

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ ANABİLİM DALI

DRONE GÖRÜNTÜLERİNDE NESNE TESPİTİ İÇİN
DERİN ÖĞRENME YÖNTEMLERİNİN İNCELENMESİ

A Survey Of Deep Learning Methods For Object Detection In Drone Images

ARAŞTIRMA MAKALESİ

SERDAR ARICI

MAYIS 2026

ÖZET

İnsansız hava araçlarından (İHA/drone) elde edilen görüntülerde nesne tespiti, son yıllarda bilgisayarlı görme ve yapay zeka alanlarının en aktif araştırma konularından biri hâline gelmiştir. Drone görüntüleri; kuş bakışı açısı, yüksek ölçek değişkenliği, küçük nesne boyutları ve karmaşık arka plan gibi kendine özgü zorluklara sahiptir. Bu çalışmada, drone görüntülerinde nesne tespiti için kullanılan derin öğrenme yöntemleri kapsamlı biçimde incelenmiştir. Geleneksel görüntü işleme yöntemlerinden başlayarak iki aşamalı dedektörler (Faster R-CNN, Mask R-CNN), tek aşamalı dedektörler (YOLO serisi, SSD, RetinaNet) ve transformer tabanlı mimariler (DETR, Swin Transformer) sistematik olarak ele alınmıştır. VisDrone, UAVDT, Stanford Drone Dataset, AU-AIR ve DOTA gibi temel benchmark veri setleri üzerindeki karşılaştırmalı sonuçlar tablolar hâlinde sunulmuştur. Mevcut yöntemlerin güçlü ve zayıf yönleri, gerçek dünya uygulanabilirliği ve hesaplama maliyeti açısından tartışılmıştır. Çalışma, bu alandaki araştırmacılara güncel bir referans kaynağı sunmayı ve gelecekteki araştırma yönlerine ilişkin önerilerde bulunmayı amaçlamaktadır.

Anahtar Kelimeler: Nesne tespiti, drone görüntüleri, derin öğrenme, YOLO, Transformer, VisDrone, küçük nesne tespiti, İHA.

ABSTRACT

Object detection in images obtained from unmanned aerial vehicles (UAV/drones) has become one of the most active research topics in computer vision and artificial intelligence in recent years. Drone imagery poses unique challenges such as bird's-eye view perspectives, high scale variability, small object sizes, and complex backgrounds. In this study, deep learning methods for object detection in drone images are comprehensively reviewed. Starting from traditional image processing methods, two-stage detectors (Faster R-CNN, Mask R-CNN), one-stage detectors (YOLO series, SSD, RetinaNet), and transformer-based architectures (DETR, Swin Transformer) are systematically examined. Comparative results on major benchmark datasets including VisDrone, UAVDT, Stanford Drone Dataset, AU-AIR, and DOTA are presented in tabular form. The strengths and weaknesses of existing methods are discussed in terms of real-world applicability and computational cost. The study aims to provide researchers in this field with an up-to-date reference and to offer suggestions for future research directions.

Keywords: Object detection, drone images, deep learning, YOLO, Transformer, VisDrone, small object detection, UAV.

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	i
İÇİNDEKİLER.....	ii
TABLolar LİSTESİ	iv
ŞEKİLLER LİSTESİ.....	v
KISALTMALAR.....	vi
1. GİRİŞ	1
1.1 Problemin Önemi ve Motivasyon	1
1.2 Drone Görüntülerinin Özellikleri ve Zorlukları	1
1.3 Makalenin Katkıları.....	1
1.4 Makalenin Organizasyonu.....	2
2. İLGİLİ ÇALIŞMALAR	3
2.1 Geleneksel Nesne Tespit Yöntemleri	3
2.2 Derin Öğrenme Tabanlı Yöntemler	3
2.3 Drone/UAV Görüntülerine Özel Çalışmalar	4
2.4 Mevcut Yöntemlerin Kısıtlamaları.....	4
3. DRONE GÖRÜNTÜLERİNİN ÖZELLİKLERİ VE ZORLUKLARI	5
3.1 Görüntü Özellikleri	5
3.1.1 Yüksek Çözünürlük ve Kuş Bakışı Açısı	5
3.1.2 Ölçek Değişkenliği.....	5
3.2 Teknik Zorluklar.....	5
3.2.1 Küçük Nesne Tespiti.....	5
3.2.2 Yoğun Nesneler ve Örtüşme.....	5
3.2.3 Hareket Bulanıklığı	6
3.2.4 Aydınlatma Değişkenleri.....	6
3.3 Veri Seti Zorlukları.....	6
3.3.1 Etiketleme Güçlüğü.....	6
3.3.2 Sınıf Dengesizliği.....	7
4. DERİN ÖĞRENME YÖNTEMLERİ	8
4.1 İki Aşamalı Dedektörler	8
4.1.1 Faster R-CNN.....	8

4.1.2	Mask R-CNN.....	8
4.2	Tek Aşamalı Dedektörler.....	9
4.2.1	YOLO Serisi.....	9
4.2.2	SSD ve RetinaNet.....	10
4.3	Transformer Tabanlı Yöntemler	10
4.3.1	DETR	10
4.3.2	Swin Transformer.....	11
4.4	Küçük Nesne Tespiti için Özel Yaklaşımlar	11
4.4.1	Feature Pyramid Networks (FPN).....	11
4.4.2	Dikkat Mekanizmaları.....	11
5.	VERİ SETLERİ VE DEĞERLENDİRME METRİKLERİ	12
5.1	Benchmark Veri Setleri	12
5.2	Değerlendirme Metrikleri.....	13
6.	DENEYSEL SONUÇLAR VE KARŞILAŞTIRMA	14
6.1	Yöntemlerin Karşılaştırmalı Analizi	14
6.2	Güçlü ve Zayıf Yönler.....	16
7.	TARTIŞMA	17
7.1	Yöntemlerin Gerçek Dünya Uygulanabilirliği.....	17
7.2	Hesaplama Maliyeti ve Doğruluk Dengesi	17
7.3	Etik ve Gizlilik Boyutu	17
8.	SONUÇ VE GELECEK ÇALIŞMALAR.....	18
	KAYNAKÇA	19

TABLÖLAR LİSTESİ

Tablo 1 Benchmark Veri Setleri Karşılaştırması.....	12
Tablo 2 VisDrone Veri Seti Üzerinde Yöntem Karşılaştırması	14
Tablo 3 UAVDT Veri Seti Üzerinde Araç Tespiti Karşılaştırması	15
Tablo 4 Yöntem Kategorileri Genel Karşılaştırması.....	15

ŞEKİLLER LİSTESİ

Şekil 4.1 İki Aşamalı ve Tek Aşamalı Dedektör Mimarilerinin Karşılaştırması	8
Şekil 4.2 YOLO Serisinin Evrimine Göre mAP@0.5 ve FPS Karşılaştırması	10
Şekil 6.1 VisDrone Veri Setinde Hız-Doğruluk Dengesi (mAP@0.5 vs FPS)	15

KISALTMALAR

AP: Average Precision (Ortalama Kesinlik)

AU-AIR: Multi-Modal UAV Dataset

BiFPN: Bidirectional Feature Pyramid Network (Çift Yönlü Özellik Piramidi Ağı)

C3k2: Çift Evrişimli Çapraz Aşama Kısmi Blok (Cross Stage Partial with two Convolutions)

CBAM: Convolutional Block Attention Module (Evrişimsel Blok Dikkat Modülü)

CNN: Convolutional Neural Network (Evrişimli Sinir Ağı)

CPU: Central Processing Unit (Merkezi İşlem Birimi)

CSP: Cross Stage Partial Network (Çapraz Aşama Kısmi Ağ)

DA: Domain Adaptation (Etki Alanı Uyarlaması)

DETR: Detection Transformer (Transformer ile Nesne Tespiti)

DOTA: Dataset for Object Detection in Aerial Images

DP: Differential Privacy (Diferansiyel Gizlilik)

E-ELAN: Extended Efficient Layer Aggregation Network (Genişletilmiş Verimli Katman Kümeleme Ağı)

F1: F1 Skoru (Harmonik Ortalama)

FPN: Feature Pyramid Network (Özellik Piramidi Ağı)

FPS: Frames Per Second (Saniye Başına Kare Sayısı)

GELAN: Genelleştirilmiş Verimli Katman Kümeleme Ağı (Generalized Efficient Layer Aggregation Network)

GFLOPS: Giga Floating-Point Operations Per Second (Giga Kayan Nokta İşlemi/Saniye)

GPS: Global Positioning System (Küresel Konumlama Sistemi)

GPU: Graphics Processing Unit (Grafik İşleme Birimi)

HOG: Histogram of Oriented Gradients (Yönlendirilmiş Gradyan Histogramı)

IMU: Inertial Measurement Unit (Atalet Ölçüm Birimi)

IoU: Intersection over Union (Kesişim/Birleşim Oranı)

İHA: İnsansız Hava Aracı

KD: Knowledge Distillation (Bilgi Damıtma)

LiDAR: Light Detection and Ranging (Işık Tabanlı Uzaktan Algılama)

mAP: Mean Average Precision (Ortalama Kesinlik Ortalaması)

MHA: Multi-Head Attention (Çok Başlıklı Dikkat Mekanizması)

