and low-level features. This approach has demonstrated competitive performance in detecting salient objects, and the performance of different attention mechanisms has been evaluated. Saliency object detection was performed using different deep learning models, including VGG-16, ResNet-50, and DenseNet-169.

Keywords: Salient object detection, deep learning, image processing

Giriş

Belirgin Nesne Tespiti (BNT) (Saliency Object Detection - SOD), bilgisayarlı görme alanında dikkat çeken araştırma konularından biridir. Burada amaç, bir görüntüdeki insan gözü tarafından en çok dikkat çeken nesnelerin veya bölgelerin tespit edilmesidir. BNT, özellikle görüntü anlama, görsel arama, robotik algılama, otonom sürüş ve tıp gibi alanlarda büyük bir öneme sahiptir. Ancak, belirgin nesnelerin tespiti; nesnelerin görüntüdeki konumları, boyutları ve kontrastları gibi çeşitli faktörler nedeniyle

zorlayıcı bir problemdir. Histogram eşikleme, bir görüntüde belirgin nesneleri arka plandan ayırmak için yaygın olarak kullanılan bir tekniktir [1]. 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 [2]. Bu 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 bu analize dayalı en iyi kararları veya tahminleri yapmak için kullanılan sistemlerdir. Rastgele Orman (RO) (Random Forest-RF), Destek Vektör Makineleri (DVM) (Support Vectör 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 ve özellikle kontrast tabanlı yaklaşımlar, bir bölgenin çevresiyle olan renk, parlaklık veya doku farklarını kullanarak belirgin nesneleri vurgular [3]. Belirgin nesne tespiti için DVM tabanlı yaklaşımlar da sunularak, hem yerel hem de küresel özellikleri kullanarak belirginlik haritaları oluşturulmuştur [4]. 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. Belirgin nesne tespiti, bir görüntüde dikkat çekici bölgelerin belirlenmesini amaçlayan önemli bir bilgisayarlı görü problemidir. DHSNet adlı çalışmada, derin öğrenme tabanlı bir hiyerarşik ağ yapısı kullanılarak, çok seviyeli özelliklerin entegrasyonu ile belirginlik haritalarının üretilmesi hedeflenmiştir [5]. Geleneksel yöntemler genellikle el ile mühendislik gerektiren özelliklere dayanırken, DHSNet doğrudan ham görüntü verisinden anlamlı gösterimler öğrenerek, daha genel ve güçlü bir saliency haritalandırması sunmaktadır. Literatürde derin öğrenme temelli yaklaşımlar, özellikle Evrişimsel Sinir Ağları (ESA) (Convolutional Neural Network-CNN) ve Tekrarlayan Sinir Ağları (TSA) (Recurrent Neural Network-RNN) tabanlı modellerle saliency tespitinde başarılı sonuçlar elde etmiş olup, DHSNet gibi hiyerarşik derin ağlar, detaylı nesne sınırları ve daha tutarlı bölgesel vurgular ile öne çıkmaktadır. DHSNet, görüntünün tamamından gelen bilgileri kullanarak kaba bir belirginlik haritası oluşturmaktadır. Ancak bu harita oldukça düşük çözünürlüklü olup, nesne sınırları ve ince detayları korumada yetersiz kalmıştır. Bu nedenle sonrasında detayların iyileştirilmesi için hiyerarşik iyileştirme gereklidir ki bu da modelin ince detayları öğrenmesiyle birlikte eğitim sürecini zorlaştırmaktadır. Derin Öğrenme bir makine öğrenme yöntemidir. Verilen bir veri kümesi ile çıktıları tahmin edecek yapay zekâyı eğitmemize olanak sağlar [6]. Derin öğrenme tekniklerinin son yıllarda hızlı gelişimi ile birlikte etkili yöntemler geliştirilmiştir. Derin öğrenme ile Evrişimli Sinir Ağları (ESA) (Convolutional Neural Networks – CNN) sayesinde, belirgin nesne tespiti yeni bir boyut kazanmıştır. Geleneksel yöntemler, daha yüzeysel özellikler ve önceden belirlenmiş kurallar kullanarak belirginliği tahmin etmeye çalışırken, derin öğrenme modelleri çok katmanlı ve hiyerarşik özellik öğrenme kapasitesiyle daha güçlü sonuçlar üretmekte, belirlenmesi zor olan kenar piksellerin doğru bir şekilde çıkarılmasını sağlamaktadır [7]. Ancak, bu modellerin bazı önemli sınırlamaları bulunmaktadır. Öncelikle, geleneksel derin öğrenme yaklaşımlarında, modelin dikkatini önemli bölgelere yönlendirme yeteneği sınırlıdır. Özellikle karmaşık sahnelerde veya düşük kontrastlı bölgelerde, modelin ilgisiz detaylara odaklanması veya belirli nesneleri kaçırması yaygın bir sorundur. Son zamanlardaki yöntemler genellikle çok ölçekli evrişimli (convolutional) özelliklerin ayrım gözetmeksizin birleştirilmesini benimsemektedir. Görüntülerdeki

