

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**KAMUFLE GÖRÜNTÜLERDEN NESNE ALGILAMA İÇİN DERİN ÖĞRENMEYE DAYALI YENİ
YAKLAŞIM**

DOKTORA TEZİ

Rabeb HENDAOUİ

**EKİM 2022
TRABZON**



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**KAMUFLE GÖRÜNTÜLERDEN NESNE ALGILAMA İÇİN DERİN ÖĞRENMEYE DAYALI YENİ
YAKLAŞIM**

Rabeb HENDAOUI

ORCID : - - -

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
“DOKTOR (BİLGİSAYAR MÜHENDİSLİĞİ)”
Unvanı Verilmesi İçin Kabul Edilen Tezdir.

Tezin Enstitüye Verildiği Tarih : 06 / 10 / 2022

Tezin Savunma Tarihi : 28 / 10 / 2022

Tez Danışmanı : : Prof. Dr. Vasif NABİYEV

ORCID : - - -

Trabzon 2022

ÖNSÖZ

Kamuflaj günlük hayatımızda yaygın bir olgudur ve kabaca doğal kamuflaj ve yapay kamuflaj olarak ikiye ayrılabilir. Kamuflajlı nesneler, boyutları son derece küçük, arka plana oldukça benzeyen veya yoğun bir şekilde gizlenmiş nesneleri ifade eder. Kamuflajlı üniforma giyen askerler veya çayırların arasında saklanan aslanlar gibi bu nesneler çevrelerinde kendilerini ustaca gizleyerek bulunmalarını zorlaştırırlar. Kamuflajlı nesne tespiti, geleneksel nesne segmentasyonundan çok daha karmaşık ve zordur. Bu yöntem, arama ve kurtarma, tür keşfi ve tıbbi görüntü analizi gibi birçok hayati uygulamayı kolaylaştırır. Bu tez çalışmasında, doğal sahnelerden kamufle edilmiş nesneleri tespit eden yeni bir derin tabanlı kamuflajlı nesne tespit modeli önerilmiştir. Yöntem, doğru tespit senaryosuna iyi uyan etkili bir dikkat stratejisi kullanmaktadır.

Doktora çalışmalarım esnasında yardım, destek ve tecrübelerini esirgemeyen danışmanım Prof. Dr. Vasif NABİYEYEV'e çok teşekkür ederim. Ayrıca bu süreçte bilgi ve yardımlarından yararlandığım Doç. Dr. Ali KÜRŞAT ERÜMİT, Fen Bilimleri Enstitüsü Sekreteri Ayşe BAYRAKTAR ve Öğr. Gör. Özlem CANPOLAT'a teşekkürü borç bilirim. Her koşulda beni destekleyen sevgili aileme, anlayış, sabır ve desteklerinden dolayı şükranlarımı sunarım.

TEZ ETİK BEYANNAMESİ

Doktora Tezi olarak sunduğum “Kamufle Edilmiş Görüntülerden Nesne Algılama İçin Derin Öğrenmeye Dayalı Yeni Yaklaşım” başlıklı bu çalışmayı baştan sona kadar danışmanım Prof. Dr. Vasıf NABİYEV’insorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 28/10/2022

Rabeb HENDAOUİ

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET.....	VIII
SUMMARY.....	IX
ŞEKİLLER DİZİNİ.....	X
TABLolar DİZİNİ.....	XIII
SEMBOLLER DİZİ.....	XIV
1. GENEL BİLGİLER.....	1
1.1. Giriş.....	1
1.2. Tez Katkısı.....	4
1.3. Tezin Kapsamı.....	5
1.4. Tezin Etkisi ve Potansiyel Uygulama Alanları.....	6
1.5. Tezin Genel Yapısı.....	6
1.6. Literatür Taraması.....	7
1.6.1. Belirgin Nesne Segmentasyonu.....	7
1.6.2. Kamufle Nesne Segmentasyonu.....	10
1.6.2.1. Düşük Seviyeli Özellikler.....	10
1.6.2.1.1. Doku Özellikleri.....	10
1.6.2.1.2. Gradyan Özellikleri.....	11
1.6.2.2. Derin Özellikler.....	12
1.7. Derin Öğrenme Algısı.....	13
1.8. Sinir Ağı Yapısı.....	15
1.8.1. Yapay Sinir Ağı.....	15
1.8.2. Evrişimli Sinir Ağları.....	18
1.9. CNN Mimarisi.....	19
1.10. CNN'deki Katmanlar.....	21
1.10.1. Evrişimsel Katman.....	21
1.10.2. Evrişim Türleri.....	22
1.10.2.1. Ters Evrişim.....	22

1.10.2.2.	Genişletilmiş Evrişim	22
1.10.2.3.	Deforme Edilebilir Evrişim	23
1.10.3.	Havuzlama Katmanı	25
1.10.4.	Tam Bağlantılı Katman	25
1.11.	Evrişimsel Sinir Ağı Eğitimi	26
1.11.1.	Kayıp Fonksiyonu	27
1.11.2.	Optimizatörler.....	27
1.11.2.1.	Stokastik Gradyan İniş	28
1.11.2.2.	Adaptif Optimizatörler	30
1.11.2.3.	Geri Yayılım	30
1.12.	Evrişimsel Sinir Ağlarında Düzenleştirme	31
1.12.1.	Ağırlık Düzenlemesi.....	31
1.12.2.	Bırakma	32
1.12.3.	Toplu Normalleştirme	32
1.12.4.	Veri Artırma.....	33
1.13.	Omurga Mimarileri	33
1.14.	Özellik Piramidi Ağ Mimarileri	38
1.15.	Veri Setleri	42
1.15.1.	COD İçin Veri Setleri.....	42
1.15.1.1.	CPD1 Veri Seti.....	42
1.15.1.2.	CPD2 Veri Seti 2.....	44
1.15.1.3.	CAMO Veri Seti	44
1.15.1.4.	CHAMELEON Veri Seti	44
1.15.1.5.	COD10K Veri Seti.....	45
1.15.2.	SOD İçin Veri Seti	45
1.15.3.	Tıbbi Segmentasyon İçin Veri setleri	46
1.16.	Performans Ölçütleri.....	47
1.16.1.	Kesinlik- Duyarlılık	47
1.16.2.	Ortalama IoU	48
1.16.3.	AUC	48
1.16.4.	F-Ölçüsü	49
1.16.5.	S-Ölçüsü	49
1.16.6.	Ortalama Mutlak Hata (MAE)	49
1.16.7.	E-Ölçüsü.....	50

2.	YAPILAN ÇALIŞMALAR.....	51
2.1.	Formalizasyon	51
2.2.	Önerilen Yöntem.....	53
2.3.	Aşağıdan Yukarıya Yöntemiyle Özellik Çıkarma.....	55
2.4.	Özellik Geliştirme	59
2.5.	Yukarıdan Aşağıya Çok Düzeyli Özellik Füzyonu	61
2.6.	Dikkat Özelliği	64
2.6.1.	Piramit Havuzlama Modülü.....	66
2.6.2.	Dönüşüm Dikkat Modülü	67
2.6.3.	Kanal Dikkat Modülü.....	69
2.6.4.	Uzamsal Dikkat Modülü	70
3.	BULGULAR VE İRDELEME.....	72
3.1.	Eğitim Verileri Detayları	72
3.2.	Değerlendirme	75
3.3.	Önerilen Yöntemin Güncel Yöntemlerle Nicel Karşılaştırması	77
3.3.1.	Kamuflajlı Nesne Algılama	78
3.3.2.	Kamuflajsız Nesne Algılama.....	86
3.3.3.	Tıbbi Görüntü Segmentasyonu.....	87
3.4.	Önerilen Yöntemin Güncel Yöntemlerle Nitel Olarak Karşılaştırılması.....	90
3.5.	Önerilen Yöntemin En Yeni Yöntemlerle Zaman Karşılaştırması.....	95
3.6.	COD Veri Setleri Genel Performans Değerlendirmesi	96
4.	SONUÇLAR.....	101
5.	ÖNERİLER.....	103
6.	KAYNAKLAR.....	105
7.	EKLER	117
ÖZGEÇMİŞ		

Doktora Tezi

ÖZET

KAMUFLE EDİLMİŞ GÖRÜNTÜLERDEN NESNE ALGILAMA İÇİN DERİN ÖĞRENMEYE DAYALI YENİ YAKLAŞIM

Rabeb HENDAOUİ

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Vasif NABİYEV
2022, 116 Sayfa, 4 Sayfa Ek

"Kamuflej nesnesi algılama" (COD) terimi genellikle bir ortamdaki gizli nesnelerin algılanmasını tanımlamak için kullanılır. Kamufleli nesnelerin tespitinde nesneler ve bulundukları ortamlar çok benzer olduğundan ve geniş veri setlerinin bulunmaması ilgili çalışmaların yapılmasını zorlaştırmaktadır. Tezde, derin öğrenmeye dayalı yeni bir COD yaklaşımı önerilmiştir. İlk aşamada, özellikler bir omurga ağından (ResNet) elde edilir. Bu özellikler daha sonra, geniş bir alıcı alandan zengin bağlam özelliklerini çıkaran ve özellik haritaları arasındaki entegrasyonu geliştiren başlangıç modülüne aktarılır. Önerilen modülün, segmentasyonu önemli ölçüde iyileştirdiği gözlemlenmiştir. İkinci adımda, düşük seviyeli öznitelikler anlamsal bilgi açısından zengin öznitelikler ile yoğun bir şekilde birleştirilir. Üçüncü adımda, kamufle edilen nesneyi arka plandan daha iyi ayırt etmek için bir dikkat mekanizması kullanılmaktadır. Dönüşüm dikkati, kanal dikkati ve uzamsal dikkat modülleri, hedef bölge hakkında daha fazla bilgi çıkarmak ve gereksiz bilgileri bastırmak için kullanılır. Modüller üzerinde yapılan deneyler, tüm kombinasyonların kamufle edilmiş nesneleri doğru bir şekilde tespit ettiğini ortaya koymuştur. Ayrıca önerilen yöntem yaygın olarak kullanılan veri setleri üzerinde test edilmiş ve performansı değerlendirilmiştir. Buna ek olarak, önerilen model tıbbi görüntü bölütlemesi için de değerlendirilmiş ve kabul edilebilir sonuçlarla kamufle edilmemiş nesne tespiti elde edilmiştir.

Anahtar Kelimeler: Kamufleli nesne tespiti, Kamuflej desenleri, Derin öğrenme, Dikkat mekanizması.

PhD. Thesis

SUMMARY

A NEW APPROACH BASED ON DEEP LEARNING FOR OBJECT DETECTION FROM CAMOUFLAGED IMAGES

Rabeb HENDAOU

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Computer Engineering Graduate Program
Supervisor: Prof. Dr. Vasif NABIYEV
2022, 116 Pages, 4 Pages Appendix

The term "camouflage object detection" (COD) is often used to describe the detection of hidden objects in an environment. Since the objects and their environments are very similar and the lack of large datasets in the detection of camouflaged objects makes it difficult to carry out related studies. In this thesis, a new COD approach based on deep learning is proposed. In the first stage, features are extracted from a backbone network (ResNet). Those features are then fed into a proposed inception module, which extracts rich context features from a large receptive field and improves integration between feature maps. It is observed that the proposed module significantly improves the segmentation. In the second step, low-level features are densely combined with features rich in semantic information. In the third step, an attention mechanism is used to better distinguish the camouflaged object from the background. Transform attention, channel attention and spatial attention modules are used to extract more information about the target region and suppress redundant information. Experiments on the modules show that all combinations accurately detect camouflaged objects. Furthermore, the proposed method is tested on widely used datasets and its performance is evaluated. In addition, the proposed model is also evaluated for medical image segmentation and uncamouflaged object detection with acceptable results are obtained.

Key Words: Camouflaged object detection, Camouflage patterns, Deep learning, Attention mechanism

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1.1. Doğal ortamlarda kamufle edilmiş hayvan örnekleri [2]	1
Şekil 1.2. Kamufle edilmiş nesne algılama yöntemleri.....	10
Şekil 1.3. Perceptron yapısı.....	15
Şekil 1.4. Çok katmanlı ANN	16
Şekil 1.5. Tanh, Sigmoid ve ReLU aktivasyon fonksiyonlarının grafiksel gösterimi	17
Şekil 1.6. CNN'nin Bileşenleri.....	19
Şekil 1.7. CNN'in tüm işlemlerini (evrişim, havuzlama, özellik çıkarma ve sınıflandırma) temsil eden CNN mimarisi	20
Şekil 1.8. Giriş özelliği haritasında evrişim işlemi	21
Şekil 1.9. Çeşitli genişleme oranlarıyla genişletilmiş evrişimin uygulanması.....	23
Şekil 1.10. Evrişimli bir 3×3 filtrede uzamsal örnekleme konumlarının gösterimi a) normal bir evrişimli filtrenin örnekleme konumları b) bir deforme edilebilir filtre	23
Şekil 1.11 Deforme edilebilir evrişim yapısı [76].....	24
Şekil 1.12. Maksimum Havuzlama ve Ortalama Havuzlama örnek	25
Şekil 1.13. Tam bağlantılı katmanın gösterimi	26
Şekil 1.14. LeNet mimarisi	34
Şekil 1.15. AlexNet mimarisi.....	35
Şekil 1.16. VGGNet mimarisi.....	36
Şekil 1.17. ResNet101 mimarisi.....	37
Şekil 1.18. Darboğaz bloğu.....	37
Şekil 1.19. FCN mimarisi.....	38
Şekil 1.20. U-Net mimarisi	39
Şekil 1.21. FPN mimarisi	41
Şekil 1.22. PANet mimarisi	41
Şekil 1.23. NAS-FPN mimarisi.....	42
Şekil 1.24. Kamufle edilmiş nesne algılama veri setlerinin görüntü örnekleri	43

Şekil 1.25.	Kamuflejiz nesne algılama veri setleri örnek gösterimi	45
Şekil 1.26.	Polip segmentasyonu veri setleri örnek gösterimi	46
Şekil 2.1.	Çalışmanın genel yapısı	51
Şekil 2.2.	a) IM ve b) MAM modülleri.....	53
Şekil 2.3.	Önerilen yöntemin genel mimarisi.....	54
Şekil 2.4.	ResNet mimarisinin artık fonksiyonu	57
Şekil 2.5.	Artık öğrenme bloğunun ve daha verimli “Darboğaz” versiyonunun karşılaştırılması	57
Şekil 2.6.	ResNet101'de özellik çıkarma	58
Şekil 2.7.	10. katmandaki kamufle edilmiş görüntüden çıkarılan özellikler örneği	59
Şekil 2.8.	(a) Basit IM ve (b) Önerilen IM yapısı	60
Şekil 2.9.	Kamufle edilmiş görüntü sonuçlarına örnekler (a) yoğun bağlantı olmadan özellik piramidi (b) yoğun bağlantı ile özellik piramidi	63
Şekil 2.10.	Yoğun özellik piramit ağı	63
Şekil 2.11.	(a) Kamufle edilmiş görüntü örnekleri ve bunlara karşılık gelen (b) dikkat haritaları	65
Şekil 2.12.	MAM mimarisi	65
Şekil 2.13.	PPM mimarisi [114].....	66
Şekil 2.14.	Standart konvolüsyon ve deforme edilebilir konvolüsyon	68
Şekil 2.15.	Kanal Dikkat Modülü	69
Şekil 3.1.	Önerilen modelin eğitilmesi için eğitim kaybı değerleri; (a) kamufle edilmiş veri setleri (b) kamufle edilmemiş veri setleri (c) polip veri setleri.....	73
Şekil 3.2.	Veri çoğaltma tekniklerine örnekler	75
Şekil 3.3.	CPD1 veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları	79
Şekil 3.4.	CPD2 veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları	81

Şekil 3.5.	CAMO veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları	82
Şekil 3.6.	CHAMELEON veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları	84
Şekil 3.7.	COD10K veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları	85
Şekil 3.8.	Kamufle edilmemiş algılama için nicel karşılaştırmalar (a) Ortalama IoU puanları (b) Kesinlik, Duyarlılık ve F-ölçü puanları	87
Şekil 3.9.	Polip segmentasyonu için çeşitli veri setlerinde nicel karşılaştırmalar.....	89
Şekil 3.10.	CPD1 veri setinde görsel karşılaştırma.....	90
Şekil 3.11.	CPD2 veri setinde görsel karşılaştırma.....	91
Şekil 3.12.	Farklı veri setlerinde kamufle edilmiş nesne tespiti için görsel karşılaştırma	92
Şekil 3.13.	Polip segmentasyonu için görsel karşılaştırma	93
Şekil 3.14.	Kamuflesiz nesne tespiti için görsel karşılaştırma.....	94
Şekil 3.15.	Yanlış tespit durumlarına örnekler.....	95
Şekil 3.16.	(a) Ortalama çalışma süresinin karşılaştırılması, (b) CPD1 veri seti üzerinde ortalama çalışma süresiyle IoU ilişkisi.....	96
Şekil 3.17.	MAE performansı	97
Şekil 3.18.	Ortalama IoU performansı	97
Şekil 3.19.	S-ölçü performansı.....	98
Şekil 3.20.	E-ölçü performansı.....	98
Şekil 7.1.	Tüm testler için S- ölçümü, F- ölçümü, E- ölçümü ve IoU ölçümü puanları	118
Şekil 7.2.	Tüm testler için MAE puanları	118

TABLÖLAR DİZİNİ

	<u>Sayfa No</u>
Tablo 2.1. Omurga ResNet 101'in mimarisi.....	55
Tablo 3.1. Veri setlerinin eğitim ve test verilerine göre dağılımı.....	74
Tablo 3.2. Bileşen analizi	75
Tablo 3.3. Genişletilmiş ve deforme edilebilir konvolüsyonun değerlendirilmesi	77
Tablo 3.4. CPD1 veri seti için karşılaştırılan modellerin nicel sonuçları.....	78
Tablo 3.5. CPD2 veri seti için karşılaştırılan modellerin nicel sonuçları.....	80
Tablo 3.6. CAMO veri seti genelinde karşılaştırılan modellerin nicel sonuçları	82
Tablo 3.7. CHAMELEON veri seti genelinde karşılaştırılan modellerin nicel sonuçları.....	83
Tablo 3.8. COD10K veri seti genelinde karşılaştırılan modellerin nicel sonuçları.....	85
Tablo 3.9. Yaygın olarak kullanılan üç nesne segmentasyonu testi veri setinde nicel sonuçlar	86
Tablo 3.10. Yaygın olarak kullanılan dört polip segmentasyonu testi veri setinde nicel sonuçlar	88
Tablo 3.11. Önerilen yöntemin ortalama çalışma süresi ve CPD1 veri setinde karşılaştırılan modeller.....	95
Tablo 3.12. Şekil 3.19'daki her yöntemin alanı	99
Tablo 3.13. Şekil 3.20'deki her yöntemin alanı	99
Tablo 7.1. CPD1 veri seti için N-kat çapraz doğrulama sonuçları.....	117
Tablo 7.2. Yeni SOD modellerinin önerilen yöntemle değerlendirme sonuçları.....	119
Tablo 7.3. COD10K veri seti için önerilen yöntem ile son COD modellerinin değerlendirme sonuçları.....	120

SEMBOLLER DİZİ

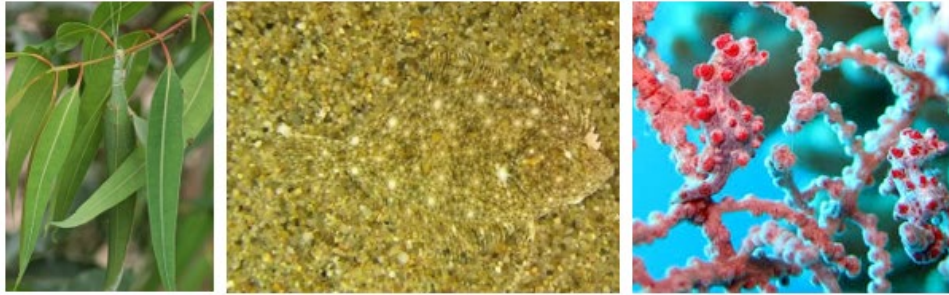
ANN	: Yapay Sinir Ağı (Artificial Neural Network)
AUC	: Eğri Altındaki Alan (Area Under Curve)
BN	: Paket Normalizasyon (Batch Normalization)
CAM	: Kanal Dikkat Modülü (Channel Attention Module)
CNN	: Evrişimsel Sinir Ağları (Convolutional Neural Networks)
COD	: Kamufle Edilmiş Nesne Algılama (Camouflaged Object Detection)
FCN	: Tam Evrişimsel Ağ (Fully Convolutional Network)
FPN	: Özellik Piramidi Ağları (Feature Pyramid Networks)
GPU	: Grafik İşleme Ünitesi (Graphics Processing Unit)
GT	: Gerçek Referans Değer (Ground Truth)
IM	: Başlangıç Modülü (Inception Module)
IOU	: Birleşimlerin Kesişimi (Intersection over Union)
MAE	: Ortalama Mutlak Hata (Mean Absolute Error)
MAM	: Çoklu Dikkat Modülü (Multi-Attention Module)
PPM	: Piramit Havuzlama Modülü (Pyramid Pooling Module)
ReLU	: Doğrultulmuş Doğrusal Birim (Rectified Linear Unit)
SAM	: Uzamsal Dikkat Modülü (Spatial Attention Module)
SGD	: Olasılıksal Gradyan İniş (Stochastic Gradient Descent)
SOD	: Belirgin Nesne Algılama (Salient Object Detection)
SVM	: Destek Vektör Makinesi (Support Vector Machine)
TAM	: Dönüşüm Dikkat Modülü (Transformation Attention Module)

1. GENEL BİLGİLER

1.1. Giriş

Bilgisayarla görme ve görüntü işlemedeki temel problemlerden biri olan görüntü bölütleme: görüntünün piksellerini, aynı küme içindeki piksellerin belirli görsel özellikleri paylaşacak şekilde birden çok kümeye ayırma işlemidir. Semantik web, akıllı video kodlama, mobil robotlar, tıbbi görüntüler ve askeri gözetim gibi çok çeşitli pratik uygulamalar, görüntü segmentasyonundan yararlanır.

Bazen nesneler, Şekil 1.1'de görüldüğü gibi kendilerini çevrelerine gizleyerek kamuflaj oluştururlar. Kamuflajın ortaya çıkması, nesne segmentasyonunu daha karmaşık bir problem haline getirir. Copeland ve Trivedi'ye [1] göre, “kamuflaj, bir hedefin imzasını arka plana gizleme girişimidir”. Başka bir deyişle kamuflaj; avın vücut desenini, dokusunu ve rengini ortamın dokusuna göre değiştirerek avcılardan saklanma yeteneğidir.



Şekil 1.1. Doğal ortamlarda kamuflaj edilmiş hayvan örnekleri [2]

Kamuflaj edilmiş bir nesne, insan görme sistemleri tarafından yeterince görülemez. Bu bağlamda, kamuflaj edilmiş nesneleri analiz etmek için bilgisayar görüşüne dayalı yaklaşımlar önerilmektedir. Kamuflajla ilgili çalışmalar kabaca iki ana alana ayrılabilir: (i) kamuflaj tasarımı ve değerlendirmesi, (ii) kamuflaj nesnesi algılama (COD). Kamuflaj tasarım teknikleri [3-6] yapay olarak kamuflaj görüntüleri oluşturur. Kamuflaj değerlendirme algoritmaları [7-10] ise kamuflaj tasarım tekniklerinin etkinliğini değerlendirir. Bir hedefi arka planından çıkarmak için bir COD kullanılır [11, 12].

Kamuflaj, doğada yaygın olarak bulunan biyolojik bir olgudur. Biyologlar, doğadaki canlıların genellikle kendi yapılarını ve fizyolojik özelliklerini kullanarak kendilerini yırtıcılardan gizlediklerini bulmuşlardır. Bu tür nesneler, aynı vücut renklerini, desenlerini ve ayrıca arka planın diğer morfolojik görünümünü, örneğin göz kamaştırıcı kamuflajı, bozucu renklendirme ve dikkat dağıtıcı işaretler gibi taklit etme yeteneği ile çevrelerinde kendilerini saklama veya gizleme konusunda büyük ustalardır. Örneğin, bir bukalemun çevre değiştiğinde vücudunun rengini değiştirebilir; bir yengeç genellikle görünümüne benzer bir yaşam alanı bulur. Bu hayvanlar, saldırıya uğramamak, iletişim kurmak ve kur yapmaktan kaçınmak gibi hayatta kalma amaçları için kendilerini kamufle ederler.

Kamuflajla ilgili en eski araştırmaların izleri geçen yüzyıla kadar götürülebilir. Araştırmacılar [13] 1918'de kamuflaj fenomenini sistematik olarak incelediler. Yüz yıl geçse de ve biyologlar bu önemli doğal olgudan inceleme tutkusunu sürdürüyorlar. Çok sayıda biyolog, kamuflajın tipik örneklerini araştırdı ve doğal seçim ve adaptasyon ilkelerini kullanarak kamuflajı açıkladı. Bu önemli doğal fenomenden ilham alan insanlar, sanat, tarım vb. birçok alanda bu modelleri taklit etmeye çalıştılar. Bilgi teknolojilerinin hızlı gelişimi ile COD, bilgisayarlı görü vizyonu topluluğundan artan araştırma çabalarını çekmiştir.

Önemli bilimsel ve pratik değeri nedeniyle, kamufle edilmiş nesneleri doğal sahnelerden algılamak/bölümlere ayırmak için önemli çabalar sarf edilmiştir, bu da etkili kamufle edilmiş nesne algılama yaklaşımlarının taleplerini arttırmaktadır. Bununla birlikte, bilgisayar görüşündeki geleneksel nesne algılama/segmentasyon görevleriyle karşılaştırıldığında, COD aşağıdaki nedenlerden dolayı önemli zorluklar göstermektedir: (i) nesne boyutlarında yüksek çeşitlilik. (ii) belirsiz nesne sınırları (iii) doku, renk ve görüş özelliklerinin nesneler ve çevreleri arasında benzerlik göstermesi. Kamufle edilmiş nesnelerin bu tür özellikleri algılamayı işlemini daha da zorlaştırır. Bu nedenle hedef nesne ile arka plan arasındaki yüksek içsel benzerlikler, COD'yi geleneksel nesne algılama görevinden çok daha zor hale getirir.

İlk çalışmalarda, COD genellikle kamufle edilmiş nesneyi ve arka planı [6-8] ayırmak için kenarlar, parlaklık, köşe noktaları, doku veya zamansal bilgiler [5] bilgiler için hesaplanan özellikleri kullanan ön plan algılama olarak ele alındı. Ancak, bu özellikler, gerçek uygulama senaryolarındaki tüm karmaşık kamuflaj stratejilerini tespit edememektedir.

Son zamanlarda, derin sinir ağlarının, özellikle evrişimli sinir ağlarının (CNN'ler) [14] benzeri görülmemiş başarısı, görüntü sınıflandırma, görüntü oluşturma ve genel nesne algılama dahil olmak üzere çeşitli bilgisayar görme alanlarında onların kullanımını yaygınlaştırmıştır. CNN tabanlı nesne algılama modellerinin çok çeşitli olmasına rağmen COD için modeller oluşturmak için özel tasarımlar gereklidir. Nesne algılama için tasarlanmış modeller, kamufle edilmiş nesneleri doğru bir şekilde algılamaya yetkin değildir. Çünkü nesnelerin belirsiz sınırlarını bulma konusunda yetersiz kalıyorlar.

Şu anda, bilim ve teknolojinin ilerlemesiyle, kamuflaj kıyafetleri ve kamuflaj ağları gibi hayvan kamuflajını simüle eden kamuflaj teknolojisi, yüksek teknoloji savaşlarında yaygın olarak kullanılmaktadır. Kamuflaj teknolojisinin kullanımıyla, sistemlerini nişancılar bir düşmanın kıdemli generallerini pusuya düşürebilir ve zırhlı araçlar, görünür dalga boylarında düşman sis aldatabilir. Bu nedenle, askeri alanlarda doğru kamufle edilmiş nesne segmentasyonu araştırmaları büyük önem taşımaktadır. COD, askeri değerine ek olarak endüstriyel algılama (örneğin ekipman kusuru tespiti), tıbbi teşhisler (örneğin akciğerlere pnömoni bulaşıp bulaşmadığının test edilmesi), izleme ve koruma (örneğin şüpheli kişi veya insansız hava aracının izinsiz giriş tespiti) ve insansız sürüş (örneğin yol engeli algılama) için uygulanabilir.

Genel olarak, askeri uygulamalarda, askerlerin ve teçhizatlarının düşman tarafından tespit edilmesini gizlemek için kamuflaj boya desenleri kullanılır, bu da onları çevreleyen doğal çevre koşullarının bir parçası ve gibi görünmesini sağlar. Geleceğin akıllı savaşı, yapay zekâ tarafından temsil edilen yaratıcı teknolojiye dayalı yeni nesil savaşı ifade eder. Tipik olanlardan biri, hedeflerin tespiti ve imhası için otonom insansız sistemlerin kullanılmasıdır [15]. Bununla birlikte, mükemmel kamuflaj teknolojisi genellikle askeri teçhizatın ve silahlı personelin çevreye iyi uyum sağlamasına izin verir. Kamufle edilen insansız hava aracının tespiti büyük zorluklar oluşturmaktadır [16].

Son yıllarda farklı türde hedef tespit algoritmaları ortaya çıkmış ve belirli ilerlemeler kaydetmiştir. Ancak kamufle edilmiş nesne segmentasyonu yaklaşımları yetersiz olduğundan hedef tespitine çok az ilgi gösterilmiştir. Örneğin askeri alanlarda, askeri kamufle nesneler genellikle kızılötesi, polarizasyon ve hiperspektral tabanlı görüntüleme ve diğer teknolojiler aracılığıyla tanımlanır. Bununla birlikte, kamufle edilmiş nesnelerin görünür ışık bandında nasıl doğru bir şekilde segmentlere ayrılacağına dair bilimsel araştırmalar göz ardı edilmiştir.

Bu tezde, askeri kamufle edilmiş kişileri tespit etme görevi ve doğal ortamlardaki kamufle edilmiş nesneleri tespit etme görevi için COD sorunu incelenmiştir.

Kamufle edilmiş kişilerin/nesnelerin tespiti günümüzde zorluğunu korumaktadır aşağıdaki nedenlerden dolayı:

- Askeri kamufle edilmiş insanların ve hedeflerin dokusu, boyutu ve konumu büyük ölçüde değişir ve bu da tespit görevini zorlaştırır. Örneğin, kar, çöl ve ormanlık alanlardaki hedeflerin uç özellikleri çevreye kolayca karışabilir.
- COD, ön plan ve arka plan genellikle bir arka plan karmaşasını, daha küçük nesne boyutunu, nesne kapanmasını, karmaşık arka planı, şekil karmaşıklığını ve dikkat dağınıklığını paylaşır.

Bu tezde hedef nesnelerin öznelik bilgilerini çıkarmak ve geliştirmek için Başlangıç Modülü (IM) ve Çoklu Dikkat Modülü (MAM) olmak üzere iki bileşenden oluşan kamufle edilmiş nesne/insan tespiti için yeni bir yaklaşım önerilmiştir: a) Kamufle edilmiş nesnenin arka plandan daha iyi ayrıştırılması, b) kamuflej görüntüsünün aşamalı olarak iyileştirilmesi.

1.2. Tez Katkısı

Tezin katkısı dört maddede özetlenebilir:

- Özellik yayılımını geliştiren, özelliklerin yeniden kullanımını teşvik eden ve küçük kamufle edilmiş nesnelerin tespitinin etkinliğini sağlayan çok ölçekli bir algılama yöntemini dayalı yoğun bağlantılardan oluşan bir özellik piramidi önerilmiştir.
- Farklı alanlardan daha etkili özellikler yakalayarak performansı artırmak için önerilen ağa yeni bir IM uygulanmıştır. Genişletilmiş ve deforme edilebilir kıvrımlar, alıcı alanı genişletmek için kullanılmıştır. Modeldeki her bir elemanın etkinliğini göstermek için önerilen IM'de bir dizi deneyler gerçekleştirilmiştir.
- Gelişmiş özellikleri iyileştirmek ve kamufle edilen nesnenin konumunu tam olarak belirlemek için bir MAM tanıtıldı. MAM'de, ilgilenilen bölge hakkında daha fazla

bilgi yakalamak ve düşük korelasyon özelliklerini bastırmak için bir dönüşüm dikkati, kanal dikkati ve uzamsal dikkat tasarlanmıştır. Onun sonucunda özellik haritalarının daha iyi bir karakterizasyonu elde edilmiştir.

- Önerilen modelin üstünlüğünü doğrulamak için; İlk olarak, yöntem kamufle edilmiş nesne tespiti için beş kıyaslama veri setinde (CPD1, CPD2, CAMO, CHAMELEON ve COD10K) diğer 11 model ile karşılaştırıldı. Önerilen model, altı değerlendirme ölçütüne dayalı olarak en iyi performansı vermektedir. İkinci olarak, önerilen yöntem, nesnenin üç veri seti (DUTS-TE, DUT-OMRON ve ECSSD) içinde kamufle olmaması durumunda test edilmiştir. Önerilen sistem, gelişmiş yöntemlerden daha rekabetçi ve iyi sonuçlar vermiştir. Bilimsel değerine ek olarak, kamufle edilmiş nesne algılama, tıbbi görüntü bölütleme alanlarındaki uygulamalar için de gerçekleştirmiştir. Böyle ki önerilen model, 4 farklı polip segmentasyonu veri setinde en gelişmiş 4 derin tabanlı tıbbi segmentasyon ağı ile karşılaştırılmıştır. Yöntem, niceliksel ve niteliksel olarak rekabetçi sonuçlar ortaya koymaktadır.

1.3. Tezin Kapsamı

Bu tez kamufle edilmiş nesne/insan tespiti üzerine odaklanmaktadır. Tez kapsamında kamufle edilen hedefi arka plandan ayıran görüntülerde kamufle edilen nesnenin tespitine yönelik bir derin öğrenme modelinin tasarımı gerçekleştirilmiştir.

Kamufle edilmiş nesneler boyut olarak son derece küçük, arka plana çok benzeyen veya çok belirsiz nesneler olabilir. Kendilerini çevrede ustaca gizlerler ve bulunmalarını zorlaştırırlar. Bu nedenle kamufle edilmiş nesneleri tespit etmek için aşağıdaki sorular ilgi çekicidir:

- 1) Küçük, yoğun nesneler ve hatta rastgele geometrik dönüşümlere sahip nesneler nasıl açıkça tespit edilir?
- 2) Göze çarpmayan görünüm ve çeşitli ölçekler koşulları altında kamufle edilmiş nesneler doğru bir şekilde nasıl bulunabilir?
- 3) Arka plandan gelen bariz gürültüler nasıl bastırılır ve kamufle edilmiş nesneler nasıl daha güvenilir bir şekilde algılanabilir.

1.4. Tezin Etkisi ve Potansiyel Uygulama Alanları

Hedef nesnelerin arka planla benzer bir görünümü paylaştığı bu uygulamaların ortak özelliği nedeniyle gizli nesne algılama sistemleri tıp, sanat ve tarım gibi alanlarda çeşitli uygulamalara sahiptir. Bu koşullar altında COD modelleri, kamufle edilmiş nesnelerin tespiti için bu uygulamaların temel bir bileşeni olarak hareket etmeye çok uygundur.