NMS: Non-Maximum Suppression (Maksimum Olmayan Baskılama)

NMS-Free: NMS Bağımsız Çıkarım (Non-Maximum Suppression Free Inference)

NPU: Neural Processing Unit (Sinir Ağı İşleme Birimi)

P: Precision (Kesinlik)

PANet: Path Aggregation Network (Yol Toplama Ağı)

PGI: Programlanabilir Gradyan Bilgisi (Programmable Gradient Information)

R: Recall (Geri Çağırma)

R-CNN: Region-based Convolutional Neural Network (Bölge Tabanlı Evrişimli Sinir Ağı)

RGB: Red-Green-Blue (Kırmızı-Yeşil-Mavi Renk Uzayı)

ROI: Region of Interest (İlgi Bölgesi)

RPN: Region Proposal Network (Bölge Önerisi Ağı)

SIFT: Scale-Invariant Feature Transform (Ölçek Değişmez Özellik Dönüşümü)

SSL: Semi-Supervised Learning (Yarı Denetimli Öğrenme)

SSD: Single Shot MultiBox Detector (Tek Geçişli Çok Kutulu Dedektör)

SURF: Speeded-Up Robust Features (Hızlandırılmış Güçlü Özellikler)

SVM: Support Vector Machine (Destek Vektör Makinesi)

Swin-B: Swin Transformer Base (Temel Kaydırmalı Pencere Transformer)

Swin-T: Swin Transformer Tiny (Küçük Kaydırmalı Pencere Transformer)

UAV: Unmanned Aerial Vehicle (İnsansız Hava Aracı)

UAVDT: Unmanned Aerial Vehicle Detection and Tracking Benchmark

VisDrone: Vision Meets Drones Dataset

VGG: Visual Geometry Group Network

WSL: Weakly-Supervised Learning (Zayıf Denetimli Öğrenme)

YOLO: You Only Look Once (Yalnızca Bir Kez Bakma)

1. GİRİŞ

1.1 Problemin Önemi ve Motivasyon

İnsansız hava araçları (İHA), son on yılda hem askeri hem de sivil alanlarda giderek yaygınlaşmıştır. Uygun maliyetli ve esnek yapıları sayesinde dronlar; tarım, arama-kurtarma, altyapı denetimi, trafik izleme, afet yönetimi ve güvenlik uygulamalarında etkin biçimde kullanılmaktadır. Bu çeşitliliğin temel itici gücü, dronların yüksek çözünürlüklü kamera sistemleriyle büyük alanları kısa sürede tarama kabiliyetidir. Ne var ki drone görüntülerinden anlamlı bilgi çıkarmak, özellikle de nesneleri otomatik olarak tespit etmek, teknik açıdan ciddi zorluklar içermektedir.

Nesne tespiti (object detection), bir görüntü içindeki nesnelerin sınıflandırılması ve konumlandırılması görevini kapsamaktadır. Derin öğrenme yöntemlerinin 2012 yılında ImageNet yarışmasında sergilediği çarpıcı başarısından bu yana, bu alandaki araştırmalar hız kazanmış; evrişimli sinir ağları (CNN) tabanlı mimariler geleneksel görüntü işleme yöntemlerinin önüne geçmiştir [1]. Ancak drone platformlarına özgü görüntü koşulları, standart nesne tespit modellerinin doğrudan kullanımını kısıtlamakta ve alana özgü çözümler geliştirilmesini zorunlu kılmaktadır.

Bu çalışmada, drone görüntülerinde nesne tespiti için literatürde önerilen derin öğrenme yöntemleri sistematik bir çerçevede incelenmektedir. Araştırmanın temel motivasyonu; alandaki mevcut yöntemleri karşılaştırmalı biçimde ortaya koymak, güçlü ve zayıf yönleri belirlemek, üzerinde çalışılan benchmark veri setlerini özetlemek ve gelecekteki araştırma yönlerine ışık tutmaktır.

1.2 Drone Görüntülerinin Özellikleri ve Zorlukları

Drone görüntüleri, geleneksel zemin düzeyi kamera görüntülerinden köklü biçimde ayrılmaktadır. Yüksek irtifa ve kuş bakışı açısı, nesnelerin görünümünü, boyutlarını ve birbirleriyle olan konumsal ilişkilerini dramatik biçimde değiştirmektedir. Tipik bir drone görüntüsünde araçlar, insanlar veya hayvanlar birkaç piksel büyüklüğünde görünebilmekte; bu durum, klasik nesne tespit modellerinin eğitildiği standart veri setlerinden ciddi bir sapma oluşturmaktadır.

Öte yandan hava koşulları, aydınlatma değişkenleri, hareket bulanıklığı ve kalabalık sahnelerdeki örtüşme (occlusion) durumları, algılama performansını olumsuz etkileyen başlıca faktörler arasında sayılabilir. Bu zorluklar, hem model mimarisine hem de eğitim verisi kalitesine yönelik özel çözüm stratejileri gerektirmektedir.

1.3 Makalenin Katkıları

- Geleneksel yöntemlerden transformer tabanlı mimarilere uzanan geniş bir literatür kapsamının sistematik incelenmesi.
- Drone görüntülerine özgü zorluklara yönelik özel yaklaşımların (FPN, dikkat mekanizmaları) detaylı analizi.

- VisDrone, UAVDT, Stanford Drone, AU-AIR ve DOTA gibi temel veri setleri üzerindeki performans karşılaştırmalarının tablolar hâlinde sunulması.
- Gerçek zamanlı uygulama senaryoları, hesaplama maliyeti ve etik boyutlar dahil olmak üzere çok yönlü bir tartışma.
- Gelecekteki araştırma yönlerine yönelik kapsamlı öneriler.

1.4 Makalenin Organizasyonu

Makalenin geri kalan bölümleri şu şekilde düzenlenmiştir: Bölüm 2'de geleneksel ve derin öğrenme tabanlı nesne tespit yöntemlerine ilişkin ilgili çalışmalar incelenmektedir. Bölüm 3'te drone görüntülerinin karakteristik özellikleri ve teknik zorluklar ele alınmaktadır. Bölüm 4'te derin öğrenme yöntemleri ayrıntılı biçimde açıklanmaktadır. Bölüm 5'te kullanılan veri setleri ve değerlendirme metrikleri tanıtılmakta, Bölüm 6'da deneysel sonuçlar karşılaştırmalı tablolar aracılığıyla sunulmaktadır. Bölüm 7'de yöntemler çeşitli boyutlarda tartışılmakta, Bölüm 8'de ise sonuçlar özetlenerek gelecek çalışma önerileri sunulmaktadır.

2. İLGİLİ ÇALIŞMALAR

2.1 Geleneksel Nesne Tespit Yöntemleri

Derin öğrenme yöntemlerinin yaygınlaşmasından önce nesne tespiti, özellik mühendisliğine (feature engineering) dayalı geleneksel yöntemlerle gerçekleştirilmekteydi. Bu yöntemler, uzmanlar tarafından tasarlanmış betimleyicilerle çalışmakta ve makine öğrenmesi sınıflandırıcılarıyla birleştirilmekteydi.

Histogram of Oriented Gradients (HOG) betimleyicisi, Dalal ve Triggs [2] tarafından 2005 yılında önerilmiş ve yaya tespitinde önemli başarılar elde edilmiştir. HOG, görüntü yamalarındaki gradyan yönelimlerinin histogramlarını hesaplayarak nesne yapısını sayısal olarak temsil etmektedir. Destek vektör makinesi (SVM) ile birleştirildiğinde oldukça etkili sonuçlar vermiş olsa da yüksek hesaplama maliyeti ve ölçek değişkenliğine olan duyarlılığı, büyük ölçekli uygulamalarda kısıtlamalar doğurmuştur.

Scale-Invariant Feature Transform (SIFT) algoritması, Lowe [3] tarafından geliştirilmiş ve ölçek ile dönüşüm değişmezliği sayesinde nesne eşleştirme ve tanıma problemlerinde geniş çaplı uygulama alanı bulmuştur. Speeded-Up Robust Features (SURF) ise Bay ve ark. [4] tarafından SIFT'in hesaplama karmaşıklığını azaltmak amacıyla türetilmiştir. Haar Cascade sınıflandırıcısı, Viola ve Jones [5] tarafından gerçek zamanlı yüz tespiti için geliştirilmiş; integral görüntü tekniği ve AdaBoost algoritmasının kombinasyonu ile yüksek hız sağlanmıştır. Ancak bu yöntem, yalnızca sınırlı kategoriler için etkili olup karmaşık arka planlarda yetersiz kalmaktadır.

Geleneksel yöntemler, belirli ve dar kapsamlı görevlerde kabul edilebilir performans sergilese de drone görüntülerinin sunduğu ölçek değişkenliği, küçük nesne boyutu ve arka plan karmaşıklığı gibi zorluklar karşısında yetersiz kalmaktadır. Bu kısıtlamalar, derin öğrenme tabanlı yaklaşımların araştırma gündemine girmesini kaçınılmaz kılmıştır.