tüm özellikler belirgin nesne tespiti için faydalı değildir ve bazıları yapay zekâ modelinin ayrıntıları öğrenip zorlanmasına neden olabilir [8]. Zhao ve Wu [9], bu sorunu çözmek için, etkili yüksek seviyeli özellikleri ve düşük seviyeli yapısal özellikleri belirlemek amacıyla Piramit Özellik Dikkat (Pyramid Feature Attention) ağını tanıtmışlardır. Piao ve ark. [10], belirginlik tespiti için önerdikleri yinelemeli sinir ağıyla (Recurrent Neural Network - RNN) önemli bir başarı elde etmişlerdir. Önerilen ağın deneysel olarak gösterdiği üç ana katkısı bulunmaktadır. İlk olarak, RGB 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 yerleştirmek için kullanmışlardır. Üçüncü olarak, yeni bir tekrarlayan dikkat modülü geliştirmişlerdir ve modelin performansını arttırmışlardır. Yapılan deneyler sonucunda, önerilen yöntemin belirgin nesneleri doğru bir şekilde tanımlayabildiğini belirlemişlerdir. Qin ve ark. [11], Sınır Farkındalıklı Belirgin (Boundary-Aware Salient) nesne tespiti için BASNet adlı bir tahmin-rafine modeli önermişlerdir. 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. BASNet'te kullanılan yöntem sınıf farkındalıklı görevler için geliştirilmiştir birden fazla modelin çıktılarını birleştirerek belirsizliği değerlendirmektedir ancak sınıf bağımsız görevlerde iyi genelleme yapamamaktadır. Zhao ve ark. [12], belirgin kenar bilgisi ile belirgin nesne bilgisi arasındaki tamamlayıcılığa odaklanarak üç adımlı bir kenar rehberlik ağı (EGNet) önermişlerdir. Önerilen EGNet modeli, ana omurga ağından bağımsızdır ve VGG tabanlı modeller temel alınmıştır. 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. Çalışmada, seçkin kenar tespiti ile belirgin nesne tespiti arasındaki tamamlayıcılığa dikkat çekmişlerdir. EGNet'te kullanılan modelin sınırlar için rehber kullanması görüntüdeki belirgin nesne sınırlarını daha düzgün ve net çıkarabilmek için bir iyileştirme yapmıştır ancak model mimarisini ve eğitim sürecini karmaşıklaştırmaktadır. Belirgin nesnelerin değişken piksel ölçekleri, bu nesnelerin tespitinde zorlu problemlerden biridir. Bu zorluklar, çok seviyeli ve çok ölçekli özelliklerin etkin kullanımıyla bağlantılıdır. Pang ve ark. [13], bitişik seviyelerdeki özellikleri entegre etmek için ESA tabanlı toplu etkileşim modülleri önermişlerdir. Bu modüllerde, sadece küçük yukarı/aşağı örnekleme oranları kullanarak daha az gürültü oluşmasını amaçlamışlardır. Wei ve ark. [14], bir etiket ayırma yöntemi kullanarak belirginlik haritasını vücut ve detay haritasına ayıran modellerini tanıtmışlardır. Çalışmalarında, omurga ağı olarak ResNet-50 modelini kullanmışlar ve özellikler arasındaki etkileşimi kullanmak için Özellik Etkileşim Ağı (Feature Interaction Network – FIN) yapısını eklemişlerdir. 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 (Distance Transformation – DT) yöntemini kullanmışlardır. Bu yöntem, merkeze yakın piksellerin daha yüksek tahmin doğruluğuna sahip olmasını kullanmaktadır. Liu ve ark. [15], belirgin nesne tespiti görevinde havuzlama tekniklerinin potansiyelini, evrişimli sinir ağlarında araştırmışlardır. Çalışmalarında, iki havuzlama tabanlı modül önermişlerdir. İlk olarak, potansiyel belirgin nesnelerin konum bilgilerini farklı özellik düzeylerindeki katmanlara yönlendirmeyi amaçlamışlardır. Daha sonra, kaba düzeydeki anlamsal bilgiyi ince düzeydeki özelliklerle birleştirmek için bir özellik toplama modülü tasarlamışlardır. Wang ve arkadaşları [16], belirgin nesne tespitinde karşılaşılan karmaşık sahne yapıları, çoklu nesneler ve dikkat dağıtıcı arka plan gibi zorlukları ele almak

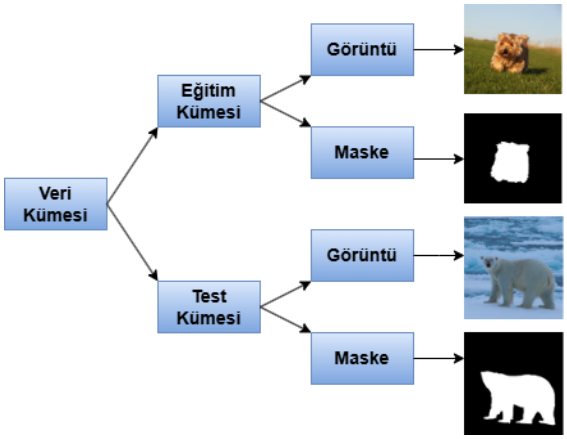
amacıyla MENet modelini önermişlerdir. MENet, insan görsel sisteminden esinlenerek geliştirilen çok aşamalı bir yapı olup, belirgin nesne sınırlarının hassasiyetini artırmak için kenar ve nesne gövdesi bilgileri ayrı dallarda işlenmiştir. Önerilen modelde her aşamada hem sınır hem de nesne bütünlüğü daha kararlı hale getirilmiştir [16]. Asheghi ve arkadaşları [17], belirgin nesne tespiti alanında karşılaşılan temel zorlukları karmaşık arka planlara sahip görüntüler, düşük kontrastlı sahneler, şeffaf yapıdaki nesneler ve birbirini örten nesneler olarak dört ana başlık altında toplamışlardır. Bu problemlerin üstesinden gelmek amacıyla, Detay Farkındalıklı Belirgin Nesne Tespiti (DASOD) adlı bir yöntem geliştirilmiştir. Önerilen DASOD yaklaşımı, maske üretimi ve maske iyileştirme olarak iki temel aşamadan oluşmaktadır. İlk aşamada; sözde maske, vücut etiketi bilgisi ve süper çözünürlük teknikleri bir araya getirilerek maske oluşturulmaktadır. İkinci aşamada ise, sözde maske iyileştirme modülü, sözde kenar çıkarıcı tarafından elde edilen detay haritasını kullanarak maske sınırlarını daha keskin ve doğru hale getirmektedir. DASOD'ta kullanılan maske sınırlarının netleştirilmesi için detay haritalarının ve kenar çıkarıcı modüllerin kullanılması, sınır doğruluğunu daha net bulunmasında etkilidir. Ancak detay odaklı iyileştirmeler, eğitim verilerine ve maske sınırlarının görünürlüğüne bağımlıdır. Eğer eğitim sırasında çok düşük kontrastlı ya da çok şeffaf nesneler türleri yeterince temsil edilmemişse model verileri öğrenmekte zorlanabilmektedir. Bu çalışmada DUTS veri kümesinde farklı Evrişimli Sinir Ağları (ESA) modelleri kullanılmıştır ve bununla birlikte modeller dikkat mekanizmalarıyla güçlendirilirmiştir [18]. Bu doğrultuda, literatürdeki modeller ile karşılaştırılarak, benzer modellere kıyasla belirgin nesnelerin sınırlarının daha keskin ve bölgesel tutarlılığın daha yüksek düzeyde elde edilmesi amaçlanmıştır. Dikkat mekanizmaları modelin öğrenme oranını ile birlikte karmaşıklılığını da arttırdığı için eğitim sürecinin daha uzun olmasına sebep olmuştur.