- Tıp uygulamaları (polip segmentasyonu)
- Yararsız tarım bölgeleri tespiti, meyve olgunluk tespiti
- Askeri uygulamalar (savaş alanında düşman tespiti)
- Arama ve kurtarma görevleri
- Gözetim uygulamaları
- İmalat yüzey hata tespiti
- Lojistikte sahte ürün üretimi tespiti
- Yaban hayatın korunması

1.5. Tezin Genel Yapısı

Tez beş bölümden oluşmaktadır ve aşağıdaki şekilde düzenlenmiştir:

Bölüm 1 araştırma alanının genel tanıtımı ile ilgilidir. Literatürdeki ilgili çalışmalar gözden geçirilmiş ve bu çalışmalar hakkında genel bilgiler verilmiştir. Ayrıca modeller, eğitim metodolojileri, önerilen modelin etkinliğini test etmek için kullanılan veri setleri bu bölümde açıklanmış ve performans ölçme için kullanılan değerlendirme metriklerine ayrıntılı olarak değinilmiştir.

Bölüm 2'de önerilen model ayrıntılı olarak sunulmaktadır. Ağın yapısı ve eğitim algoritmaları detaylı olarak verilmiştir.

Bölüm 3'te tüm deneylerin detayları, önerilen model bileşenlerinin değerlendirme, önerilen sinir ağı yapılarının en son yöntemlerle karşılaştırılması, zaman değerlendirmesi ve çeşitli COD veri setlerinde performans sonuçları verilmiştir.

Tezin 4. bölümünde, önerilen tüm çalışmaların ana sonuçları yer almaktadır. Ayrıca, öneriler ve gelecekte yapılacak çalışmalara Bölüm 5. te değinilmektedir.

1.6. Literatür Taraması

Belirgin nesne bölütleme, kamufle edilmiş nesne bölütlemeden farklı olsa da birbirini tamamlaması nedeniyle taraflardan birinin teknolojik gelişimi diğerine fayda sağlayabilir.

Kamufle edilmiş nesneler, belirgin nesnenin yalnızca özel durumları olduğundan dolayı, bu bölüm ilk olarak kamufle edilmiş nesne bölütleme ile bazı benzerlikleri paylaşan belirgin nesne bölütleme üzerine mevcut çalışmaları gözden geçirmektedir. Daha sonra kamuflaj nesne bölütlemenin mevcut yöntemleri tartışılmıştır.

1.6.1. Belirgin Nesne Segmentasyonu

Belirgin nesne algılama (SOD) modelleri, insan davranışını taklit eder ve görüntülerden veya sahnelerden en belirgin bölgeyi/nesneyi yakalar, bu alan hem bilgisayarla görme hem de örüntü tanıma görevlerinde birçok önemli uygulamayı içerir.

Bir SOD yöntemi, belirgin bölgelere kesin olarak ulaşarak ve belirgin kaçırarak hata metrik değerlerini en düşük seviyede tutmalıdır. Ayrıca doğru belirgin nesne lokalizasyonu için yüksek çözünürlüklü bir belirginlik haritasının üretilmesinde SOD yönteminin hesaplama açısından ucuz olması beklenmektedir. Son yirmi yılda aktif bir araştırma alanı olarak görüntü tabanlı SOD için minimum gereksinimleri karşılamak adına çok sayıda model denenmiştir.

Belirginlik tespiti için erken çabalar odak tahminine dağılı olmaktadır [17,18]. Odak tahmini, bir gözlemcinin birkaç saniyelik serbest görüntüleme içinde sabitlenebileceği uzamsal konumlara katılmayı amaçlar. SOD, sabitleme tahmininden farklıdır çünkü birincisi için modeller sahnedeki belirgin bölgelerin/nesnelerin tüm kapsamını algılamalı ve bölümlere ayırmalıdır. Bu amacı gerçekleştirmek için geleneksel SOD modelleri tarafından benimsenen genel bir yaklaşım, bir belirginlik haritası üretirken sahnedeki belirgin öğelere yüksek olasılık değerleri atamaktır. Tespit edildikten sonra tüm belirgin nesneyi parçalara ayırmak için eşikleme gibi teknikler kullanılabilir.

Belirginlik tespiti literatürü, buluşsal tabanlı ve öğrenme tabanlı yaklaşımlar olarak sınıflandırılmıştır. Belirginlik tespitinde kontrast, belirgin bölge tespiti için çok önemli bir

faktördür [19,20]. Beyin, bir görüntüdeki yüksek kontrastlı nesnelere/bölgelere karşı çok hassastır. Belirginlik tespitinin geleneksel sezgisel yaklaşımları çoğunlukla düşük seviyeli görsel özelliklere dayanır ve hesaplama yöntemlerinin çoğu denetimsizdir. Bu geleneksel aşağıdan yukarıya yöntemler belirginlik tespiti sırasında buluşsal özellikler yaklaşımını (yani kontrast, konum ve doku gibi) takip eder. Sezgisel özelliklere genellikle görsel öncelikler veya belirginlik tespiti için ipuçları denir [21,22]. Kontrast önceliği çok önemli bir özelliktir ve en çok kullanılan önceliklerden biridir. Somut olarak, kontrast öncelikleri ve global önceliklerden oluşur. Kontrast önceliği, belirgin bölgelerin mahallelerinden veya sahnelerinden her zaman benzer olmadığını varsayar [23]. Kontrast önceliğinin yanı sıra konum öncelikleri merkez öncelikler ve arka plan önceliklerinden oluşur. Merkez öncelikleri, belirgin nesnenin görüntünün ortasında görüldüğünü tanımlarken arka plan öncelikleri, bir görüntünün kenarlığının arka planın parçası olma şansının daha yüksek olduğunu belirtir.

Belirginlik tespiti için geleneksel olarak buluşsal tabanlı yaklaşımlar, içsel düşük seviyeli ipuçlarını kullanır ve denetimsiz tekniklere dayanır. Bu teknikler, özellikle görüntü karmaşık olduğunda ve ortak görsel özellikleri paylaştığında, belirgin hedefleri doğru bir şekilde tespit etmek için bazen yetersizdir. Bu sorunların üstesinden gelmek, karmaşık arka plan görüntüsünde belirgin bir nesne bulmak için eğitim verileriyle öğrenmeye dayalı yöntemler kullanılır.

Klasik öğrenmeye dayalı etkinlik saptama yöntemleri, üst düzey özelliklerin ve denetlenen bilgilerin derecesini artırmak için entegre edildiği, veriye dayalı yaklaşımlar olarak da adlandırılan denetimli veya yarı denetimli öğrenmeye dayalı etkinlik saptama yöntemleridir. Judd ve ark. [24], on beş izleyicinin sabitleme konumlarını içeren eğitim veri setine dayalı bir Destek Vektör Makinesi (SVM) sınıflandırıcısı aracılığıyla göz sabitleme tahmini için bir model önerdi. [25]'de, Liu ve ark. koşullu rastgele alana dayalı bir ikili belirginlik tahmin şeması sunmuşlar. Yang ve ark. [26], belirginlik tespiti için koşullu rastgele alan ve ayırt edici bir sözlükle eğitilen bir yöntem önermektedir. Tasarlanan yöntem, yukarıdan aşağıya devam eden, yapılandırılmış denetim altında eğitilen ve daha sonra verimli öğrenme için bir maksimum marj mekanizması izleyen katmanlı bir yapı içerir. [27]'te, Borji ve ark. göz sabitlemeleri için AdaBoost sınıflandırıcı aracılığıyla doğrudan bir haritalama yaklaşımı eğitimi için yüksek seviyeli görsel özelliklerle (örn., insanlar, yüzler ve arabalar, vb.) düşük seviyeli özellikleri (ör. yön, renk ve yoğunluk) entegre eklemişler.

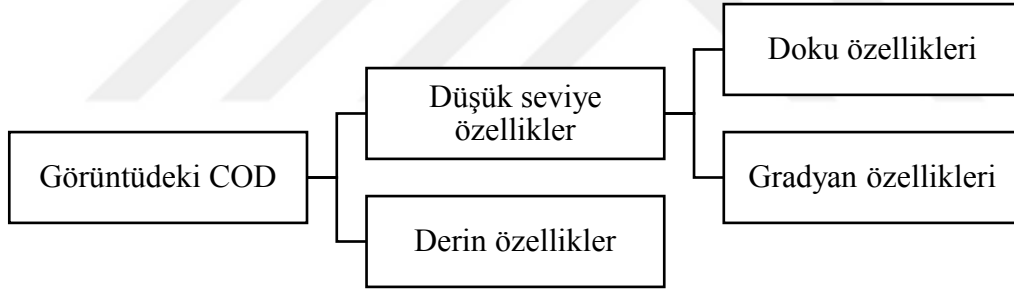
Wang ve ark. [28], belirgin nesne tespiti için düşük seviyeli, orta seviyeli ve yüksek seviyeli özelliklerin entegre edildiği çoklu örneğe dayalı öğrenme yöntemi önermektedirler. Jiang ve ark. [29,30] belirginlik tespiti için regresyon modeli önermekte ve belirginlik değerlerinin üretilmesi amacıyla regresyon orman sınıflandırıcısını eğitmektedirler. Tong ve ark. [31], büyük bir veri setinde yalnızca bir sınıflandırıcıyı eğitmek yerine, önyüklemeye öğrenme tekniği ile belirgin bir nesne algılama modeli ortaya koydu. Ayrıca çok çekirdekli güçlendirme yöntemiyle zayıf sınıflandırıcılardan güçlü bir sınıflandırıcı elde etmek için bir grup zayıf SVM'nin eğitimini gerçekleştirmektedirler. Klasik öğrenmeye dayalı yöntemler ön bilgileri kullandığından ve bazen geleneksel özellik tabanlı belirginlik tespit tekniklerinden daha iyi performans gösterdiğinden bu yöntemler tercih edilmektedir. Klasik öğrenme temelli yaklaşımlarda özenle seçilmediği takdirde bulunan özellikler modellerin performansını düşürebilmektedir.

Ancak, bu alanda derin öğrenmeye dayalı tekniklerin son zamanlardaki hızlı gelişimi yukarıda bahsedilen sorunların çoğunun üstesinden gelmede oldukça başarılı olmuştur. Tam evrimsel sinir ağı (FCN), derin öğrenme tabanlı SOD'nin merkezinde yer alır [32]. FCN'nin güçlü hiyerarşik çok ölçekli özellik gösterimi, kaba bir belirginlik tahmini için çeşitli şekillerde ve veriye dayalı bir şekilde sınır kesinliği haritası için rafine edilmesi sebebiyle kullanılmıştır. Bununla birlikte SOD için geleneksel modeller, gerçek zamanlı performans sağlama avantajına sahiptir ve doğal ortamda uygulanabilir.

Bazı derin modelleri çok katmanlı özelliklerin temsil kabiliyetini geliştirmek ve eğitim sürecini hızlandırmak için belirginlik önceliklerinden faydalanmaktadır. Wang ve ark. [33] belirginlik tespitine rehberlik etmek için belirgin bölgeler hakkında bilgilendirici ön bilgi olarak çoklu geleneksel yöntemlerin belirginlik tahminlerini birleştirdi. Chen ve ark. [34], belirginlik iyileştirmesi için bir ilk tahmin olarak belirginlik önceliklerini kullandılar. Zhang ve ark. [35], birden fazla geleneksel SOD'dan gürültülü denetim ile derin bir denetimsiz belirginlik tespiti sistemi tasarladı. [36]'deki kontrast gibi basit sezgisel operatör, [37,38]'deki çok ölçekli özelliklerin kontrast modellemesi için benimsenmiştir. Bu uyarlamalar, derin öğrenme tabanlı SOD tarafından gösterilen muazzam ilerlemeye ve üstün performansa rağmen geleneksel belirginlik algılama araçlarının derin modellerin performans çitasını daha da yükseltmek için yararlı olabileceğini düşündürmektedir.

1.6.2. Kamufle Nesne Segmentasyonu

Kamufle edilmiş bir nesnenin görsel özellikleri arka plana çok benzer: (i) kamufle edilen nesnenin yoğunluğu veya rengi çevrelerine yakındır, (ii) doku, arka planla birleşecek şekilde yok edilir ve (iii) kamufle edilen nesnenin sınırı bulanıklaşır. Kamufle edilmiş nesnelerin bu tür görsel özellikleri tespit görevini daha da zorlaştırmaktadır. Bu karmaşıklık nedeniyle literatürde görsel kamuflej tespiti girişimi için daha az çalışma yapılmıştır. Ancak araştırmacılar kamufle edilmiş nesneleri çevrelerinden algılamak için çeşitli görsel özellikleri (örn. yoğunluk veya renk, doku, hareket, gradyan vb.) kullanan farklı algoritmalar geliştirdiler. Burada mevcut yöntemler Şekil 1.2'de görüldüğü gibi gruplandırılmıştır. Kamufle edilmiş cisimlerin tespiti için görsel özellikler önem taşımaktadır. Aşağıdaki alt bölümde, görsel özelliklerin her birinin dikkate alınmasına ilişkin güncel teknikler tartışılmıştır.



Şekil 1.2. Kamufle edilmiş nesne algılama yöntemleri

1.6.2.1. Düşük Seviyeli Özellikler

1.6.2.1.1. Doku Özellikleri

Bazen nesnenin rengi arka plana benzer olmasını rağmen farklı doku desenleri içermektedir. Dokunun, nesneyi çevresinden ayırt ettiği kabul edilir. Galun ve ark. [39], doku elemanlarının yapısal özelliklerini filtre tepkileriyle birleştiren aşağıdan yukarıya toplama yöntemi kullanarak kamufle edilmiş nesneleri tespit etmektedir. Doku elemanlarının şeklini uyarlanabilir bir şekilde tanımlar ve onları boyutlarına, en boy

oranlarına, yönelimlerine, parlaklıklarına vb. göre karakterize eder. Daha sonra farklı dokuları ayırt etmek için bu özelliklerin çeşitli istatistiksel ölçümleri dikkate alınır. Yaklaşım çeşitli doku türlerini içeren görüntülere uygulanır. Bu yöntem, nesneler ve arka planlar için farklı dokular içeren görüntüler için iyi çalışır. Ancak nesne ve arka plan benzer bir doku içeriyorsa teknik iyi sonuçlar vermeyebilir. [40]'ta Bhajantri ve Nagabhushan, kamuflle edilmiş kusuru tespit etmek için bir teknik önerdiler. Burada birlikte oluşum matrisi [41] tabanlı doku özellikleri küçük bir görüntü bölgesi içinde hesaplanır. Kusurlu kısım küme analizi ve havza segmentasyonu ile tespit edilir. Bu yöntemin doğruluğu doku özelliğine bağlıdır. Nesnelerin ve arka planın benzer türde doku içerdiği diziler için iyi çalışmayabilir ve gerçek yaşam verilerine uygulanmaz. Ayrıca bu yöntemi uygulamadan önce arka plan bilgilerinin bilinmesi gerekir. Sengottuvelan ve ark. [42], nesnenin kamuflle edilmiş kısmını tespit etmek için geliştirile teknikte gri seviyeli birlikte oluşum matrisi tabanlı doku özelliği ve dendrogram kullanılmaktadırlar. Bu teknik, verilen görüntünün birkaç bloğa veya daha küçük bölgelere bölünmesi nedeniyle çok zaman alıcıdır. Gölgeleme efektleri içeren görüntüler ve benzer dokular içeren nesne ve arka plan için çalışmaz. Song ve Geng [9], kamuflaj dokusunu tasarlamak ve değerlendirmek için ağırlıklı yapısal benzerliğe dayalı bir teknik önermektedirler. Chennamsetty ve ark. [43] kamuflle edilmiş nesne algılama için doğrusal denklemlere ve istatistiksel özniteliklere dayalı istatistiksel karakterizasyon yaklaşımı kullanılmaktadırlar.

Tartışma: Renk ve yoğunluk bilgisinden doku özelliği çıkarımı maliyetli olmaktadır. Ancak kamuflle edilmiş nesneleri tespit etmek için daha etkilidir. Nesne ile arka plan renk benzerliği nedeniyle kamuflaj oluşturduğunda doku özelliklerinin kullanılması benzeri durumlarda umut verici sonuçlar ortaya koymaktadır.

1.6.2.1.2. Gradyan Özellikleri

Bir nesne, arka planla benzer bir renge veya dokuya sahip olduğunda bu özellikleri kullanarak o nesneyi algılamak zordur. Gradyan bilgisi, görüntülerdeki nesneyi arka plandan ayırt etmeye yaramaktadır. Bu doğrultuda, kamuflle edilen nesneyi tespit etmek için bazı yöntemler geliştirilmiştir. Tankus ve Yeshurun [44,45] benzer özelliklere (renk ve doku gibi) sahip düz bir arka plandan ilgile alanı ayırmak için Darg operatörü önermiştir. Darg, doğrudan görüntünün gri seviye bilgisine uygulanmaktadır. Nesnelerdeki pürüzsüz 3D

dışbükey veya içbükey yamalara uygulanmış ve herhangi bir özel ışık kaynağı veya yansıtma işlevi ile sınırlı değildir. Darg operatörü kullanılarak elde edilen sonuçlar büyük ölçüde eşik değerlerine bağlıdır. Pan ve ark. [46], Darg operatörünün yardımıyla karmaşık bir arka planda nesnenin kamufle edilmiş kısmını tespit etmek için bir algoritma geliştirdiler. Bu yöntemin en büyük dezavantajı uygun eşik değerinin seçilmesindeki zorluktur. Uygun bir eşik seçimi önemli bir konudur. Ayrıca bu yöntem içbükey arka plan ve koyu renkli nesneler içeren bir ortamda verimli çalışmamaktadır.

1.6.2.2. Derin Özellikler

Son yıllarda, Evrimsel Sinir Ağlarının (CNN'ler) büyük hesaplama gücünden ve derin özellik bilgilerinin doğru konumlandırma yeteneğinden yararlanarak, derinlik evrimsel sinir ağları, verilen görüntülerdeki kamufle edilmiş nesneleri doğru ve hızlı bir şekilde tespit etmek için algoritma kullanılmaktadır.

Li ve ark. [47] görüntü iyileştirmeye dayalı bir kamufle edilmiş renkli hedef tespit algoritması sunmuştur. Bu model, hedef ve arka plan arasındaki renk kontrastını kullanır ve kamufle edilmiş nesneleri tespit etmek için görüntü iyileştirme yoluyla kontrastı genişletir. Bununla birlikte, model görüntüyü iyileştirdiğinde, karmaşık arka planları da iyileştirir ve kamufle edilmiş nesnelerin kenarını bulanıklaştırır. Zheng ve ark. [48], temel olarak karmaşık doğal ortamlardan kamufle edilmiş görüntüye sahip insanların nasıl tespit edileceğine ve segmentlere ayrılacağına odaklanan COD görevi için bir model DDCN önermiştir. DDCN, arka plandaki gizlenmiş kişiyi doğru ve hızlı bir şekilde tanımlamak için ön plan özellikleriyle bir araya getirilen anlamsal bilgi özelliğini kullanır. Bu model yalnızca taklitçileri tespit etmede tatmin edici sonuçlar elde ettiğini kanıtlayabilir, ancak diğer yönlerden aynı etkiye sahip olup olamayacağını kanıtlayamaz. Bu doğrultuda, Fang ve ark. [49], çeşitli katmanların özelliklerini entegre etmek için eklemeyi kullanarak anlamsal bilgiyi geliştirmek ve girdinin kabul alanını genişletmek için genişletilmiş konvolüsyondan yararlanan bir model tasarlamıştır. Model yalnızca belirli senaryolarda kamuflej modundaki insanları tespit edebilmektedir, bu nedenle modelin uygulama kapsamı sınırlıdır. Yan ve ark. [50], orijinal ve çevrilmiş görüntüye karşılık gelen ana akış ve ayna akışı gibi iki segmentasyon akışına sahip olan yeni bir ağ önerdi. Ana akış örnek segmentasyonu, ayna

akışı ise kamufle edilmiş nesne segmentasyonu için kullanılmaktadır. Ayna aksısının çıkışı, COD doğruluğunu artırmak için ana akışın çıkışıyla birleştirilir. Le ve ark. [2], verilen sahnelerden hedef nesneyi belirlemek ve parçalamak için sınıflandırma akışı ve segmentasyon akışı özelliğini kullanan CNN'lere dayalı uçtan uca bir ağ olan Anabranh Network (ANet) sunmuştur. ANet, kamuflej nesnelerinin var olma olasılığını hesaplamak için dal ağlarını kullanır, ancak dal ağlarının ve segmentasyon ağlarının kamuflej nesnelerini tespit etmek için etkili bir şekilde birleştirildiğini garanti edemez. Fan ve ark. [51]'de kamuflejli nesne tespiti için arama modülü ve tanımlama modülü olmak üzere iki ana bileşenden oluşan bir yapı-SINet önermişlerdir. SINet, hedef nesneyi kabaca bulmak için arama modülünü etkin bir şekilde kullanır ve verilen görüntülerdeki kamufle edilmiş nesneleri tam olarak bulmak ve tespit etmek için tanımlama modülünden yararlanır. Bununla birlikte, SINet ayrıca çıkarılan sıkı bilgiler üzerinde bilgi fazlalığına neden olabilecek bir genişletme konvolüsyon işlemi gerçekleştirir. Daha yakın zamanda, Wang ve ark. [52] kamuflej modundaki tankı tespit etmek için YOLO ağından oluşan bir model sunmuştur. Ancak makaledeki F-skora göre, bu model daha fazla iyileştirmeye ihtiyaç duymaktadır.

1.7. Derin Öğrenme Algısı

Bilgisayarlı görme alanındaki; el hareketi tanıma, insan etkinliği tanıma, konuşma tanıma, el yazısı karakter tanıma, eylem etiketleme, görüntü sınıflandırma ve etiketleme gibi pek çok uygulamada son zamanlarda büyük bir değişim görülmüştür. Derin öğrenme olarak adlandırılan makine öğrenimi teknolojisinin ortaya çıkışı ve başarılı bir şekilde geliştiği bilgisayarlı görme alanlarında da görülmektedir. 2010'dan bu yana bilgisayarlı görme alanındaki birçok araştırmacı geleneksel el yapımı özellik tanımlayıcısından genellikle veriye dayalı algoritmalar olarak adlandırılan öğrenme tabanlı özellik tanımlayıcısına geçmiştir. Şu anda derin öğrenme bilgisayarlı görme alanındaki aktif araştırma alanıdır.

Derin öğrenme, insanın müdahalesi olmadan bilgisayara, ne yapılacağına ve nasıl yapılacağına olanak sağlayan makine öğreniminin bir alt kümesidir. Derin öğrenmedeki hızlı gelişme ve güç tüketimi, bir görüntü sensörünün çözünürlüğü, yüksek bilgi işlem gücü, bellek depolama kapasitesi dahil olmak üzere sistem kapasitesindeki iyileştirmeler, görme tabanlı öğrenmenin yayılmasını daha da hızlandırarak performansı ve maliyet etkinliğini arttırmıştır. Anlamsal bölütleme, görüntü sınıflandırma, eşzamanlı lokalizasyon ve

haritalama, nesne tanıma, nesne algılama, el yazısı karakter tanıma, konuşma tanıma ve eylem tanıma gibi çeşitli problemlerde ürettiği mükemmel performans nedeniyle derin öğrenme teknikleri büyük ilgi görmüştür. Derin öğrenmeye dayalı modeller bazı problemlerde, insan düzeyindeki beklentilerin ötesinde en gelişmiş sonuçları elde edebilir. Tüm derin öğrenme tabanlı modeller, özellikleri öğrenmek ve çıkarmak için birçok derin sinir ağı katmanını, eğitim ve test içinse birçok etiketli veri gerektirir.

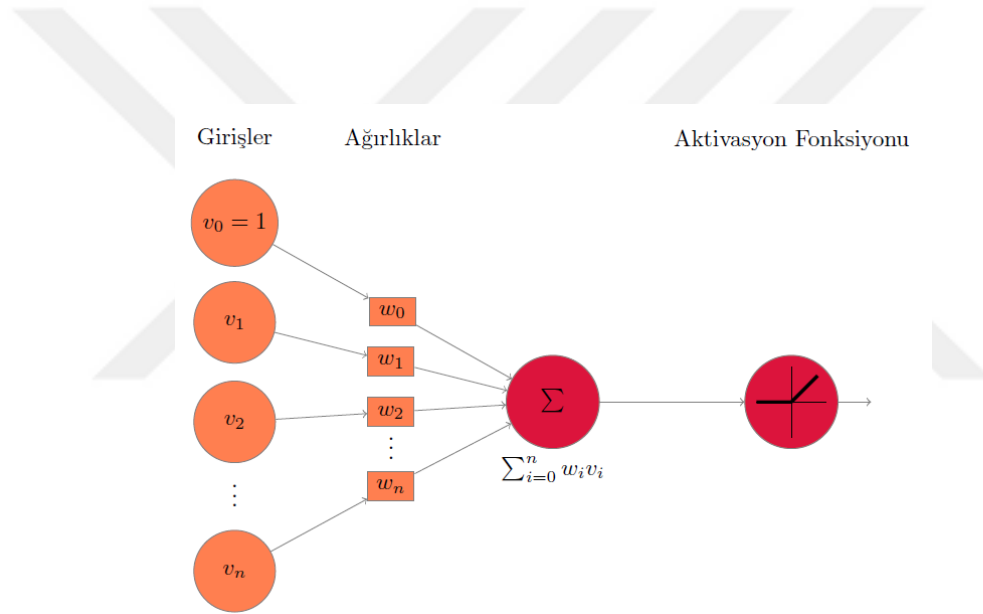
Derin öğrenme mimarilerinin ortak bir özelliği, temsilleri doğrudan verilerden öğrenebilmeleridir, bu da genellikle el yapımı özellikleri kullanan diğer tekniklerden daha iyi performans sağlar. Hinton ve ark. çalışması [53] pek çok gizli katmana sahip bir ağın etkili bir şekilde eğitilmesinin mümkün olduğunu gösteren çalışmalardan biridir. Bu araştırma, daha önce önemli olan daha büyük sinir ağlarının eğitilmesi sorununun üstesinden gelinmesine yardımcı olmuştur [54]. Bu tür derin katman ağlarını eğitmek için bazen yüksek performanslı GPU, bilgi işlem için iyi performans sağlar. Son yirmi yılda farklı yapı taşları kullanan çeşitli türlerde derin öğrenme mimarileri mevcuttur. Sorunlu sinir ağı adresinin sayısını ve türünü büyük ölçüde genişletmiştir. En popüler derin öğrenme mimarileri; Oto kodlayıcılar [54], derin inanç ağları [55,56], derin oto kodlayıcılar [57], evrimsel kısıtlı boltzmann makinesi [58], tekrarlayan sinir ağı [59], uzun kısa süreli bellek ağıdır [60], evrimsel sinir ağı [61] ve derin yığınlama ağıdır [62].

Benzere mimari biçimleri oluşturmak için derin öğrenme [63,64] teknikleri kullanılmaktadır. Nesne algılama için en çok tercih edilen mimari, evrimsel sinir ağlarıdır. Derin öğrenme modelleri için bu en önemli özelliktir. Denetimli ve denetimsiz bir şekilde çeşitli özellik düzeylerini ve başlangıç seviyesindeki veya en düşük seviyedeki basit özellikleri öğrenir, katman sayısı arttıkça daha karmaşık özellikleri çıkarır. İkinci en önemli mimari, bir zaman serisi verilerini ve diğer sıralı verileri denetimli bir şekilde öğrenen ve işleyen tekrarlayan bir sinir ağıdır. Cho ve ark. [65] çalışmasında, denetimli bir şekilde geri yayılım kullanarak büyük miktarda etiketlenmiş veri üzerinde başlayıp bitecek şekilde eğitilebilen modellerin potansiyelini göstermiştir. Bu sinir ağlarının hakimiyetini geri kazanmaya yardımcı olan ve derin öğrenme devrimini başlatan çalışmalardan biridir. Son yıllarda literatürde derin öğrenme kullanarak nesne tespiti için yapılan birçok araştırma kapsamlı bir şekilde incelenmiştir.

1.8. Sinir Ağı Yapısı

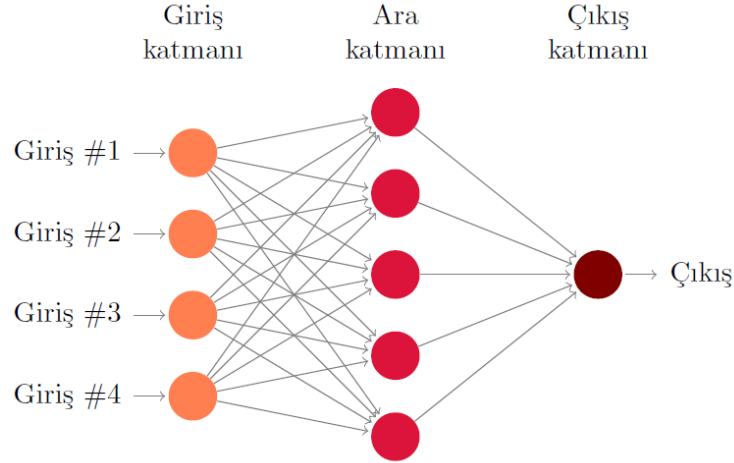
1.8.1. Yapay Sinir Ağı

Yapay Sinir Ağı (ANN), girdileri çıktılara eşlemek, verilerden kalıpları öğrenmek için kullanılan biyolojik sinir hücresinden ilham allanarak oluşturulan bir makine öğrenimi modelleri ailesidir. Böyle bir modelin en basit şekli algılayıcıdır. Bir algılayıcı ağırlık ve girdi öğeleri vektörü tarafından tanımlanan doğrusal bir fonksiyona dayalı olarak tahminde bulunmaktadır (Şekil 1.3). Bu, algılayıcının karar sınırının doğrusal olduğu ve karmaşık kalıpları öğrenmek için uygulanabilirliğini yasakladığı anlamına gelir.



Şekil 1.3. Perceptron yapısı

Doğrusal fonksiyonlar kullanılarak modellenemeyen karmaşık desenlerin modellenmesinde algılayıcının bu sınırlaması gizli bir katman tanıtılarak çözülür (Şekil 1.4). Gizli katman, önceki ve ardışık katmandaki her nörona bağlı bir nöron topluluğudur. Gizli katmanlar, girdinin doğrusal dönüşümünü ve ardından doğrusal olmayan bir aktivasyonu kullanarak girdiyi yeni bir forma dönüştürür. Bu yeni dönüşüm, istenen çıktıyı elde etmek için ardışık katmandaki nöronlar tarafından daha fazla işlenebilir.



Şekil 1.4. Çok katmanlı ANN

Bir dizi gizli katman, çıktı katmanında kolayca değerlendirilebilen (örneğin, dönüşümlerin öğelerinin doğrusal bir kombinasyonuna dayalı olarak sınıflandırılan) girdinin giderek daha karmaşık bir dönüşümünü gerçekleştirerek bu dönüşümü birçok kez gerçekleştirir. Bu süreç genellikle temsil veya özellik öğrenme olarak adlandırılır.

Çok katmanlı bir ANN kullanılarak bir girdinin bir çıktıya eşlenmesi matematiksel olarak şu şekilde yazılabilir (örneğin, yalnızca bir gizli katmanı olan bir ANN için) [66]:

$$y_k(V; \theta) = f^2 \left(\sum_{j=1}^M W_{kj}^{(2)} f^1 \left(\sum_{i=1}^D W_{ji}^{(1)} v_i + b_j^{(1)} \right) + b_k^{(2)} \right) \quad (1.1)$$

Burada V , D boyutlarının girdi vektörüdür; $\theta = W^{(1)}, W^{(2)}, b^{(1)}, b^{(2)}$, her katmandaki parametrik ağırlıklar ve sapmadır ve y_k , çıktı vektörünün k . öğesidir.

$f^2(\cdot)$ ve $f^1(\cdot)$, ilgili katmanlarda doğrusal olmayan aktivasyonlardır. Eşitlik 1.1'den görülebileceği gibi, ANN'lar girdinin doğrusal ve doğrusal olmayan dönüşüm zincirini kullanarak girdinin çıktıya eşlenmesini modellemektedir.

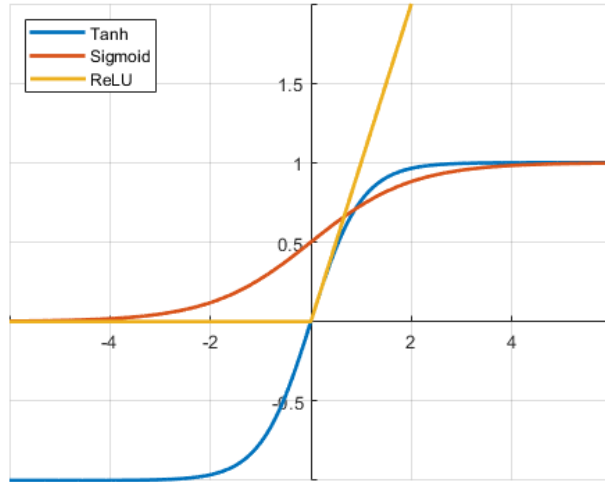
Eşitlik 1.2, üç baskın aktivasyon fonksiyonunu gösterir. Sigmoid aktivasyon fonksiyonu, geleneksel makine öğrenimi mimarilerinde alışılmış bir şekilde kullanılmıştır. Modern sinir ağlarında düzeltilmiş doğrusal birim (ReLU) [67] varsayılan bir öneridir.

Basitliğine ek olarak, matematiksel optimumluğu kanıtlanmıştır. Bununla birlikte sigmoid aktivasyon fonksiyonu, herhangi bir değeri $[0, 1]$ aralığıyla sınırladığı ve yorumlana bilirliliği için (ağın sonunda) tek nesne kategorisi sınıflandırma problemlerinden yaygın olarak kullanılmaktadır. Tanh aktivasyon fonksiyonu, tüm değerleri $[-1, 1]$ aralığıyla sınırlayarak sigmoid'e oldukça benzer davranır. Eşitlik 1.2'deki aktivasyon fonksiyonları, belirli problemlere uyarlamak için daha da genelleştirilebilir.

$$\text{sigmoid}(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{1 + e^{-x}} \quad (1.2)$$

$$\text{ReLU}(x) = \max(0, x)$$



Şekil 1.5. Tanh, Sigmoid ve ReLU aktivasyon fonksiyonlarının grafiksel gösterimi

Bir ANN'in tasarlaması genişliğin (bir katmandaki nöron sayısı) ve gizli katmanların derinliğinin (sayısı) uygun seçimini ve uygun doğrusal olmayan aktivasyon fonksiyonunu gerektirir. Model parametreleri, eğitim verilerinin hata fonksiyonunu en aza indiren bir optimizasyon süreci aracılığıyla öğrenilir; burada L , model çıktısı ile belirli bir girdi için istenen çıktı arasındaki fark olarak tanımlanır.

$\nabla L(W)$ parametre setindeki hata gradyanına dayanan farklı algoritmalar kullanılarak yinelemeli şekilde yapılır. Her katmandaki gradyan, hata geri yayılım algoritması [66] ile hesaplanabilir. Her katmandaki tüm ağırlıklar için hata gradyanı hesaplandıktan sonra, uygun parametreler, aşağıdakiler kullanılarak belirli bir hata eşiğine veya yineleme sayısına yakınsayana kadar güncellenebilir:

$$W^{(t+1)} = W^{(t)} - \alpha \nabla L(W^{(t)}) \quad (1.3)$$

Burada α öğrenme katsayı ve t yineleme sayısını temsil eder. Stokastik, mini toplu gradyan inişi ve momentumyineleme sürecini iyileştirmek için kullanılan çeşitli yaklaşımlardır [68].