2.2 Derin Öğrenme Tabanlı Yöntemler

Girshick ve ark. [6] tarafından 2014 yılında önerilen R-CNN (Region-based Convolutional Neural Network), nesne tespiti alanındaki derin öğrenme devriminin başlangıcını simgelemektedir. R-CNN, seçici arama algoritmasıyla yaklaşık 2.000 bölge önerisi üretmek her bir bölgeyi ayrı ayrı CNN'den geçirmekte ve destek vektör makinesiyle sınıflandırmaktadır. Bu yaklaşım etkileyici doğruluk sonuçları üretmiş olsa da yoğun hesaplama gereksinimleri nedeniyle gerçek zamanlı kullanıma uygun değildir.

Girshick [7] tarafından geliştirilen Fast R-CNN, ortak evrişim hesaplaması ve ROI Pooling mekanizması aracılığıyla R-CNN'nin hızını önemli ölçüde artırmış ve çok görevli kayıp fonksiyonu ile eğitimi basitleştirmiştir. Ren ve ark. [8] ise Faster R-CNN'i önererek bölge önerisi ağırlığı (Region Proposal Network - RPN) modelin içine entegre etmiş ve böylece uçtan uca eğitilebilir ilk iki aşamalı dedektörü hayata geçirmiştir. He ve ark. [8] tarafından geliştirilen Mask R-CNN ise Faster R-CNN altyapısına piksel düzeyinde segmentasyon başlığı ekleyerek nesne tespiti ile örnek segmentasyonunu birlikte ele almıştır.

Liu ve ark. [9] tarafından 2016 yılında önerilen SSD (Single Shot MultiBox Detector), tek geçişte farklı ölçeklerdeki nesneleri tespit edebilen ilk gerçek zamanlı yöntemlerden biridir.

YOLO (You Only Look Once) mimarisi, Redmon ve ark. [10] tarafından aynı yıl önerilmiş ve nesne tespiti için tek bir bütünleşik sinir ağı regresyon problemi olarak formüle edilmiştir. YOLOv2 [11], YOLOv3 [12], YOLOv4 [13], YOLOv7 [14] ve YOLOv8 [15], YOLOv9 [16], YOLOv10 [17] ve YOLO11 [18] gibi ardışık versiyonlar, her nesilde belirgin doğruluk ve hız iyileştirmeleri sunmuştur.

Lin ve ark. [19] tarafından geliştirilen RetinaNet, Focal Loss fonksiyonunu tanıtarak sınıf dengesizliği sorununu ele almış ve tek aşamalı dedektörlerin iki aşamalı yöntemlerle rekabet edebilir hâle gelmesini sağlamıştır. Aynı grubun önerdiği Feature Pyramid Network (FPN) [20] ise çoklu ölçek özellik haritalarını birleştirerek küçük nesne tespitinde önemli başarılar elde etmiştir.

2.3 Drone/UAV Görüntülerine Özel Çalışmalar

Drone görüntülerine yönelik nesne tespiti çalışmaları, 2017 yılından itibaren belirgin biçimde artış göstermiştir. VisDrone veri setinin kamuya açık hâle getirilmesiyle birlikte [21], [22] bu alandaki benchmark çalışmaları ivme kazanmıştır. Du ve ark. [22] tarafından düzenlenen VisDrone Challenge yarışması, yıllar içinde önemli bir referans noktasına dönüşmüştür.

UAVDT veri seti [22], kentsel trafik senaryolarında drone tabanlı araç tespitine odaklanmakta ve iklim koşullarının nesne tespiti üzerindeki etkisini sistematik olarak incelemektedir. Stanford Drone Dataset [23], Robicquet ve ark. tarafından çok kategori nesne izleme görevleri için oluşturulmuş olup yayalar, bisikletler ve golf arabaları gibi çeşitli nesne sınıflarını içermektedir.

DOTA veri seti, uydu ve hava görüntüleri üzerinde dönel kutu (oriented bounding box) tespitini destekleyen kapsamlı bir benchmark sunarken DroneVehicle, termal ve RGB görüntü çiftleri içermesiyle modalite füzyonu araştırmalarına zemin hazırlamaktadır. Bu veri setleri, drone görüntüleri alanındaki araştırmaların yönünü büyük ölçüde belirlemiştir.

2.4 Mevcut Yöntemlerin Kısıtlamaları

Mevcut derin öğrenme tabanlı nesne tespit yöntemlerinin drone görüntüleri bağlamında çeşitli kısıtlamaları bulunmaktadır. Her şeyden önce, standart veri setleri üzerinde eğitilmiş modeller drone görüntülerinin kendine özgü özelliklerini yakalayamamakta ve sınırlı genelleme kabiliyeti sergilemektedir. Küçük nesne tespiti hâlâ çözüme kavuşturulamamış temel sorunlardan biri olarak öne çıkmaktadır; mevcut yöntemler 10×10 pikselin altındaki nesnelerde ciddi performans düşüşü yaşamaktadır.

Hesaplama kaynakları açısından ise Transformer tabanlı modeller ve büyük iki aşamalı dedektörler, gömülü sistemler veya kenar hesaplama platformlarında gerçek zamanlı çalışmayı kısıtlayan yoğun bellek ve işlemci gereksinimlerine sahiptir. Sınıf dengesizliği ve veri etiketleme maliyeti de pratik uygulamaların önündeki önemli engellerdir. Son olarak, drone görüntülerinde gizlilik ve etik boyutlar, bu alandaki araştırmaların dikkat etmesi gereken kritik bir konu olarak giderek daha fazla önem kazanmaktadır.

3. DRONE GÖRÜNTÜLERİNİN ÖZELLİKLERİ VE ZORLUKLARI

3.1 Görüntü Özellikleri

3.1.1 Yüksek Çözünürlük ve Kuş Bakışı Açısı

Modern drone kameraları, 4K ve üzeri çözünürlüklerde görüntü üretebilmektedir. Bu durum, bir yandan zengin mekânsal detay sağlamakta, öte yandan veri depolama, aktarım ve işleme süreçlerinde ciddi yük oluşturmaktadır. Kuş bakışı açısı ise nesnelerin perspektif görünümünü ve birbirlerine göre konumsal ilişkilerini kökten değiştirmektedir. Zemin düzeyindeki kameralarla eğitilmiş standart modeller bu değişime hızlı uyum sağlayamamaktadır.

Kuş bakışı açısından bakıldığında araçlar dikdörtgen şeklinde görünürken insanlar nokta benzeri küçük lekelerle dönüşmektedir. Nesne görünüm sınıfları (object appearance classes) bu nedenle yeniden tanımlanmak zorundadır. Ayrıca kamera açısı ve yüksekliğe bağlı olarak perspektif bozulması ortaya çıkmakta, bu da olası alanı (anchor) ve özellik çıkarma stratejilerini etkilemektedir.

3.1.2 Ölçek Değişkenliği

Drone'un yere olan mesafesi değiştiğinde görüntüdeki nesnelerin görünür boyutları dramatik biçimde farklılaşmaktadır. Örneğin 20 metre irtifada bir araç 200×100 piksel olarak görünürken, 200 metre irtifada aynı araç 20×10 piksele kadar küçülebilmektedir. Bu aşırı ölçek değişkenliği, sabit boyutlu özellik haritaları üreten modellerin genelleme kapasitesini zorlamaktadır.

Çok ölçekli özellik hiyerarşileri (FPN, PANet gibi) bu problemi kısmen çözmektedir. Ancak gerçek zamanlı çalışma gereksinimi ile yüksek ölçek değişkenliğine dayanıklılık arasında sürekli bir denge kurulması gerekmektedir. Eğitim sırasında veri büyütme tekniklerinin (ölçek değişimi, mozaik augmentation) etkin biçimde kullanılması, modelin geniş bir ölçek aralığında genelleme yapabilmesi açısından kritik önem taşımaktadır.

3.2 Teknik Zorluklar

3.2.1 Küçük Nesne Tespiti

Drone görüntülerindeki en yaygın ve kritik zorluk, küçük nesne tespitinin güçlüğüdür. VisDrone veri setindeki nesnelerin büyük çoğunluğu, toplam görüntü alanının yüzde birinden daha az bir kısmını kaplamaktadır. Bu boyuttaki nesneler için ayırt edici özellikler son derece sınırlıdır ve standart evrişim filtrelerinin algılama kapasitesi yetersiz kalmaktadır.

Küçük nesneleri tespit etmeye yönelik başlıca yaklaşımlar şunlardır: süper çözünürlük (super resolution) ön işleme teknikleri, yüksek çözünürlüklü özellik haritaları üretmeye yönelik ağ modifikasyonları, odak kaybı (focal loss) gibi kayıp fonksiyonu optimizasyonları ve küçük nesnelere odaklanan örnekleme stratejileri. Bununla birlikte, bu tekniklerden hiçbirisi tek başına tatmin edici bir çözüm sunamamakta; pratikte bunların kombinasyonu tercih edilmektedir.

3.2.2 Yoğun Nesneler ve Örtüşme

Drone görüntülerindeki kalabalık sahnelerde nesneler sıklıkla birbiriyle örtüşmektedir. Araçların yoğun trafik akışında, yayaların kalabalık alanlarda ya da hayvanların sürü hâlinde

bulunduğu ortamlarda örtüşme oranı yükselmekte ve bu durum tespit doğruluğunu olumsuz etkilemektedir.