Materyal ve Metot

Bu çalışmada Evrişimli Sinir ağları (ESA) modelleri kullanılarak belirgin nesne tespiti gerçekleştirilmiştir.

Veri Kümesi

Bu çalışmada DUTS veri kümesi kullanılmıştır [18]. DUTS, belirgin nesne tespiti alanında kullanılan bir veri kümesidir. Bu veri kümesi, görüntülerdeki belirgin nesnelerin tespiti ve segmentasyonu için geliştirilen algoritmaların eğitim ve değerlendirilmesi amacıyla oluşturulmuştur. Eğitim kümesinde 10553, test kümesinde 5019 adet görüntü bulunmaktadır. Ayrıca veri seti içinde tüm görüntü verilerinin segmentasyon maskesi de bulunmaktadır. Görüntülerde doğal sahnelerden nesneler, hayvanlar, bitkiler ve insanlar gibi birbirinden farklı içerikler bulunmaktadır (Şekil 1).



Şekil 1. *DUTS veri kümesinin eğitim-test görüntülerinden örnekler*

Çalışma kapsamında eğitim görüntülerinde veri artırma uygulanmıştır. Bu anlamda 10553 adet

görüntünün her birine rastgele kırpma ve yatay çevirme, rastgele döndürme, parlaklık konstrat bozulması, dolgulama (padding) ve yeniden boyutlandırma görüntü artırma teknikleri uygulanmıştır. Böylelikle veri setindeki her bir görüntü için toplamda 73871 adet görüntü elde edilmiştir.

Segmentasyon Modeli

Segmentasyon; görüntü işleme ve bilgisayarla görme alanında önemli bir adımdır ve bir görüntüyü anlamlı bölgelere ayırma işlemini ifade eder [19]. Bu işlem, her bir piksele ait sınıf veya nesne etiketini belirleyerek, görüntüdeki farklı nesnelerin veya alanların ayrıştırılmasını sağlar. Segmentasyon; tıbbi görüntü analizi, otonom araçlar, yüz tanıma, uydu görüntü analizi gibi birçok alanda kullanılmaktadır. Derin öğrenme tabanlı yaklaşımlar, özellikle ESA ve Unet gibi mimariler kullanılarak, segmentasyon görevlerinde büyük başarılar elde etmiştir [20]. Evrişim Katmanı (Convolutional Layer), ESA'nın ilk katmanıdır ve giriş verisinden anlamlı özellikler çıkarmak için kullanılır. Bu katmanda, belirli bir filtre (kernel) kullanılarak giriş verisi üzerinde kaydırmalı işlemler gerçekleştirilir ve yerel özellikler algılanır. Modelin daha derin katmanlarında, nesnelerin genişliklerinin ve yüksekliklerinin korunmasını sağlamak ve parametre sayısını azaltarak hesaplama maliyetini düşürmek amacıyla havuzlama katmanları (Pooling Layers) kullanılır. Bu katmanlar, genellikle maksimum havuzlama (Max Pooling) veya ortalama havuzlama (Average Pooling) işlemlerini gerçekleştirerek özellik haritalarını daha kompakt hale getirir. Belirginlik tespiti yapılırken literatürde yaygın olarak kullanılan yöntemlerden biri Tam Bağlantılı Ağlardır (Fully Connected Network) [21]. Bu ağlar, her bir nöronun önceki katmandaki tüm nöronlarla bağlantılı olduğu yapıları içerir ve özellikle derin öğrenme modellerinde karar aşamasında etkin bir rol oynar. Ağdaki nöronların toplam fonksiyonunda üretilen çıktılar, belirli değişimlerden geçerek işlenir. Bu değişimleri belirlemek ve modele doğrusal olmayan özellikler kazandırmak için aktivasyon fonksiyonları kullanılır. Aktivasyon fonksiyonları, ağın öğrenme kapasitesini artırarak karmaşık ilişkileri modellemesine olanak tanır. En yaygın kullanılan aktivasyon fonksiyonları arasında ReLU (Rectified Linear Unit), sigmoid ve tanh gibi fonksiyonlar bulunmaktadır [22]. Bu çalışmada, nesnelerin belirginlik tespiti gerçekleştirilmiştir. Belirginlik tespiti için Tam Bağlantılı Ağlar (Fully Connected Network - FCN) kullanılmıştır. Bu sayede çıkarılan öznitelikleri işleyerek daha güçlü ve soyut temsiller oluşturulmuştur. FCN, ağın farklı seviyelerde öğrenilen uzamsal ve yapısal bilgileri birleştirerek belirgin alanları daha etkili şekilde sınıflandırmasına olanak tanır. Ayrıca, modelin belirgin piksellerle arka planı ayırt edebilmesi için güçlü karar mekanizmaları sağlar. Görsel olarak nesneleri ilişkilendirme, belirginlik tespiti için oldukça önemlidir. Literatürde yapılan çalışmalarda nesne özelliklerini öğrenmek için bir dizi evrişimli katman ve havuzlama katmanlarını bir araya getiren Evrişimli Sinir Ağı (Convolutional Neural Network) modelleri kullanılmıştır. Belirgin nesnelerin tespiti; ölçü, şekil ve konumlarında büyük varyasyonlar içerir. Bu, nesnelerin farklı boyutlarda, şekillerde ve konumlarda görüntülenebileceği anlamına gelir. Örneğin, aynı nesne farklı görüntülerde küçük, büyük, deforme olmuş veya farklı açılarda olabilir. Bu varyasyonlar, modelin nesneleri doğru şekilde tespit etmesini zorlaştırır. Küçük nesneler, belirginlik haritalarında yeterince vurgulanamayabilirken, farklı şekiller ve deforme olmuş nesneler de yanlış tespitlere neden olabilir. Ayrıca, nesneler görüntü içinde farklı konumlarda yer alabilir, bu da modelin her türlü konumda doğru tahmin yapmasını zorlaştırır. Bu tür varyasyonlar, belirgin nesne tespiti için büyük zorluklar teşkil eder. Bu değişikliklere uyum sağlamak için bu çalışmada literatürde önerilen "Piramid Özellik Dikkat Ağı" kullanılmıştır (Şekil 2) [9]. Piramid Özellik Dikkat Ağı; Bağlam Farkındalıklı Piramit Özellik Çıkarımı, Dikkat Mekanizmaları ve Kayıp Fonksiyonu olmak üzere üç temel bileşenden oluşmaktadır.