1.8.2. Evrişimli Sinir Ağları

ANN'larda girdi, bireysel girdi öğelerinin düzenlenmesinin herhangi bir bilgi iletmediği bir vektör biçimindedir. Bu gibi durumlarda, her nöron bir önceki katmandaki her nörona bağlıdır. Giriş verilerinin çeşitliliği ağın konfigürasyonunu etkilemektedir.

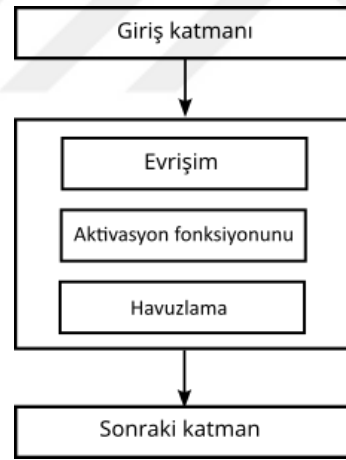
CNN, örneğin 2 boyutlu görüntü verilerindeki özellikleri tanımak için özel olarak tasarlanmış çok katmanlı ileri beslemeli ağıdır. CNN'nin mimarisi, Hubel ve Wiesel'in kedinin görsel korteksindeki nörobiyolojik sinyal işleme çalışmasından esinlenmiştir [69]. CNN'nin tipik bir uygulaması, görüntülerdeki çeşitli nesnelerin tanınmasından oluşur. Bununla birlikte, evrişimli ağlar çeşitli farklı görevler için de başarıyla kullanılmıştır [70,71].

CNN'lerdeki nöronlar, görüntünün alt görüntü adı verilen küçük bir bölümünü dikkate alarak çalışır. Alt görüntüler daha sonra ağ tarafından tanınabilecek özellikler açısından incelenir. Basit bir örnek olarak, özellik; dikey bir çizgi, bir yay veya bir daire olabilir. Bu özellikler daha sonra ağın ilgili özellik haritaları tarafından tespit edilir. Daha sonra görüntüyü sınıflandırmak için bir özellikler kombinasyonu kullanılır. Ayrıca, ağı değişen kontrast, parlaklık, renk doygunluk seviyeleri, gürültü vb. seviyeleri dayanıklı hale getirmek için birden fazla farklı özellik haritası kullanılır.

Özellik haritalarından oluşan iki tür katman vardır. Evrişim katmanının amacı, giriş görüntüsündeki özellikleri tanımdır. Genellikle her bir harita belirli bir özelliği tanıyan birkaç özellik haritasından oluşur. Alt görüntüler orijinal görüntünün örtüşen bölgelerini kasıtlı olarak kapsar. Böyle bir tasarım, öteleme, döndürme, eğiltme vb. gibi görüntü bozulmalarını tolere etmek için önemlidir.

Alt örnekleme katmanı olarak adlandırılan diğer katman türü, her zaman bir evrişim katmanından sonra gelir. Aynı sayıda özellik haritasından oluşur, burada evrişim katmanından gelen her harita, aşağıdaki alt örnekleme katmanında karşılık gelen özellik haritası için bir girdi olarak kullanılır. Alt örnekleme katmanındaki nöronlar tarafından kapsanan alt görüntüler genellikle örtüşmez.

Ağın derinliğine bağlı olarak, evrişimsel ve alt örnekleme katmanları, son alt örnekleme katmanına ulaşılan kadar değişir. Son alt örnekleme katmanından sonra herhangi bir sayıda tam bağlantılı katman olabilir, bunlardan sonuncusu çıktı katmanıdır [72].



Şekil 1.6. CNN'nin Bileşenleri

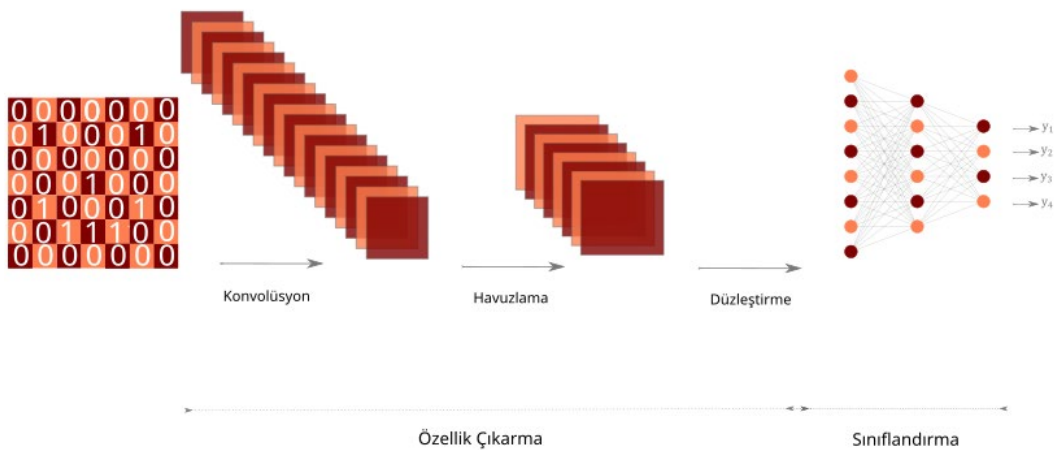
1.9. CNN Mimarisi

Evrişimli ağın mimarisini açıklamak için giriş katmanında bir görüntü verisi vardır. Bir evrişimsel ağ, mimari olarak iki önemli bölüme ayrılmıştır. Her parça farklı bir amacı yerine getirmek için tasarlanmıştır. Ağın ilk kısmı özellik çıkarımı gerçekleştirir, evrişim ve alt örnekleme katmanlarından oluşur. İkinci kısım,

Şekil 1.7'de görüldüğü gibi çıkarılan özneliklere dayalı olarak sınıflandırma gerçekleştirir.

Özellik çıkarma işlemi sunulan girdi görüntüsü ile başlar, her piksel ilk evrişim katmanının özellik haritalarında gruplandırılmış nöronlar için girdiyi temsil eder. Özellik haritasındaki nöronlar 2 boyutlu bir ızgarada düzenlenmiştir. Aynı özellik haritasındaki tüm nöronlar ağırlıklarını paylaşır. Bu daha az bellek gerektirerek ve daha iyi bir performans sağlayarak uygulamanın optimize edilmesini sağlar. Daha da önemlisi mimari bir özellik olmasıdır. Belirli bir özellik haritasındaki her nöronun aynı özelliği tanınması beklenir. Bu özellik esas olarak nöral girdileri filtreleyen ağırlıkların bir kombinasyonu ile tanınır. Verilen özellik haritasındaki tüm nöronlar için aynı ağırlıkların paylaşılması her piksel için aynı filtrenin kullanılmasını sağlar.

Giriş görüntüsü sıfırıncı indekse sahip katmanın çıktısı olarak kabul edilir. Giriş katmanını daha sonra dönüşümlü evrişimsel ve alt örnekleme katmanları takip eder. Evrişim katmanındaki özellik haritasının nöronları, bir önceki alt örnekleme katmanında bulunan özellik haritaları kümesinden girdi alır. Konvolüsyonel katmanı daha sonra özellik haritalarının boyutunu azaltan bir alt örnekleme katmanı takip eder. Son alt örnekleme katmanı özellik çıkarma işlemini sonlandırır. Bu noktada ağıncı kısmı sınıflandırma sürecini başlatır. Sınıflandırma girdileri çıkarılan özelliklere göre sınıflandıran tam bağlantılı bir katman kullanılarak gerçekleştirilir.



Şekil 1.7. CNN'in tüm işlemlerini (evrişim, havuzlama, özellik çıkarma ve sınıflandırma) temsil eden CNN mimarisi

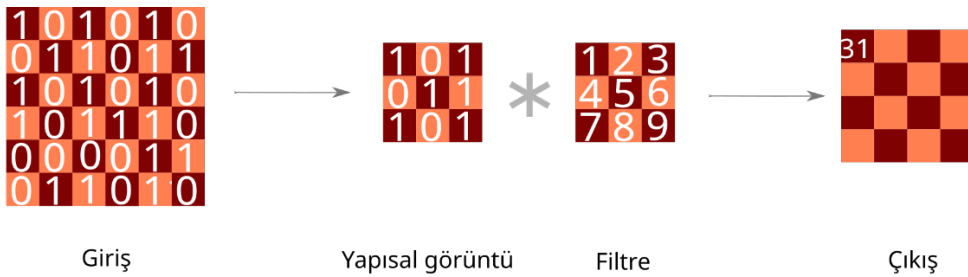
1.10. CNN'deki Katmanlar

1.10.1. Evrişimsel Katman

Evrişimsel katmanların amacı sunulan görüntülerdeki özellikleri tespit etmektir. Her biri belirli bir özelliği tanıyan çoklu özellik haritalarından oluşur. Özellik tanıma alt görüntünün bir filtreden geçirilmesi olarak düşünülebilir. Filtreleme esas olarak ağırlık ayarlamaları ile yapılır. İşlem Şekil 1.8'de gösterilmektedir.

Bir evrişim, girdi ile ağırlıkların çarpımı ve ardından bir sapma eklenmesidir. Lineer bir işlemdir. Filtre veya çekirdek olarak bilinen ağırlık dizisi her bir öge üzerinde evrişim gerçekleştiren giriş dizisinde kayar. Her işlemin sonucu özellik haritası olarak bilinen çıktı dizisine eklenir. Görüntüden özellik çıkarma işlemi CNN'nin evrişim katmanları boyunca gerçekleşir. Evrişim işlemi Eşitlik 1.4'te tanımlanmıştır. $m \times n$ boyutundaki bir F filtresini, i satırı ve j sütunundaki bir 2 boyutlu I giriş verisinin tek bir konumu üzerinde evritirirken, filtrenin ağırlıkları girişle eleman bazında çarpılır ve ardından tek bir çıkış değeri üretmek için toplanır:

$$S(i, j) = (I * F)(i, j) = \sum_m \sum_n I(i + m, j + n) F(m, n) \quad (1.4)$$



Şekil 1.8. Giriş özelliği haritasında evrişim işlemi

1.10.2. Evrişim Türleri

1.10.2.1. Ters Evrişim

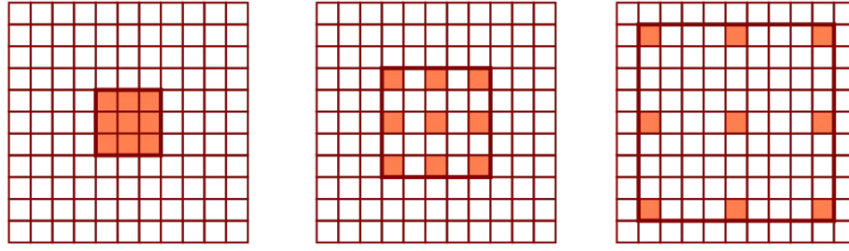
Ters Evrişim (dekonvolüsyon), çoğunlukla bölümlleme ağlarında kullanılan evrişim operatörünün bir versiyonudur. Havuzlama, görüntünün alt örnekleme için uygun bir operatör iken, üst örnekleme için transpoze edilmiş evrişim kullanılır. Görüntülerin öğrenilmiş üst örnekleme bilgisayarla görme için güçlü bir araçtır ve FCN [32] ve U-net [73] gibi segmentasyon modellerinin hayati bir parçasıdır.

$$X = (Y * F)(i, j) \quad (1.5)$$

Eşitlik 1.5, $X \cdot Y$ bir diziyi (ağdaki önceki katmandan çıktı), F bir konvolüsyon çekirdeğini ifade etmektedir ve konvolüsyon işlemi Eşitlik 1.4'te açıklanmıştır. Uygun bir adım, konvolüsyon çekirdeği ve Y 'ye sıfır dolgu yapılarak, ters konvolüsyon yardımıyla örneklerin öğretilmesi gerçekleştirilebilir.

1.10.2.2. Genişletilmiş Evrişim

Genişletilmiş evrişim [74], parametrelerde artış olmadan alıcı alanı büyüten bir evrişim yapısıdır. Bunu çekirdek boyutunu artırmadan boşlukları doldurarak yapmaktadır. Genişletilme, filtrenin boyutunu büyütür ve elde edilir. Genişletilme adımı r , küçük boyutlu çekirdek k ise filtre $[k + (k-1)(r-1)]$ 'e genişletilir. Giriş matrisinde filtre elemanlarının eşleştiği yere olan mesafe, *Genişletme katsayısı* D tarafından belirlenir. $D = 1$ ise, standart evrişimden bahsedilmektedir. Genişletilmiş evrişim ile büyütülmüş alıcı alan, bağlam özümsemesinde yardımcı olur. Görüntüden daha bağlamsal ve çok ölçekli bilgi yakalayarak görüntünün daha geniş bir görünümünü sağlar, ancak uzaysal çözünürlüğü değiştirmez. Yöntem daha az parametre gerektirdiğinden verimlidir. Şekil 1.9'da, çekirdek boyutu 3×3 olduğunda, en soldaki bölüm 1 genişleme oranına, ortadaki 2 genişleme oranına ve en sağdaki 4 genişleme oranına sahip genişletilmiş evrişimi göstermektedir. Boşluklar hücrelerle doldurularak alıcı alanın büyütüldüğü görülebilir. Bu, parametre sayısının önemli ölçüde artmadığını açıkça göstermektedir.

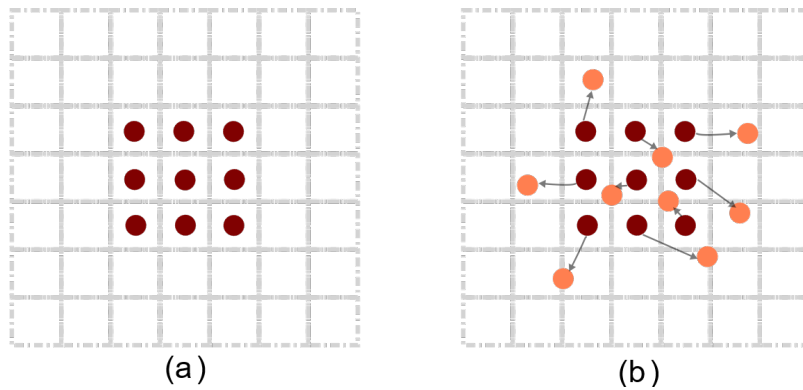


Şekil 1.9. Çeşitli genişleme oranlarıyla genişletilmiş evrişimin uygulanması

1.10.2.3. Deforme Edilebilir Evrişim

Bu konvolüsyonel yapıdaki ağlarda deforme edilebilir modüller, geometrik dönüşümlere karşı değişmezliği öğrenmek için herhangi bir nesne algılama sistemindeki normal konvolüsyonel katmanların yerini almak için kullanılabilir.

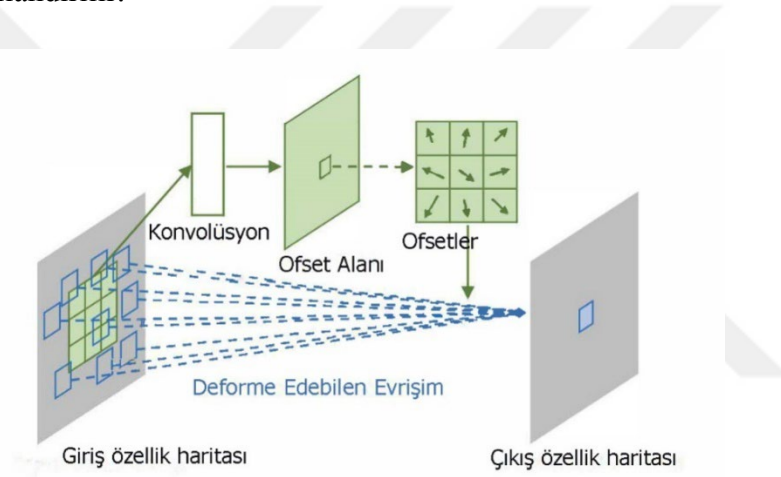
Ölçek, döndürme, öteleme ve diğer çarpıtma türleri gibi geometrik dönüşümler, nesne algılamanın ana zorluklarından biridir. Bu sorunla başa çıkmanın bir yolu, yeterli çeşitlilikte veri toplamak veya oluşturmaktır. Başka bir yol, ölçek değişmez özellik dönüşümü gibi dönüşümlere değişmez özellik algılama algoritmalarını kullanmaktır [75]. Deforme edilebilir evrişim [76], parçalar arasındaki deformasyonu, nesnenin parçası olmanın tek tek parçaları puanına karşı tartarak uzamsal deformasyonu öğrenerek bu sorunu çözer.



Şekil 1.10. Evrişimli bir 3×3 filtrede uzamsal örnekleme konumlarının gösterimi a) normal bir evrişimli filtrenin örnekleme konumları b) bir deforme edilebilir filtre

Deforme edilebilir evriřimler, global bir dönüşüm yerine özellik haritalarının yerel ve yoğun dönüşümlerini kullanır. Deforme edilebilir evriřimler, evriřimli filtrelerin uzamsal örnekleme konumlarına öğrenilebilir ofset parametreleri ekler (Şekil 1.10 ve Şekil 1.11).

Örnekleme konumlarını deęiřtirmenin ardındaki motivasyon, evriřimli bir aęın alıcı alanının deęiřtirilebilmesidir. Alıcı alan, bir nöronun bilgi aldıęı girdi alanıdır. Derin sinir aęları için alıcı alan görüntünün tamamı olabilir. Ancak merkeze yakın pikseller daha fazla bilgi saęlar ve etkin alıcı alan çok daha küçüktür. Deforme edilebilir evriřimler nöronun merkezlendięi nesneyle eşleşen etkili bir alıcı alan saęlamak için girdinin özellik haritalarında koşullandırılır.



Şekil 1.11. Deforme edilebilir evriřim yapısı [76]

Düzenli evriřim için daha önce bahsedilen 1.4 Eşitlikle karşılaştırıldığında, deforme edilebilir evriřim aşağıdaki şekilde verilebilir:

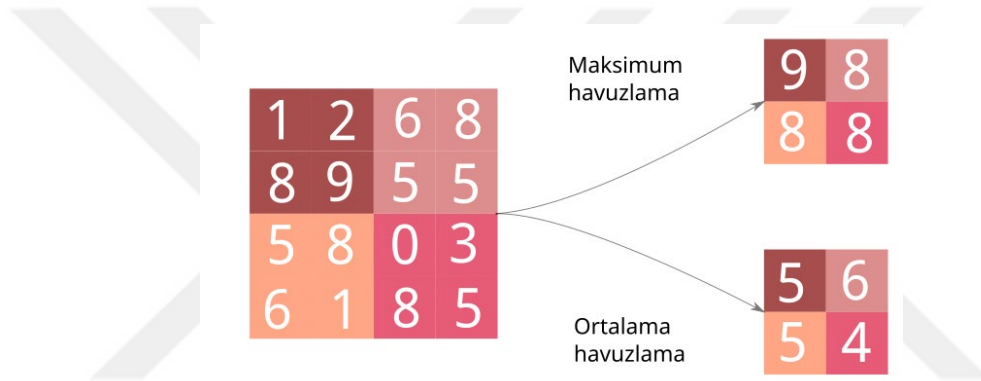
$$S(i, j) = \sum_m \sum_n I(i + m + \delta_m, j + n + \delta_n) F(m, n) \quad (1.6)$$

Burada, δ_m ve δ_n ofsetlerdir.

1.10.3. Havuzlama Katmanı

Havuzlama operatörünün amacı çıktı değişmezliğini elde etmek için girdilere doğrusal olmayan bir dönüşüm uygulamaktır. Bir CNN'de havuzlama katmanı genellikle 2 çeşittir: Maksimum Havuzlama ve Ortalama Havuzlama. Bu iki tür katmanın işlevselliği Şekil 1.12'de gösterilmektedir. Maksimum havuzlama, kernel tarafından kapsanan resmin segmentinden maksimum değeri geri yükler. Oysa ortalama havuzlama, kernel tarafından kapsanan resmin bitinden çok sayıda değerın ortalamasını geri yükler.

Konvolüsyonel katman ve havuzlama katmanı, birlikte bir evrişimsel sinir ağının I. katmanını yapılandırır.



Şekil 1.12. Maksimum Havuzlama ve Ortalama Havuzlama örnek Gösterimi

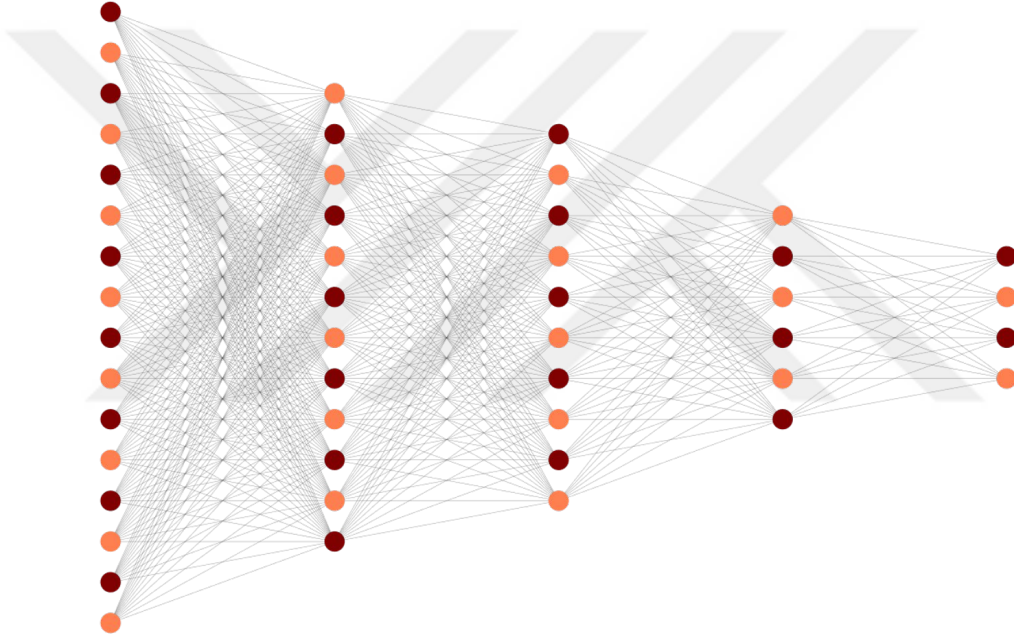
1.10.4. Tam Bağlantılı Katman

Tam bağlantılı katman önceki havuzlama katmanına tam bağlantılar kurar. Tam bağlantılı katman, sınıflandırma yapan ANN ile benzer bir yapıya sahiptir (Şekil 1.13). Son tam bağlı katmandan sonra çıktı vektörlerini SoftMax işlevi aracılığıyla normalleştirmek yaygındır.

SoftMax şu şekilde tanımlanabilir:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (1.7)$$

Burada z_i , SoftMax fonksiyonuna giriş vektörünün elemanlarıdır. i . nörondaki çıkış vektörünü üstel bir fonksiyon aracılığıyla sıfır ile bir arasında bir değere taşır. Daha sonra, i . nörondaki çıktı vektörünü katmandaki tüm çıktı vektörlerinin toplamına bölerek, i . nörondaki normalleştirilmiş çıktı vektörü oluşturulur ve $\hat{y}_i$ olarak adlandırılır. Normalleştirilmiş çıktı vektörü, tüm olası kategoriler arasındaki olasılık dağılımını temsil eder.



Şekil 1.13. Tam bağlantılı katmanın gösterimi

1.11. Evrimsel Sinir Ağı Eğitimi

CNN'yi eğitmek, eğitilmiş modelin eğitim verilerine uyması için evrimsel katmanlarda ve tam bağlantılı katmanlarda öğrenilebilir parametreleri optimize etmeye karşılık gelir. CNN'nin çıktısı, SoftMax işlevi aracılığıyla sınıfların olasılık dağılımını ürettiği için belirli bir sınıfa ait olan girdi özellikleri iyi eğitilmiş modeli ileri beslemeli

yayma yoluyla en yüksek olasılığa sahiptir. CNN'nin standart en uygun şekle sokma yöntemi, fonksiyonunun kayıplarını en aza indirmektir.

1.11.1. Kayıp Fonksiyonu

Herhangi bir sinir ağını eğitmek için bir kayıp fonksiyonu minimize edilir. Sezgisel olarak, bu kayıp fonksiyonu bir modelin eğitim veri seti üzerinde yaptığı hatalar olarak görülebilir. Modele, tahminin istenen çıktıdan ne kadar uzak olduğuna dair bir fikir verir ve ardından model ağırlıklarını buna göre günceller. Kullanım durumuna ve göreve bağlı olarak CNN'i eğitmek için kullanılabilecek çok çeşitli kayıp fonksiyonları vardır.

Sınıflandırma görevleri için en yaygın kayıp çapraz entropi kaybıdır. Çapraz entropi, iki dağılım arasındaki farkı bulmayı amaçlayan bilgi teorisinden bir ölçüdür. Burada sezgiseldik P temel olasılık dağılım ile Q hedef dağılımın tahminine dayalıdır. Bu durumda P 'den Q 'nun P 'den çapraz entropisi, P yerine Q kullanarak bir olayı temsil etmek için ek bit sayısıdır. Eşitlik 1.8 çapraz entropi kaybı L 'nin matematiksel tanımıdır; burada $p(y_i)$ gerçek referans değer sınıfı y_i 'nin olasılığıdır. Ağ çıktıları z_i 'den verilen görüntü için gerçek referans değer sınıfının olasılığını hesaplamak için Eşitlik 1.7'teki SoftMax fonksiyonu kullanılır. Gerçek referans değer sınıfının daha yüksek olasılığı çapraz entropi kaybını azaltır. Semantik segmentasyon bağlamında, her pikseli sınıflandırmak için, genellikle her piksel için gerçek referans değer sınıfının olasılığı hesaplanır ve ağı optimize etmek için çapraz entropi kaybı kullanılır.

$$L = \sum_{i=1}^n y_i(p(y_i)) \quad (1.8)$$

1.11.2. Optimizatörler

Bir kayıp işlevi modelin eğitim verilerine ne kadar iyi uyduğunun kalitesini ölçer. Düşük kayıplar elde etmek için kayıp fonksiyonunun minimum olması için öğrenilebilir parametrelere ince ayar yapmak için uygun bir optimize edici uygulamak çok önemlidir

böylece kayıp işlevi en aza indirilir. Eğitim verilerini uydurmanın genel şekli geniş bir rastgele değerler koleksiyonu oluşturmaktır ve en iyi model deneme yanılma yoluyla elde edilir. En iyi öğrenilebilir parametreleri kapsamlı bir şekilde aramak yerine kayıp fonksiyonunun gradyanı (Eşitlik 1.3) minimum kayba doğru bir yön sağlar.

Çok büyük miktarda eğitim verisinin mevcut olduğu durumlarda, bellek sınırlamaları nedeniyle tüm örnekleri kullanarak tek bir güncelleme adımı gerçekleştirmek mümkün değildir. Bu nedenle model parametrelerini güncellemek için çevrimiçi öğrenme kullanılabilir. Çevrimiçi bir kurulumda, $\nabla L(W)$ gradyanı N eğitim örneğinin bir alt kümesiyle tanımlanır.

$$\nabla L(W) = \sum_{n=1}^N \nabla L_n(W) \quad (1.9)$$

Çapraz entropi fonksiyonu, dışbükey bir fonksiyon olan Kullback-Leibler diverjansının [77] bir parçası olduğu için nispeten büyük bir öğrenme oranı uygulamak modelin optimal noktasını kaçırma olasılığının yüksek olmasına yol açar. Tersine nispeten küçük bir öğrenme oranı kayıpta sürekli bir azalmaya yol açar. Ancak modelin yakınsama açısından en uygun noktaya ulaşması zaman alır.

Öte yandan böyle bir gradyan iniş algoritması, tüm bir veri setinin gradyanlarını hesaplayarak öğrenilebilir parametreleri günceller. Eğitim verilerinin boyutu büyük olduğunda tek bir güncelleme hala uzun bir zaman gerektirir [78].

1.11.2.1. Stokastik Gradyan İniş

CNN'leri eğitmek için yaygın bir optimizasyon yöntemi momentumlu stokastik gradyan inişidir (SGD). SGD, her eğitim örneğine veya verilere dayalı olarak kayıp işlevinin gradyanını günceller. Tüm eğitim veri seti arasında fazla hesaplamalar gerçekleştiren geleneksel gradyan inişinin aksine SGD, sınırlı veri boyutunda sık hesaplamalar gerçekleştirerek çok daha hızlıdır. SGD, tüm eğitim veri seti için kaybın mutlak minimum noktasını bulmak yerine her eğitim numunesi veya paket (batch) için yerel minimum noktalara yakınsayarak modeli optimize eder. Aktivasyon katmanı doğrusallık

sağlamadığından çapraz entropi kaybı dışbükey olmayan bir yüzeye sahip olmaktadır. Dışbükey olmayan yüzey ise genellikle yerel minimum noktaları içermektedir.

Bu nedenle SGD'deki yerel gradyanların sık sık güncellenmesi modelin daha iyi bir kayıpla yeni bir yerel minimuma atlamasını sağlar ve model sonunda mutlak minimuma yakınsar. SGD'nin güncelleme kuralı şu şekilde tanımlanır:

$$\begin{aligned} V^{(t+1)} &= \beta V^t - \alpha \nabla L_n(W^t) \\ W^{(t+1)} &= W^t + V^{(t+1)} \end{aligned} \tag{1.10}$$

Momentum terimi V , önceki güncellemeleri hatırlar ve yerel minimum veya eyer noktalarından kaçmaya yardımcı olur.

SGD, toplu paket (batch) gradyanları ifade eder ve kayıplarını matematiksel olarak dışbükey olmayan bir yüzey şeklinde temsil edilir, böylece SGD aşırı yüklenmeye devam eder. Gradyanın yönünü kontrol etmek için kademeli bir öğrenme oranı uygulanmaktadır. Bu oran eğitimin erken aşamasında nispeten büyük bir öğrenme oranı uygular, bu nedenle yakınsama hızlıdır. Birkaç eğitim yinelemesinden sonra öğrenme oranı büyüklük sırasına göre azaltılır. Öğrenme oranındaki azalma, en düşük kayıplı en iyi yerel minimumun tutarlı bir şekilde bulunmasına yol açar.

Öğrenme hızının düzgün bir şekilde azaltılması için üstel bir azalma algoritması, her dönemden sonra öğrenmeyi kademeli olarak azaltır ve her dönem modelin tüm eğitim veri setini eğitmesi için toplam yineleme sayısını ifade eder. Öğrenme oranının üstel düşüş şeklinde ifade edilebilir:

$$\mu = \mu \cdot e^{-\gamma t} \tag{1.11}$$

Burada γ , düşüş oranını (decay rate) ifade eden bir hiper parametredir ve t yineleme sayısıdır.

1.11.2.2. Adaptif Optimizatörler

Adaptif Optimizatörler (Adam [79], AdaGrad [80] veya RMSProp [81]), gradyanların [78] büyüklüğüne bağlı olarak öğrenme hızını uyarlanabilir şekilde ayarlar. Uyarlanabilir iyileştiriciler, SGD gibi uyarlanabilir olmayan iyileştiricilere kıyasla daha verimli gradyan güncellemeleri gerçekleştirmek için öğrenme hızını dinamik olarak değiştirir. Bununla birlikte uyarlanabilir optimize ediciler öğrenme oranının yerel minimum noktalardan birinde küçülmesine neden olur, böylece model hiçbir zaman en iyi optimal kayba ulaşmaz.

1.11.2.3. Geri Yayılım

SGD gibi optimize ediciler, kayıp fonksiyonunda minimum kayıp elde etmek için uygun yönü gösterir.

Algoritma1	Geri Yayılım
while	mevcut iterasyon < maksimum iterasyon VE kayıp fonksiyonunun gradyanı < istenen kriter ise do
	for Eğitim veri setinden girdi örneği do
	Optimize edici aracılığıyla kayıp fonksiyonunun gradyanını hesaplayın
	for Her katman çıkış katmanından geriye doğru ilerler do
	Katmanın gradyanını hesaplayın
	endfor
	for Her katman çıkış katmanından geriye doğru ilerler do
	Katmanın gradyanını hesapla
	Önyargı parametresini güncelle
	for Katmandaki her bir gizli nöron için do
	Gizli nöronun gradyanını hesapla
	Ağırlık parametresini güncelle
	endfor
	endfor
	endfor
	endwhile

Her katmandaki ağırlıkları ve önyargıları değiştirmek için bir geri yayılım [59], her katmanın gradyanlarını bulmak için etkili bir yol sağlar ve kayıp fonksiyonundan CNN'nin ilk katmanına gradyanları yinelemeli olarak hesaplar. Her katman için gradyan, türetmenin zincir kuralı ile tahmin edilir (Eşitlik 1.12).

$$\frac{d}{dx}[f(u)] = \frac{d}{du}[f(u)] \frac{du}{dx} \quad (1.12)$$

Algoritmalar, geri yayılım algoritmasını sunar. Geri yayılım, SGD gibi optimize edici aracılığıyla kayıp fonksiyonunun gradyanını hesapladıktan sonra kullanılır.

1.12. Evrimsel Sinir Ağlarında Düzenleştirme

Bir sinir ağının eğitilmesi, bir eğitim setine göre maliyet fonksiyonunu en aza indirir. Ancak nihai hedefimiz, görünmeyen örnekler üzerinde çıkarım süresi boyunca iyi performans göstermektir. Eğitim hatasını azaltmak yerine bir modelin daha iyi genelleştirilmesine izin vermek için kullanılan tekniklere düzenleştirme teknikleri denir.

1.12.1. Ağırlık Düzenlemesi

Yaygın bir düzenleştirme tekniği ağırlık bozunumunu kullanmaktır. Ağırlık düzenlestirmesi, derin öğrenme modelindeki model parametrelerinin (θ) bağımsız olarak değer bakımından çok büyük olmasını kısıtlar ve sıfır olmayan parametreleri cezalandırır. Abartılı büyük bir parametre değeri, tipik olarak, modelin veri setini (veri dağılımını değil) öğrendiğini gösterir, bu da zayıf genelleştirilebilirliğe neden olur ve bu tür değerleri bastırarak sınırlandırılabilir. Sıfır olmayan parametreleri cezalandırarak, seyrek ve yapısal ağlar geliştirilir. Pratikte, ağırlık düzenlemesi yaygın olarak, sıfır olmayan parametreleri cezalandıran L1 normunu ve büyük parametre değerlerini cezalandıran L2 normunu $J(\theta)$ amaç fonksiyonuna dahil ederek elde edilir [82].

1.12.2. Bırakma

Bırakma düzenlemesi (Dropout), eğitim sırasında ağıın nöronlarını p olasılığı ile düşürerek aşırı uyum ve birlikte adaptasyonu önleyen bir tekniktir. Bir nöronun kapatılması, sinir ağı katmanlarının nöronları arasındaki ara bağlantının değiştirilmesiyle sonuçlanır. Böylece kalan birimlerin her birinin diğer komşu nöronlar tarafından öğrenilen değerlere ulaşmasına zorlanır. Bunun yerine çıkarım zamanında tüm nöronlar kullanılır, ancak dış bağlantılarla ilişkili ağırlıklar, eğitim sırasında kullanılan azaltılmış birim sayısını hesaba katmak için seçilen birimlerin düşme olasılığı ile çarpılır.