Non-Maximum Suppression (NMS) algoritması bu sorunu kısmen ele almaktadır; ancak ağır örtüşme koşullarında standart NMS yetersiz kalmaktadır. Soft-NMS ve DIOU-NMS gibi geliştirilmiş NMS varyantları daha iyi sonuçlar sunmaktadır. Ayrıca nesne ilişki modelleme mekanizmaları ve grafik sinir ağları, örtüşen nesneler arasındaki bağlamsal ilişkileri yakalama konusunda umut verici sonuçlar göstermektedir.

3.2.3 Hareket Bulanıklığı

Hızlı hareket eden dronelerden ya da süratle hareket eden nesnelerden kaynaklanan hareket bulanıklığı, görüntü kalitesini düşürmekte ve nesne tespitini zorlaştırmaktadır. Video tabanlı drone görüntülerinde kare hızı (frame rate) ve pozlama süresi, bulanıklık miktarını doğrudan belirlemektedir.

Bulanıklıkla başa çıkmak için çeşitli yaklaşımlar mevcuttur: görüntü düzeltme (deblurring) ön işleme adımı, video temelli tespit yöntemleri (temporal bilgi kullanımı) ve bulanık görüntüler üzerinde veri büyütme ile eğitim. Optik akış (optical flow) bilgisinin yardımcı girdi olarak sinir ağına beslenmesi de hareket bulanıklığına karşı dayanıklılığı artıran bir strateji olarak öne çıkmaktadır.

3.2.4 Aydınlatma Değişkenleri

Drone görüntüleme koşulları, gün içinde ve farklı hava durumlarında önemli ölçüde değişmektedir. Alacakaranlık saatlerindeki düşük aydınlık düzeyi, şiddetli güneş ışığının neden olduğu aşırı pozlama ve bulutlu havalarda ortaya çıkan düz aydınlatma koşulları, nesne tespiti modellerinin performansını olumsuz yönde etkileyebilmektedir.

Bu zorluğa karşı geliştirilen çözüm yolları arasında görüntü normalleştirme ve kontrast geliştirme teknikleri, düşük ışıktaki görüntü iyileştirme mimarileri ve farklı aydınlatma koşullarını temsil eden zengin veri setleriyle eğitim yer almaktadır. Termal kameralar ile görünür ışık kameralarının birleştirildiği çok modlu (multimodal) yaklaşımlar da aydınlatma sorunlarını azaltmada etkili bir strateji olarak değerlendirilmektedir.

3.3 Veri Seti Zorlukları

3.3.1 Etiketleme Güçlüğü

Drone görüntülerindeki etiketleme süreci, standart görüntülere kıyasla çok daha zahmetlidir. Küçük nesnelerin hassas biçimde sınırlanması (bounding box annotation) hem deneyim hem de dikkat gerektirmekte; binlerce görüntü için bu işlemi gerçekleştirmek ciddi iş gücü maliyeti doğurmaktadır. Görüntü başına ortalama nesne sayısının yüksek olduğu VisDrone gibi veri setlerinde etiketleme süreci özellikle ağırlaşmaktadır.

Yarı otomatik ve aktif öğrenme tabanlı etiketleme araçları bu sürecin hızlandırılmasına yardımcı olmaktadır. Bununla birlikte, zayıf denetimli öğrenme (weakly-supervised learning) ve sıfır atışlı öğrenme (zero-shot learning) yöntemleri, etiketli veri gereksinimini azaltmaya yönelik araştırma gündeminin önemli başlıkları arasında yer almaktadır.

3.3.2 Sınıf Dengesizliđi

Drone görüntüleme senaryolarının büyük çoğunluğunda belirli nesne sınıfları (örneğin araçlar) diğerlerine (örneğin hayvanlar veya sporcular) kıyasla çok daha sık karşılaşılmaktadır. Bu sınıf dengesizliđi, modelin nadir sınıfları öğrenmesini güçleştirmekte ve genel performansı olumsuz etkilemektedir.

Sınıf dengesizliđiyle mücadelede kullanılan başlıca teknikler şunlardır: ağırlıklı kayıp fonksiyonları (weighted loss), aşırı örnekleme (oversampling) ve az örnekleme (undersampling) stratejileri, Focal Loss [19] ve veri büyütme tekniklerinin kombinasyonu. Ancak aşırı dengeleme yöntemleri, model genelleme kapasitesini olumsuz etkileyebileceğinden dikkatli biçimde uygulanmalıdır.

4. DERİN ÖĞRENME YÖNTEMLERİ

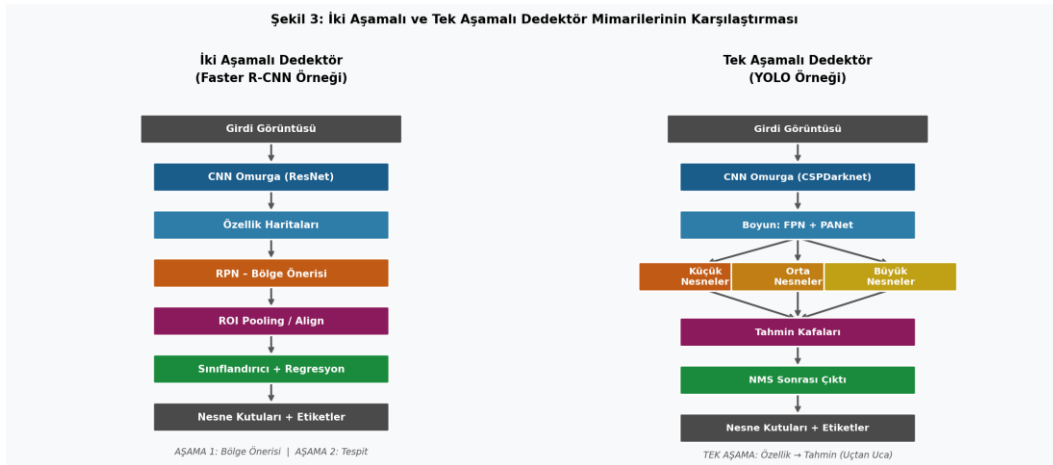
4.1 İki Aşamalı Dedektörler

4.1.1 Faster R-CNN

Faster R-CNN, Ren ve ark. [7] tarafından 2015 yılında önerilmiş ve nesne tespiti alanında önemli bir kırılma noktası oluşturmuştur. Önceki iki aşamalı yöntemlerin aksine, Faster R-CNN bölge önerisi ağını (Region Proposal Network - RPN) evrişim omurgasıyla paylaşarak hesaplama verimliliğini büyük ölçüde artırmıştır.

Mimarinin çalışma prensibi iki temel aşamada özetlenebilir: birinci aşamada RPN, görüntüdeki nesne aday bölgelerini (proposals) üretmektedir; ikinci aşamada ise bu bölgeler sınıflandırma ve konum iyileştirme için tam bağlantılı katmanlardan geçirilmektedir. Anchor (çapa) mekanizması, farklı boyut ve en-boy oranlarına sahip nesnelerin tespitini mümkün kılmaktadır. ResNet-50 veya ResNet-101 omurgasıyla elde edilen yüksek doğruluk, bu yöntemi benchmark çalışmalarında referans model olarak konumlandırmaktadır. Ancak 5-15 FPS aralığındaki hız değeri, gerçek zamanlı drone uygulamaları için yetersiz kalmaktadır.

Drone görüntüleri bağlamında Faster R-CNN'in küçük nesneleri tespit etme kabiliyetini artırmak amacıyla FPN ile birleştirilen çeşitli versiyonlar geliştirilmiştir. Bu modifikasyonlar, özellikle VisDrone veri setinde standart mimariye kıyasla belirgin doğruluk kazanımları sağlamaktadır.



Şekil 4.1 İki Aşamalı ve Tek Aşamalı Dedektör Mimarilerinin Karşılaştırması

4.1.2 Mask R-CNN

Mask R-CNN, He ve ark. [8] tarafından 2017 yılında geliştirilmiş ve Faster R-CNN mimarisini piksel düzeyinde örnek segmentasyonu ile genişletmiştir. ROI Align mekanizması, ROI Pooling'in ayrıklaştırma hatasını ortadan kaldırmakta ve özellikle küçük nesnelerin konumsal doğruluğunu iyileştirmektedir.

Drone görüntülerinde Mask R-CNN'in segmentasyon başlığı, araç yoğunluğu haritalama ve kalabalık analizi gibi görevlerde değerli bilgi sunmaktadır. Bununla birlikte, segmentasyon maskesi üretimi ek hesaplama yükü getirmekte ve modelin gerçek zamanlı uygulamadaki kullanımını daha da kısıtlamaktadır. Çeşitli çalışmalarda hafifletilmiş omurgalar (MobileNet, EfficientNet) ile Mask R-CNN'in birleştirilerek hız-doğruluk dengesi iyileştirilmeye çalışılmıştır.

4.2 Tek Aşamalı Dedektörler

4.2.1 YOLO Serisi

YOLO (You Only Look Once) ailesi, nesne tespitini tek bir ileri besleme geçişiyle gerçekleştiren yöntemler grubunu tanımlamaktadır. Redmon ve ark. [10] tarafından önerilen orijinal YOLOv1, görüntüyü bir ızgaraya bölerek her hücrenin olası nesnelerin konum ve sınıfını tahmin etmesini sağlamıştır. Yüksek hızı (45 FPS), gerçek zamanlı uygulamalar açısından son derece çekici kılmıştır.