Bağlam Farkındalıklı Piramit Özellik Çıkarımı (Context-aware Pyramid Feature Extraction - CPFE)

Bu modül, farklı ölçeklerde bağlamsal bilgiler elde etmek amacıyla yapay sinir ağının belirli

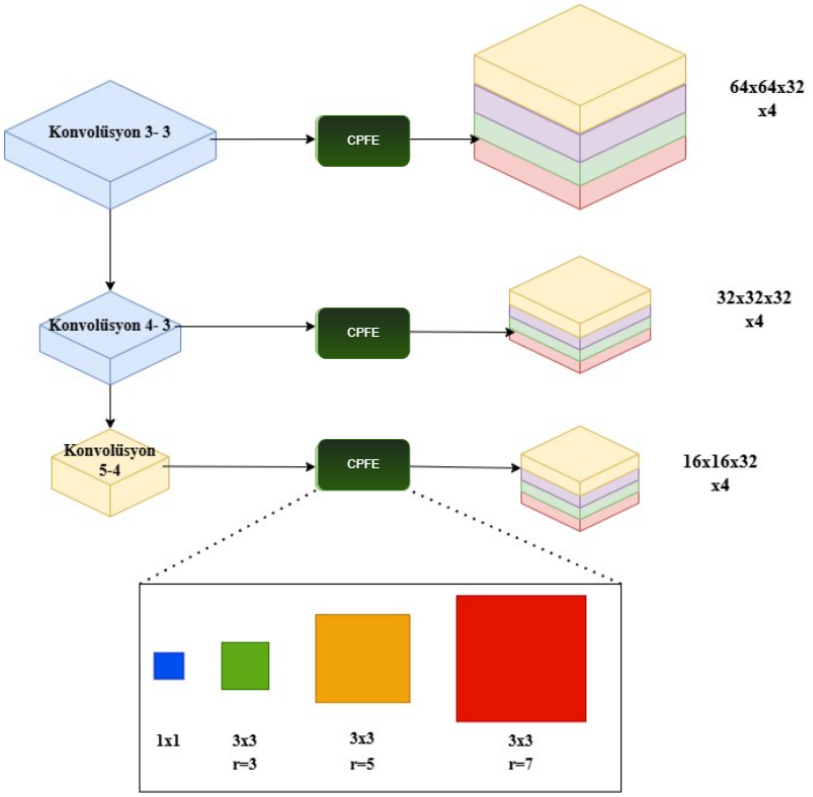
katmanlarından (VGG-16 için conv3-3, conv4-3 ve conv5-3) çıkarılan yüksek seviyeli özellik haritalarına odaklanır. Özellik haritaları, farklı genişleme oranlarına sahip "atrous" konvolüsyonlar ile işlenerek çoklu ölçekli ve farklı alıcı alanlarına sahip hale getirilir. Böylece model, nesnelerin ölçek ve şekil farklılıklarına karşı daha dirençli hale gelir ve belirgin nesnelerin daha tutarlı bir şekilde tespit edilmesi sağlanır.

Dikkat Mekanizmaları

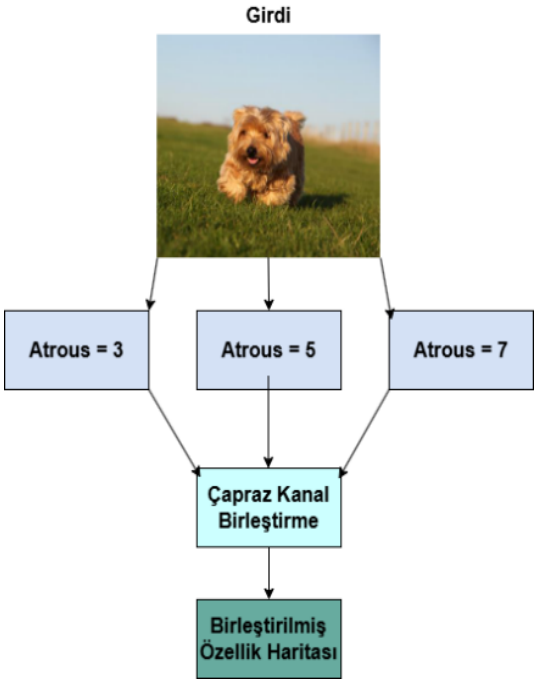
Dikkat mekanizmaları, modelin önemli bilgilere odaklanmasını ve dikkat dağıtıcı unsurları filtrelemesini sağlamak için Kanal Bazlı Dikkat ve Mekânsal Dikkat olmak üzere iki ana bileşenden oluşur. Kanal Bazlı Dikkat (Channel-wise Attention- CA): Bu mekanizma, yüksek seviyeli özellik haritalarında yer 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 [9]. Mekânsal Dikkat (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 [9].

Kayıp Fonksiyonu (Loss Function)

Modelin doğruluğunu artırmak için geleneksel çapraz-entropi kaybına ek olarak, kenar bölgelerindeki detayların daha iyi öğrenilmesini sağlamak amacıyla Kenar Koruma Kaybı (KKK) (edge preservation loss) 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. 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 sağlar. Bağlam farkındalıklı özellik çıkarım modülü, ağın ara katmanlarından alınan bir özellik haritasını giriş olarak kullanır. Bu modül, farklı genişleme oranlarına sahip üç adet 3×3 konvolüsyon katmanı ile bir adet 1×1 konvolüsyon katmanından oluşur. Her bir konvolüsyon katmanında 32 çıkış kanalı bulunur ve bu yapı, farklı ölçeklerde bağlamsal bilgi yakalamayı amaçlar [9]. VGG-16'daki konvolüsyon katmanlarından 3-3 (üçüncü bloktaki üçüncü evrişimli katman), 4-3 (dördüncü bloktaki üçüncü evrişimli katman) ve 5-4 (beşinci ve son bloktaki dördüncü evrişimli katman) temel yüksek seviyeli özellikler olarak kullanılmıştır [9]. 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. 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 (Şekil 2, Şekil 3) [9].