1.12.3. Toplu Normalleştirme

Toplu normalleştirme, her katmanın aktivasyonlarına uygulanan bir ağı girdileri standart hale getirmek için kullanılan bir tekniktir. Bir sinir ağıının eğitimi sırasında, her katmandaki ağırlıkların dağılımı değişmeye devam eder, buna iç değişken kayması denir. Bu olguya, eğitimin yavaş yakınsamasına neden olur. Bu nedenle ağırlıkların dağılımını her yinelemede aynı tutmak ağıın eğitimi hızlandırmaya yardımcı olabilir. Daha yüksek öğrenme oranına izin verir ve ağırlıkların zayıf başlatılması üzerindeki etkiyi azaltır. Normalleştirme, paket genelinde ortalama ve varyans hesaplanarak yapılır. Pratikte, her katmanın aktivasyonlarını kesinlikle sıfır ortalama ve birim varyans olarak sınırlamak ağıın ifade gücünü sınırlayabilir. Bu nedenle pratikte toplu normalleştirme ağıın ortalama ve varyans için arzu ettiği değerleri öğrenmesine izin verebilecek normalleştirilmiş değerlerin kayması ve ölçeği için parametreleri öğrenmesine izin verir.

Her katman girişine Eşitlik 1.13 uygulanır, burada $\hat{x}^{(i)}$ 'nin toplu normalleştirilmiş öznitelik vektörüdür ve $\mathbb{E}_{\mathbb{B}}[.]$ ve $\text{Var}_{\mathbb{B}}[.]$, B kümesine göre beklenti ve varyansı gösterir.

$$\hat{x}^{(i)} = \frac{x^i - \mathbb{E}_{\mathbb{B}}[x^i]}{\sqrt{\text{Var}_{\mathbb{B}}[x^i]}} \quad (1.13)$$

1.12.4. Veri Artırma

Bir modelin genelleme özelliklerini artırmak için test süresi boyunca karşılaşılabilecek tüm farklı senaryolar üzerinde onu eğitmek avantajlıdır. Bu orijinal olarak gözlemlenen verilerle ilgili verileri yapay olarak üreterek yapılabilir. Görüntüler için bu genellikle dikey çevirme, yatay çevirme, döndürme, ölçeklendirme veya parazit veri ekleyerek yapılır. Eğitim verilerinin miktarını yapay olarak artırmak için veri büyütme kavramı, özellikle sınırlı büyüklükteki veri setleriyle ilgilidir.

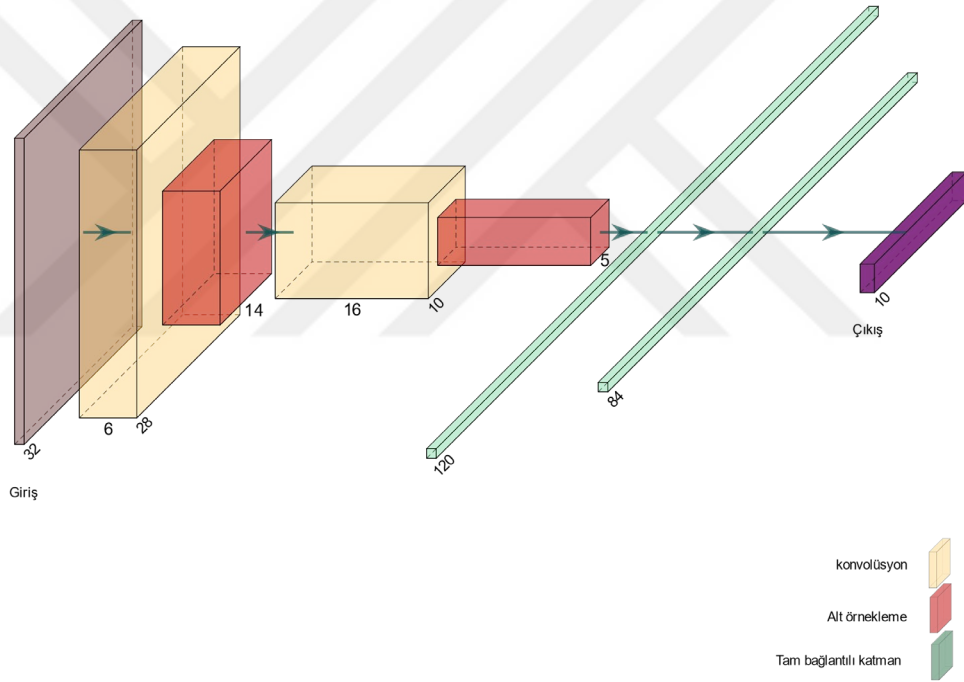
1.13. Omurga Mimarileri

Derin öğrenme modelleri, bilgisayarlı görme literatüründe temel bir teknik haline gelmiştir. AlexNet derin mimarisi [14], 2012'de ImageNet tanıma zorluğunda [83] geleneksel bilgisayarlı görme yöntemlerinden daha iyi performans gösterdiğinden beri derin öğrenme modelleri bilgisayarlı görme araştırmasının temel bir tekniği haline gelmiştir. Bununla birlikte öncü çalışmalar ve sinir ağlarının potansiyelinin ilk kanıtları, derin ağların öğrenmesini sağlayan geri yayılım [59] algoritmasının icadıyla 1986'ya kadar uzanır.

Bu bölümde, çok çeşitli öğrenme problemleri için özellik çıkarıcı olarak kullanılan derin öğrenme mimarileri hakkında bir literatür taraması bulunmaktadır.

Hinton'un 1986'da ve Lecun'un 1998'de sırasıyla geri yayılım [59] ve el yazısı tanıma evrişim ağı [72] hakkındaki öncü çalışmalarından bu yana, sinir ağlarının geliştirilmesine yönelik ana tasarım felsefesi, bunları katman benzeri bir yapı izleyerek uygulamak olmuştur. Her katman genellikle bir dizi matematiksel işlem ve isteğe bağlı olarak optimize edilecek bir dizi parametre ile temsil edilir. Katmanlar üst üste sıralandığından işlemlerin gerçekleştirilme sırası doğal olarak tanımlanır. Bu tür bir hesaplama sırası hesaplamalar arasındaki tüm bağımlılıkların mantıksal bir temsilini de sağlayan bir hesaplama grafi ile matematiksel olarak temsil edilir. Bu nedenle bir dizi X girişi ve bunlarla ilişkili gerçek referans değerler T verildiğinde ağ optimize edilmiş parametre kümesine dayalı olarak bir Y tahmini çıktısı verecek şekilde inşa edilir. Bu optimizasyon, gerçek referans değerler ile ağın tahminleri arasındaki hatanın bir ölçüsünü sağlayan L kayıp fonksiyonunu (seçilen öğrenme probleminin özelliği) en aza indirecek şekilde gerçekleştirilir. Günümüzde sinir ağları, dizi

[60,84, 85], evrişim ve graf [86] tabanlı sinir modelleri üzerine yapılan araştırmalarla çok çeşitli alanlarda uygulama bulmuştur, ancak hala yapıya benzer aynı katmanı paylaşmaktadırlar. Nitekim benzeri bir yapı 1989 yılında Yann Lecun tarafından el yazısı rakam tanıma için önerilen LeNet evrişimli ağlar üzerine yapılan ilk çalışmalardan bu yana derin bilgisayarlı görme modelleri üzerine araştırmaların merkezinde yer almıştır. LeNet-5'te evrişimli katmanlar, alt örnekleme katmanlarına değiştirilir. Her çiftten sonra çıktı değerlerini $[-A, A]$ aralığında ölçeklemek için bir hiperbolik tanjant işlevi kullanılır; burada A , hiperbolik tanjantın genliğini düzenleyen bir ölçekleme katsayısıdır. Tahminler, Şekil 1.14'te görüldüğü gibi, son alt örnekleme katmanına eklenen tamamen bağlı katmanların sonuçlarıdır.

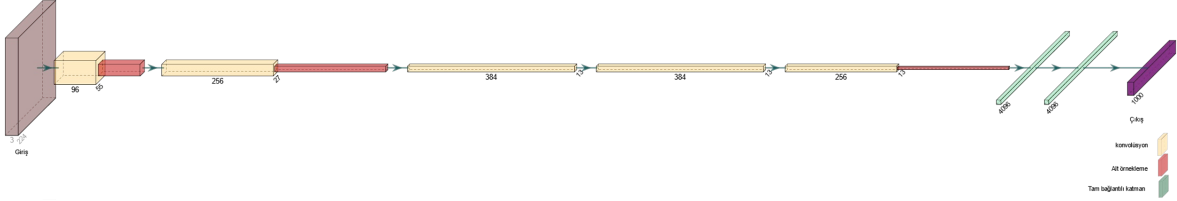


Şekil 1.14. LeNet mimarisi

Sinir ağı mimarisinde konvolüsyon-alt örnekleme-aktivasyon örüntüsünün bir blok olarak birden fazla kez tekrarlanması, o zamandan beri aşağıdaki çalışmalarda da kullanılan bir tasarım örüntüsü haline gelmiştir.

Ağ katmanlarının birbiri ardına eklenecek şekilde özel bir tasarım biçimi olan AlexNet mimarisinde, Krizhevsky ve ark. [14] benzer tasarım ilkelerini izlemiştir. Şekil 1.15'te görüldüğü gibi yalnızca birinci, ikinci ve beşinci katmanlar maksimum havuzlama ile alt örnekleme izlediği beş evrişim katmanından oluşur. Birinci ve ikinci maksimum

havuzlama katmanını ayrıca yeni bir yerel normalleştirme katmanı takip eder. Tüm evrişimli katmanları ve tamamen bağlı olanları, hiperbolik tanjant işlevini kullanan LeNet-5'e göre bir değişiklik olan ReLU doğrusal olmama durumu izler.



Şekil 1.15. AlexNet mimarisi

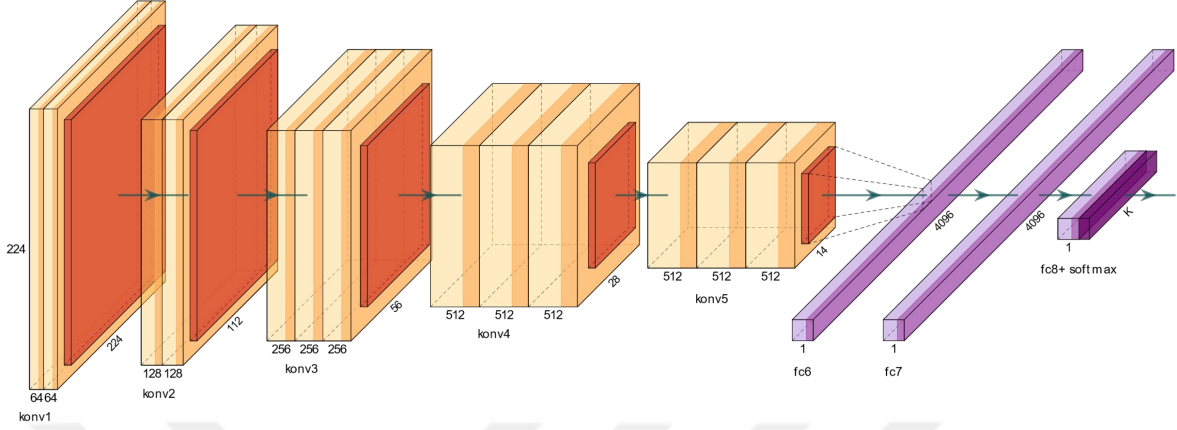
Bu aktivasyon fonksiyonlarından ilkinin, [14]'de belirtildiği gibi, öğrenme prosedürüne yardımcı olduğu bulunmuştur. Çünkü sınırsız olan ReLU ile girdilerin doygunluğu gerçekleşemezken, hiperbolik tanjant, değer aralığı $[-A, A]$ arasında sınırlı olduğu için girdileri çok büyük olursa bilgi kaybına yol açabilir.

AlexNet ile eş zamanlı olarak Zeiler ve ark. 2013'te yayımlanan makalelerinde [87] sinir ağlarının iç işleyişi hakkında yeni bilgiler sağlamıştır. Bu bilgiler derin öğrenme mimarisi araştırmaları üzerine sonraki çalışmalar için temel olmuştur.

Araştırma, girişlerden çıkışlara kadar konvolüsyonel katmanların, ağı eğitildiği veri setinden giderek daha karmaşık görsel kalıplar öğrendiğini göstermiştir. Bu bilgi, her bir evrişimsel katmandaki filtrelerin öğrenilen parametreleri içinde temsil edilir. Her iki çalışmada da belirtildiği gibi, konvolüsyonel filtrelerin üst üste yerleştirilmesi, önceki konvolüsyonel katmanların çıktılarıyla konvolüsyonlanan filtrelerde daha geniş bir alıcı alan elde etme etkisine sahiptir. Bu nedenle Zeiler ve ark. [87] araştırması derin modellerin derinliğinin artırılmasının daha da karmaşık görsel örüntülerin öğrenilmesini sağlayacağını ve böylece muhtemelen daha yüksek sınıflandırma performansı elde edileceğini ortaya koymuştur. Burada derinlik, konvolüsyonel ağı katman sayısı olarak düşünülmüştür.

Beklendiği gibi, aşağıdaki en iyi performans gösteren model VGGNet [88] (Şekil 1.16), daha yüksek sınıflandırma doğruluğu elde etmek için sinir ağı derinliğini artırma fikrine odaklanarak oluşturulmuştur. Bu modelin AlexNet'e göre bir diğer iyileştirmesi, daha fazla ağı derinliği elde etmek ve ağı eğitilebilir parametrelerinin sayısında büyük bir azalma

sağlamak için en fazla 3×3 evrişimli filtrelerin kullanılmasıdır: AlexNet 11×11 , 5×5 ve 3×3 boyutlarında daha geniş çeşitlilikte konvolüsyonel filtreler kullanmıştır.



Şekil 1.16. VGGNet mimarisi

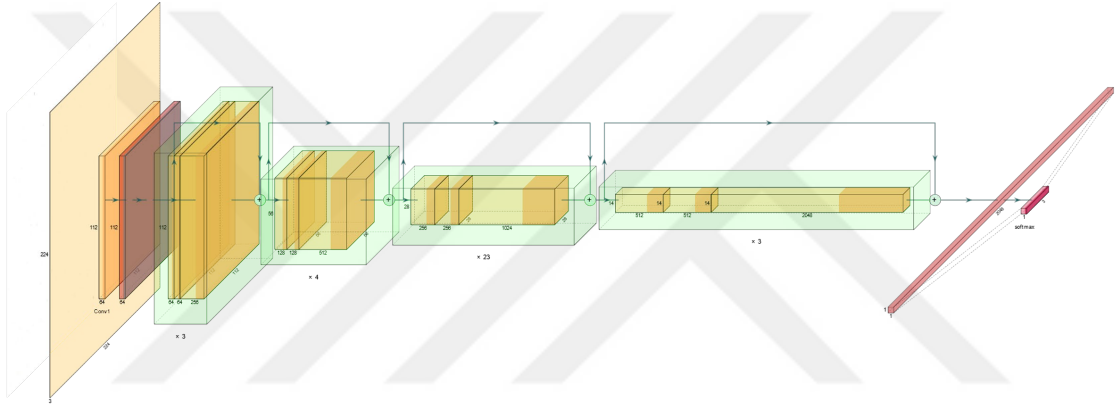
GoogleNet [89], ağdaki bağlantıları atlayarak mini bir modül oluşturan ve bu modülün ağ boyunca tekrarlandığı başlangıç modülleri adı verilen bir modül önermiştir. Başlangıç modülleri, farklı çekirdek boyutuna sahip birkaç evrişimsel katmandan oluşmaktadır. GoogleNet, başlangıç modüllerinin dahil edilmesiyle çok daha az parametre ile önemli ölçüde daha yüksek performans değerleri elde etmiştir.

Derinliğin artmasıyla birlikte, gradyanın kaybolması/dışarı taşması ve ağı bozulması sorunları, sonraki araştırmaların odak noktası haline gelmiştir. Ağ derinliği arttıkça, eğitim sırasında hataların geri yayılması, geri yayılan gradyanların sayısal düşük akış nedeniyle sıfıra düşme veya çok sayıda yığılmış katmana sahip ağlarda gradyanların geçtiği çok sayıda çarpma nedeniyle sonsuza kadar patlama olasılığının artmasına neden olur. Ayrıca, yüzlerce kat derin mimarinin bozulmuş performans olgusunu gösterdiği, bunun yerine daha sığ, ancak derin muadillerinde görünmediği bulunmuştur [90].

He ve diğerleri tarafından yayınlanan artık öğrenme çatısı (residual learning framework) yönteminin getirdiği paradigma kayması ile [90] 2015 yılında, yukarıda bahsedilen sorunların çoğu çözülmüştür. Önerilen artık yöntem, evrişimli katman gruplarının öğrenme problemini, X girdilerinin genel bir $H(X)$ fonksiyonunun yaklaşıklığından artık bir fonksiyonun öğrenilmesine kaydırır.

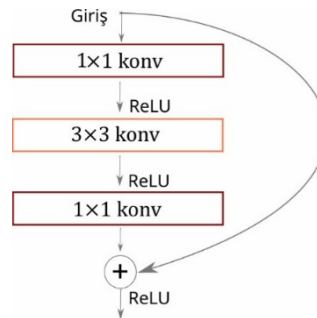
$$F(X) = H(X) - X \quad (1.14)$$

Böylece, konvolüsyonel katmanların her bir alt ağı, yüksek ağ derinliğinde ortaya çıkabilecek keyfi karmaşık örüntüler yerine, girdilerine atıfta bulunan temsilleri öğrendiği için öğrenme kolaylaşır. Aslında, bu nedenle artık öğrenme kullanımı önceki çalışmalara kıyasla çok daha derin ağlara izin vermiştir. VGG-16 ve VGG-19'dan ResNet, büyük ölçüde artan hesaplama maliyetlerine rağmen 200 katmanla uygulandığında bile istikrarlı ve marjinal olarak artan bir performans sergilemiştir.



Şekil 1.17. ResNet101 mimarisi

ResNet mimarisinin (Şekil 1.17) hesaplama karmaşıklığını modülerleştirmek ve azaltmak için bir araç olarak He'nin [90] çalışması, Şekil 1.18 'de gösterilen yeni darboğaz (bottleneck) bloğunu da sergilemektedir.

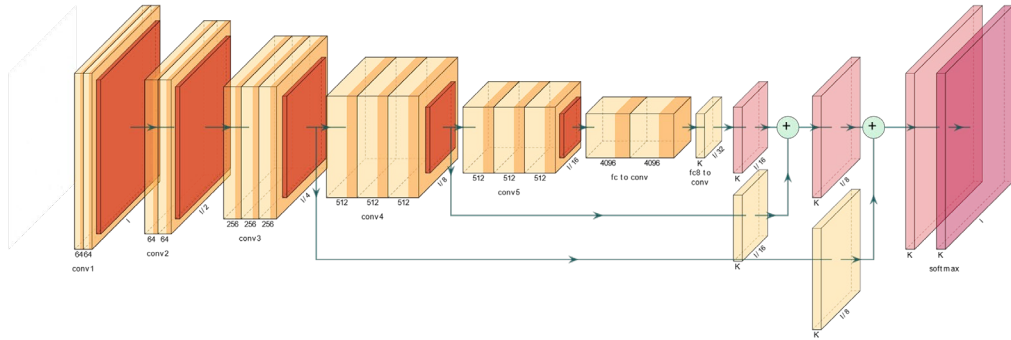


Şekil 1.18. Darboğaz bloğu

Hesaplama açısından daha masraflı olan 3×3 filtrelerin uygulanmasından önce 1×1 konvolüsyonlar aracılığıyla girişlerinin kanal boyutunu azaltan benzeri bir blok içinde artık öğrenme uygular. Daha sonra bunlar, blok çıkışları hesaplanmadan önce tekrar orijinal kanal boyutuna genişletilir. Dolayısıyla, giriş ve çıkış kanal-boyut uyumluluğu çözüldüğünden, bu yapı darboğaz bloklarının seri veya paralel işlenmesi yoluyla derin öğrenme modellerinin modüler eştirilmesine uygundur.

1.14. Özellik Piramidi Ağ Mimarileri

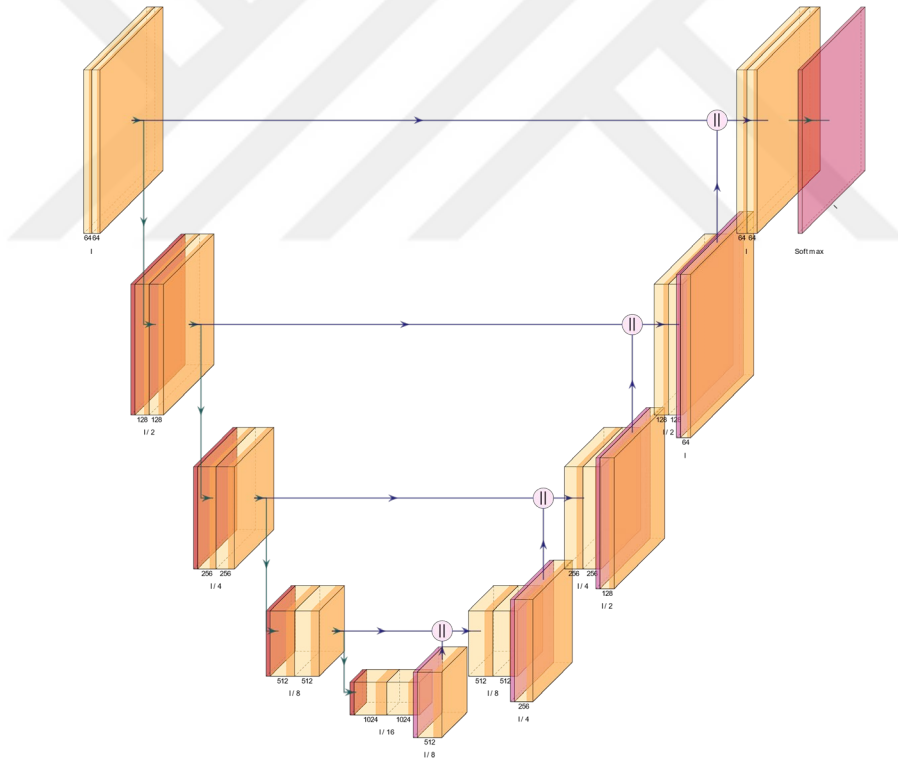
Görüntü piramitleri konsepti tek bir görüntüden çoklu ölçeklerde özellik çıkarma için erken bilgisayar görme yöntemlerine kadar uzanır. Bu tür tekniklerin öncü çalışmaları, görüntü işleme için piramit yöntemleri [91] ve anlamsal özelliklerin çıkarılması için SIFT algoritmasıdır [75]. Sonuncusu görüntü piramitlerinden özellik haritaları oluşturmanın ilk örneklerinden birini sergilemektedir. Çünkü bu tür bir algoritmanın ilk adımlarından biri girdi görüntüsünün birden çok ölçeğinde girdinin iki kopyasının farklılığından kaynaklanan özellik haritaları üretmektir. Görüntü aynı ölçekte ancak her biri farklı standart sapmalara sahip Gauss filtreleriyle konvolüsyona tabi tutulmuştur. Tek bir girdiden oluşturulan görüntü piramidi veri yapısının, dijital görüntü işleme ve analiz tekniklerinin performansını artıran farklı desenlerin bulunmasına izin verdiği bilinmektedir [91]. Bu nedenle, aynı kavramlar bilgisayarla görme için derin öğrenme modellerine de aktarılmıştır.



Şekil 1.19. FCN mimarisi

He ve arkadaşları tarafından 2014 yılında önerilen SPPnet [92], çok ölçekli temsillerden yararlanarak sınıflandırma ve algılama sorunlarının üstesinden gelmek için konvolüsyonel katmanlar tarafından çıkarılan görüntü özelliklerinin piramitlerini kullanan ilk mimarilerden biridir. Benzer yaklaşımlar kısa bir süre sonra Şekil 1.19 'da gösterilen tam evrışimli ağ FCN mimarisi [32] ile anlamsal segmentasyon için uygulanmıştır.

Temelde, semantik segmentasyon performansı üzerindeki yeniliği, en derin katmandan başlayarak farklı sinir ağı derinliklerinde hesaplanan özellik haritalarını, kendileri de daha düşük ağ derinliğindeki öncüllerleriyle birleştirilen yeni ara özellik haritalarına aşamalı olarak birleştirme kabiliyeti ile temsil edilir. Son hesaplamanın sonucu, semantik segmentasyon maskesini tahmin etmek için kullanılan semantik açıdan zengin bir çıktı özellik haritasını temsil eder, böylece tahminlerinin kalitesini artırmak için çok ölçekli özellik bilgilerini kullanır.



Şekil 1.20. U-Net mimarisi

Semantik segmentasyonun temellerinden biri olan U-Net [73] mimarisi (Şekil 1.20), yeni birçok katmanlı kodlayıcı-kod çözücü tam evrışimli ağ aracılığıyla semantik segmentasyon çıktıları üretmek için FCN ile aynı tasarım ilkesini kullanmıştır. Şekil 1.19'de

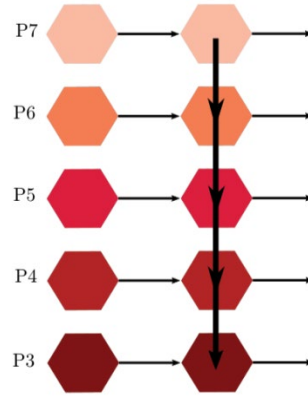
U-Net, kod çözücünün farklı seviyelerindeki yukarı örneklenmiş özellik haritalarını, kodlayıcı-kod çözücü ağ derinliği açısından aynı seviyedeki kodlayıcının özellik haritası ile birleştiren özellik birleştirme mekanizması nedeniyle FCN'ye benzer. FCN'den ayırt edici özelliği, U-Net küçük biyomedikal görüntü veri setlerine uygulanmak üzere tasarlandığından, nispeten sınırlı sayıda örnekle eğitilebilmesidir.

Daha sonra, özellik haritalarının piramitleri kavramı nesne algılama modelleri için de kullanılmıştır. MS-CNN [93] gibi çok ölçekli nesne algılama yaklaşımları, konvolüsyonel ağların yalnızca en derin ve aynı zamanda semantik açıdan zengin çıkış katmanını kullanan ya da giriş görüntüsünü yukarı örnekleyerek aynı görüntünün farklı ölçekteki kopyaları üzerinde algılama yapan önceki yöntemlerin aksine, algılama için özellik piramitlerini kullanmanın yollarını önermiştir [94,95]. Bu yöntemlerden herhangi biri aracılığıyla çok ölçekli temsil kullanma ihtiyacı, bu yöntemlerin çok geniş bir alıcı alana sahip en derin konvolüsyonel katmanların tek başına kullanılması nedeniyle küçük ve orta ölçekli nesneleri tespit etmede etkisiz olmasından [87] veya tespitlerin aynı girdi görüntüsünün farklı ölçeklerdeki birden fazla kopyası üzerinden hesaplanmasının hesaplama açısından verimsiz olmasından kaynaklanmıştır. Yine de FCN gibi bir özellik füzyon mekanizması henüz geliştirilmemiştir.

Konvolüsyonel detektörlerin hesaplama verimsizliğinin yanı sıra tüm ölçeklerdeki nesnelerin tespitindeki düşük performansı çözmek için bir adım daha ileri gidilerek Tek Atışlı Çoklu Kutu Detektörü (Single-Shot multibox Detector) SSD [96] mimarisi elde edilmiştir. Ana ilerlemesi, VGG16 omurga özellik çıkarıcısının çok ölçekli özellik haritaları aracılığıyla tek aşamalı nesne tespiti olmuştur [88].

Derin öğrenme nesne algılama modellerindeki en büyük ilerleme Lin ve ark. [97] tarafından geliştirilen özellik piramidi ağları (FPN) ile sağlanmıştır. Sağlam çok ölçekli nesne algılama yeteneğine sahip bir nöral mimari geliştirmek amacıyla FPN mimarisi, Şekil 1.19 ve Şekil 1.20'de gösterilen FCN [32] ve U-net [73] modelini yansıtan bir dizi tasarım ilkesiyle oluşturulmuştur. Aşağıdan yukarıya ve yukarıdan aşağıya yol olmak üzere iki ana bileşen tarafından tanımlanmıştır. Aşağıdan yukarıya yol, yüksek çözünürlüklü genel ayrıntılardan düşük çözünürlüklü ancak nesne düzeyindeki örüntülere kadar çoklu ölçeklerde çıktı özellik haritaları oluştururken, yukarıdan aşağıya yol, orijinal özellik piramidinin tüm seviyelerinde anlamsal olarak zengin özellikler oluşturmak için bunları birleştirir.

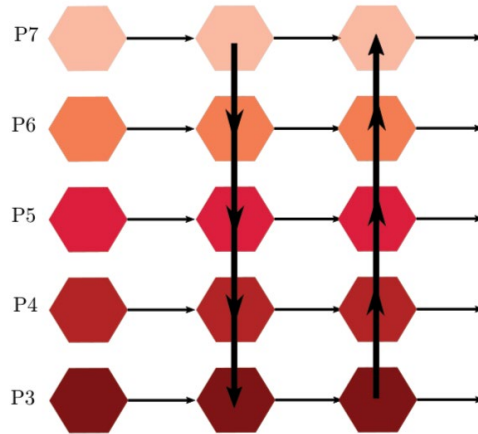
Aşağıdan yukarıya yol, tercih edilen bir omurga mimarisinin seçilen çıktı özellik haritalarından oluşurken, yenilikçi yukarıdan aşağıya özellik hiyerarşisi, aşağıdan yukarıya özellik haritalarının seçimine dayalı olarak tasarlanmıştır. Özellik piramitleri ağını uygulamak için, adından da anlaşılacağı üzere, iki yol arasındaki yanal bağlantılar ve yukarıdan aşağıya yolun komşu seviyeleri arasındaki yukarıdan aşağıya bağlantılar Şekil 1.21'de gösterildiği gibi tasarlanmış ve kullanılmıştır.



Şekil 1.21. FPN mimarisi

FPN [97] tarafından gösterilen umut verici sonuçlar, bu tür modeller üzerinde daha fazla araştırma geliştirilmesini motive etmiştir.

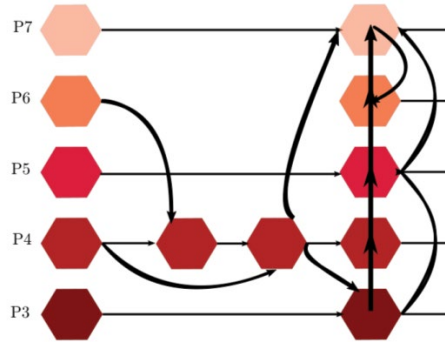
Aşağıdan yukarıya-yukarıdan aşağıya FPN mimarisinin temel bir uzantısı, daha da zengin birleştirilmiş özellik haritaları oluşturmak için önceki yukarıdan aşağıya yoldan inşa edilen başka bir aşağıdan yukarıya yol geliştiren PANet'tir [98].



Şekil 1.22. PANet mimarisi

FPN (Şekil 1.21) ve PANet (Şekil 1.22) arasındaki karşılaştırmada, aşağıdan yukarıya yolun omurga özelliklerini temsil eden bu özellikler, soldaki standart FPN modelinde olduğu gibi, aynı yukarıdan aşağıya yol tarafından birleştirilir. PANet mimarisinin yeniliği, PANet mimarisinin ortasındaki yukarıdan aşağıya yolla elde edilen birleşik özellikleri daha da birleştiren en sağdaki aşağıdan yukarıya yolda gözlemlenebilir.

PANet mimarisinde uygulanan yukarıdan aşağıya aşağıdan yukarıya yolunun uygun şekilde yapılandırılmasıyla, iki yolu içeren bu yeni blok, bu tür bir yapının tekrar tekrar kullanılmasına elverişlidir. Gerçekten de Şekil 1.23'te gösterilen NAS-FPN [99] takip modeli, bu tür bir yapıyı, P3'den P7'ye kadar olan ilk aşağıdan yukarıya yolun seçilen özellik haritalarına eklenen yeni bir derin sinir ağı katmanı türü olarak ele almaktadır. Görülebileceği gibi, yeni katman, tekrarlanan yukarıdan aşağıya aşağıdan yukarıya katmanların önceki birleştirilmiş özellik haritalarının sağ tarafına basit bir şekilde bağlanmasıyla birden çok kez tekrarlanabilir.



Şekil 1.23. NAS-FPN mimarisi

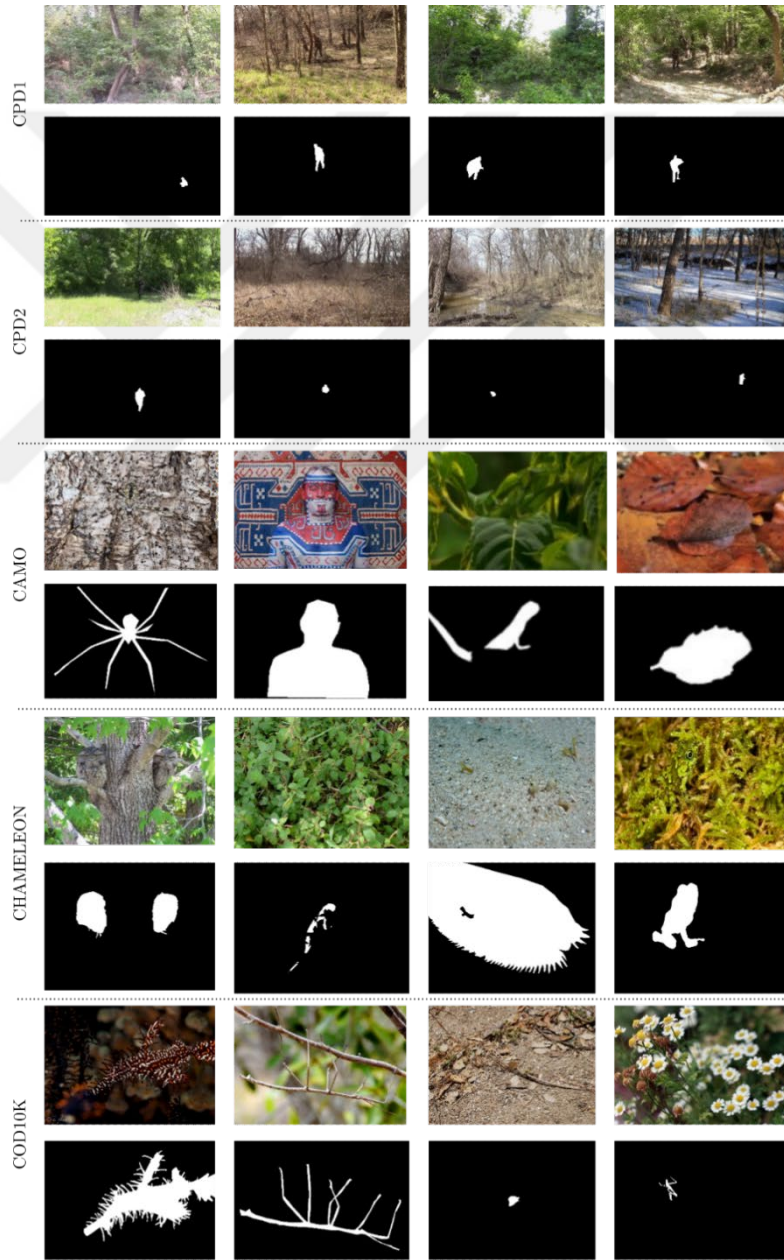
1.15. Veri Setleri

1.15.1. COD İçin Veri Setleri

1.15.1.1. CPD1 Veri Seti

Kamuflajlı insanlar veri seti 1 (CPD1) [48], farklı kamuflaj desenlerine sahip insanları içeren 20 grubun video kliplerinin toplanmasıyla oluşturulmuştur. Her bir doğal sahnede

farklı kamuflaj desenlerinin kamuflaj performansını değerlendirmek için sabit bir kamera bu videolar elde edilmiştir. Veri seti 1000 görüntü içermekte ve farklı kamuflaj desenleri, bol miktarda doğal sahne, farklı aydınlatma ve diğer koşulları, insanların farklı ölçekleri ve duruşlarını içermektedir. Son olarak, video kliplerden 480×854 boyutunda 1000 görüntü seçilmiştir. Tüm görüntüler piksel düzeyinde etiketlenmiştir. Veri seti 1000 görüntü içermektedir. Veri setindeki görüntülerin %80'i eğitim için, kalan %20'si ise test için kullanılmıştır. Şekil 1.24'te karşılık gelen gerçek referans değer etiketleriyle birlikte CPD1'dan birkaç örnek verilmiştir.