YOLOv2 [11], anchor mekanizmasını tanıtmış ve eğitim stratejilerini geliştirerek doğruluğu önemli ölçüde artırmıştır. YOLOv3 [12], çok ölçekli tahmin kafaları ekleyerek küçük nesne tespitinde kayda değer ilerleme sağlamıştır. YOLOv4 [13], CSPDarknet omurgası, PANet boyun mimarisi ve Mosaic veri büyütme gibi çeşitli iyileştirmelerle hem doğruluk hem de hız açısından önceki versiyonları geride bırakmıştır.

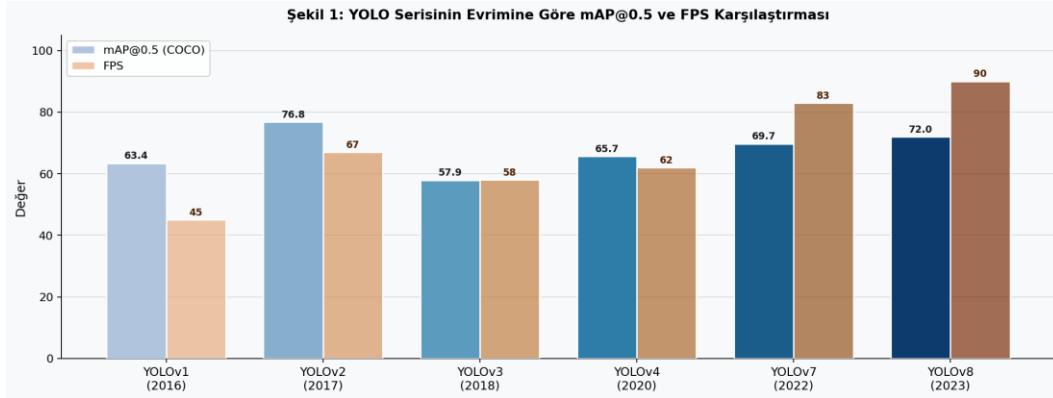
YOLOv7 [14], E-ELAN (Extended Efficient Layer Aggregation Network) mimarisini benimseyerek o döneme kadarki en yüksek doğruluk-hız dengesine ulaşmıştır. Ultralytics tarafından geliştirilen YOLOv8 [15] ise tamamen yeniden tasarlanmış bir mimari sunmuş; ayrıştırılmış tespit başlıkları (decoupled heads), geliştirilmiş veri büyütme ve yeni omurga yapısı ile hem hız hem doğruluk açısından çita yükseltilmiştir. VisDrone veri setinde YOLOv8-m ile elde edilen mAP@0.5 değeri 32,5 olarak raporlanmış ve bu sonuç, serinin tüm önceki versiyonlarının üzerinde kalmıştır.

YOLOv9, Wang ve ark. [16] tarafından 2024 yılında önerilmiş ve Programlanabilir Gradyan Bilgisi (Programmable Gradient Information - PGI) mekanizmasını tanıtmıştır. PGI, derin ağlarda bilgi kaybını önlemek amacıyla tasarlanmış olup Genelleştirilmiş Verimli Katman Kümeleme Ağı (GELAN) omurgasıyla birleştirildiğinde YOLOv8'e kıyasla belirgin doğruluk artışı sağlamaktadır. VisDrone veri setinde YOLOv9-c ile elde edilen mAP@0.5 değeri 34,1 olarak raporlanmıştır.

YOLOv10, Wang ve ark. [17] tarafından 2024 yılında önerilmiş ve NMS (Non-Maximum Suppression) bağımlılığını ortadan kaldıran çift atamaya dayalı (dual-assignment) eğitim stratejisini benimsemiştir. Bu yaklaşım, model çıkarım süresini kısaltmakta ve kenar hesaplama donanımlarında konuşlandırmayı kolaylaştırmaktadır. CSP-GELAN omurgasıyla donatılan YOLOv10-m, yalnızca 15,4 milyon parametre ile yüksek doğruluk sunmakta ve parametre verimliliği açısından önceki versiyonları geride bırakmaktadır.

YOLO11, Khanam ve Hussain [18] tarafından 2024 yılında sunulan en güncel YOLO neslidir. C3k2 (Cross Stage Partial with two Convolutions) omurgası ve geliştirilmiş veri büyütme stratejileriyle donatılan YOLO11-m, VisDrone veri setinde 36,8 mAP@0.5 ve 88 FPS ile

serinin tüm önceki versiyonlarının üzerinde bir performans sergilemiştir. Hem doğruluk hem de hız açısından YOLO serisinin şu anki zirvesini temsil etmektedir.



Şekil 4.2 YOLO Serisinin Evrimine Göre mAP@0.5 ve FPS Karşılaştırması

4.2.2 SSD ve RetinaNet

SSD (Single Shot MultiBox Detector) [9], VGG-16 omurgası üzerine inşa edilmiş ve farklı çözünürlüklerdeki özellik haritalarından tahmin üreten çok ölçekli bir mimari benimsemiştir. Tek aşamalı yapısı sayesinde yüksek kare hızları (46 FPS civarı) elde edilmektedir; ancak küçük nesnelerin tespitinde Faster R-CNN'e kıyasla daha zayıf performans sergilemektedir.

RetinaNet [19], Lin ve ark. tarafından 2017 yılında önerilmiş ve Focal Loss adlı yeni bir kayıp fonksiyonu sunmuştur. Focal Loss, eğitim sürecinde kolayca sınıflandırılan arka plan örneklerinin ağırlığını azaltarak zor örneklerle odaklanmaktadır. Bu yaklaşım, tek aşamalı dedektörlerde sınıf dengesizliği problemini önemli ölçüde hafifletmiş ve RetinaNet'in iki aşamalı yöntemlerle rekabet edebilir doğruluk seviyelerine ulaşmasını sağlamıştır. FPN omurgasıyla birleştirilen RetinaNet, drone görüntülerinde çok ölçekli nesne tespitinde etkili bir model olarak öne çıkmaktadır.

4.3 Transformer Tabanlı Yöntemler

4.3.1 DETR

DETR (Detection Transformer), Carion ve ark. [24] tarafından 2020 yılında önerilmiş ve nesne tespitini bir küme tahmini (set prediction) problemi olarak ele almıştır. Standart Transformer encoder-decoder mimarisini kullanan DETR, el ile tasarlanmış bileşenleri (NMS, anchor tasarımı) tamamen ortadan kaldırmaktadır. Encoder, CNN omurgasından gelen özellik haritalarını dönüştürürken, decoder nesne sorgularını (object queries) görüntü özellikleriyle ilişkilendirerek sınıf ve konum tahminleri üretmektedir.

DETR'nin temel avantajı; uçtan uca eğitilebilir, el ile ayar gerektirmeyen temiz bir mimari sunmasıdır. Ancak yavaş yakınsama (yaklaşık 500 epoch), yüksek hesaplama maliyeti ve küçük nesnelerde görece düşük performans, drone uygulamaları açısından önemli kısıtlardır.

Deformable DETR gibi geliřtirmeler, dikkat mekanizmalarını daha verimli hâle getirerek bu sorunları kısmen gidermeye çalışmaktadır.

4.3.2 Swin Transformer

Swin Transformer, Liu ve ark. [25] tarafından 2021 yılında önerilmiş ve hiyerarşik pencere dikkat mekanizması ile görüntü tanıma görevlerinde CNN'lere güçlü bir alternatif sunmuştur. Pencere tabanlı öz-dikkat (window-based self-attention) hesaplama karmaşıklığını doğrusal olarak tutarak büyük çözünürlüklü görüntülerde verimli çalışmayı mümkün kılmaktadır. Shifted window mekanizması ise pencerelerin birbirleriyle etkileşimini sağlamaktadır.

Swin-T (küçük) ve Swin-B (temel) versiyonları, drone görüntülerinde nesne tespiti görevlerinde umut verici sonuçlar sergilemiştir. VisDrone veri setinde Swin-B ile elde edilen mAP@0.5 değeri 31,7 olarak raporlanmıştır. Bununla birlikte, yüksek parametre sayısı (Swin-B için ~88 milyon) ve sınırlı FPS değeri (yaklaşık 22), bu modellerin gömülü sistemlerde konuşlandırılmasını güçleştirmektedir.

4.4 Küçük Nesne Tespiti için Özel Yaklaşımlar

4.4.1 Feature Pyramid Networks (FPN)

FPN [20], Lin ve ark. tarafından geliştirilen ve çok ölçekli nesne tespiti için temel bir referans mimarisi hâline gelen bir yapıdır. Evrişim ağlarının doğal olarak oluşturduğu özellik hiyerarşisini, üst-aşağı yol boyunca anlam zenginleştirme ve alt-üst yol boyunca yerel detay aktarımı ile birleştiren FPN; hem yüksek düzey anlam hem de yüksek çözünürlük bilgisini içeren çok ölçekli özellik haritaları üretmektedir.