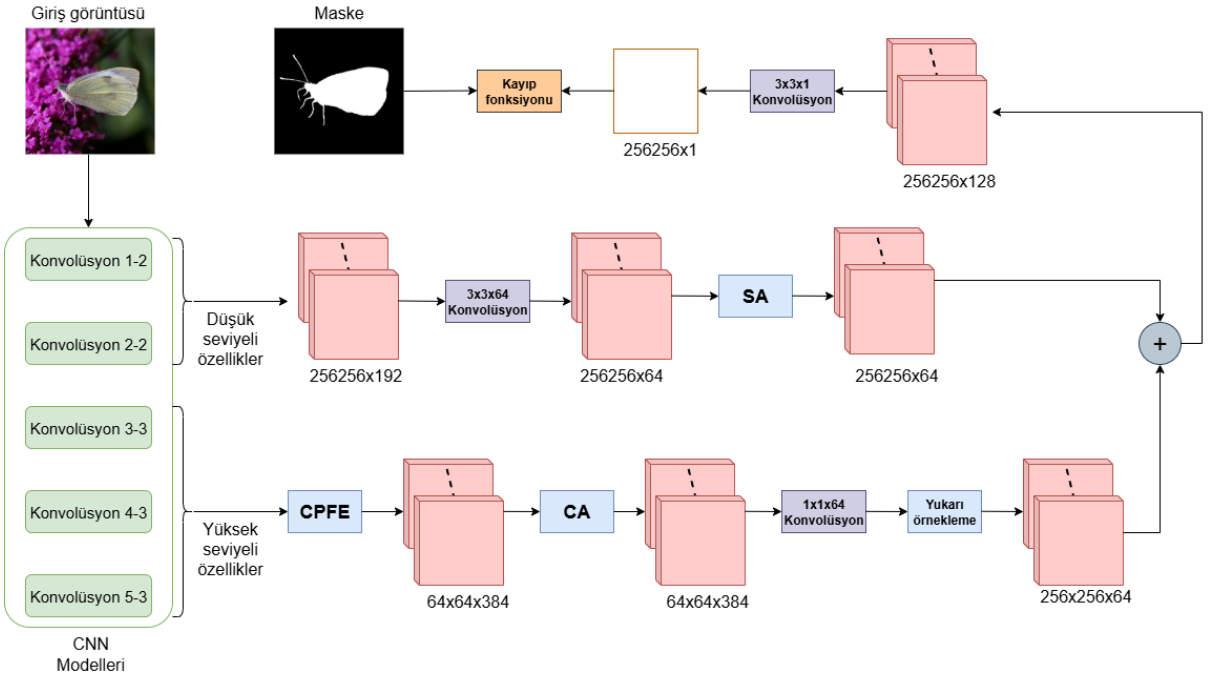
Şekil 2. *Bağlam farkındalıklı özellik çıkarım modülü [9]*



Şekil 3. *Girdi görüntüsünün, üç farklı genleşme oranına sahip (3, 5 ve 7) atrous evrişim katmanına uygulanması*

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. Çalışmanın genel şeması Şekil 4'te gösterilmiştir. Şekil 4'te, CPFE; bağlam farkındalıklı piramit özellik çıkarımı, SA

mekânsal dikkat ve CA kanal bazlı dikkat anlamına gelmektedir. Yüksek düzey özellikler modellerin 3-3, 4-3 ve 5-3 katmanlarından elde edilirken, düşük düzey özellikler 1-2 ve 2-2 katmanlarından alınmakta ve 1-2 boyutuna yukarı örneklenmektedir. Bu çalışmada VGG-16 modeline ek olarak, ResNet-50 ve DenseNet-169 modelleri de kullanılmış ve her biri için aynı kanallar arası birleşme işlemi uygulanmıştır. Ancak, bu modeller birlikte değil, bağımsız olarak değerlendirilmiştir.



Şekil 4. *Çalışmanın genel şeması [4]*

Tüm eğitim süreci, 190 epoch boyunca, 16'lık topluluk boyutu (batch size) ve 4×10^{-4} öğrenme katsayısı kullanılarak yürütülmüştür (Tablo 1). Eğitimler, NVIDIA GeForce RTX 4060 Ti grafik işlem birimi (GPU) üzerinde gerçekleştirilmiştir. Eğitim 190 epok'ta tamamlanmıştır, çünkü eğitim ve doğrulama kaybı belirli bir noktadan sonra düşmemiş, modelde anlamlı bir iyileşme gözlenmemiştir. Öğrenme ilerlemediği için erken durdurma stratejisi eğitim sürecini otomatik olarak sonlandırmıştır.

Tablo 1. *Eğitim hiperparametreleri*

Parametreler	**Değerleri**
Bölüm	190
Topluluk boyutu	16
Öğrenme katsayısı	4 × 10-4

Karşılaştırma Metrikleri

Ortalama Mutlak Hata (OMH) (Mean Absolute Error – MAE): Tahmin edilen değerler ile gerçek değerler arasındaki ortalama mutlak farkı ölçen bir hata metriğidir (Denklem 1) [9].

$$OMH = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1)$$

n, toplam örnek sayısını ifade eder. Yani, değerlendirme yapılan veri kümesindeki örneklerin (gözlemlerin) sayısı; y_i, gerçek (gerçek dünya veya gözlemlenen) değer; $\hat{y}_i$, model tarafından tahmin edilen değerdir. F-Skor (F-Score): F-Skor, modelin doğruluk (precision) ve duyarlılık (recall) arasında bir denge kurarak, bu iki metriğin bir birleşimini sunar. F-Skor, beta parametresi ile doğruluk ve

duyarlılık arasında bir ağırlıklandırma yapar (Denklem 2). Bu çalışmada literatürde kullanıldığı gibi $\beta = 0.3$ olarak kullanılmıştır [21].