Şekil 1.24. Kamufle edilmiş nesne algılama veri setlerinin görüntü örnekleri

1.15.1.2. CPD2 Veri Seti 2

CPD2-Kamufle edilmiş insanlar veri seti 2 [49], CPD1 kamufle edilmiş insanlar veri setinin [44] bir uzantısıdır. Veri seti 1000 görüntüden 2600 görüntüye (854×854) genişletilmiştir, her biri 100 görüntü içeren 26 farklı kamuflej deseninden oluşmaktadır. Veri setinden 400 görüntü test için kullanılmıştır.

1.15.1.3. CAMO Veri Seti

Kamufle Edilmiş Nesne (CAMO) veri seti [2] farklı boyutlarda 1250 görüntü içermektedir. Veri seti, genellikle sırasıyla gerçek dünyadaki hayvanlara ve insanlara karşılık gelen doğal olarak kamufle edilmiş nesneler ve yapay olarak kamufle edilmiş nesneler olmak üzere iki kategoriye odaklanmaktadır: kamufle edilmiş hayvanlar amfibiler, kuşlar, böcekler, memeliler, sürüngenler ve çeşitli ortamlardaki su altı hayvanlarından oluşur; yani yer, su altı, çöl, orman, dağ ve kar.

CAMO veri seti (Şekil 1.24), birden fazla nesne içerir, ayrı tek nesneler ve uzamsal olarak bağlı / örtüşen nesneler içerir, bu da veri setini kamufle edilmiş nesne segmentasyonu için daha zor hale getirir. Burada da veri setindeki görüntülerin %80'i eğitim için, kalan %20'si ise test için kullanılmıştır.

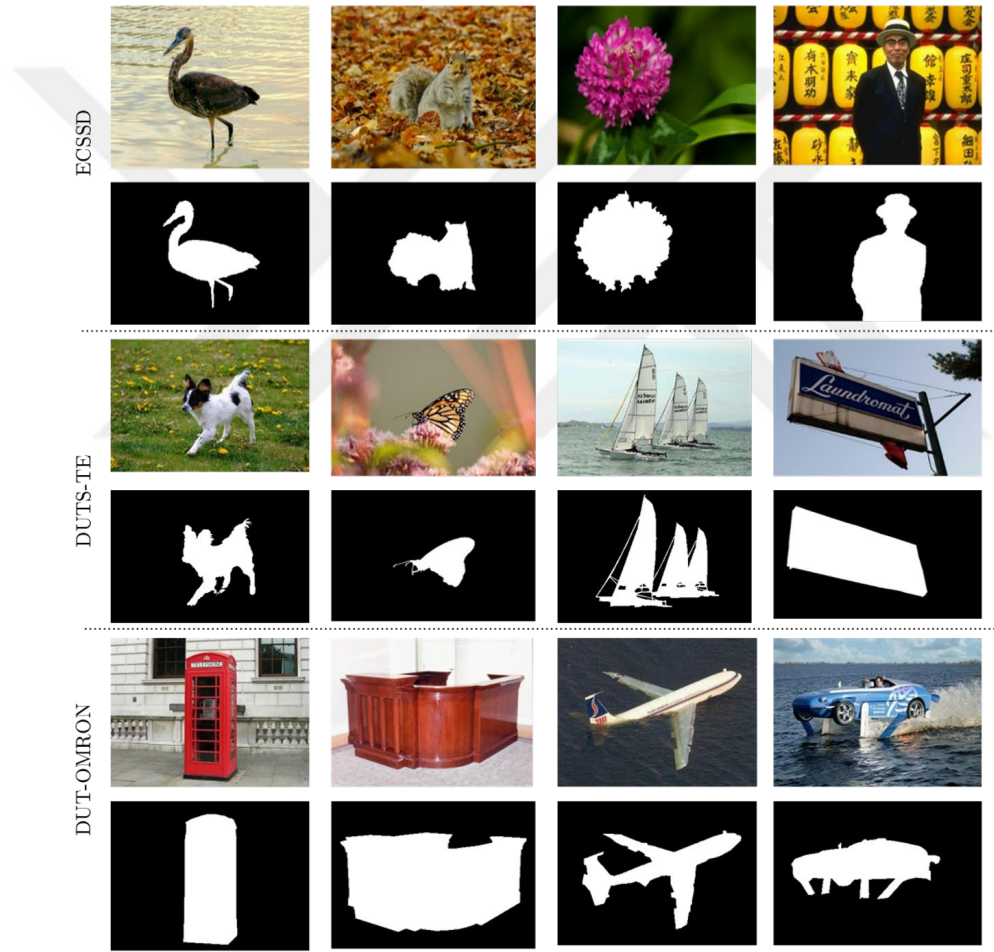
1.15.1.4. CHAMELEON Veri Seti

CHAMELEON veri seti [100] internetten kamufle edilmiş hayvan etiketli 76 yüksek çözünürlüklü görüntü toplamaktadır. Her görüntü, nesne düzeyinde mekâna uygun olarak elle maskelenmiştir. Bu görüntüler, doğal ortamda kamufle olmak için avantajlarını kullanan hayvanlarla ilgilidir. Veri seti yalnızca test için kullanılmıştır.

1.15.1.5. COD10K Veri Seti

COD10K [51] en yeni ve en kapsamlı kamufle edilmiş nesne algılama veri setidir. COD10K, kamufle edilmiş nesneleri arasında karasal canlılar, deniz canlıları, amfibiler vb. bulunan 10.000 görüntü (5.066 kamufleli, 3.000 arka planlı, 1.934 kamufleli) içerir.

1.15.2. SOD İçin Veri Seti

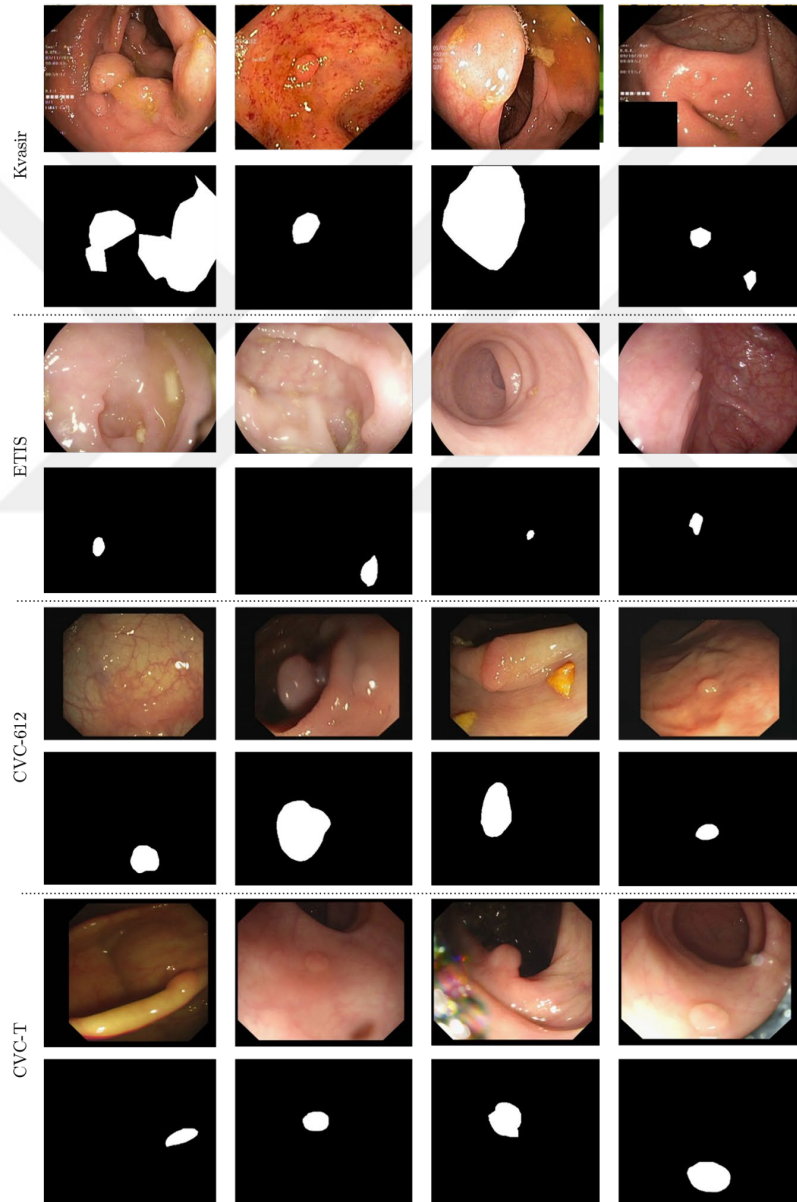


Şekil 1.25. Kamufle nesne algılama veri setleri örnek gösterimi

Tezde SOD için değerlendirmelerin yapılması amacıyla test aşamasında normal nesne görüntülerini içeren ESCCD, DUT-OMRON ve DUTS-TE veri tabanı kullanılmıştır. ESCCD veri setinde [101] semantik olarak anlamlı fakat yapısal olarak karmaşık 1000

seçilmiş görüntü vardır. DUT-OMRON [102], birden çok nesne, yüksek görüntü içeriği varyasyonları ve karmaşık arka planlar içeren 5168 görüntü içerir. DUTS ise [103], eğitim ve test kümelerinde sırasıyla 10.553 ve 5019 görüntü içeren en son yayınlanan büyük ölçekli veri setidir. Hem eğitim hem de test setleri karmaşık sahneler içermektedir.

1.15.3. Tıbbi Segmentasyon İçin Veri Setleri



Şekil 1.26. Polip segmentasyonu veri setleri örnek gösterimi

Tezde önerilen yaklaşımın doğruluğunun kanıtlanması için farklı veri tabanlarında sistemin performansı değerlendirilmiştir; kamufle görüntülerin yanı sıra tıbbi görüntülerin segmentasyonu da test edilmiştir. Bu amaçla polip segmentasyonu ile ilgili 4 çeşit veri seti kullanılmıştır (Şekil 1.26).

CVC-ClinicDB (CVC-612) [104], her biri farklı bir polip gösteren 31 kolonoskopi video sekansından seçilen 612 polip görüntüsü içerir. Kvasir [104], polip segmentasyonu için 1000 görüntü içeren en büyük veri setidir. ETIS [106] ve EndoScene (CVC-T) [107] model performansını değerlendirmek için kullanılmıştır, çünkü bu küçük veri setleri sırasıyla 196 ve 60 görüntü içermektedir.

1.16. Performans Ölçütleri

Önerilen modelin performans değerlendirmesi belirginlik tahmin sonuçları ile farklı veri setleri için gerçek referans değer açıklamaları arasındaki uygunluk derecesinin ölçülmesini gerektirir. Bu bölümde incelenen çeşitli değerlendirme ölçütleri arasında kesinlik-duyarlılık ve birleşme üzerinden kesişim (IoU) gibi örtüşmeye dayalı değerlendirme ölçütleri nesne segmentasyonu üzerine önceki bazı çalışmalarda kullanılmışken, diğerleri ise; hizalamayı geliştirme ölçüsü (E-ölçü) ve yapısal ölçü (S-ölçü) maskelenmiş haritaları daha kapsamlı bir şekilde değerlendirmek için ancak son zamanlarda tanıtılmıştır.

1.16.1. Kesinlik- Duyarlılık

Bu yöntem, bir kabul eşiğini değiştirerek kesinlik ve duyarlılık arasında değiş tokuş yapılmasına izin verir. Daha yüksek bir eşik, daha doğru sınıflandırılmış piksellere yol açar ancak alınan ilgili örneklerin sayısını azaltır.

Kesinlik keşfedilen sınıf etiketlerinden kaçının doğru olduğunu tanımlar:

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (1.15)$$

Duyarlılık, gerçekte bulunan doğru etiketlerin sayısını belirler:

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (1.16)$$

Burada doğru pozitifler için TP, yanlış pozitifler için FP ve yanlış negatifler için FN parametreleri kullanılmıştır.

1.16.2. Ortalama IoU

IoU [108] anlamsal görüntü bölütleme görevleri için standart bir ölçüdür. Her sınıf için, o sınıfa ait olduğu tahmin edilen tüm piksellerin yanı sıra aslında o sınıfa ait olan tüm piksellerin alanının bir kesri olarak belirli bir sınıf için doğru sınıflandırılmış piksellerin alanının bir ölçüsüdür. Bu metrik tüm sınıflar için tüm veri seti üzerinden değerlendirilir. IoU'ların ortalaması tüm sınıflar üzerinden hesaplanır. Eşitlik 1.17'deki ortalama IoU modelin tüm sınıflar için ne kadar iyi segmentlere ayrıldığına dair daha iyi bir fikir verir.

$$\text{ortIoU} = \frac{1}{N} \sum_{i=1}^N \frac{TP(C_i)}{TP(C_i) + FP(C_i) + FN(C_i)} \quad (1.17)$$

Kesinlik ve duyarlılık ile karşılaştırıldığında, IoU puanı doğru pozitifleri, yanlış pozitifleri ve yanlış negatifleri dikkate aldığı için daha düşük puanlarla sonuçlanır.

1.16.3. AUC

Alıcı işletim eğrisi AUC altındaki bir alan, bir sınıflandırma analizi, tüm alıcı işletim eğrisinin alanını ölçer ve AUC, tüm olası sınıflandırma eşiklerinde gerçek pozitif oranı ve yanlış pozitif oranı temel alan sınıflandırma performansını sağlar.

1.16.4. F-Ölçüsü

Bu metrik [109], negatif olmayan bir β ağırlığı altında geri çağırma ve hassasiyetin ağırlıklı harmonik ortalamasını temsil eder. Deneylerde, model performansını değerlendirmek için ortalama F-Ölçümü (ortalama F) kullanılmıştır. Matematiksel olarak F_β şu şekilde tanımlanır.

$$F_\beta = \frac{(1 + \beta^2) \text{ Kesinlik} \times \text{ Duyarlılık}}{\beta^2 \text{ Kesinlik} + \text{ Duyarlılık}} \quad (1.18)$$

[109]'u izleyerek, $\beta^2 = 0.3$ 'ü belirledi. F-ölçü puanı ne kadar yüksek olursa modelin o kadar iyi performans gösterdiğine dikkat edilmelidir.

1.16.5. S-Ölçüsü

Bu, tahmin edilen belirginlik haritası ile gerçek harita arasındaki yapısal benzerliği ölçmek için kullanılan bir değerlendirme metriğidir. [110]'a göre, S_α şu şekilde tanımlanır:

$$S_\alpha = (1 - \alpha)S_o + \alpha S_r \quad (1.19)$$

Burada S_α nesneye duyarlı yapısal benzerliği ve S_r bölgeye duyarlı yapısal benzerliği ifade eder. [110]'u takiben, $\alpha = 0.5$ 'tir. S-ölçüt puanı ne kadar yüksek olursa model o kadar iyi performans gösterir.

1.16.6. Ortalama Mutlak Hata (MAE)

Bu, tahmin edilen belirginlik haritası ile yer gerçeği arasındaki ortalama mutlak hatayı doğrudan hesaplamak için kullanılan bir ölçüdür [111]. *MAE* olarak tanımlanır:

$$MAE = \frac{1}{H \times W} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)| \quad (1.20)$$

Burada H ve W sırasıyla belirginlik haritasının yüksekliğini ve genişliğini gösterir. S tahmin edilen belirginlik haritasını, G ise ilgili zemin gerçeğini göstermektedir. x ve y her bir pikselin koordinatını ifade etmektedir. *MAE* ne kadar düşükse model performansı o kadar iyidir.

1.16.7. E-Ölçüsü

Geliştirilmiş hizalama ölçüsü [112], bir ikili haritanın görüntü seviyesi istatistiklerini ve yerel piksel eşleştirmelerine belirleyen yakın zamanda önerilen başka bir ölçüdür.

$$E = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H \Phi_{FM}(x, y) \quad (1.21)$$

Burada Φ bir ikili haritanın iki özelliğini (piksel düzeyinde eşleşme ve görüntü düzeyinde istatistikler) yakalamak için kullanılan geliştirilmiş hizalama matrisidir. W ve H sırasıyla görüntünün yüksekliği ve genişliğidir; FM ise ön plan haritasıdır.

Model performansını değerlendirmek için ortalama E-ölçümü kullanılmaktadır. E-ölçümü puanı ne kadar yüksekse, model o kadar iyi performans gösterir.

2. YAPILAN ÇALIŞMALAR

Bu bölümde, önerilen modelin modülleri ve algoritmaları ayrıntılı olarak verilmektedir. Çalışmanın genel yapısı Şekil 2.1’de verilmiştir.



Şekil 2.1. Çalışmanın genel yapısı

Bu tezde, COD için temel olarak dört bölümden oluşan yeni bir uçtan uca ağ sunulmuştur: İlk bölüm, giriş görüntüsünün konvolüsyonel özelliklerinin modifiye edilmiş 101 katmanlı bir ResNet [90] tarafından belirlenen özellik çıkarma işleminden sorumlu olan Omurga Ağıdır. İkinci kısım Başlangıç Modülüdür ve çok ölçekli alıcı alanların yeni özellik haritalarını üretmek amacıyla geliştirilmiştir. Üçüncü bölüm ise çok düzeyli anlamsal bilgilerle kaynaştırılmış özellik haritalarını etkin bir şekilde üreten Yoğun Özellik Piramididir. Son olarak, dördüncü bölüm Çoklu Dikkat Modülüdür (MAM). Arka plan konum bilgisinden kamufle edilmiş hedefi belirlemeye odaklanır. MAM tüm özellik haritalarına uygulanır ve ardından tahmin edilen haritalar birleştirilerek nihai çıktı elde edilir.

Bu modele özgü ayrıntılar, aşağıdaki biçimlendirmeden sonra sunulmaktadır.

2.1. Formalizasyon

Bir sinir ağı, her biri sıfır veya daha fazla optimize edilebilir parametre ile ifade edilen $\mathcal{F}_i()$ $i \in \{1, \dots, N\}$ fonksiyonlar kümesi olarak gösterebilir; $W_i \in \mathbb{R}^{d_{i-1} \times d_i}$ ağırlıkları ve $b_i \in \mathbb{R}^{d_{i-1} \times d_i}$ biaslar ve hiperparametreler olarak gösterilir. Bu tür

fonksiyonların sıralı biçimde $X \in \mathbb{R}^{d_0 \times d_1}$ girişlerine uygulanması ağı temsil eden bir bileşik fonksiyonu dolaylı tanımlar.

$$\mathcal{F}_i(\cdot) = \mathcal{F}_n \circ \mathcal{F}_{n-1} \circ \dots \circ \mathcal{F}_1(\cdot) \quad (2.1)$$

Dolayısıyla, $\theta = \{W_1, \dots, W_N, b_1, \dots, b_N\}$ parametre kümesine bağlı olan bileşik fonksiyon $\mathcal{F}_\theta, X$ girişlerini Y çıktılarına aşağıdaki şekilde eşler

$$\mathcal{F}_{W_1, \dots, W_N, b_1, \dots, b_N}: \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_n} \quad (2.2)$$

Bu tür karmaşık yapıları doğrusal olmayan bileşik fonksiyonun amacı, optimize edilebilir parametreleri vasıtasıyla, doğrudan gözlemlenemeyen veri dağılımı $P(X, Y)$ 'den M adet (x_m, x_n) $m \in \{1, \dots, n_örnekler\}$ olarak örneklenen X girdileri ile Y gerçek referans değerleri arasındaki eşleşmeye yaklaşmaktır.

Sinir ağlarının tasarımı, girdi verilerinden ilgili kalıpları otomatik olarak öğrenmek için bir veya daha fazla omurga özellik çıkarıcı (çok modlu veriler veya bir model topluluğu durumunda ikincisi) ve omurga çıktılarına keyfi bir şekilde bağlanan bir veya daha fazla göreve özgü dal kullanmak için ortak yol gösterici ilkeyi tercih etmektedir.

Çok görevli öğrenme ortamında, seçilen tüm görevler için arzu edilen temsilleri geri yayılım yoluyla otomatik olarak öğrenmekle görevlendirilen ortak bir omurga özellik çıkarıcısı tanımlamak yaygın bir uygulamadır.

Böyle bir durumda, sinir ağının daha önce bahsedilen bileşik fonksiyon temsili bu genel ayara uyum sağlayacak şekilde güncellenebilir.

Aşağıdaki tanım verilebilir:

- $\theta_{\mathcal{F}}$ omurganın öğrenilebilir parametreleri olsun
- $\theta_{\mathcal{G}_i}$ i . göreve özgü dalın öğrenilebilir parametreleri olsun

Ayrıca:

- $\mathcal{F}(\cdot; \theta_{\mathcal{F}})$ omurga modelini
- $\mathcal{G}_i(\cdot; \theta_{\mathcal{G}_i})$ $i \in \{1, \dots, n_görev\}$ göreve özgü başlıkları temsil etsin

Sinir ağı çıktıları, hesaplamalar arasındaki ilişkili bağımlılıklarla birlikte tanımlanabilir:

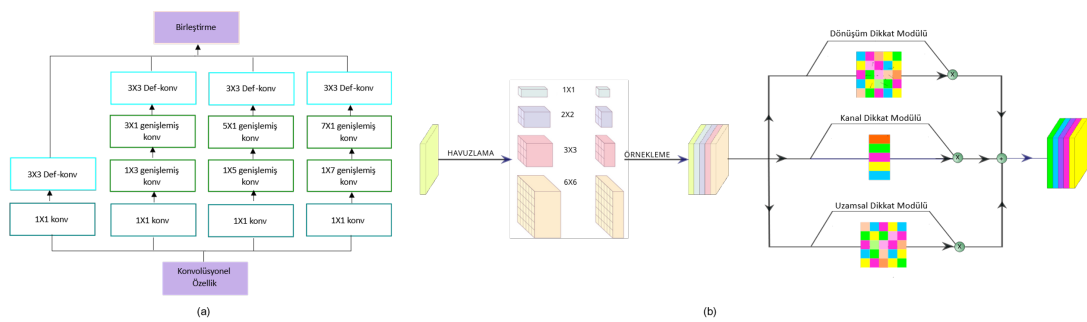
$$\begin{aligned} Z &= \mathcal{F}(X; \theta_{\mathcal{F}}) \\ Y_i &= \mathcal{G}_i(\mathcal{F}(X; \theta_{\mathcal{F}}); \theta_{\mathcal{G}_i}); i \in \{1, \dots, n_{-}\} \end{aligned} \quad (2.3)$$

Burada Z , ilgili tahminleri oluşturmak için her bir $\mathcal{G}_i$ göreve özel kafayı besleyen omurga modelinin ara çıktısıdır.

2.2. Önerilen Yöntem

İnsanlar belirli sahnelerde hedef nesneyi ararken, çoğu ilk olarak hedef nesnenin menzil içinde var olup olmadığına karar verir ve yaklaşık konumunu belirlerler. İkinci adım, hedef nesnenin etrafındaki sahneyi analiz etmek ve nesnenin tamamını daha doğru bir şekilde tespit etmek için hedefi iyileştirmektir. Görsel mekanizmadan esinlenen bu çalışma, kamufle edilmiş nesne tespiti amacına ulaşmak için “insan gözünü” taklit eden ResNet101'i omurga olarak kullanmaktadır. Gözlem sürecinde, farklı aşamalarındaki özellik haritaları görsel geri alma sürecini temsil eder.

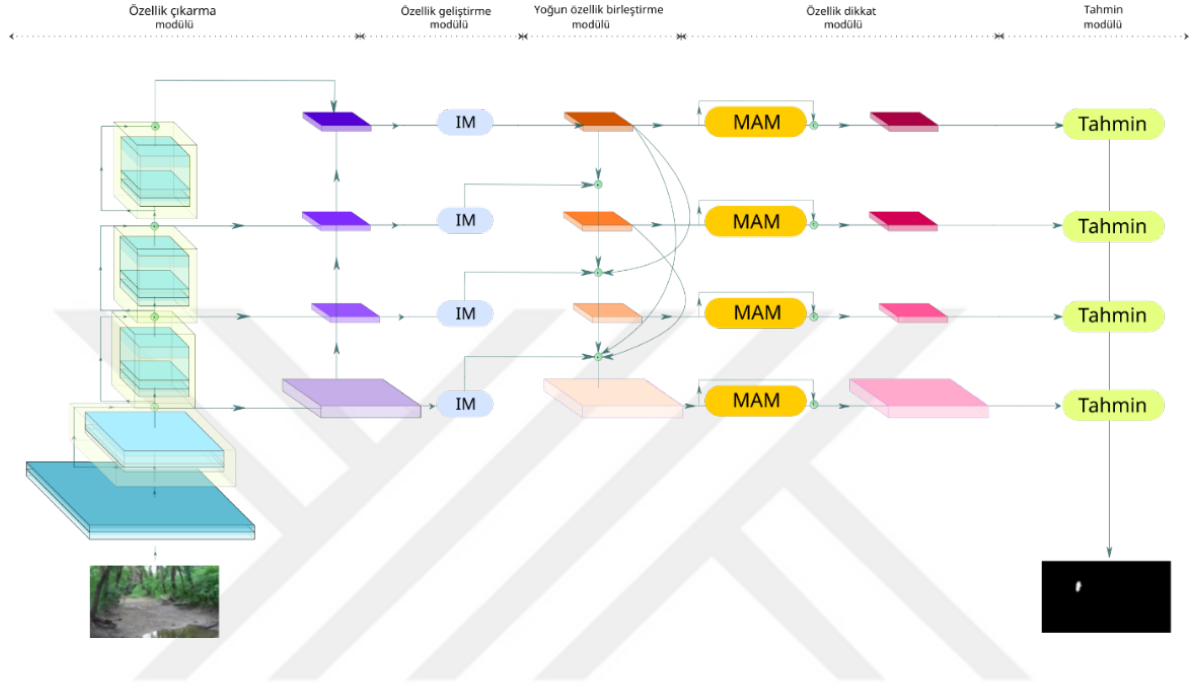
İnsan görsel gözlem mekanizmasına benzer şekilde, önerilen modelin kamufle edilmiş nesneleri tespit etme süreci de kabaca iki ana adıma ayrılabilir.



Şekil 2.2. a) IM ve b) MAM modülleri

İlk adım, elde edilen özellik bilgilerinin başlangıç modülüne (IM- Inception Module) girilerek özellik bilgilerinin geliştirilmesi, görüntülerde kamufle edilmiş nesnelerin olup olmadığına karar verilmesi ve kamufle edilmiş nesnelerin kabaca konumlandırılmasıdır.

İkinci adım, geliştirilmiş bilgi hala gürültü ve parazit bilgisine sahip olduğundan, MAM, görüntülerdeki kamufle edilmiş nesneleri doğru bir şekilde tespit etmek, segmentlere ayırmak ve hedef nesneleri tamamen belirlemek için çeşitli katmanların bilgilerinin entegrasyonu için güçlendirilmiş ipuçlarını kullanır.



Şekil 2.3. Önerilen yöntemin genel mimarisi

Önerilen yöntem, giriş görüntüsünün konvolüsyonel özelliklerinin modifiye edilmiş çok katmanlı bir ResNet101 tarafından çıkarıldığı Omurga Ağından özelliklerin elde edilmesiyle başlar. Daha sonra bu özellikler bir başlangıç modülüne dayalı olarak geliştirilir. Bu modül, çok ölçekli alıcı alanların yeni özellik haritalarını üretmek amacıyla geliştirilmiştir. Çıktı olarak alınan özellikler yoğun bir şekilde birbirine bağlanır ve üst düzey özellik haritalardan düşük düzey özelliklere daha fazla bağlamsal bilgi eklemek için yukarıdan aşağıya bir yolla birleştirilir. Arka plan konum bilgisinden kamufle edilmiş hedefi vurgulamaya odaklanmak amacıyla, daha hassas özellikler oluşturmak üzere Çoklu Dikkat Modülü tanıtılmıştır. Üç dikkat bloğu (dönüşüm, kanal ve uzamsal) paralel olarak uygulanmaktadır. Dönüşüm dikkat modülü küçük ölçekli gizli nesnelerin tespit edilmesinde, kanal dikkat modülü ise çok ölçekli zengin anlamsal özelliklerin çıkarılmasında önemli ölçüde fayda sağlar. Uzamsal dikkat modülü faydalı derin özellikler çıkarır. Tüm bloklardan

elde edilen rafine özellikler nihai bir dikkat haritasında birleştirilir. Sonuncusu ise evrimsel özelliklerle birleştirilir. Son olarak, bölümlenmiş maske her katmanda tahmin edilir ve tahmin edilen tüm haritalar ağ çıktısı olarak nihai bir bölümlenmiş maske için birleştirilir. Önerilen model, her katmanda kamufle edilmiş bir haritayı art arda tahmin eder ve nihai kamufle edilmiş haritayı ağ çıktısı olarak alır (Şekil 2.3).

2.3. Aşağıdan Yukarıya Yöntemiyle Özellik Çıkarma

Önerilen yöntem mimarisinin omurgası bir FPN modeli olduğundan, bu modelin ana bileşenlerini oluşturmak için kullanılan özellik çıkarıcısının tanımlanmasını gerektirir. Bu özellik çıkarıcılar aşağıdan yukarıya ve yukarıdan aşağıya olmak üzere 2 şekilde yapılmaktadır. İlki, seçilen FPN özellik çıkarıcısı tarafından üretilen çıktı özellik haritalarının bir seçimi ile tanımlanır. Bunun yerine yukarıdan aşağıya yol, aşağıdan yukarıya yolun tanımlanma şekline göre oluşturulur.

Tablo 2.1. Omurga ResNet 101'in mimarisi

<i>Orijinal</i>		<i>Değiştirilmiş</i>	
<i>Katmanlar</i>	<i>Omurga Ağı</i>	<i>Çıkış Boyutu</i>	<i>Çıkış Boyutu</i>
<i>C1</i>	$7 \times 7, 64$ 3×3 , max havuzlama,	$1/2$	$1/2$
<i>C2</i>	$\begin{Bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{Bmatrix} \times 3$	$1/4$	$1/4$
<i>C3</i>	$\begin{Bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{Bmatrix} \times 4$	$1/8$	$1/8$
<i>C4</i>	$\begin{Bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{Bmatrix} \times 23$	$1/16$	<i>Genişletilmiş konvolüsyon</i> $1/8$
<i>C5</i>	$\begin{Bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{Bmatrix} \times 3$	$1/32$	<i>Genişletilmiş konvolüsyon</i> $1/8$
	FC	<i>Kaldırıldı</i>	

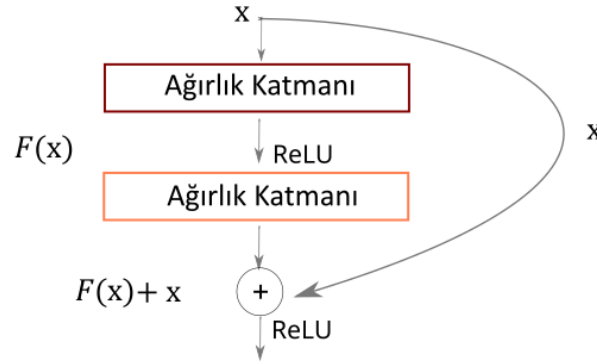
Aşağıdan yukarıya yöntemi omurga modeli kullanan ileri beslemeli bir hesaplama ve ResNet, [90] yönteminin ilkelerini izleyerek tasarlanmış, değişken derinlikte bir derin öğrenme modelleri ailesidir. Bu çalışmada, farklı seviyelerden özellikler çıkarmak için omurga olarak ResNet-101 yapısı kullanılmıştır. Burada, yapıdaki son ek numara katmanlar açısından ağı derinliğini temsil eder. ResNet'i tanımlayan katmanlar, her biri farklı sayıda katmandan oluşan C1, C2, C3, C4 ve C5 olarak adlandırılan 5 aşamada gruplandırılmıştır. Değişken ağı derinliğine sahip ResNet'ler genel yapıyı koruyarak istenen derinliğe ulaşmak için yukarıda belirtilen aşamaların her birine daha fazla katman ekler. Bu tür modeller derinlikleriyle ayırt edilse de temel yapı taşları, evrimsel katman gruplarıyla temsil edilen darboğazlardır. Darboğaz bloklarının her biri, ayrıntıları kısaca verilen artık öğrenme (residual learning) tasarım ilkelerine göre uygulanır.

ResNet mimarisi artık öğrenmenin temel yapı taşı artık bloğudur. İç evrimsel katmanların bir veya daha fazla alt kümesi tarafından ele alınan öğrenme probleminin, X girdilerinin genel bir $H(X)$ fonksiyonuna yaklaştırılmasından, her bir evrimsel katman tarafından $F(X) = H(X) - X$ artık fonksiyonunun (Residual function) öğrenilmesine kaydırılmasını sağlar. (Eşitlik 1.14)

Bu tür bir formülasyon, artık öğrenme paradigmasını benimseyen her bir ağı bileşenindeki öğrenme probleminin karmaşıklığını azaltır. Çünkü öğrenilecek fonksiyon F artık keyfi olarak karmaşık bir fonksiyon değil, katmanın girdilerine göre artık eşlemedir. Böylece konvolüsyonel katmanlar tarafından gerçekleştirilen işlem aşağıdaki şekilde verilebilir.

$$H(X) = F(X) + X \quad (2.4)$$

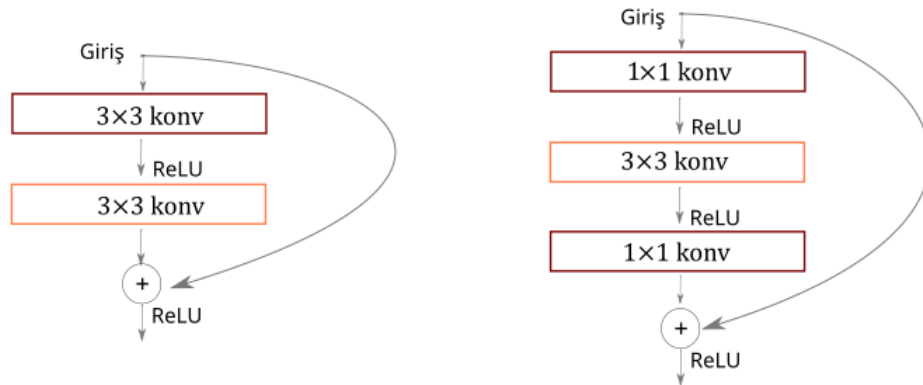
Bu tür değiştirilmiş fonksiyon H bir artık bloğunda hesaplanan işlemi temsil eder [90], F artık bloğundaki bir veya daha fazla katman tarafından öğrenilen artık fonksiyonu temsil ederken, X ise F ile toplanan referans eşlemesidir. Pratikte, Eşitlik 2.4'ü temsil eden artık bloğu Şekil 2.4'te görüldüğü gibi uygulanır. Dolayısıyla, referans eşlemesinin eleman bazında toplamı olarak uygulanan ve iki kez tekrarlanan artık blok ResNet mimarisinin temel bileşeni haline gelir.