Drone görüntüleri bağlamında FPN, küçük nesnelerin tespit oranını önemli ölçüde artırmaktadır. PANet [26], FPN'e ek bir alt-üst yol (bottom-up path augmentation) ekleyerek düşük düzey konum bilgisinin daha verimli aktarılmasını sağlamaktadır. BiFPN (EfficientDet [27] içinde kullanılan) ise çift yönlü ağırlıklı özellik piramidi yaklaşımıyla FPN'in sınırlamalarını aşmaktadır.

4.4.2 Dikkat Mekanizmaları

Dikkat mekanizmaları (attention mechanisms), modelin görüntünün belirli bölgelerine veya özellik kanallarına seçici olarak odaklanmasını sağlamak ve küçük nesnelerin tespitinde önemli bir rol oynamaktadır. Kanal dikkat mekanizmaları (SENet gibi) her özellik kanalının göreceli önemini ağırlıklandırırken, uzamsal dikkat mekanizmaları görüntünün hangi bölümlerinin daha önemli olduğunu belirlemektedir.

Transformer'ın öz-dikkat mekanizması, uzun menzilli bağımlılıkları modelleyerek kalabalık sahnelerdeki nesne ilişkilerini yakalamada üstünlük sağlamaktadır. CBAM (Convolutional Block Attention Module) gibi hafif dikkat modülleri, mevcut CNN mimarilerine entegre edilerek belirgin doğruluk iyileştirmeleri sağlanmaktadır. Drone görüntülerinde arka plan baskınlığını azaltmak ve nesnelere odaklanmayı artırmak için dikkat mekanizmalarının kullanımı giderek yaygınlaşmaktadır.

5. VERİ SETLERİ VE DEĞERLENDİRME METRİKLERİ

5.1 Benchmark Veri Setleri

Drone görüntülerinde nesne tespiti araştırmaları, yıllar içinde oluşturulan çeşitli benchmark veri setleri üzerinde değerlendirilmektedir. Bu veri setleri, araştırmacıların yöntemlerini standart koşullar altında karşılaştırmasını mümkün kılmaktadır. Aşağıda temel benchmark veri setleri incelenmiştir.

VisDrone [21], Tianjin Üniversitesi tarafından hazırlanmış ve UAV platformlarında çeşitli hava koşulları ile irtifada toplanmış 10.209 görüntü içermektedir. Pedestrian (yaya), car (araç), van (minibüs), truck (kamyon), bus (otobüs), motor (motosiklet), bicycle (bisiklet) ve tricycle (üç tekerlekli araç) dahil olmak üzere 10 nesne kategorisi içermektedir. Görüntü başına ortalama 52 nesne ile yoğun sahneler barındıran bu veri seti, küçük nesne tespiti araştırmaları için kritik bir referans noktası oluşturmaktadır.

UAVDT veri seti [22], kentsel trafik senaryolarında drone görüntüleme uygulamalarına yönelik olarak tasarlanmıştır. 80.000 video karesi ve yaklaşık 850.000 bounding box ile araç, otobüs ve kamyon sınıflarında nesne takibi ve tespiti görevlerini kapsayan bu veri seti, farklı hava koşullarını (güneşli, bulutlu, gece) ve kamera açılarını yansıtmaktadır.

Stanford Drone Dataset [23], Robicquet ve ark. tarafından Stanford Üniversitesi kampüsü üzerinde uçan dronelerden elde edilmiştir. Yaya, bisiklet, araba, golf arabası, elektrikli bisiklet ve tekerlekli sandalye olmak üzere altı sınıf içermektedir. Sosyal güç modeli araştırmaları ve çok nesne izleme görevlerinde yaygın biçimde kullanılmaktadır.

AU-AIR veri seti, farklı irtifa ve uçuş koşullarında çok modlu sensör verilerini (RGB kamera, IMU, GPS) birleştirmektedir. 32.823 görüntü ve 8 nesne sınıfıyla birden fazla algılama modalitesini destekleyen bu veri seti, sensor füzyonu araştırmaları için değerli bir kaynak sunmaktadır.

DOTA (Dataset for Object deTection in Aerial images), uçak, araç, basketbol sahası, depolama tankı gibi 15 nesne kategorisini kapsayan ve dönel kutu (oriented bounding box) desteği sunan kapsamlı bir benchmark sunmaktadır. Yüksek çözünürlüklü uydu ve hava görüntülerinden oluşan bu veri seti, geniş alan tarama uygulamalarına yönelik araştırmalarda öne çıkmaktadır.

Tablo 1 Benchmark Veri Setleri Karşılaştırması

Veri Seti	Yıl	Görüntü Sayısı	Nesne Sayısı	Sınıf	Çözünürlük	Platform
VisDrone	2018	10.209	~540.000	10	Değişken	UAV
UAVDT	2018	80.000 kare	~850.000	3	1080×540	UAV
Stanford Drone	2016	60+ video	>11.000	6	Yüksek	UAV
AU-AIR	2019	32.823	>132.000	8	1920×1080	UAV
DOTA	2018	2.806	>188.000	15	800–13000px	Uydu/UAV
DroneVehicle	2022	56.878 çift	~980.000	5	RGB+Termal	UAV

5.2 Değerlendirme Metrikleri

Nesne tespiti modellerinin performansını ölçmek için çeşitli metrikler kullanılmaktadır. Bu metriklerin doğru yorumlanması, farklı yöntemlerin adil biçimde karşılaştırılması açısından büyük önem taşımaktadır.

Kesinlik (Precision), tespit edilen nesneler arasında gerçekten var olan nesnelerin oranını ifade etmektedir. Geri çağırma (Recall) ise gerçekte var olan tüm nesneler arasında doğru tespit edilenlerin oranını göstermektedir. F1-Score, kesinlik ve geri çağırmanın harmonik ortalaması olarak tanımlanmakta ve tek bir denge ölçütü sunmaktadır.

Ortalama Kesinlik (Average Precision - AP), farklı güven eşiklerinde elde edilen kesinlik-geri çağırma eğrisinin altında kalan alanı ifade etmektedir. mAP (mean Average Precision) ise AP değerlerinin tüm nesne sınıfları üzerinden ortalamasını hesaplamaktadır. mAP@0.5, IoU eşiği 0,5 olarak ayarlandığında elde edilen mAP değerini gösterirken; mAP@0.5:0.95, IoU eşiği 0,5 ile 0,95 arasında farklı değerlerde hesaplanan AP'lerin ortalamasını temsil etmektedir ve daha katı bir değerlendirme ölçütüdür.

Kare başına işlem sayısı (Frames Per Second - FPS), modelin gerçek zamanlı kullanıma uygunluğunu ölçmektedir. Drone uygulamaları genellikle en az 25-30 FPS gerektirirken, gerçek zamanlı uygulamalarda 50+ FPS tercih edilmektedir. Parametre sayısı (milyon cinsinden) ve GFLOPS (giga kayan nokta işlemleri/saniye), modelin hesaplama karmaşıklığını ve bellek gereksinimini nitelendirmektedir.

6. DENEYSEL SONUÇLAR VE KARŞILAŞTIRMA

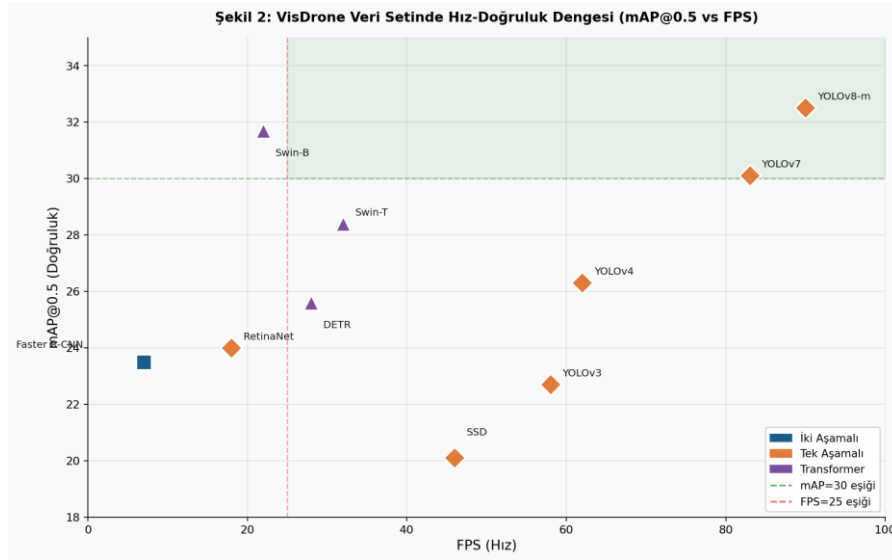
6.1 Yöntemlerin Karşılaştırmalı Analizi

Bu bölümde, literatürde raporlanan sonuçlar esas alınarak farklı derin öğrenme yöntemlerinin VisDrone ve UAVDT veri setlerindeki karşılaştırmalı performansları sunulmaktadır. Sonuçlar, orijinal çalışmalarda veya kapsamlı karşılaştırma çalışmalarında bildirilen değerleri yansıtmaktadır.