$$F - Skoru = \frac{(1+\beta^2)\ X\ Kesinlik\ X\ Duyarlılık}{\beta^2\ Kesinlik + Duyarlılık} \quad (2)$$

Burada kesinlik (precision), tahmin edilen pozitif örneklerin kaçının gerçekten pozitif olduğunu ve duyarlılık (recall), gerçek pozitif örneklerin kaçının doğru tahmin edildiğini gösteren değerlerdir (Denklem 3-4).

$$Kesinlik = \frac{Doğru\ Pozitif}{Doğru\ Pozitif + Yanlış\ Pozitif} \quad (3)$$

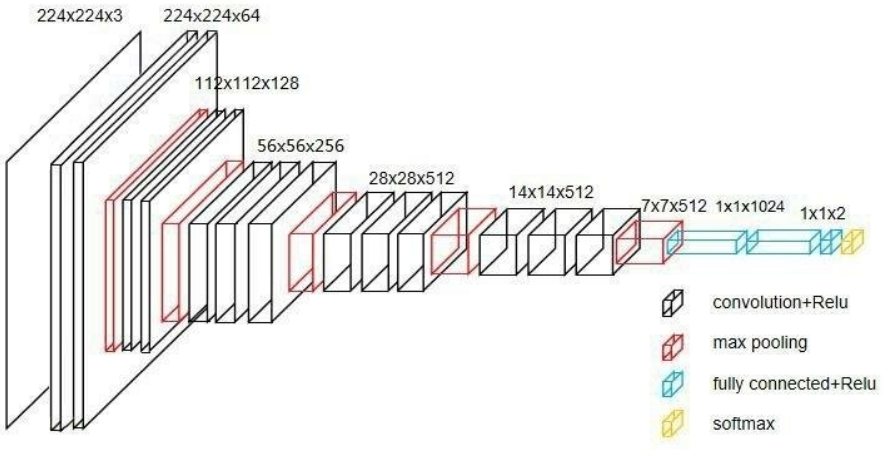
$$Duyarlılık = \frac{Doğru\ Pozitif}{Doğru\ Pozitif + Yanlış\ Negatif} \quad (4)$$

Sα (Yapısal Benzerlik Ölçütü), görüntü segmentasyonu alanında, elde edilen segmentlerin yapısal açıdan ne derece doğru olduğunu belirlemek amacıyla kullanılır. Bu metrik, segmentasyon çıktısının nesnenin özgün yapısal özelliklerini ne kadar başarılı bir şekilde yansıttığını değerlendirmeye olanak tanır [19]. Eϕ (Geliştirilmiş Hizalama Ölçütü), segmentasyon çıktılarının görsel olarak ne kadar algılanabilir olduğunu ve tahmin edilen maske ile gerçek maske arasındaki hizalanma kalitesini ölçen bir metriktir. Bu ölçüt, hem genel doğruluk düzeyini hem de yerel detayların ne ölçüde korunabildiğini değerlendirir [19] (Denklem 5).

$$E_\phi = \frac{1}{N} \sum_{i=1}^{N} \frac{2xP_\phi(i)xG_\phi(i)}{P_\phi(i)+G_\phi(i)} \quad (5)$$

Derin Öğrenme Modelleri

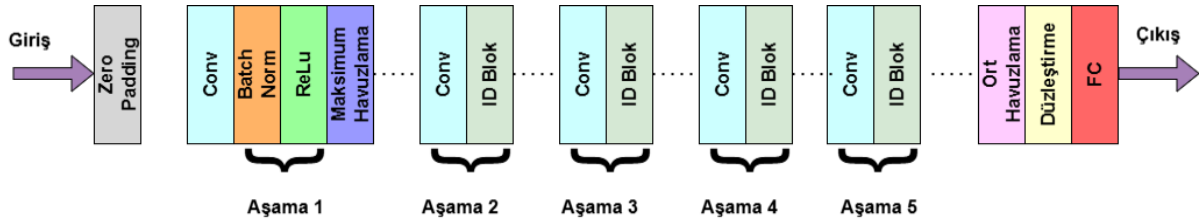
Bu çalışmada farklı omurga ağları kullanılarak belirgin nesne tespiti sonuçları karşılaştırılmıştır. VGG, 2014 yılında geliştirilmiş olan bir ESA mimarisidir. En çok bilinen versiyonu olan VGG-16, adını 16 katmandan oluşmasından alır. VGG mimarisi, özellikle basit ve tekrarlayan yapısı sayesinde birçok görüntü sınıflandırma ve nesne tespiti görevinde yüksek performans göstermiştir [23]. VGG-16, VGG-19 gibi versiyonları bulunmaktadır (Şekil 5) [24].



Şekil 5. *VGG-16 model mimarisi [24]*

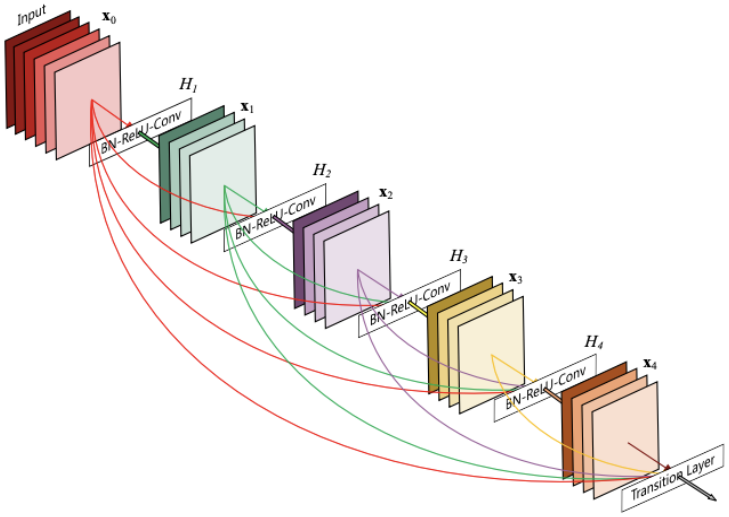
ResNet, 2015 yılında Microsoft Research tarafından ImageNet yarışmasında tanıtılmıştır. ResNet'in temel yeniliği, "residual bloklar" adı verilen özel yapılardır. Bu bloklar, giriş sinyalini doğrudan çıkışa ekleyerek (skip connection) gradyan akışını iyileştirir. Bu sayede, ResNet mimarisi yüzlerce katman derinliğindeki ağların bile etkin bir şekilde eğitilmesine olanak tanır. ResNet-18, ResNet-34 ve ResNet-50 gibi çeşitli versiyonları bulunan bu mimari, görüntü sınıflandırma, nesne tespiti ve segmentasyon

gibi birçok bilgisayarla görme görevinde yaygın olarak kullanılmaktadır (Şekil 6) [25].