Şekil 2.4. ResNet mimarisinin artık fonksiyonu

Uygulamada derin ResNet'lerin hesaplama karmaşıklığını azaltmak için şekilde gösterilen darboğaz modülü tasarlanmıştır [90].

Şekil 2.5'te görülebileceği gibi, artık bloklar ile darboğaz modülü (bottleneck) arasındaki temel fark, 3×3 maliyetli konvolüsyonlar uygulanmadan önce filtrelerin sayısını azaltmak için 1×1 konvolüsyonların boyut azaltma tekniği olarak kullanılmasıdır. Pahalı işlemler hesaplandıktan sonra sonuç, eleman bazında toplamadan önce eklenen başka bir aşağı akış 1×1 konvolüsyon katmanı tarafından orijinal boyuta yeniden ölçeklendirilir. Bu tür konvolüsyon işlemlerinin her birini her zaman toplu normalizasyon ve ReLU aktivasyonu takip eder.

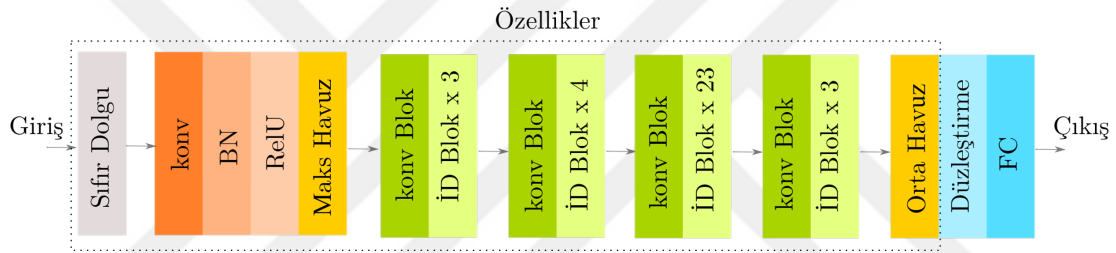


Şekil 2.5. Artık öğrenme bloğunun ve daha verimli “Darboğaz” versiyonunun karşılaştırılması

ResNet mimarisini gösteren Tablo 2.1'de gösterildiği gibi başlangıçta kullanılanlar (C1) hariç her bir darboğaz aynı yapıyı (C2, C3, C4 ve C5) takip etmektedir. Her bir darboğaz temel olarak bir dizi evrimsel katman ve paralel bir referans eşlemesinden oluşmaktadır. Darboğaz tasarımı, ResNet mimarileri oluşturmak için He ve ark. [90] tarafından önerilmiştir.

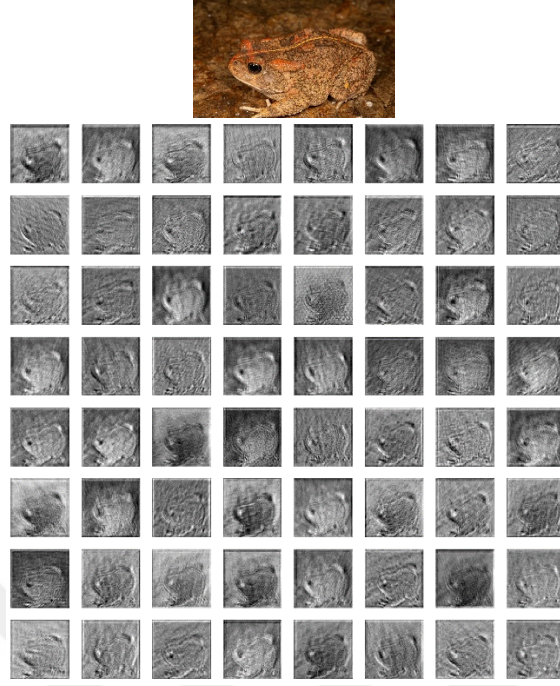
Orijinal ResNet'te 101 konvolüsyonel katman vardır ve katmanlar aynı özellik haritası boyutunda bir seviye olarak kabul edilir. ResNet'te

Şekil 1.17'de gösterildiği gibi toplam beş seviye vardır. Seviye 1 hariç, her seviye birkaç artık bloktan oluşur. Örneğin seviye 2'de üç blok ve seviye 3'te dört blok vardır (Şekil 2.6). Çalışmada, her seviyenin son artık bloğunun çıktısı {C1, C2, C3, C4, C5} ve giriş görüntüsüne göre boyutları $\{1/2, 1/4, 1/8, 1/16, 1/32\}$ olarak belirtilmiştir.



Şekil 2.6. ResNet101'de özellik çıkarma

Kamufle edilmiş nesneyi tahmin alanına uygulamak için, Tablo 2.1'de gösterildiği gibi üzerinde bazı değişiklikler yapılmıştır. Üç ana değişiklik bulunmaktadır. İlk olarak, sınıflandırma görevleri için özel olarak tasarlanmış olan tam bağlı katmanlar kaldırılmıştır. Bu aynı zamanda parametre sayısını da büyük ölçüde azaltmaktadır. İkincisi, orijinal ResNet'in nihai özellik haritası boyutu girdiden 32 kat daha küçük olduğundan, üzerinde doğrudan üst örnekleme yapılırsa sonuçlar çok kaba olacaktır. Bu eksikliği çözmek için, özellik haritasının boyutunu azaltmadan ve herhangi bir ek parametre gerektirmeden aynı alıcı alanın korunmasını sağlayan genişletilmiş konvolüsyon [74] seviye 4 ve seviye 5'e uygulanmıştır. Sonuçta, bu iki seviyedeki özellik haritası boyutu girişten 8 kat daha küçük olacak şekilde değişmektedir.



Şekil 2.7. 10. katmandaki kamufle edilmiş görüntüden çıkarılan özellikler örneği

Algoritma 2 Önceden eğitilmiş CNN modeli kullanılarak özellik vektörü çıkarımı

Giriş Görüntü

For Her bir girdi görüntüsü **do**

Giriş görüntüsünü $H \times W \times 3$ piksele yeniden boyutlandırın

Önceden eğitilmiş Resnet-101'den özellikler çıkarın

Özellik haritasının ortalama değerini hesaplayın

Özellikleri bir sütun vektörüne dönüştürme

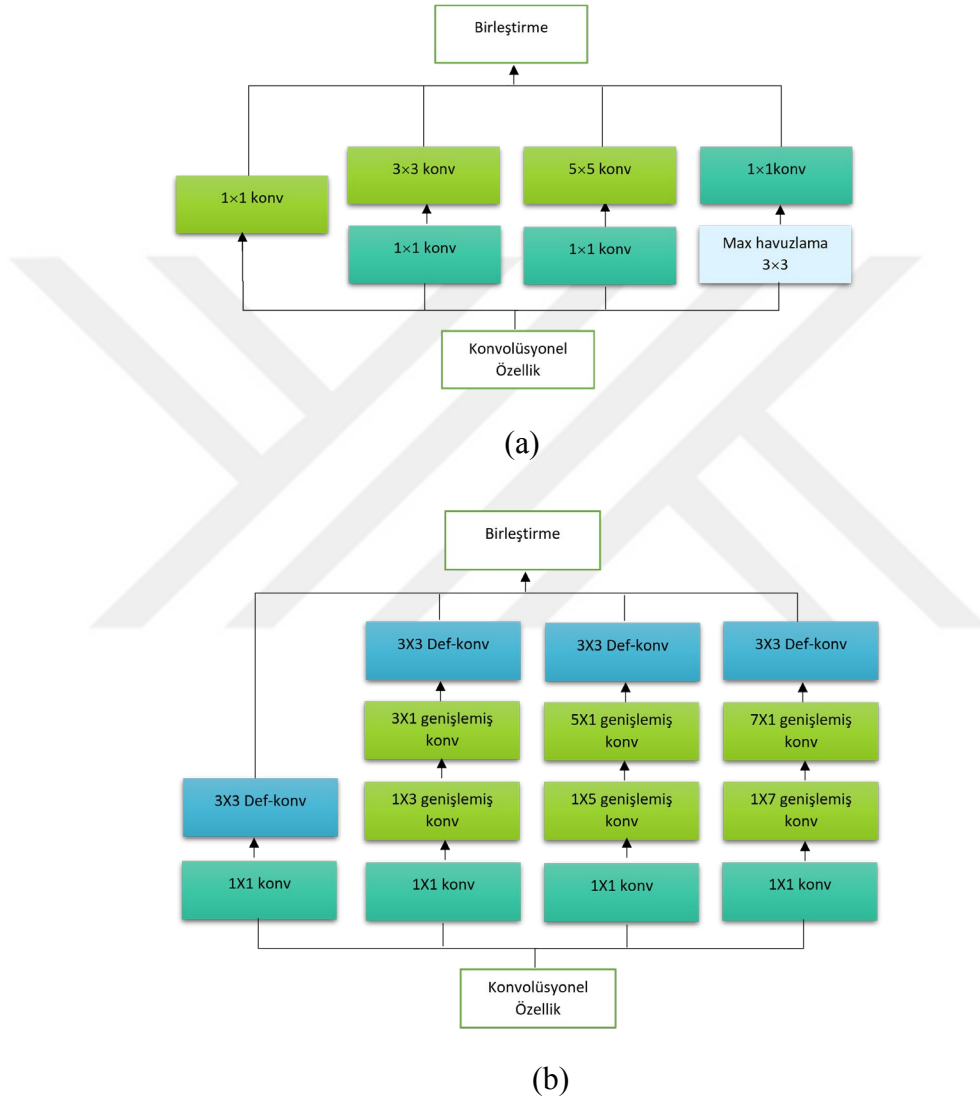
EndFor

Çıkış ResNet-101 mimarisinden çıkarılan özellik vektörü

2.4. Özellik Geliştirme

Kamufle edilmiş nesne segmentasyonu için hedef özelliklerin bilgisini daha da geliştirmek amacıyla bir başlangıç modülü önerilmiştir. Tasarlanan modül, girdilerden çok

ölçekli bağlamsal bilgiyi yakalamayı amaçlayan başlangıç modülünden türetilmiştir [89]. Orijinal başlangıç modülünün ana fikri, Şekil 2.8'de gösterildiği gibi, farklı çekirdek boyutlarına sahip çoklu konvolüsyonel katmanların kullanılmasıdır. Bu konvolüsyonel katmanlar, çeşitli alıcı alan boyutlarına sahip çok ölçekli özellik çıkarıcılar davranışı göstermektedirler.



Şekil 2.8. (a) Basit IM ve (b) Önerilen IM yapısı

Önerilen IM'de, öncelikle giriş özellikleri 4 dala ayrılacak, özellikler 1×1 konvolüsyondan geçecek ve daha sonra düzenli 3×3 konvolüsyon yerine düzensiz $1 \times n$ ve $n \times 1$ genişletilmiş konvolüsyon uyarlanacaktır. Genişletilmiş konvolüsyonlar, kodlanmış

özellikleri çeşitli ve kapsamlı görüş alanlarıyla geliştirmek için paralel bir şekilde kullanılır. Burada $n = 3, 5, 7$ olarak seçilir, 1×1 konvolüsyonun özellikleri $1 \times n$ genişletilmiş konvolüsyona tabi tutulur ve ardından üç dal için $n \times 1$ genişletilmiş konvolüsyondan geçer. Her bir genişletilmiş konvolüsyonun genişletilme oranı 3'tür. $n \times n$ konvolüsyonun iki konvolüsyon çarpanlarına ayrılması çift yönlüdür. Çünkü [113]'e göre bu iki yapı aynı alıcı alanlara sahiptir ve çarpanlara ayırma daha düşük bir hesaplama maliyeti gerektirir.

Standart IM ile karşılaştırıldığında, bir diğer önemli fark, önerilen IM'de her dalın sonuna 3×3 deforme edilebilir konvolüsyon katmanının eklenmesidir. Deforme edilebilir konvolüsyon katmanı ilk olarak [76]'da tanıtılmıştır ve burada uzamsal örnekleme konumları verilerden öğrenilen ek ofsetlerle artırılmıştır. Bilindiği gibi, kamufle edilmiş nesne bölütlemesi zorlu bir görevdir. Buna ek olarak, karmaşık arka plan ile hedef küçük olabilir ve bazı durumlarda tespit yetersiz kalabilir. Bu sorunu çözmek için, deforme edilebilir konvolüsyon kullanılır. Çünkü alıcı alanını dinamik olarak değiştirebildiği ve geometrik dönüşümler tarafından engellenen algılama performansını iyileştirdiği için küçük nesneleri tespit etmede geleneksel konvolüsyon katmanlarından daha iyi sonuçlar verir. Tüm konvolüsyon dallarından gelen birleştirilmiş özellikler 1×1 konvolüsyondan geçerek giriş özellikleriyle artık bir bağlantıya sahip olacaktır. Bu ise optimizasyon işlemini kolaylaştırmaktadır.

Basit IM ve önerilen IM arasında evrimsel katmanların etkisinin araştırılması ve farkları deneysel olarak Bölüm 3'te verilmiştir. Elde edilen sonuçlara göre $n \times n$ çarpanlara ayrılmış evrimsel blok yardımıyla, değiştirilmiş IM daha çeşitli ve daha geniş görüş alanı özellikleri çıkarabilir. Öte yandan önerilen IM, birincil genişletilmiş veya orijinal IM'den daha güçlüdür. Böyle ki, genişletilmiş konvolüsyon ve deforme edilebilir konvolüsyondan sonra çıktıların alıcı alanları çeşitlidir ve nispeten büyüktür. Bu da farklı ölçeklerde çeşitli bağlamsal bilgilerin dahil edilmesine katkıda bulunur.

2.5. Yukarıdan Aşağıya Çok Düzeyli Özellik Füzyonu

FPN mimarisi dört ana bileşenden oluşur: yanal bağlantılar, yukarıdan aşağıya bağlantılar, aşağıdan yukarıya ve yukarıdan aşağıya yollar.

Aşağıdan yukarıya yol, tamamen konvolüsyonel bir omurga elde etmek için standart şablon modelinin son tam bağlı katmanı olmayan bir omurga mimarisinden oluşur. Bu modelde girdi görüntüleri, girdi katmanından ağın son katmanını temsil eden en üste doğru iletilir, bu nedenle aşağıdan yukarıya yol olarak adlandırılır.

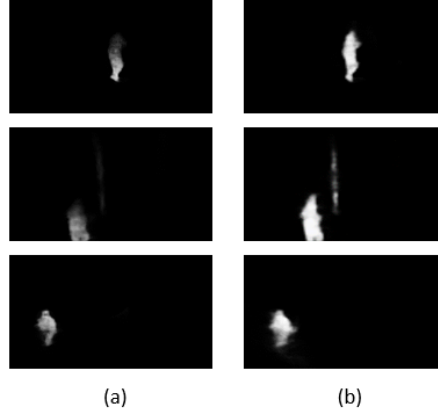
Omurga katmanları tarafından üretilen tüm özellik haritalarından bir alt küme FPN mimarisinin geri kalanını geliştirmek için seçilir. Doğal bir seçim, öğrenilen çok ölçekli özelliklerin en geniş çeşitliliğini korumak için farklı özellik semantiğine ve ölçeğine sahip anlamsal güçlü çıktı özelliklerinin seçimine dayalı olmaktadır [97]. Omurgaya bağımlı olarak üretilen aşağıdan yukarıya $C_i \in \{1, \dots, N\}$ olarak gösterilen özellik haritası seçilir. Seçilen aşağıdan yukarıya özellik haritaları göz önüne alındığında, yukarıdan aşağıya $P_i \in \{1, \dots, N\}$ sayıda birleştirilmiş özellik oluşturulur.

Üretildikleri i seviyesine bağlı olarak, aynı seviyede aşağıdan yukarıya doğru özellik haritasının anlamsal olarak en iyi durumunu temsil ederler.

Kamufle edilmiş nesne tespitindeki zorluklardan biri saklı nesnenin farklı bir görünüme sahip olması veya bütünlük taşınamamasıdır. Bu ise gizli nesnenin tespit edilmesinde başarısızlıklara yol açmaktadır.

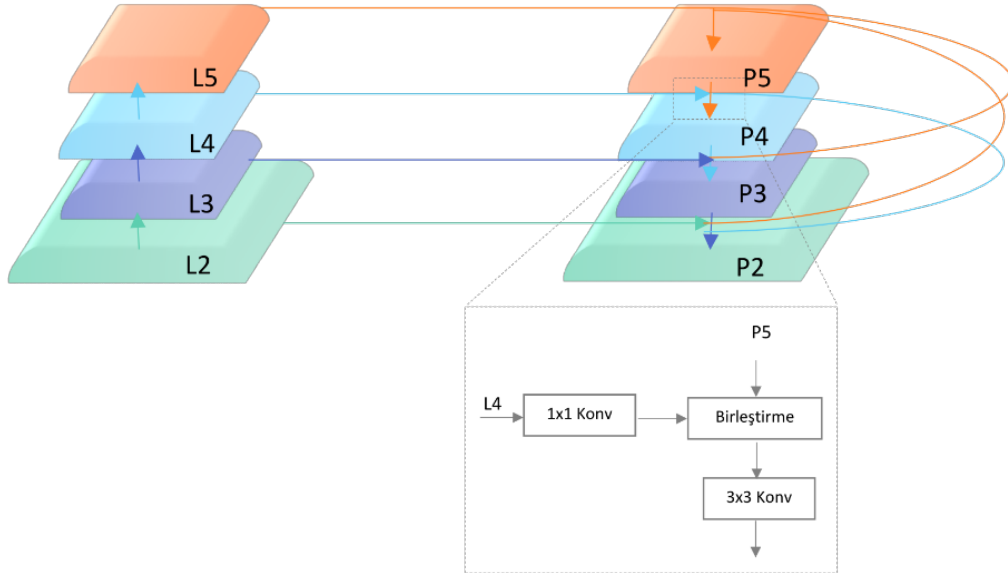
Diğer bir sorun kamufle olan nesnenin ölçeğinin değişkenliğidir. Derin ağlardaki büyük işlem adımları ve havuzlama etkisi nedeniyle küçük bölgeler genellikle gözden kaçır. Ayrıca, sadece üst düzey soyut anlamsal özelliklerin kullanılması bazı uzamsal bilgilerin kaybına yol açmakta ve bu durumda nesne sınırları tam belirlenememektedir. Çalışmada bu sorunların çözümü için, geliştirilmiş özellikleri birden fazla seviyeden yukarıdan aşağıya doğru birleştirerek yoğun özellik piramidi tasarlanmıştır. Özellik piramidi ağı aracılığıyla yoğun bağlantıların eklenmesi, özelliklerin yayılmasını geliştirmektedir.

Bilindiği gibi, yüksek seviyeli özellikler güçlü anlamsal bilgiye ancak zayıf bağlamlara, düşük seviyeli özellikler ise güçlü bağlamlara fakat zayıf anlamsallıklara sahiptir. Özellik piramidi, çok seviyeli bilgileri birleştirmek için etkili çok ölçekli bir yöntemdir. Yukarıdan aşağıya yol ve yanal bağlantı yoluyla bağlanan özellik piramidini kullanır. Özelliklerin birden fazla seviyeden yoğun bir şekilde kaynaştırılması, Şekil 2.9'da görülebileceği gibi ağın performansını büyük ölçüde artırabilir.



Şekil 2.9. Kamufle edilmiş görüntü sonuçlarına örnekler
(a) yoğun bağlantı olmadan özellik piramidi
(b) yoğun bağlantı ile özellik piramidi

Şekil 2.10'dan görüldüğü gibi ve Eşitlik 1.5'e göre, başlangıç modülünden çıkan özelliklere 1×1 konvolüsyon katmanından oluşan yanıl bağlantılar uygulandığında, dört yeni geliştirilmiş özellik haritası ile aşağıdan yukarıya özellikleri güçlendirebilir.



Şekil 2.10. Yoğun özellik piramit ağı

Özetle, yukarıdan aşağıya yolun işleyişi, bitişik seviye bilgilerini eleman bazında toplama yoluyla birleştirmek ve ardından nihai özellik haritasını elde etmek için 3×3 konvolüsyon katmanını uygulamaktır. Önerilen yöntemde aynı uzamsal boyuta sahip L5, L4 ve L3 için, P4 doğrudan P5 ve L4'e eleman bazında eklenerek elde edilir, ardından P5, P4 ve L3 eklenerek P3 elde edilir. L2, P5, P4 ve P3'ün füzyonu için, P5, P4 ve P3'ün kaba bir uzaysal yukarı örneklemeinden sonra yukarıda açıklanan işlemlerin benzeri olarak P2 elde edilir.

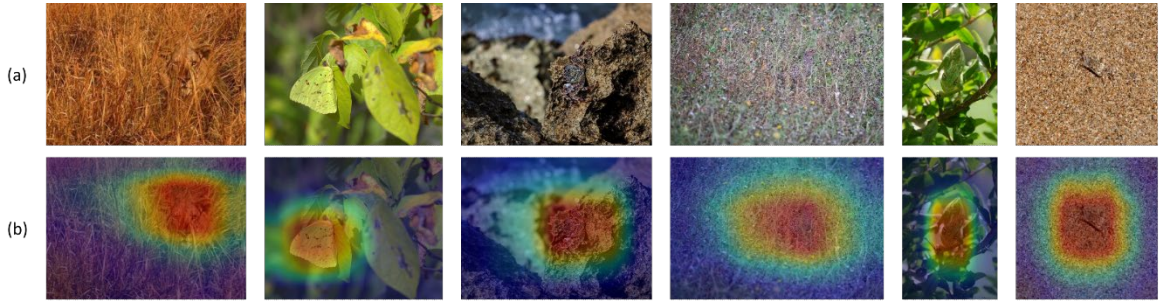
Çok seviyeli özelliklerin yukarıdan aşağıya doğru yoğun füzyonu aşağıdaki bağıntıya uygun olarak gerçekleştirilmektedir:

$$P_i = \begin{cases} Konv_{1 \times 1}(L_i) & i = 5 \\ Konv_{3 \times 3} \left\{ \sum_{j=i+1}^5 P_j + Konv_{1 \times 1}(L_i) \right\} & i = 4, 3 \\ Konv_{3 \times 3} \left\{ Konv_{1 \times 1}(L_i) + \sum_{j=i+1}^5 \text{örnekleme}(P_j) \right\} & i = 2 \end{cases} \quad (2.5)$$

2.6. Dikkat Özelliği

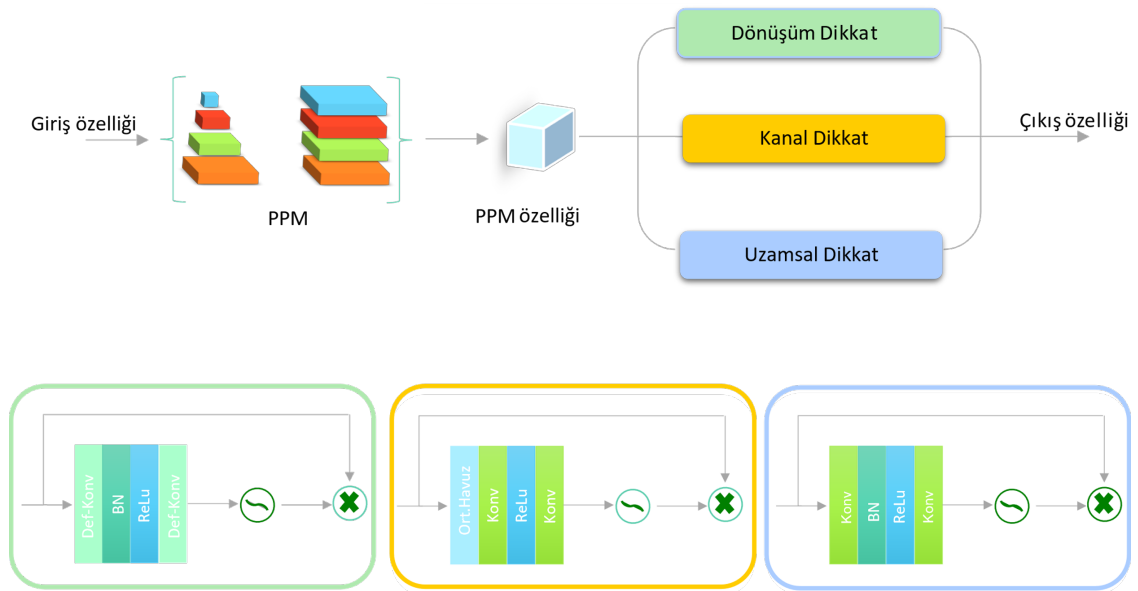
Sezgisel olarak, doğal ortamdaki avcılar ve hayvanlar çok fazla dikkat çekmemek için kamuflaj kullanırlar. Buradan yola çıkarak, sahnedeki belirsiz veya kamuflle edilmiş nesneleri doğru bir şekilde bulmak için çalışmada çoklu dikkat modülü önerilmiştir. Adından da anlaşılacağı gibi, dikkat mekanizması; görsel bilgi işlemenin verimliliğini ve doğruluğunu artırmak için ilgi alanına giren bilgilere daha fazla dikkat etmek ve diğer yararsız bilgileri önemli ölçüde bastırmaktır.

Bu olguya dayanarak, gizli hedefe dikkat çekmek ve dikkati çekmek amacıyla, kamuflle edilmiş nesneyi arka plandan daha iyi bulmak ve daha ince ayrıntılar üretmek için bir MAM tasarlanmıştır. MAM, ilgili olmayan ayrıntıları göz ardı ederek görüntülerin önemli kısımlarını vurgular (Şekil 2.11).



Şekil 2.11. (a) Kamufle edilmiş görüntü örnekleri ve bunlara karşılık gelen (b) dikkat haritaları

Dikkat mekanizması insan görsel sisteminde önemli bir rol oynar. İnsanlar karmaşık sahnelerle karşılaştıklarında dikkatlerini hızlı bir şekilde önemli noktalara odaklayabilir ve bu noktalara öncelik verebilirler. İnsan görsel sisteminde olduğu gibi, bilgisayar tabanlı bir dikkat mekanizması, ağırlıklı bilgi işlem gücünü temel özelliklere odaklayabilir. Böylece kamufle edilmiş görüntülerden gereksiz bilgiler bir kenara bırakılarak önemli özellikler daha etkili bir şekilde çıkarılabilir. Tespiti daha zor görüntüler için yerel özellik çıkarma yeteneğini geliştirmek amacıyla tezde bir MAM geliştirilmiştir.

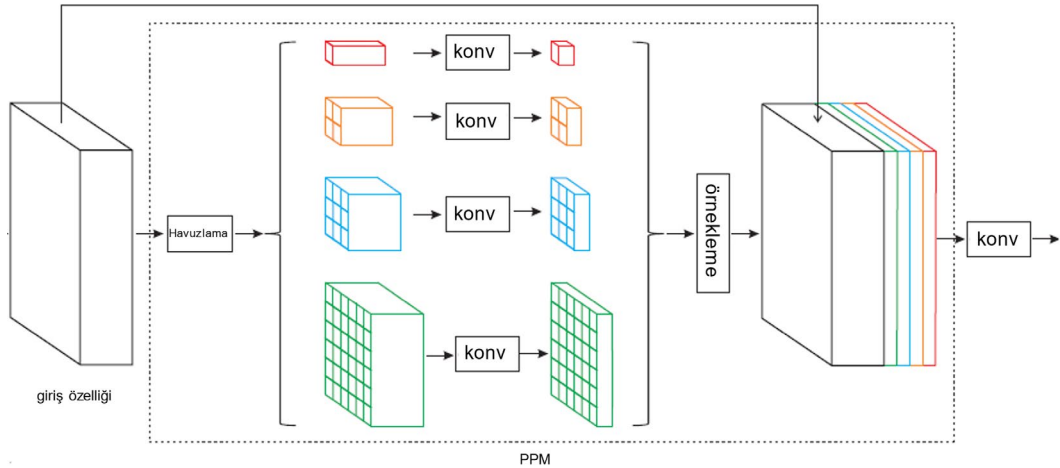


Şekil 2.12. MAM mimarisi

Şekil 2.12'den gösterildiği gibi, MAM ilk olarak giriş özelliğine bir Piramit Havuzlama Modülü (PPM) [114] uygular ve PPM özelliğinin çıktısını doğrudan yeni giriş özelliği olarak kullanır, ardından Dönüşüm Dikkat Bloğu, Kanal Dikkat Bloğu ve Uzamsal Dikkat Bloğundan paralel olarak üç dikkat haritası oluşturur. Son olarak, tüm dikkat haritaları, nihai ağırlıklı dikkat özelliği haritasını elde etmek için eleman bazında toplama yoluyla birleştirilir. MAM hem düşük hem de yüksek seviyeli anlamsal bilgilerde eşit olarak yayılan bağlam özelliklerine uygulanır ve bu da daha doğru tahminlerle sonuçlanır.

2.6.1. Piramit Havuzlama Modülü

Gizlenmiş görüntülerde hedef, arka plan ile benzer sınırlara sahiptir. Kamufle edilmiş nesne ile arka plan arasındaki güçlü bağlamsal ilişki göz önünde bulundurulduğunda, derin ağa uygun bağlamsal önceliğin eklenmesi kamufle edilmiş nesne tespitinin performansını artırabilir. Ortalama havuzlamanın global bağlamsal öncelik olarak iyi bir model olduğu kanıtlanmıştır. Ancak bir özellik haritasındaki tüm pikselleri tek bir vektör oluşturmak için doğrudan birleştirmek uzamsal ilişkiyi kaybeder ve belirsizliğe neden olabilir. PSPNet'ten [114] esinlenerek oluşturulan piramit havuzlama modülü (PPM), Şekil 2.13'te gösterildiği gibi, bağlam bilgisi kaybını azaltmaktadır.



Şekil 2.13. PPM mimarisi [114]

Piramit havuzlama modülü, hem global kapsamdan hem de yerel kapsamdan çok seviyeli bağlam bilgilerini içeren hiyerarşik bir önceliktir. Bu modülde en kaba seviye, tek bir özellik çıktısı oluşturmak için kullanılan global ortalama havuzlamadır. Diğer piramit seviyeleri, özellik haritasını farklı boyutlarda (1×1 , 2×2 , 3×3 , 6×6) alt bölgelere ayırır ve sırasıyla ortalama havuzlama kullanır. Bunun sonucunda farklı seviyelerden farklı çıktı boyutları elde ediliyor. Kanal boyutlarını azaltmak için her seviyeden sonra 1×1 konvolüsyon katmanı kullanılır. Ardından, özellik haritaları, doğrusal olmayan interpolasyon yoluyla orijinal girişle aynı boyutu elde etmek için örneklenir. Son olarak, farklı seviyelerdeki özellikler birleştirilir.

Piramit havuzlama modülünün küçük sahnelerin çıkarılmasında faydalı olduğunu belirtmek gerekir. Derin bir konvolüsyon ardından sonra, özellik haritasındaki sadece birkaç alt bölge küçük kamufle edilmiş nesneler hakkında daha az geçerli bilgi içerir. Bu nedenle, özellik haritası doğrudan tahmin yapmak için kullanılırsa, parçalı tahmin sonuçları elde etmek kolaydır. Buna karşılık, piramit havuzlama modülü, farklı çekirdek boyutları kullanarak özellik haritasını ortalama havuzlama katmanıyla birkaç farklı alt bölgeye böler. Bir alt bölgenin ihtiyaç duyulan küçük gizli nesne bilgilerini içermesi muhtemeldir. Daha sonra doğrusal olmayan interpolasyon yoluyla orijinal özellik haritasıyla aynı boyutta olacak şekilde örneklenir.

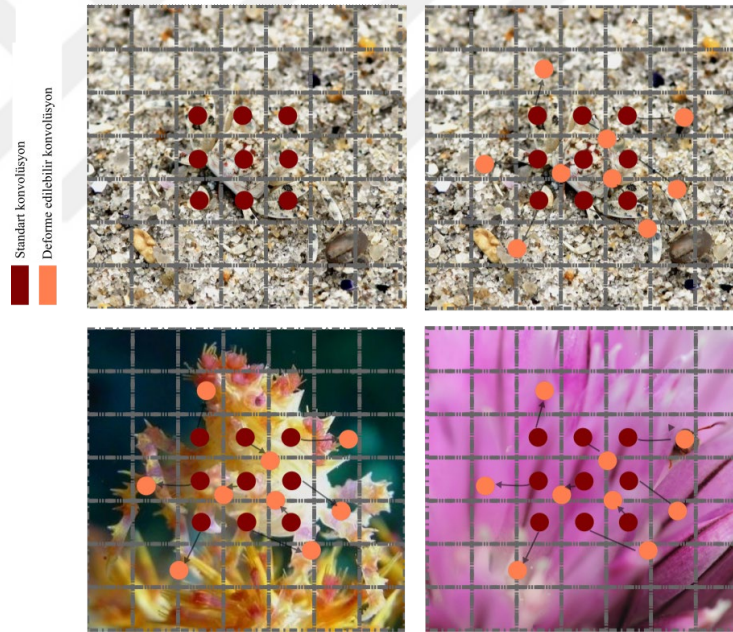
2.6.2. Dönüşüm Dikkat Modülü

Kamufle edilmiş nesneler genellikle ölçek ve şekil açısından önemli ölçüde farklılık gösterir ve bu da COD'de zorlu bir sorundur. Önceki derin öğrenme tabanlı modeller, birden fazla standart konvolüsyonel katmanı kullanarak farklı özellikler elde etmeye çalışmaktadır; bu da bu karmaşık sahneleri, özellikle de eşit olmayan şekilde dağılmış kamufle edilmiş nesneleri ele almak için yetersizdir. Deforme edilebilir konvolüsyon, düzensiz özellikleri yakalayabildiğinden, dönüşüm dikkat modülü özelliklerin ölçek ve şekil değişimini yakalamak için tasarlanmıştır.

Deforme edilebilir konvolüsyon [76], özellik katmanlarındaki uzamsal örnekleme konumlarını ek ofsetlerle artırabilir ve ofsetleri hedef görevlerden öğrenebilir. Bununla birlikte, esas olarak deforme edilebilir konvolüsyon katmanının ekstra bir ofset katmanı

aracılığıyla nasıl oluşturulacağına odaklanmaktadır. Literatürde, belirli bilgisayarlı görme görevlerinde deforme edilebilir konvolüsyon katmanının nasıl oluşturulacağı tartışılmamaktadır. Örneğin, deforme edilebilir konvolüsyon katmanı sadece yüksek seviyeli özellik katmanlarında kullanılırken düşük seviyeli özellikler göz ardı edilmektedir. Bu çalışmada, deforme edilebilir konvolüsyon katmanının hem düşük hem de yüksek seviyeli özellik katmanlarında uygulanabileceği ve bu sayede daha ikna edici kamufrajlı harita tahminleri üretebileceği bulunmuştur (Tablo 3.2).