Tablo 2 VisDrone Veri Seti Üzerinde Yöntem Karşılaştırması

Yöntem	Kategori	mAP@0.5	mAP@0.5:0.95	FPS	Param (M)	Omurga
Faster R-CNN	İki Aşamalı	23,5	12,3	7	41,4	ResNet-50
SSD	Tek Aşamalı	20,1	9,8	46	24,1	VGG-16
RetinaNet	Tek Aşamalı	24,0	12,7	18	36,4	ResNet-50+FPN
YOLOv3	Tek Aşamalı	22,7	11,1	58	61,9	Darknet-53
YOLOv4	Tek Aşamalı	26,3	14,2	62	64,4	CSPDarknet-53
YOLOv7	Tek Aşamalı	30,1	16,8	83	36,9	E-ELAN
YOLOv8-m	Tek Aşamalı	32,5	18,1	90	25,9	CSPDarknet v2
YOLOv9-c	Tek Aşamalı	34,1	19,3	78	25,3	GELAN
YOLOv10-m	Tek Aşamalı	35,2	20,1	85	15,4	CSP-GELAN
YOLO11-m	Tek Aşamalı	36,8	21,2	88	20,1	C3k2
DETR	Transformer	25,6	13,9	28	41,3	ResNet-50+Transformer
Swin-T	Transformer	28,4	15,3	32	28,3	Swin Transformer-T
Swin-B	Transformer	31,7	17,6	22	87,8	Swin Transformer-B

Tablo 2 incelendiğinde, YOLO11-m'nin 36,8 mAP@0.5 ve 88 FPS ile hız-doğruluk dengesi açısından serinin zirvesine yerleştiği görülmektedir. YOLOv10-m, yalnızca 15,4 milyon parametre ile en verimli model konumundadır. Transformer tabanlı Swin-B, 31,7 mAP@0.5 değeriyle yüksek doğruluk sunmakta; ancak yalnızca 22 FPS ve 87,8 milyon parametre gerektirerek hesaplama yükü açısından dezavantajlıdır. Faster R-CNN, düşük hızına (7 FPS) karşın rekabetçi doğruluk sunmaktadır.



Şekil 6.1 VisDrone Veri Setinde Hız-Doğruluk Dengesi (mAP@0.5 vs FPS)

Tablo 3 UAVDT Veri Seti Üzerinde Araç Tespiti Karşılaştırması

Yöntem	Kategori	AP (Araç)	AP50	FPS	Donanım
Faster R-CNN	İki Aşamalı	22,3	34,7	9	GTX 1080 Ti
SSD300	Tek Aşamalı	18,5	28,4	52	GTX 1080 Ti
RetinaNet	Tek Aşamalı	23,8	36,1	20	GTX 1080 Ti
YOLOv4	Tek Aşamalı	27,1	42,5	65	RTX 3090
YOLOv7	Tek Aşamalı	31,4	48,2	85	RTX 3090
YOLOv8-m	Tek Aşamalı	34,6	52,1	92	RTX 3090
EfficientDet-D3	Tek Aşamalı	29,7	45,3	38	RTX 3090
DETR	Transformer	26,4	41,0	30	RTX 3090

UAVDT veri setinde araç tespiti sonuçlarına göre YOLOv8-m, AP50 metriğinde 52,1 ve 92 FPS ile en iyi performansı sergilemektedir. EfficientDet-D3 ise yüksek doğruluğu (45,3 AP50) ve orta hızı (38 FPS) ile gömülü sistemler için cazip bir seçenek olarak öne çıkmaktadır.

Tablo 4 Yöntem Kategorileri Genel Karşılaştırması

Kategori	Yöntemler	Doğruluk	Hız (FPS)	Küçük Nesne	Bellek	Gerçek Zamanlı
İki Aşamalı	Faster R-CNN, Mask R-CNN	Yüksek	Düşük (5-15)	Orta	Yüksek	Hayır
Tek Aşamalı	YOLO serisi, SSD, RetinaNet	Orta-Yüksek	Yüksek (45-100)	Orta-İyi	Orta	Evet
Transformer	DETR, Swin-T, Swin-B	Yüksek	Orta (20-35)	İyi	Çok Yüksek	Kısıtlı
Hibrit/Özel	EfficientDet, PANet	Yüksek	Orta (30-60)	İyi	Orta-Yüksek	Kısıtlı

6.2 Güçlü ve Zayıf Yönler

Karşılaştırmalı analizden elde edilen temel bulgular şu şekilde özetlenebilir: İki aşamalı dedektörler genel olarak yüksek doğruluk sunmakta, ancak hız açısından dezavantajlıdır. Tek aşamalı dedektörler, özellikle YOLOv9 [16], YOLOv10 [17] ve YOLO11 [18] ile günümüzde hem hız hem doğruluk açısından tatmin edici sonuçlar vermektedir. Transformer tabanlı yöntemler, özellikle Swin-B [25], yüksek doğruluk sağlamakta; ancak hesaplama maliyeti diğer yaklaşımların üzerindedir.

Küçük nesne tespitinde FPN [20] ile desteklenen yöntemler belirgin avantaj sunmaktadır. Gerçek zamanlı uygulamalar için YOLO serisi, hesaplama kısıtlı senaryolar için ise hafif omurgalarla birleştirilen tek aşamalı yöntemler tercih edilmelidir.

7. TARTIŞMA

7.1 Yöntemlerin Gerçek Dünya Uygulanabilirliği

Drone tabanlı nesne tespiti gerçekte dünya uygulamalarında benimsenmesi, laboratuvar koşullarındaki başarıdan çok daha karmaşık gereksinimler doğurmaktadır. Gerçek bir drone sisteminde nesne tespit modeli, sınırlı hesaplama gücüne sahip gömülü bir platform üzerinde çalışmak zorundadır. NVIDIA Jetson serisindeki gibi GPU destekli kenar hesaplama (edge computing) donanımları, orta ölçekli modellerin gerçek zamanlı çalıştırılmasına imkân tanısa da büyük ölçekli Transformer modelleri için bu donanımlar yeterli kalmamaktadır.

Model sıkıştırma teknikleri (quantization, pruning, knowledge distillation) bu açığı kapatmada kritik bir rol üstlenmektedir. YOLOv8-nano gibi kompakt modeller, yeterli doğruluğu korurken belirgin hız kazanımları sağlamaktadır. Öte yandan farklı senaryolara yönelik bağlamsal uyarlama (domain adaptation) kapasitesi, modellerin çeşitli ortamlarda güvenilir biçimde çalışabilmesi açısından önem taşımaktadır. Sabit bir ortamda eğitilen model, farklı iklim koşulları veya kentsel/kırsal ortam değişimine uğradığında performansında belirgin düşüşler yaşanabilmektedir.

7.2 Hesaplama Maliyeti ve Doğruluk Dengesi

Drone uygulamalarında hesaplama maliyeti ile tespit doğruluğu arasındaki denge, tasarım sürecinin belirleyici bir eksenini oluşturmaktadır. Görev kritikliğine göre bu dengenin farklı biçimlerde kurulması gerekmektedir. Örneğin arama-kurtarma operasyonlarında kaçırılan nesne sayısını (false negative) minimize etmek hayati öneme sahipken, trafik izleme uygulamalarında yüksek hız ve düşük gecikme ön planda olabilmektedir.

Literatürde bu dengeyi ele almak için çeşitli stratejiler önerilmektedir: adaptif çözünürlük işleme (görüntünün yalnızca ilgili bölgelerini yüksek çözünürlükte işleme), birden fazla modelin kaskad biçimde kullanılması ve görev bağlamına duyarlı model seçimi. Heterojen hesaplama mimarileri (CPU+GPU+NPU) de bu dengenin optimize edilmesine katkı sağlamaktadır.

7.3 Etik ve Gizlilik Boyutu

Drone tabanlı görüntüleme ve nesne tespiti sistemlerinin yaygınlaşması, önemli etik ve hukuki sorular doğurmaktadır. Kalabalık alanlarda yüzlerce bireyin kimliğini tespit edebilen sistemler, mahremiyet hakkı açısından ciddi kaygılar uyandırmaktadır. Özellikle yüz tanıma veya kişi yeniden tanımlama (person re-identification) uygulamalarıyla birleştiğinde drone gözetimi, bireylerin onayı olmaksızın kişisel verilerin toplanmasına yol açabilmektedir.

Bu bağlamda veri anonimleştirme teknikleri ve gizlilik koruyan öğrenme yöntemleri (federated learning, differential privacy) araştırma gündemine giderek daha fazla girmektedir. Hukuki çerçeve açısından ise farklı ülkelerin ve bölgelerin drone operasyonlarına ilişkin ayrı yasal düzenlemelere sahip olduğu görülmektedir. Araştırmacıların bu konularda sorumluluk bilinciyle hareket etmesi ve geliştirilen sistemlerin kullanım kısıtlamalarını açıkça tanımlaması büyük önem taşımaktadır.

8. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, drone görüntülerinde nesne tespiti için kullanılan derin öğrenme yöntemleri kapsamlı biçimde incelenmiştir. Geleneksel görüntü işleme yöntemlerinden başlayarak iki aşamalı dedektörler, tek aşamalı dedektörler ve transformer tabanlı mimariler sistematik olarak ele alınmış; bu yöntemlerin temel benchmark veri setleri üzerindeki karşılaştırmalı performansları tablolar hâlinde sunulmuştur.