Şekil 6. *ResNet-50 model mimarisi*

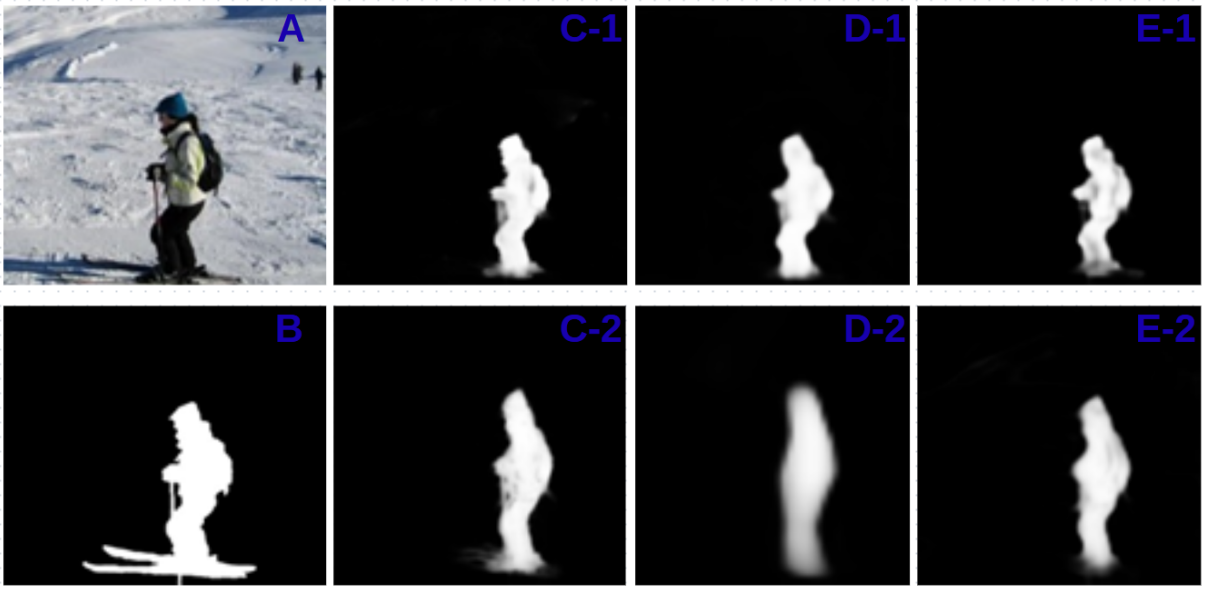
DenseNet, 2017 yılında Gao Huang ve arkadaşları tarafından geliştirilmiş, derin öğrenme mimarilerinde bilgi akışını optimize etmek için yoğun bağlantıları kullanan yenilikçi bir modeldir [7]. DenseNet'in temel özelliği, her katmanın, önceki tüm katmanlardan gelen özellik haritalarını giriş olarak alması ve kendi çıktılarını tüm sonraki katmanlara iletmesidir (Şekil 7).



Şekil 7. *DenseNet model mimarisi [7]*

Tartışma ve Sonuç

Bu çalışmada ESA yöntemleri kullanılarak BNT gerçekleştirilmiştir. ESA mimarilerinde farklı omurga ağları kullanılmış ve temel derin öğrenme omurga ağlarına dikkat mekanizması eklenmesinin sonuca etkileri gözlemlenmiştir Modellere ait sonuçlar Şekil 8'de gösterilmiştir. Şekil 8'deki sonuçlara bakıldığında ResNet-50 ve DenseNet-169 için dikkat mekanizması eklendiğinde görüntülerdeki nesnenin daha doğru boyutta ve yapıda tespit edildiği anlaşılmaktadır. Ancak dikkat katmanlarının eklenmesi modelin karmaşıklığını artırarak uzamsal ayrıntıların bir kısmının kaybolmasına ve kenar bölgelerinde düzensiz sonuçların ortaya çıkmasına neden olmuştur. Bu nedenle modellerde sonuç görüntülerinde bir miktar bozulma olmaktadır. Buna rağmen VGG-16 için dikkat mekanizması eklendiğinde elde edilen sonuç görüntüsünde hem nesnenin daha doğru boyutta ve yapıda tespit edildiği, hem de görüntüde bozulma olmadığı görülmektedir.



Şekil 8. *Farklı derin öğrenme modellerine ait sonuçlar. A) Orijinal görüntü, B) Gerçek maske, (C-1) VGG-16 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*

Tablo 2'de literatürdeki BNT sonuçları ve bu çalışmada yapılan uygulamaların sonuçları karşılaştırma metrikleri ile sayısal sonuçlar ile de verilmiştir. Çalışmalar DUTS veri kümesinde yapılmıştır.

Tablo 2. *Belirginlik tespiti sonuçlarının karşılaştırılması*

Modeller	OMH	F-Skor (Beta=0.3)	Sα	Eϕ
[9]	0.0405	0.8702	X	X
[11]	0.0470	0.8600	X	X
[13]	0.0390	0.8130	0.8840	0.9170
[14]	0.0340	X	X	0.9100
[15]	0.0390	0.8940	0.8900	X
VGG-16	0.0646	0.8136	0.8017	0.8178
ResNet-50	0.0539	0.8355	0.8213	0.8443
DenseNet-169	0.0563	0.8405	0.8245	0.8493
VGG-16 (dikkat mekanizması ile)	**0.0472**	**0.8164**	0.8453	0.8401
ResNet-50 (dikkat mekanizması ile)	0.0495	**0.8192**	0.8308	0.8446
DenseNet-169 (dikkat mekanizması ile)	0.0579	0.7754	0.8183	0.8094

X: Belirtilen çalışmada sonuç değeri verilmemiştir.