Şekil 2.14 standart ve deforme edilebilir konvolüsyonların farklı etkilerini göstermektedir. Standart konvolüsyonun özelliğin tüm yönleri için sabit olduğu, deforme edilebilir konvolüsyonun ise nesnelerin ölçeğine ve şekline göre uyarlanabilir bir şekilde ayarlandığı ve bunun da daha iyi sonuçlar elde edilmesine yardımcı olduğu görülebilir.



Şekil 2.14. Standart konvolüsyon ve deforme edilebilir konvolüsyon

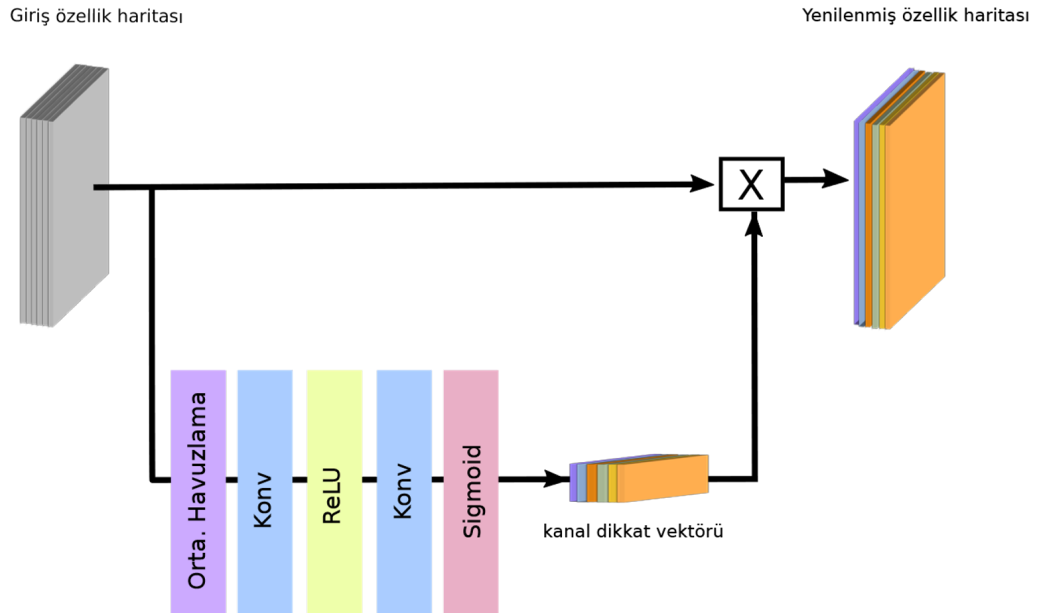
Dönüşüm dikkat modülü, önemli özellikleri vurgulayarak ve istenmeyenleri bastırarak daha iyi nesne temsili için önerilmiştir. Dolayısıyla, bu blok ağın ön plan bölgelerine odaklanmasını güçlendirebilir. Şekil 2.12 gösterildiği gibi, özellik haritası 3×3 deforme edilebilir konvolüsyonel katmanlar tarafından işlenir, ardından bir toplu norm düzeltilmiş doğrusal birim ve ikinci bir 3×3 deforme edilebilir konvolüsyonel katman gelir. Son olarak,

arındırılmış özellik ile giriş özelliği arasında bir eleman çarpımı gerçekleştirilerek çıkış özelliği hesaplanır.

2.6.3. Kanal Dikkat Modülü

COD'deki ortak sorun, arka planın değişen karmaşıklığıdır. Bir model eğitildiğinde, ağ derinleştikçe anlamsal bilgi giderek daha karmaşık hale gelir.

Bu çalışmada, aynı görüntü konumunda meydana gelen benzer değerlere sahip özellik kanallarını geliştirmek için bir kanal dikkat modülü önerilmektedir (Şekil 2.15). Bir görüntüdeki aynı konumun farklı kanallar için benzer değerlere sahip olması, aralarında çok az fark olan veya hiç fark olmayan en az iki tür özellik olabileceği anlamına gelir. Modülün çıktısı, benzer kanallar arasındaki ilişkiyi daha belirgin hale getirmeyi amaçlar. Üst düzey anlamsal özellik haritasındaki farklı kanallarla ilgili kamufle edilmiş görüntüdeki farklı türdeki önemli bilgileri yakalayabilir. Bu, uyarlanabilir özellik çıkarma işlemindeki dengeyi iyileştirerek temel özelliklerin çıkarılmasını kolaylaştırır.



Şekil 2.15. Kanal Dikkat Modülü

CNN'lerdeki özelliklerin farklı kanalları, farklı anlamsal bilgiler için farklı yanıtlar üretir. Kanallar, düşük seviyeli özellikler için çeşitli yapısal detaylar ve yüksek seviyeli özellikler için farklı anlamlar içerir. Bu nedenle, önemli özelliklere odaklanmak ve önemsiz olanları zayıflatmak gerekir. Kanal dikkat modülü, kamufle edilmiş nesnelere yüksek tepki gösteren kanallara daha büyük ağırlıklar verir ve aşağıdaki gibi gösterilebilir:

İlk olarak, global uzamsal bilginin kanal başına uygulandığı giriş özellik haritasına ortalama havuzlama uygulanır:

$$F_{PL} = \frac{1}{T \times L} \sum_{k=1}^T \sum_{m=1}^L F_h(k, m) \quad (2.6)$$

Burada h kanal numarasını, $F_h(k, m)$, (k, m) konumundaki h . kanalının F_h değerini ifade etmektedir. $T \times L$ ise F_h 'nin h . elemanının uzamsal boyutlarını temsil etmektedir. Havuzlama işleminden sonra, üretilen kanal tanımlayıcısı bir ReLu ve bir sigmoid özellik eşliğinde iki konvolüsyon işlemine tabi tutulur.

$$C_A = \alpha \left(C \left(\varphi \left(C(F_{PL}) \right) \right) \right) \quad (2.7)$$

Burada C konvolüsyon işlemini, α sigmoid fonksiyonunu, φ ReLu fonksiyonunu ifade eder. Son olarak, F_P giriş özelliği ve kanal dikkatinin ağırlıkları elemanları çarpılmaktadır.

$$F_{C_A} = C_A \otimes F_P \quad (2.8)$$

2.6.4. Uzamsal Dikkat Modülü

Doğal görüntüler genellikle çok sayıda ön plan ve karmaşık arka plan ayrıntısı içerir. Düşük seviyeli özellikler çeşitli ayrıntılar içerirken, yüksek seviyeli özellikler daha düşük

sonuçlara yol açabilecek arka plan gürültüleri içerebilir. COD'de amaç, insan dikkatini dağıtabilecek diğer dokular olmadan gizli nesneler ve arka plan arasındaki ayrıntılı sınırları belirlemektir. Bu nedenle, tüm uzamsal konumları eşit olarak değerlendirmek yerine, kamufle edilmiş bölgelere daha fazla odaklanmak için bir uzamsal dikkat modülü önerilmiştir. Bu modül sayesinde tahmin için etkili özellikler oluşturulabilmektedir.

Uzamsal dikkat modülünde giriş özellik haritası Konv 1×1 , BN ve ReLu'ya tabi tutulur. Özellik haritası yine 1×1 konvolüsyon işleminden geçerek sigmoid fonksiyonu tarafından etkinleştirilir ve Eşitlik 2.9 'da gösterildiği gibi F_P giriş özellik haritasıyla çarpılarak nihai arındırılmış özellik elde edilir. Burada S_A uzamsal dikkat modülünü temsil etmektedir.

$$F_{S_A} = S_A \otimes F_P$$

(2.9)

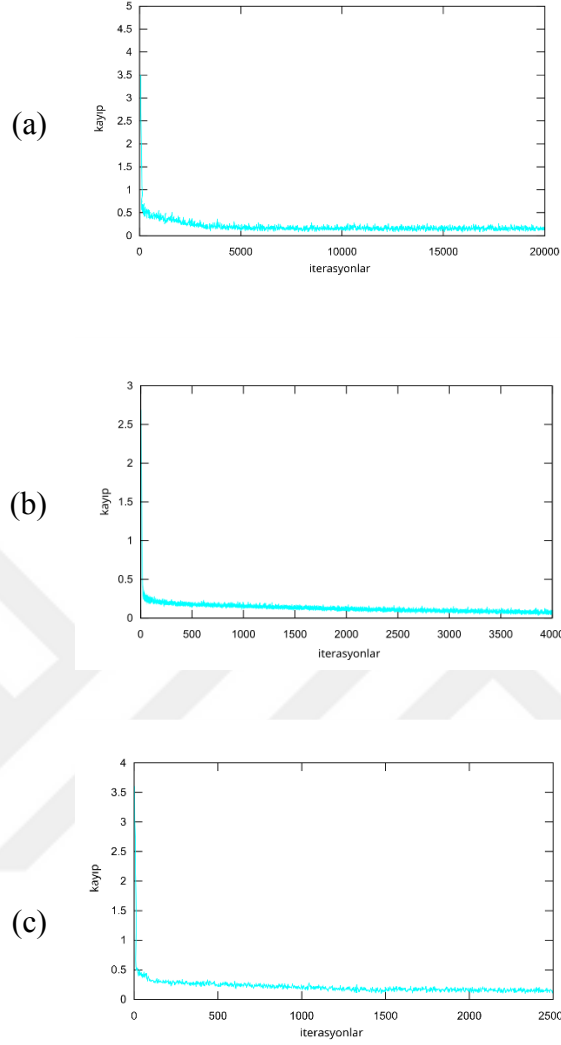
3. BULGULAR VE İRDELEME

Bu bölümde, tüm deneyler ve deney sonuçları ayrıntılı olarak tartışılmakta ve önerilen modelin uygulama detayları sunulmaktadır. Ayrıca modeldeki her bileşenin genel ağa katkısı detaylandırılmaktadır. Bu model üç farklı durumda değerlendirilmiştir: (i) nesne kamufle edildiğinde, (ii) nesne kamufle edilmediğinde, (iii) yöntemin uygulanabilirliğinin ortaya konulması için tıbbi klinik teşhis örneklerinde. (Genel olarak, tıbbi klinik teşhis ve görüntü işlemede doğal kamuflaj desenleri bol miktarda bulunur. Bu nedenle, modelin tıbbi görüntülerde segmentasyonu, polip görüntüleri üzerindeki değerlendirilmiştir).

Önerilen model 12 farklı veri setinde (5 kamufle edilmiş görüntülerden nesne algılama, 4 polip segmentasyonu ve 3 belirgin nesne algılama için kıyaslamalar) çeşitli gelişmiş yöntemlerle karşılaştırılarak test edilmiştir. Deneysel sonuçlar, modelin diğer veri setleri ile kıyaslanmasında daha iyi performans gösterdiğini ortaya koymuştur.

3.1. Eğitim Verileri Detayları

Önerilen yöntem, GPU Nvidia GTX 1080 kullanılarak Caffe çerçevesine [115] uygulanmıştır. İlk aşamada, ağ kamufle edilmiş nesne görevi için eğitilmiştir. 0.9 momentum değeri ve 0,0005 ağırlık azalması ile eğitim için bir stokastik gradyan iniş optimizasyonu (SGD) algoritması kullanılmıştır. Temel öğrenme oranı, mini-paket boyutu bir olacak şekilde 10^{-10} olarak belirlenmiş ve belirli hata değerine erişildikten sonra yaklaşık 20k iterasyonda durdurulmuştur. İkinci aşamada, model kamufle edilmemiş nesneler ve tıbbi görüntü bölütleme için yeniden eğitilmiştir.



Şekil 3.1. Önerilen modelin eğitilmesi için eğitim kaybı değerleri; (a) kamufle edilmiş veri setleri (b) kamufle edilmemiş veri setleri (c) polip veri setleri

Ağı eğitmek için çapraz entropi kaybı kullanılmıştır. Ağımızdaki farklı katmanlar üzerinde birden fazla tahmin olduğundan, toplam kayıp L , öngörülen tüm belirginlik haritaları üzerindeki çapraz entropi kaybının toplamı olarak tanımlanır:

$$L = - \sum_s \sum_{i,j} t_{i,j} \log(c_{i,j}^s) - (1 - t_{i,j}) \log(1 - c_{i,j}^s) \quad (3.1)$$

Burada s ağıdaki katman indeksidir, $t_{i,j}$ pikseldeki (i,j) gerçek referans değeridir (ground truth) (1, kamufle edilmiş bölgeler için ve 0, aksi halde) ve $c_{i,j}^s \in [0,1]$, ağıın s . katmanındaki pikselde (i,j) tahmin edilen kamufle değeridir.

Tablo 3.1. Veri setlerinin eğitim ve test verilerine göre dağılımı

	Veri Setleri	Toplam	Eğitim	Test
COD	CAMO	1250	1000	250
	CHAMELEON	76	-	76
	COD10K	5066	-	2026
	CPD1	1000	800	200
	CPD2	2000	-	400
SOD	DUTS	15.553	10.553	5019
	DUT-OMRON	5168	-	5168
	ECSSD	1000	-	1000
Polip segmentasyonu	KVASIR	1000	900	100
	CVC-612	612	550	62
	ETIS	196	-	196
	CVC-T	60	-	60

Çalışmada CAMO, kamufllajlı insan tespiti eğitim veri setleri birleştirilmiş ve kamufllaj nesne tespiti için eğitim veri seti olarak kullanılmıştır. Kamufle edilmemiş nesneler için eğitim veri seti olarak DUTS-TR [103], polip segmentasyonu için Kvasir [105] ve CVC-ClinicDB [104] setleri kullanılmıştır (Tablo 3.1). Seçilen eğitim görüntülerine ayna yansıması ve döndürme işlemleri uygulanarak veri çoğaltması gerçekleştirilmiştir (Şekil 3.2). COD tespiti için en gelişmiş yöntemlerin CAMO ve COD10K veri kümelerine dayalı olarak eğitildiğini ve DDCN modelinin CPD1 veri kümesine dayalı olarak tarandığını belirtmek gerekir. SOD ve polip segmentasyonu için Tablo 3.1’de belirtilen aynı eğitim veri kümeleri tüm yöntemler için kullanılmıştır.



Şekil 3.2. Veri çoğaltma tekniklerine örnekler

3.2. Değerlendirme

Önerilen yöntemdeki farklı modüllerin etkisini araştırmak için bir değerlendirme çalışması yapılmıştır.

Tablo 3.2. Bileşen analizi

Versiyon	Yöntem	$S_{\alpha} \uparrow$	$F_{\beta}^{ort} \uparrow$	$E_{\phi}^{ort} \uparrow$	$M \downarrow$
1	Temel (VGG)	0.896	0.846	0.927	0.11
2	Temel (ResNet 50)	0.911	0.864	0.943	0.06
3	Temel (ResNet 101)	0.922	0.874	0.955	0.01
4	Temel (B) + IM	0.930	0.879	0.968	0.008
5	Temel (B) + IM + TAM	0.933	0.880	0.970	0.006
6	Temel (B) + IM + TAM + CAM	0.934	0.883	0.973	0.005
7	Temel (B) + IM + TAM + CAM + SAM	0.937	0.887	0.976	0.004

Burada S_{α} , F_{β}^{ort} , E_{ϕ}^{ort} , M uygun olarak S, F, E, ortalama mutlak hata ölçütleridir.

İlk olarak, ağ farklı omurgalar üzerinde çalıştırılmıştır. Deneylerde, temel (B) olarak ResNet-101'i [90] seçilmiş ardından sırasıyla Başlangıç Modülü (IM), Dönüşüm Dikkat

Modülü (TAM), Kanal Dikkat Modülü (CAM) ve Uzamsal Dikkat Modülü (SAM) ağı eklenmiştir.

Tablo 3.2'de gösterildiği gibi, modüllerin eklenmesiyle test performansı kademeli olarak iyileşmiştir. Tüm bu modüller model performansını arttırmış ve birleştğinde en iyi sonuçları alınmıştır. Bulgular, tüm bileşenlerin önerilen model için gerekli olduğunu göstermiştir.

Omurgaların etkinliği: Tablo 3.2, daha iyi omurga kullanılarak ağı performansının önemli ölçüde artırılabilirliğini göstermektedir. Bu çalışmada, VGG [88] ve ResNet50 [90]'dan daha iyi sonuç verdiği için yöntem için temel olarak ResNet-101 [90] mimarisi kullanılmıştır.

Bileşenlerin etkinliği: IM'nin önemi araştırılmış ve Tablo 3.2'den, Temel (B) + IM 'in (4 no'lu satır) Temel ResNet 101'den (No.3) daha iyi performans gösterdiği gözlemlenmiştir bu da başlangıç modülünün performansını iyileştirmek için gerekli olduğunu açıkça göstermiştir.

Tezde Dikkat modüllerinin katkısı da araştırılmıştır. Deneysel çalışmalarda Temel + IM + TAM (No.5)'in Temel + IM (No.4)' den ve Temel + IM + TAM + CAM (No.6) 'nın Temel + IM + TAM (No.5) 'den daha iyi performans sergilediği gözlenmiştir. Bu iyileştirmeler, dönüşüme Dikkatin dahil edilmesinin, önerilen modelin nesneleri doğru bir şekilde algılamasını sağlayabileceğini göstermiştir. Temel (B)+ IM + TAM + CAM + SAM yaklaşımı (No.7) genellikle diğer modellerden daha iyidir ve F-ölçüsü, S-ölçüsü ve E-ölçüsü açısından sırasıyla %1.3, %1.5 ve %2.1'lik bir kazançla performansı önemli ölçüde iyileştirmiştir. Tablo 3.1'de gösterildiği gibi, modüllerin eklenmesiyle test performansı kademeli olarak iyileşmiştir ve tüm modüllerin kullanımı model performansını arttırmıştır. Bu da önerilen model için tüm bileşenlerin gerekli olduğunu göstermiştir.

İkinci olarak, omurgaya ve IM'ye dayalı, genişlemiş konvolüsyonun ve deforme edilebilir konvolüsyonun etkinliğini doğrulamak için bir dizi deneysel seçenek tasarlanmıştır. Tablo 3.3'te sonuçlar ayrıntılı olarak göstermektedir.

Basit IM [89], 3×3 , 5×5 ve 7×7 konvolüsyonlarla modife edilmiş IM'ye uyarlanmıştır. Ardından, farklı konvolüsyonlara 2 ve 3 genişleme oranları uygulanmış ve son olarak, genişletilmiş başlangıç modülüne deforme edilebilir konvolüsyon eklenmiştir. Değiştirilen IM'nin basit IM'den daha iyi performans gösterdiği gözlemlenebilir. Ayrıca, genişletilmiş

konvolüsyon, her bir genişleme oranında, üç ölçünün tamamı açısından değiştirilmiş IM'ye karşılık gelen konvolüsyondan tutarlı bir şekilde önemli performans artışı sağlamıştır. Bu, genişletilmiş konvolüsyonun ve konvolüsyon boyutunun algılamayı iyileştirdiği anlamına gelir.

Son olarak, daha doğru tespiti için mimariye deforme edilebilir konvolüsyon eklenmiştir. Bu süreçte, genişlemiş konvolüsyon ve deforme edilebilir konvolüsyon birbirlerini uyumlu şekilde tamamlamaktadır.

Tablo 3.3. Genişletilmiş ve deforme edilebilir konvolüsyonun değerlendirilmesi

Versiyon	Yöntem		$S_{\alpha} \uparrow$	$F_{\beta}^{ort} \uparrow$	$E_{\phi}^{ort} \uparrow$	$M \downarrow$
1	IM (Basit)		0.924	0.875	0.957	0.01
2	IM (modifiye)		0.926	0.877	0.960	0.01
3	IM + genişleme konv	R=2	0.927	0.878	0.963	0.09
4		R=3	0.928	0.878	0.965	0.09
5	IM + genişleme konv + deforme edilebilir konv		0.930	0.879	0.968	0.008

3.3. Önerilen Yöntemin Güncel Yöntemlerle Nicel Karşılaştırması

Tezde kamufle edilmiş nesne tespiti için beş genel kıyaslama veri seti üzerinde deneyler yapılmıştır: CPD1 [48], CPD2 [49] CHAMELEON [100], CAMO [2] ve COD10K [51]. Önerilen yaklaşım kamufle edilmemiş nesne tespiti için de değerlendirilmiş, DUTS-TE [103], DUT-OMRON [102] ve ECSSD [101] veri setleri kullanılmıştır. Ayrıca önerilen yöntemin performansı, polip segmentasyonu için dört genel veri seti üzerinde (Kvasir [105] ve CVC-ClinicDB/CVC-612 [104] test bölümü, ETIS test bölümü [106] ve EndoScene/CVC-T [107] dahil olmak üzere) değerlendirilmiştir.

3.3.1. Kamuflajlı Nesne Algılama

CPD1 veri seti [48] için, karşılaştırılan yedi model: Amulet [116], UCF [117], SFCN [118], HDFN [119], HRSOD [120], derin tabanlı belirginlik segmentasyonu için AFNet [121] ve DDCN [48] derin tabanlı kamuflaj segmentasyonudur.

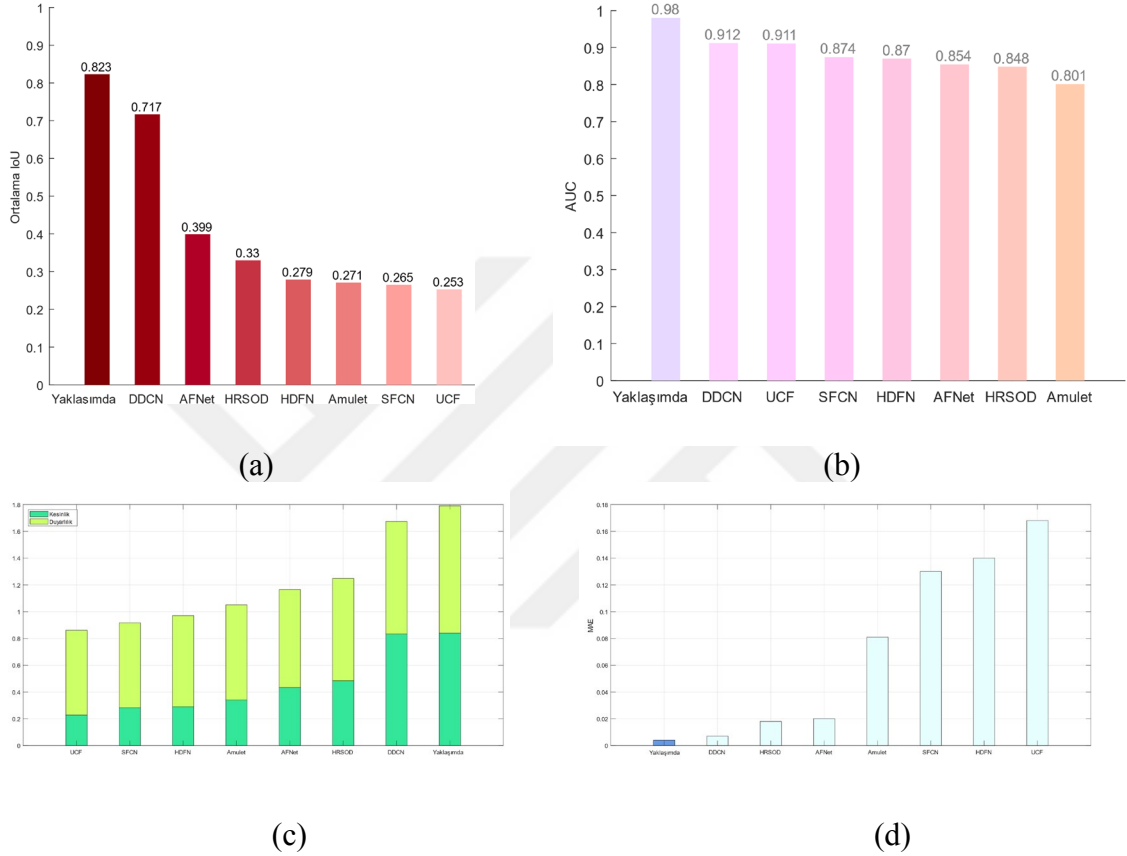
Tablo 3.4, dört değerlendirme S-ölçümü, F-ölçümü, E-ölçümü ve ortalama mutlak hata ölçütleriyle tüm yöntemlerin karşılaştırma sonuçlarını göstermektedir. Önerilen yöntemin, tüm değerlendirme ölçütlerinde rakip yaklaşımlardan daha iyi performans gösterdiği açıktır. DDCN yöntemi [48] ile karşılaştırıldığında, önerilen yöntem F-ölçümünü ve E-ölçümünü sırasıyla %7 ve %3,4 oranında iyileştirmektedir. Ayrıca, yaklaşım MAE değerini önemli ölçüde düşürmüştür. Bu, modelin öngörülen hedef bölgelere daha fazla odaklandığını ve daha doğru maske haritaları ürettiğini gösterir. En son değerlendirme göstergesi S-ölçümü, geleneksel değerlendirme göstergelerinin sınırlamalarını vurgulamak için kullanılmıştır. Geleneksel yaklaşımlar, maske haritaları değerlendirilirken genel yapısal bilgiyi yeterince sağlamayan pikselleri kullanır. Çalışmada, önerilen model %6,5'lik bir gelişme ile S-ölçümü performansını korumaktadır ve tüm nicel sonuçlar, diğer çalışmalarla kıyaslandığında önerilen modelin gelişmiş performans sağladığını göstermektedir.

Tablo 3.4. CPD1 veri seti için karşılaştırılan modellerin nicel sonuçları

Yöntem	$S_{\alpha} \uparrow$	$F_{\beta}^{ort} \uparrow$	$E_{\phi}^{ort} \uparrow$	$M \downarrow$
Amulet [116]	0.630	0.347	0.618	0.081
UCF [117]	0.589	0.310	0.557	0.168
SFCN [118]	0.617	0.329	0.581	0.130
HDFN [119]	0.618	0.337	0.577	0.140
HRSOD [120]	0.690	0.456	0.679	0.018
AFNet [121]	0.728	0.486	0.742	0.020
DDCN [48]	0.872	0.817	0.942	0.007
Önerilen Yöntem	0.937	0.887	0.976	0.004

Şekil 3.3 incelendiğinde modelin, sırasıyla %84 ve %95 ile hem en yüksek kesinlik hem de duyarlılık puanlarına sahip olduğu görülür. Bu örneklerde daha doğru segmentasyon

işleminin yapıldığını kanıtlamaktadır. Şekil 3.3'ten önerilen model en iyi ortalama IoU puanını elde etmekte olup, DDCN [48] ile karşılaştırıldığında yöntemin ortalama IoU puanlarını %10,6 oranında iyileştirdiği görülmektedir. Önerilen yaklaşım en düşük hata puanı ile en yüksek AUC değerine ulaşmıştır.



Şekil 3.3. CPD1 veri seti üzerinde nicel karşıştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları

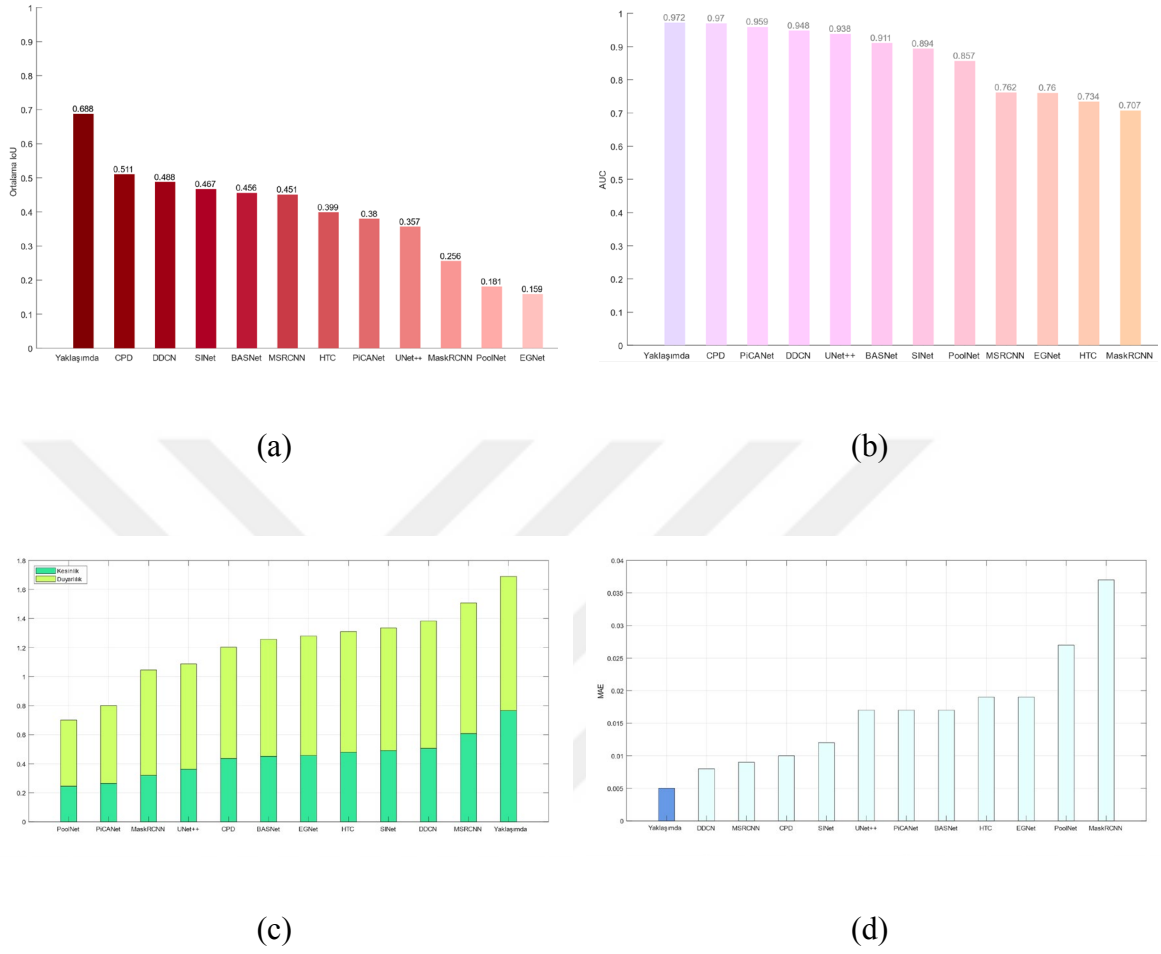
CPD2 veri seti [49] için, 11 model karşılaştırılmıştır: MaskRCNN [122], UNet++ [123], PiCANet [124], MSRCNN [125], PoolNet [126], BASNet [127], CPD [128], HTC [129], EGNNet [130], DDCN [48] ve SINet [51]. Karşılaştırılan yöntemlerin performansları Tablo 3.5'te gösterilmektedir. Önerilen kamufle nesne algılama ağının son güncel yöntemlerin bazıları ile karşılaştırılması modelin üstünlüklerini ortaya koymaktadır. DDCN [48] ve SINet [51] üzerindeki iyileştirmeler, en çok ortalama mutlak hata değerlendirmesinde önemlidir.

Belirgin nesne algılama yöntemleri arasında CPD [128], özellikle kamufle edilmiş nesne algılama görevi için tasarlanmasa da SINet [51] ve DDCN [48] ile karşılaştırmalı performans elde eder. Bununla birlikte, tüm değerlendirme metriklerinde önerilen yöntem daha iyi performans göstermektedir.

Şekil 3.4'te önerilen modelin diğer modellerle karşılaştırma sonuçları verilmiştir. Ortalama IoU, AUC, kesinlik ve duyarlılık ölçütleri sırasıyla %68, %95, %76 ve %84 değerleri ile en iyi sonuçları vermiştir.

Tablo 3.5. CPD2 veri seti için karşılaştırılan modellerin nicel sonuçları

Yöntem	$S_{\alpha} \uparrow$	$F_{\beta}^{ort} \uparrow$	$E_{\phi}^{ort} \uparrow$	$M \downarrow$
MaskRCNN [122]	0.614	0.320	0.671	0.037
UNet++ [123]	0.714	0.467	0.704	0.017
PiCANet [124]	0.754	0.525	0.751	0.017
MSRCNN [125]	0.743	0.575	0.789	0.009
PoolNet [126]	0.632	0.300	0.575	0.027
BASNet [127]	0.756	0.548	0.817	0.017
CPD [128]	0.828	0.655	0.829	0.010
HTC [129]	0.703	0.461	0.664	0.019
EGNet [130]	0.588	0.283	0.501	0.019
SINet [51]	0.767	0.583	0.835	0.012
DDCN [48]	0.815	0.671	0.801	0.008
Önerilen Yöntem	0.869	0.786	0.955	0.005



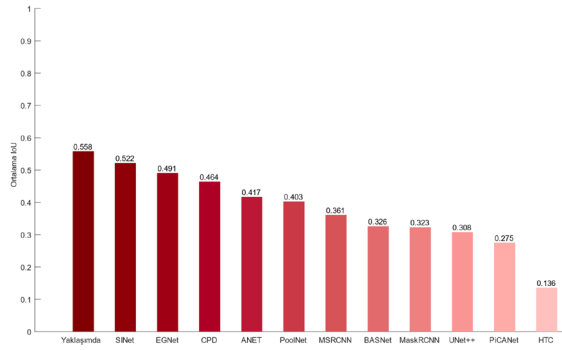
Şekil 3.4. CPD2 veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları

CAMO [2] veri seti için, karşılaştırılan 11 model şunlardır: MaskRCNN [122], UNet++ [123], PiCANet [124], MSRCNN [125], PoolNet [126], BASNet [127], CPD [128], HTC [129], EGNNet [130], ANET [2] ve SINet [51].

Tablo 3.6'da önerilen yöntemin tüm değerlendirme ölçütlerinde en iyi puanlara sahip olduğu görülür. SINet [51] ile karşılaştırıldığında, F-ölçümünü, E-ölçümünü sırasıyla %2,8, %2,6 oranında arttırmıştır. S-ölçümü için modelimiz %0,3'lük bir artışla biraz daha iyi sonuç vermektedir. Ayrıca, MAE puanları %0,9 oranında azalmıştır.