Yapılan inceleme, aşağıdaki temel sonuçları ortaya koymaktadır: YOLO11 [18] serisi, günümüzde drone uygulamaları için en iyi hız-doğruluk dengesini sunan yöntemler arasında öne çıkmaktadır. YOLOv9 [16] ve YOLOv10 [17], sırasıyla PGI mekanizması ve NMS-bağımsız çıkarım stratejileriyle bu alandaki teknik ilerlemenin devam ettiğini göstermektedir.

Transformer tabanlı modeller (özellikle Swin Transformer [25]), doğruluk açısından rekabetçi sonuçlar sağlamakla birlikte hesaplama maliyeti gerçek zamanlı gömülü uygulamalar için hâlâ aşılması gereken bir engel oluşturmaktadır. Küçük nesne tespiti, drone görüntüleme senaryolarında en kritik açık araştırma problemi olmayı sürdürmekte ve FPN [20], dikkat mekanizmaları ile süper çözünürlük teknikleri en umut verici yaklaşımlar olarak öne çıkmaktadır.

Gelecekteki araştırma yönlerine ilişkin önerilerimiz şu şekilde sıralanabilir: Birincisi, hafif ve gömülebilir model mimarilerinin geliştirilmesi. Kenar hesaplama cihazlarında gerçek zamanlı çalışabilecek, bilgi damıtma ve nöron budama gibi sıkıştırma tekniklerinden yararlanan kompakt modeller, saha uygulamaları için kritik önem taşımaktadır. İkincisi, çok modlu veri füzyonunun yaygınlaştırılması. Termal, LiDAR ve RGB görüntülerin birleştirilmesi, gece görüşü ve düşük görünürlük koşullarındaki tespit doğruluğunu artırabilmektedir. Üçüncüsü, zayıf denetimli ve yarı denetimli öğrenme yaklaşımlarının araştırılmasıdır. Etiketleme maliyetini azaltacak bu yöntemler, veri kıtlığı sorununu aşmada belirleyici bir rol üstlenebilir. Dördüncüsü, etki alanı uyarlaması (domain adaptation) ve sürekli öğrenme (continual learning) tekniklerinin geliştirilmesidir. Farklı çevre koşullarına hızlı uyum sağlayabilen modeller, gerçek dünya senaryolarında güvenilirliği artıracaktır. Son olarak, etik çerçeveler ve gizlilik koruyan sistemlerin araştırılması, bu alandaki teknolojik gelişmelerin toplumsal sorumlulukla buluşmasını sağlayacaktır.

Drone görüntülerinde nesne tespiti, yapay zeka ve bilgisayarlı görme alanlarının en dinamik araştırma konularından biri olmaya devam etmektedir. Bu çalışmanın, konuyla ilgilenen araştırmacılar ve uygulayıcılar için kapsamlı bir başvuru kaynağı işlevi göreceği umulmaktadır.

KAYNAKÇA

- [1] R. Girshick, J. Donahue, T. Darrell, ve J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation”, içinde *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA: IEEE, Haz. 2014, ss. 580-587. doi: 10.1109/CVPR.2014.81.
- [2] N. Dalal ve B. Triggs, “Histograms of Oriented Gradients for Human Detection”, içinde *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, San Diego, CA, USA: IEEE, 2005, ss. 886-893. doi: 10.1109/CVPR.2005.177.
- [3] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, c. 60, sy. 2, ss. 91-110, Kas. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
- [4] H. Bay, T. Tuytelaars, ve L. Van Gool, “SURF: Speeded Up Robust Features”, içinde *Computer Vision – ECCV 2006*, c. 3951, A. Leonardis, H. Bischof, ve A. Pinz, Ed., içinde *Lecture Notes in Computer Science*, vol. 3951. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, ss. 404-417. doi: 10.1007/11744023_32.
- [5] P. Viola ve M. Jones, “Rapid object detection using a boosted cascade of simple features”, içinde *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA: IEEE Comput. Soc, 2001, s. I-511-I-518. doi: 10.1109/CVPR.2001.990517.
- [6] R. Girshick, “Fast R-CNN”, içinde *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, Ara. 2015, ss. 1440-1448. doi: 10.1109/ICCV.2015.169.
- [7] S. Ren, K. He, R. Girshick, ve J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, *IEEE Trans. Pattern Anal. Mach. Intell.*, c. 39, sy. 6, ss. 1137-1149, Haz. 2017, doi: 10.1109/TPAMI.2016.2577031.
- [8] K. He, G. Gkioxari, P. Dollar, ve R. Girshick, “Mask R-CNN”, içinde *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Eki. 2017, ss. 2980-2988. doi: 10.1109/ICCV.2017.322.
- [9] W. Liu vd., “SSD: Single Shot MultiBox Detector”, içinde *Computer Vision – ECCV 2016*, c. 9905, B. Leibe, J. Matas, N. Sebe, ve M. Welling, Ed., içinde *Lecture Notes in Computer Science*, vol. 9905. , Cham: Springer International Publishing, 2016, ss. 21-37. doi: 10.1007/978-3-319-46448-0_2.
- [10] J. Redmon, S. Divvala, R. Girshick, ve A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, içinde *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Haz. 2016, ss. 779-788. doi: 10.1109/CVPR.2016.91.
- [11] J. Redmon ve A. Farhadi, “YOLO9000: Better, Faster, Stronger”, içinde *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Tem. 2017, ss. 6517-6525. doi: 10.1109/CVPR.2017.690.
- [12] J. Redmon ve A. Farhadi, “YOLOv3: An Incremental Improvement”, 08 Nisan 2018, *arXiv*: arXiv:1804.02767. doi: 10.48550/arXiv.1804.02767.
- [13] A. Bochkovskiy, C.-Y. Wang, ve H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, 23 Nisan 2020, *arXiv*: arXiv:2004.10934. doi: 10.48550/arXiv.2004.10934.
- [14] C.-Y. Wang, A. Bochkovskiy, ve H.-Y. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors”, içinde *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada: IEEE, Haz. 2023, ss. 7464-7475. doi: 10.1109/CVPR52729.2023.00721.

- [15] Glenn Jocher, *Ultralytics YOLOv8*. (2023). [Çevrimiçi]. Erişim adresi: <https://github.com/ultralytics/ultralytics>
- [16] C.-Y. Wang, I.-H. Yeh, ve H.-Y. Mark Liao, “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information”, içinde *Computer Vision – ECCV 2024*, c. 15089, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, ve G. Varol, Ed., içinde Lecture Notes in Computer Science, vol. 15089. , Cham: Springer Nature Switzerland, 2025, ss. 1-21. doi: 10.1007/978-3-031-72751-1_1.
- [17] A. Wang vd., “YOLOv10: Real-Time End-to-End Object Detection”, 2024, *arXiv*. doi: 10.48550/ARXIV.2405.14458.
- [18] R. Khanam ve M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements”, 23 Ekim 2024, *arXiv*: arXiv:2410.17725. doi: 10.48550/arXiv.2410.17725.
- [19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, ve P. Dollar, “Focal Loss for Dense Object Detection”, içinde *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Eki. 2017, ss. 2999-3007. doi: 10.1109/ICCV.2017.324.
- [20] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, ve S. Belongie, “Feature Pyramid Networks for Object Detection”, içinde *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Tem. 2017, ss. 936-944. doi: 10.1109/CVPR.2017.106.
- [21] P. Zhu, L. Wen, X. Bian, H. Ling, ve Q. Hu, “Vision Meets Drones: A Challenge”, 23 Nisan 2018, *arXiv*: arXiv:1804.07437. doi: 10.48550/arXiv.1804.07437.
- [22] D. Du vd., “The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking”, içinde *Computer Vision – ECCV 2018*, c. 11214, V. Ferrari, M. Hebert, C. Sminchisescu, ve Y. Weiss, Ed., içinde Lecture Notes in Computer Science, vol. 11214. , Cham: Springer International Publishing, 2018, ss. 375-391. doi: 10.1007/978-3-030-01249-6_23.
- [23] A. Robicquet, A. Sadeghian, A. Alahi, ve S. Savarese, “Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes”, içinde *Computer Vision – ECCV 2016*, c. 9912, B. Leibe, J. Matas, N. Sebe, ve M. Welling, Ed., içinde Lecture Notes in Computer Science, vol. 9912. , Cham: Springer International Publishing, 2016, ss. 549-565. doi: 10.1007/978-3-319-46484-8_33.
- [24] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, ve S. Zagoruyko, “End-to-End Object Detection with Transformers”, içinde *Computer Vision – ECCV 2020*, c. 12346, A. Vedaldi, H. Bischof, T. Brox, ve J.-M. Frahm, Ed., içinde Lecture Notes in Computer Science, vol. 12346. , Cham: Springer International Publishing, 2020, ss. 213-229. doi: 10.1007/978-3-030-58452-8_13.
- [25] Z. Liu vd., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows”, içinde *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Eki. 2021, ss. 9992-10002. doi: 10.1109/ICCV48922.2021.00986.
- [26] S. Liu, L. Qi, H. Qin, J. Shi, ve J. Jia, “Path Aggregation Network for Instance Segmentation”, içinde *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Haz. 2018, ss. 8759-8768. doi: 10.1109/CVPR.2018.00913.
- [27] M. Tan, R. Pang, ve Q. V. Le, “EfficientDet: Scalable and Efficient Object Detection”, içinde *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Haz. 2020, ss. 10778-10787. doi: 10.1109/CVPR42600.2020.01079.