Beyanlar

Teşekkür: Bu makale Gönül Sinem Özdoğan'ın Yüksek Lisans tez çalışması kapsamında yapılmıştır.

Fon/Finansman Bilgileri: Bu çalışma için herhangi bir kurum ve/veya kuruluştan destek alınmamıştır.

Etik Kurul Onay ve İzinler: Yazarlar çalışmanın, etik kurul izni veya herhangi bir özel izin gerektirmediğini beyan etmektedir.

Çıkar Çatışmaları/Çatışan Çıkarlar: Herhangi bir çıkar çatışması yoktur.

Yazarların Katkısı: Yazarlar çalışmaya eşit oranda katkı sağlamıştır.

Yapay Zekâ (AI) Beyanı: Yazarlar, bu makalenin hazırlanması sürecinde yapay zekâ araçlarının (ChatGPT ve Gemini gibi) yalnızca dil düzeltme, çeviri ve okunabilirliği artırma amacıyla kullanıldığını beyan ederler. Yazarlar, yapay zekâ destekli tüm dil çıktıları da dâhil olmak üzere, metnin tamamının özgünlüğü, doğruluğu ve bütünlüğü konusunda tüm sorumluluğu üstlenirler.

Kaynaklar

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

[2] Sur, A., Sagar, S.S., Pal, R., Mitra, P., & Mukhopadhyay, J. (2009). A new image watermarking scheme using saliency based visual attention model. *2009 Annual IEEE India Conference*, 1-4. https://doi.org/10.1109/INDCON.2009.5409402

[3] Ren, Z., Gao, S., Chia, L. T., & Tsang, I.W.H. (2014). Region-based saliency detection and its application in object recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *24*(5), 769-779. https://doi.org/10.1109/TCSVT.2013.2280096

[4] Zhang, L., Zhang, D., Sun, J., Wei, G., & Bo, H. (2019). Salient object detection by local and global manifold regularized SVM model. *Neurocomputing*, *340*, 42-54. https://doi.org/10.1016/j.neucom.2019.02.041

[5] Liu, N., & Han, J. (2016). Dhsnet: Deep hierarchical saliency network for salient object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2016),* 678-686. https://doi.org/10.1109/CVPR.2016.80

[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444. https://doi.org/10.1038/nature14539

[7] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017)*, 4700-4708. https://doi.org/10.1109/CVPR.2017.243

[8] Kim, T., Kim, K., Lee, J., Cha, D., Lee, J., & Kim, D. (2023). Revisiting image pyramid structure for high resolution aalient object detection. *Proceedings of the Asian Conference on Computer Vision (ACCV-2022)*, 257-273. https://doi.org/10.1007/978-3-031-26293-7_16

[9] Zhao, T., & Wu, X. (2019). Pyramid feature attention network for saliency detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR-2019)*, 3080-3089. https://doi.org/10.1109/CVPR.2019.00320

[10] Piao, Y., Jiang, Y., Zhang, M., Wang, J., & Lu, H. (2023). PANet: Patch-aware network for light field salient object detection. *IEEE Transactions on Cybernetics*, *53*(1), 379-391. https://doi.org/10.1109/TCYB.2021.3095512

[11] Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., & Jagersand, M. (2019). Basnet: Boundary-aware salient object detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR-2019)*, 7479-7489. https://doi.org/10.1109/CVPR.2019.00766

[12] Zhao, J., Liu, J-J., Fan, D-P., Cao, Y., Yang, J., & Cheng, M-M. (2019). EGNet: Edge guidance network for salient object detection. *IEEE/CVF International Conference on Computer Vision*

(ICCV-2019), 8778-8787. https://doi.org/10.1109/ICCV.2019.00887

[13] Pang, Y., Zhao, X., Zhang, L., & Lu, H. (2020). Multi-scale interactive network for salient object detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR-2020)*, 9410-9419. https://doi.org/10.1109/CVPR42600.2020.00943

[14] Wei, J., Wang, S., & Huang, Q. (2020). F3Net: fusion, feedback and focus for salient object detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*, 12321-12328. https://doi.org/10.1609/aaai.v34i07.6916

[15] 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

[16] Wang, Y., Wang, R., Fan, X. Wang, T., & He, X. (2023). Pixels, regions, and objects: Multiple enhancement for salient object detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR-2023)*, 10031-10040.https://doi.org/10.1109/CVPR52729.2023.00967

[17] Asheghi, B., Salehpour, P., Khiavi, A. M., Hashemzadeh, M., & Monajemi, A. (2024). Dasod: Detail-aware salient object detection. *Image and Vision Computing*, *148*, 105154. https://doi.org/10.1016/j.imavis.2024.105154

[18] Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., & Ruan, X. (2017). Learning to detect salient objects with image-level supervision. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017)*, 3796-3805. https://doi.org/10.1109/CVPR.2017.404.

[19] Xiong, X., Wu, Z., Tan, S., Li, W., Tang, F., Chen, Y., Li. S., Ma, J., & Li, G. (2026). Sam2-unet: Segment anything 2 makes strong encoder for natural and medical image segmentation. *Visual Intelligence*, *4*(1), 2. https://doi.org/10.1007/s44267-025-00106-w

[20] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI-2015)*, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[21] Lu, Y., Chen, Y., Zhao, D., & Chen, J. (2019). Graph-FCN for image semantic segmentation. *International Symposium on Neural Networks (ISNN-2019)*, 97-105. https://doi.org/10.1007/978-3-030-22796-8_11

[22] Parhi, R., & Nowak, R. D. (2020). The role of neural network activation functions. *IEEE Signal Processing Letters*, *27*, 1779-1783. https://doi.org/10.1109/LSP.2020.3027517

[23] Üzen, H., & Fırat, H. (2024). Deep learning-based adaptive ensemble learning model for classification of monkeypox disease. *Konya Journal of Engineering Sciences (Konjes)*, *12*(4), 822-837. https://doi.org/10.36306/konjes.1471289

[24] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR-2015)*, 1-14. https://doi.org/10.48550/arXiv.1409.1556

[25] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2016)*, 770-778. https://doi.org/10.1109/CVPR.2016.90