Tablo 3.6. CAMO veri seti genelinde karşılaştırılan modellerin nicel sonuçları

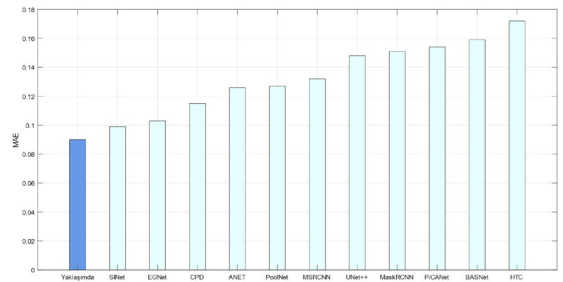
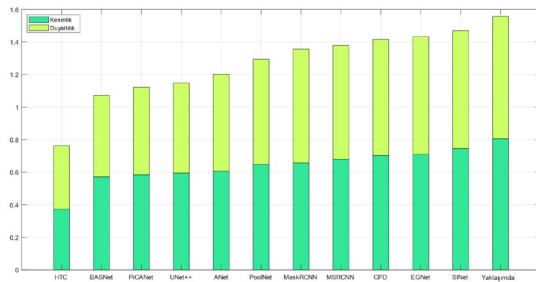
Yöntem	$S_\alpha \uparrow$	$F_\beta^{ort} \uparrow$	$E_\phi^{ort} \uparrow$	$M \downarrow$
MaskRCNN [122]	0.574	0.519	0.714	0.151
UNet++ [123]	0.599	0.461	0.654	0.148
PiCANet [124]	0.608	0.419	0.584	0.154
MSRCNN [125]	0.617	0.526	0.668	0.132
PoolNet [126]	0.702	0.563	0.698	0.127
BASNet [127]	0.618	0.475	0.661	0.159
CPD [128]	0.725	0.614	0.729	0.115
HTC [129]	0.477	0.206	0.441	0.172
EGNet [130]	0.731	0.647	0.768	0.103
ANET [2]	0.682	0.541	0.686	0.126
SINet [51]	0.751	0.676	0.772	0.099
Önerilen Yöntem	0.754	0.704	0.798	0.090



(a)



(b)



Şekil 3.5. CAMO veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları

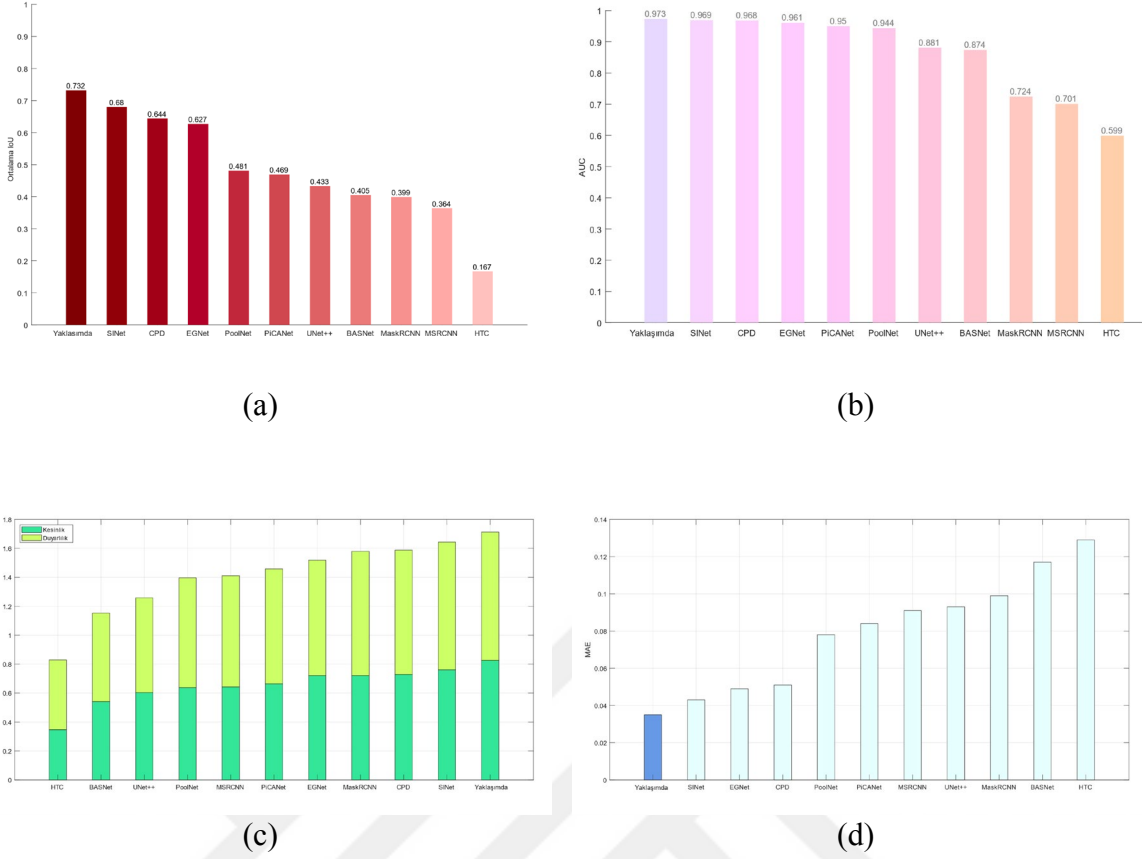
Şekil 3.5'ten görüldüğü gibi, önerilen model sırasıyla %80 ve %64 ile hem en yüksek kesinlik hem de duyarlılık puanlarına sahiptir ve en iyi ortalama IoU puanı elde edilmiştir. Ayrıca SINet'le [51] kıyasla ortalama IoU puanlarını %3,6 artmıştır, en düşük hata puanını korurken %94 ile en yüksek AUC değerine ulaşmıştır.

Tablo 3.7, CHAMELEON [100] veri seti için farklı yöntemlerin nicel sonuçlarını özetlemektedir. Sonuçlardan, en son COD dedektörü olan SINet'in [51] diğer rakiplerinden daha üstün metrik performansı olduğu görülmektedir. SINet [51] ile karşılaştırıldığında, önerilen yöntemin daha yüksek S-ölçüsü, E- ölçüsü ve F-ölçüsü puanlarına ve daha küçük MAE puanlarına sahip olduğu ve kamufle edilmiş nesne algılamayı daha doğru bir şekilde tanımlayabildiği anlaşılmaktadır. Özellikle, E- ölçüsü üzerinde %1,3 ve F-ölçüsü üzerinde %1,7 iyileştirmeye sahiptir.

Tablo 3.7. CHAMELEON veri seti genelinde karşılaştırılan modellerin nicel sonuçları

Yöntem	$S_{\alpha} \uparrow$	$F_{\beta}^{ort} \uparrow$	$E_{\phi}^{ort} \uparrow$	$M \downarrow$
MaskRCNN [122]	0.643	0.610	0.778	0.099
UNet++ [123]	0.695	0.558	0.762	0.093
PiCANet [124]	0.768	0.615	0.748	0.084
MSRCNN [125]	0.637	0.504	0.686	0.091
PoolNet [126]	0.776	0.632	0.779	0.078
BASNet [127]	0.687	0.528	0.721	0.117
CPD [128]	0.853	0.753	0.867	0.051
HTC [129]	0.516	0.236	0.488	0.129
EGNet [130]	0.848	0.750	0.870	0.049
SINet [51]	0.868	0.791	0.893	0.043
Önerilen Yöntem	0.876	0.808	0.906	0.035

Şekil 3.6 incelendiğinde, önerilen modelin, SINet'e göre %97 ile en iyi AUC puanına, %85 ile duyarlılık puanına, %82 ile kesinlik puanına ve %5,2'lik bir iyileşme ile ortalama IoU puanına ulaştığı gözlenmiştir [51]. Diğer 10 yöntem ile karşılaştırıldığında daha üstün bir performansa sahip olduğunu ve bunun da kamufle edilmiş nesneleri daha doğru bir şekilde algılamasına olanak sağladığını göstermektedir.



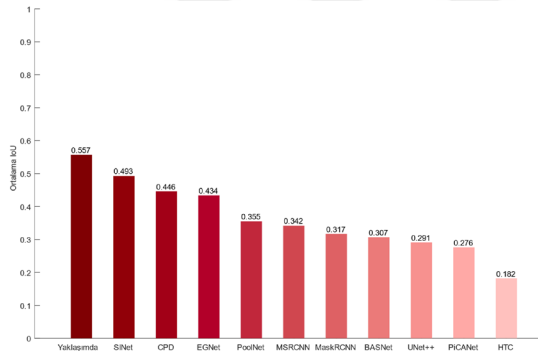
Şekil 3.6. CHAMELEON veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları

COD10K [51] veri seti için, Tablo 3.8'de görüldüğü gibi, dört değerlendirme göstergesinin tümü önceki yöntemlerden daha iyi performansa sahiptir. Önerilen yöntemin en iyi performansı gösterdiği, en zorlayıcı yöntem olması nedeniyle tüm yöntemlerin bu veri setinde daha düşük performansta kaldığı görülmektedir. Özellikle, kamufle edilmiş nesne algılama görevleri için özel olarak tasarlanmış SINet [51] ile karşılaştırıldığında, E_{ϕ}^{ort} %3,3 ve F_{β}^{ort} %4,5 artmış ve MAE %1 azalmıştır.

Şekil 3.7'de görüldüğü gibi, önerilen model sırasıyla %69 ve %71 ile hem en yüksek kesinlik hem de duyarlılık puanlarına ve en iyi ortalama IoU puanına sahiptir ve ayrıca SINet'e [51] kıyasla ortalama IoU puanlarını %6,4 oranında artırmıştır. En düşük hata puanını korurken %95 ile en yüksek AUC değerine ulaşmıştır.

Tablo 3.8. COD10K veri seti genelinde karşılaştırılan modellerin nicel sonuçları

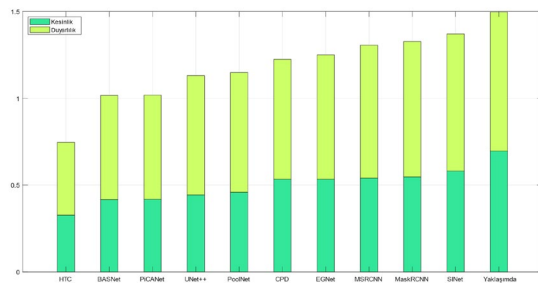
Yöntem	$S_\alpha \uparrow$	$F_\beta^{ort} \uparrow$	$E_\phi^{ort} \uparrow$	$M \downarrow$
MaskRCNN [122]	0.613	0.468	0.748	0.080
UNet++ [123]	0.623	0.409	0.673	0.085
PiCANet [124]	0.649	0.411	0.642	0.082
MSRCNN [125]	0.641	0.477	0.705	0.073
PoolNet [126]	0.705	0.500	0.712	0.069
BASNet [127]	0.634	0.417	0.678	0.105
CPD [128]	0.746	0.582	0.771	0.058
HTC [129]	0.547	0.252	0.519	0.087
EGNet [130]	0.737	0.574	0.779	0.055
SINet [51]	0.771	0.635	0.808	0.051
Önerilen Yöntem	0.772	0.680	0.867	0.041



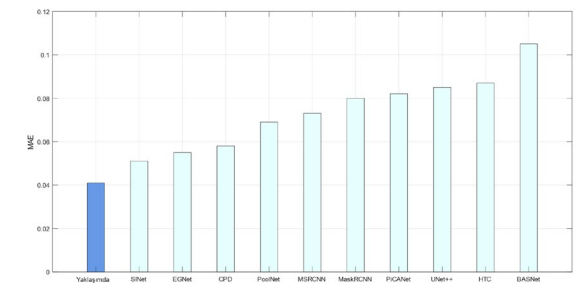
(a)



(b)



(a)



(b)

Şekil 3.7. COD10K veri seti üzerinde nicel karşılaştırmalar (a) ortalama IoU puanları (b) AUC puanları (c) kesinlik ve duyarlılık puanları ve (d) MAE puanları

3.3.2. Kamuflajsız Nesne Algılama

Tablo 3.9’da, önerilen modelin ve güncel SOD ağlarından bazıları (PiCANet [124], BASNet [127], CPD [128] ve EGNet [130]), yaygın olarak kullanılan 3 veri seti DUTS-TE [103], DUT-OMRON [102] ve ECSSD [101] üzerinde karşılaştırmaları gösterilmektedir.

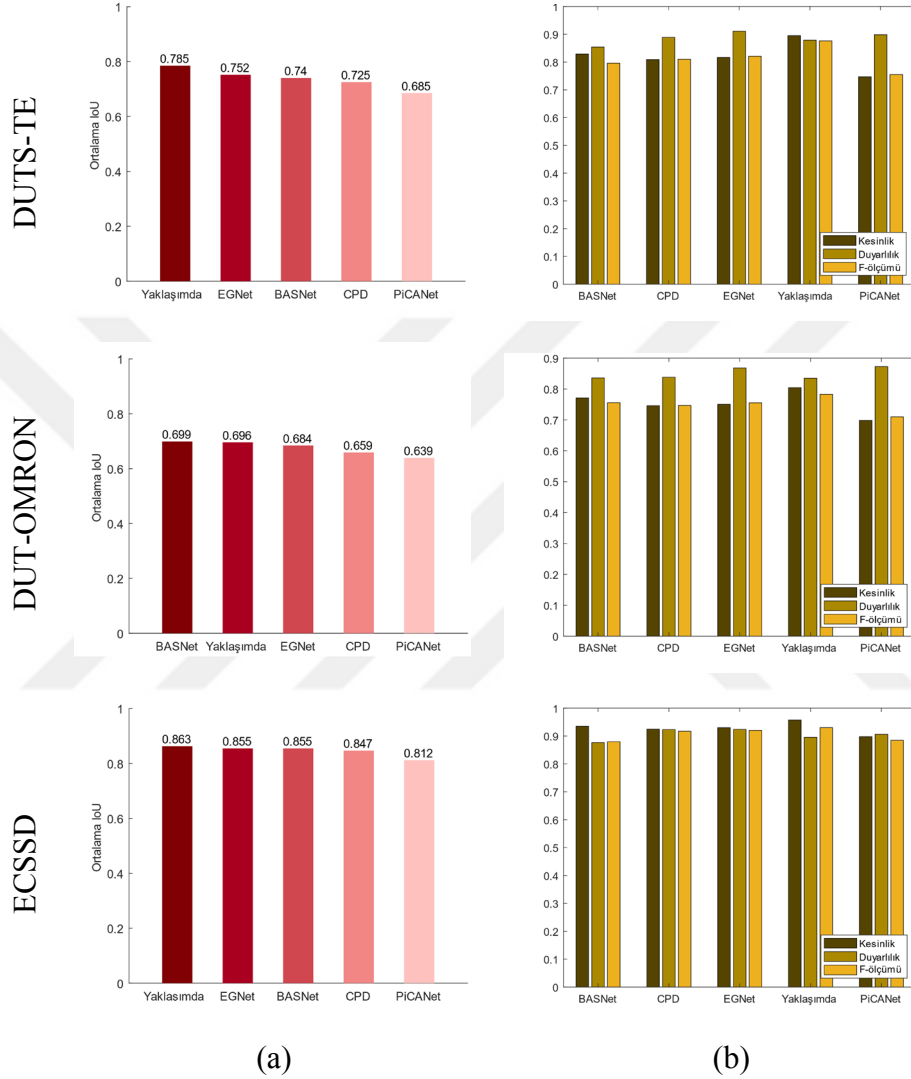
Önerilen modelin tüm veri setleri için MAE, F-ölçümü ve E-ölçümü çoklu ölçümler açısından diğer yöntemlere göre daha iyi performans gösterdiği görülmektedir. Bu da önerilen modelin kamuflaj olmayan nesnelerde de sistemin etkinliğini ortaya koymaktadır.

Tablo 3.9. Yaygın olarak kullanılan üç nesne segmentasyonu testi veri setinde nicel sonuçlar

	Yöntem	$S_\alpha \uparrow$	$F_\beta^{ort} \uparrow$	$E_\phi^{ort} \uparrow$	$M \downarrow$
DUTS-TE	PiCANet [124]	0.861	0.782	0.872	0.054
	BASNet [127]	0.866	0.823	0.896	0.048
	CPD [128]	0.869	0.821	0.898	0.043
	EGNet [130]	0.887	0.839	0.907	0.039
	Önerilen Yöntem	0.889	0.873	0.926	0.032
DUT-OMRON	PiCANet [124]	0.826	0.722	0.833	0.068
	BASNet [127]	0.836	0.767	0.865	0.056
	CPD [128]	0.825	0.742	0.847	0.056
	EGNet [130]	0.841	0.760	0.857	0.053
	Önerilen Yöntem	0.832	0.775	0.865	0.050
ECSSD	PiCANet [124]	0.914	0.887	0.923	0.046
	BASNet [127]	0.916	0.917	0.943	0.037
	CPD [128]	0.918	0.913	0.942	0.037
	EGNet [130]	0.925	0.918	0.943	0.037
	Önerilen Yöntem	0.916	0.930	0.945	0.035

Sonuçlardan elde edilen temel gözlemler şu şekildedir: (i) önerilen model, tüm büyük ölçekli veri setlerinde önemli bir farkla F-ölçümünü iyileştirmiştir ve genellikle MAE'yi azaltmıştır. Bu da modelin tahmin edilen bölgelere daha fazla odaklandığını ve daha doğru

maske haritaları sağladığını göstermektedir. (ii) yalnızca DUTS-TR [103] veri seti üzerinde eğitilmiş olmasına rağmen, önceden eğitilmiş diğer algoritmalarından önemli ölçüde daha iyi performans göstermektedir. Üstün performansı diğer büyük ölçekli veri setleri üzerinde iyi genelleme yeteneklerine sahip olduğunu doğrulamaktadır.



Şekil 3.8. Kamufle edilmemiş algılama için nicel karşılaştırmalar (a) Ortalama IoU puanları (b) Keskinlik, Duyarlılık ve F-ölçü puanları

3.3.3. Tıbbi Görüntü Segmentasyonu

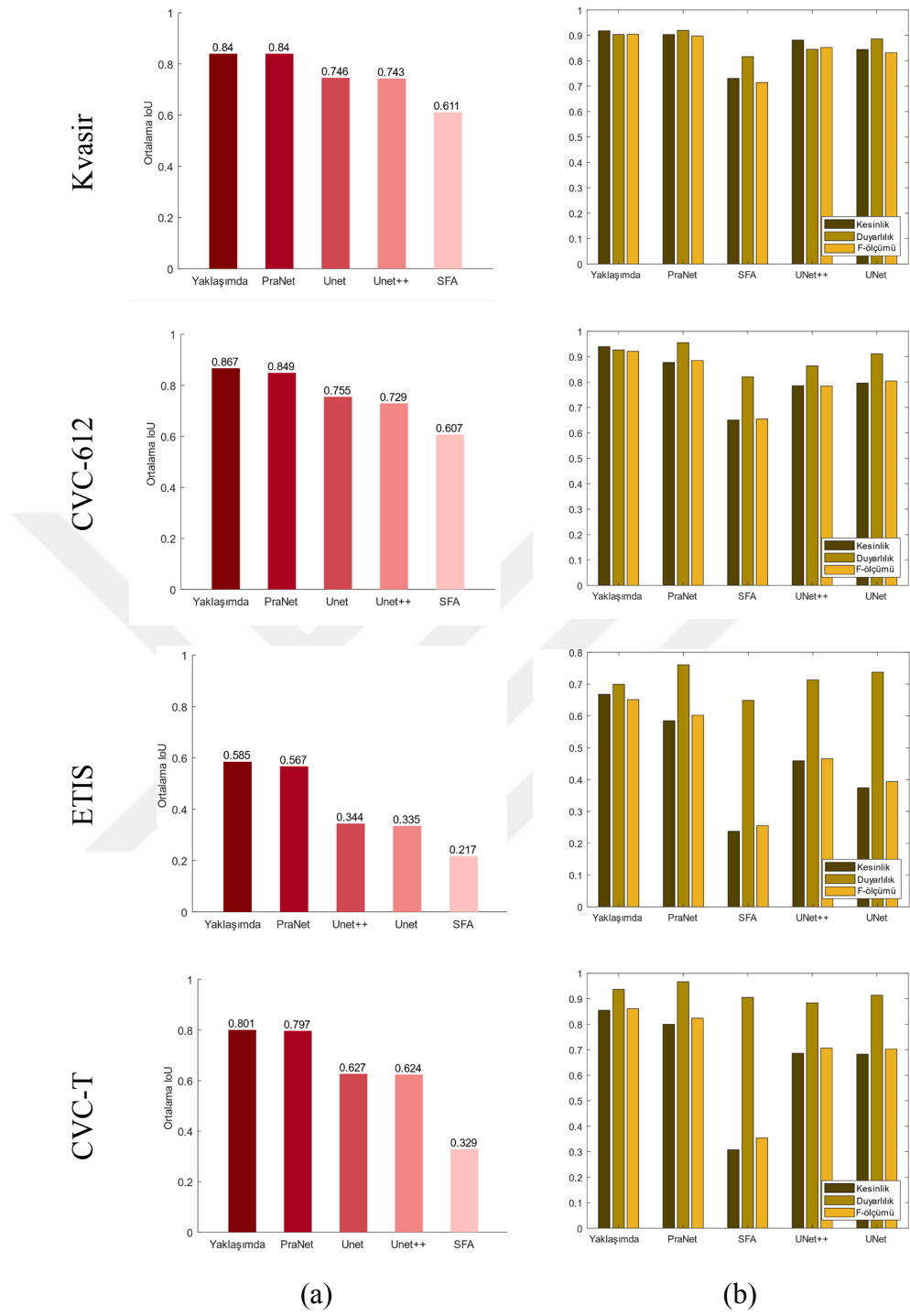
Önerilen yöntemin genelliğinin ve doğruluğunun kanıtlanması için tezde derin tabanlı en gelişmiş 4 tıbbi segmentasyon ağı, U-Net [73], U-Net++ [123], SFA [131] ve PraNet [132] ile karşılaştırılması yapılmıştır.

Nicel karşılaştırma sonuçları Tablo 3.10'da gösterilmektedir. Görülebileceği gibi, önerilen model Kvasir [105] ve CVC-T [107] veri setleri için karşılaştırılabilir performansa sahiptir ve hatta CVC-612 [104] ve ETIS [106] veri setleri için son yöntemlerden daha üstündür.

Ortalama IoU puanları Şekil 3.9'da sunulmuştur. Görüldüğü gibi, önerilen yöntem tüm rakiplerini dört veri setinde de tutarlı bir şekilde geride bırakmaktadır.

Tablo 3.10. Yaygın olarak kullanılan dört polip segmentasyonu testi veri setinde nicel sonuçlar

	Yöntem	$S_\alpha \uparrow$	$F_\beta^{ort} \uparrow$	$E_\phi^{ort} \uparrow$	$M \downarrow$
Kvasir	U-Net [73]	0.858	0.818	0.881	0.055
	U-Net++ [123]	0.862	0.821	0.886	0.048
	SFA [131]	0.782	0.723	0.834	0.075
	PraNet [132]	0.915	0.898	0.944	0.030
	Önerilen Yöntem	0.909	0.895	0.939	0.029
CVC-612	U-Net [73]	0.889	0.823	0.913	0.019
	U-Net++ [123]	0.873	0.794	0.891	0.022
	SFA [131]	0.793	0.700	0.840	0.042
	PraNet [132]	0.936	0.899	0.963	0.009
	Önerilen Yöntem	0.939	0.915	0.968	0.007
ETIS	U-Net [73]	0.684	0.398	0.643	0.036
	U-Net++ [123]	0.683	0.401	0.629	0.035
	SFA [131]	0.557	0.297	0.531	0.109
	PraNet [132]	0.794	0.628	0.808	0.031
	Önerilen Yöntem	0.799	0.649	0.824	0.014
CVC-T	U-Net [73]	0.843	0.710	0.848	0.022
	U-Net++ [123]	0.839	0.707	0.834	0.018
	SFA [131]	0.640	0.467	0.644	0.065
	PraNet [132]	0.925	0.871	0.950	0.010
	Önerilen Yöntem	0.922	0.870	0.944	0.009

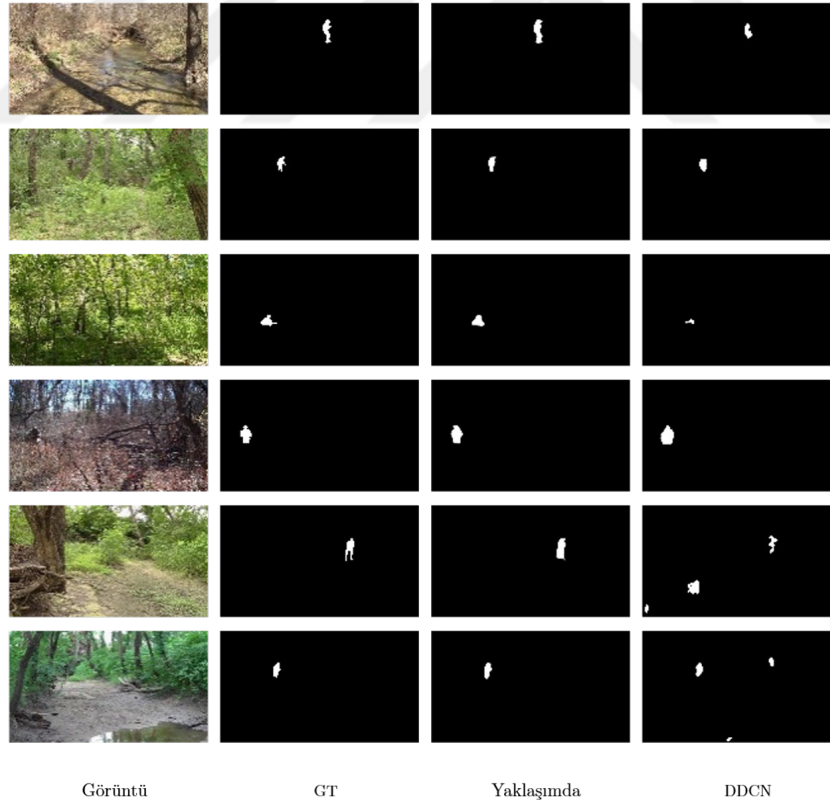


Şekil 3.9. Polip segmentasyonu için çeşitli veri setlerinde nicel karşılaştırmalar
 (a) Ortalama IoU puanları (b) kesinlik, duyarlılık ve F-ölçü puanları

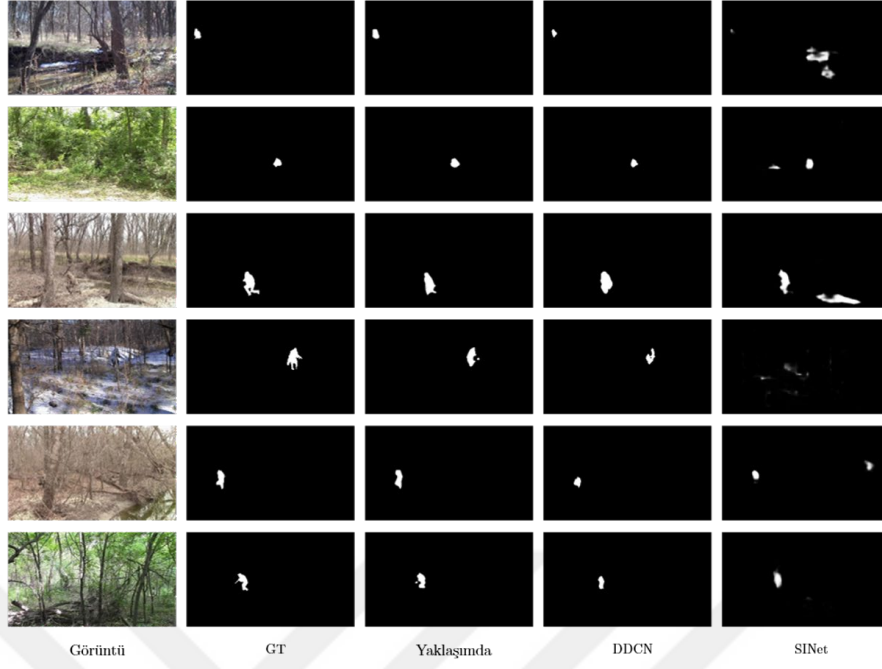
3.4. Önerilen Yöntemin Güncel Yöntemlerle Nitel Olarak Karşılaştırılması

Tezde önerilen yöntemin üstünlüklerini göstermek için modelin tahmin sonuçları güncel son yöntemlerle farklı senaryolarda görselleştirilmiştir. Şekil 3.13'de gösterildiği gibi, önerilen yöntemin gizli hedef tamamlayıcıyı vurguladığını ve daha kesin olduğu gözlemlenmiştir. Çeşitli zorlu senaryolarda, farklı ölçekler ve insanların duruşu (1. ve 3. sıra), küçük nesneler (3. sıra), tikanıklık (2. ve 4. sıra) ve ayrıca gizli hedeflerin (5. ve 6. sıra) doğru bir şekilde yerini tespit etmede daha üstündür. Karşılaştırma sonuçlarından, önerilen yöntem tarafından üretilen segmentasyon haritalarının daha keskin ve daha doğru olduğu anlaşılmaktadır. Aynı zamanda

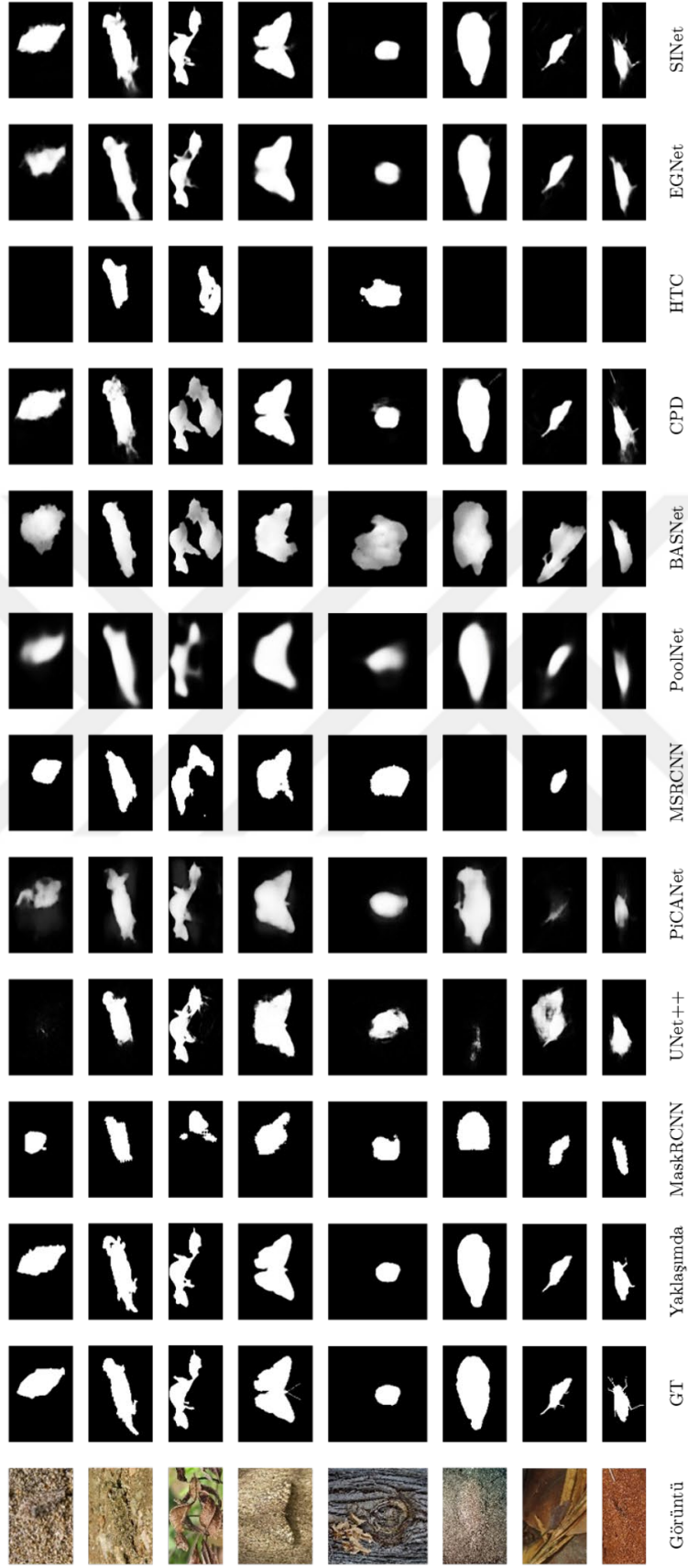
Şekil 3.10'dan ve Şekil 3.11'den görüldüğü üzere DDCN'den ve SINet'ten hep daha iyi performansa sahiptir. Bu aynı zamanda önerilen yaklaşımın etkinliğini de göstermektedir. Önerilen yöntem, ilgili GT ile en yüksek benzerliğe sahip kamuflaj haritasını elde etmektedir.



Şekil 3.10. CPD1 veri setinde görsel karşılaştırma



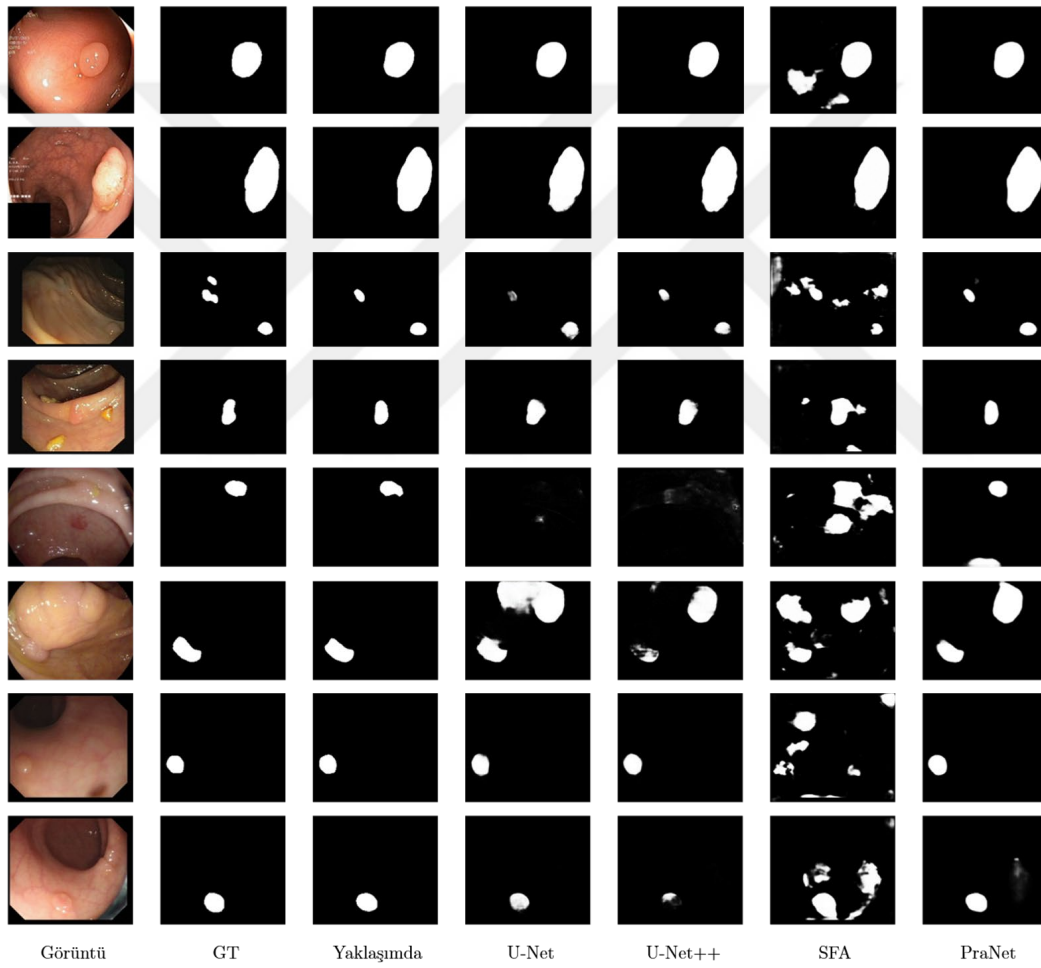
Şekil 3.11. CPD2 veri setinde görsel karşılaştırma



Şekil 3.12. Farklı veri setlerinde kamufle edilmiş nesne tespiti için görsel karşılaştırma

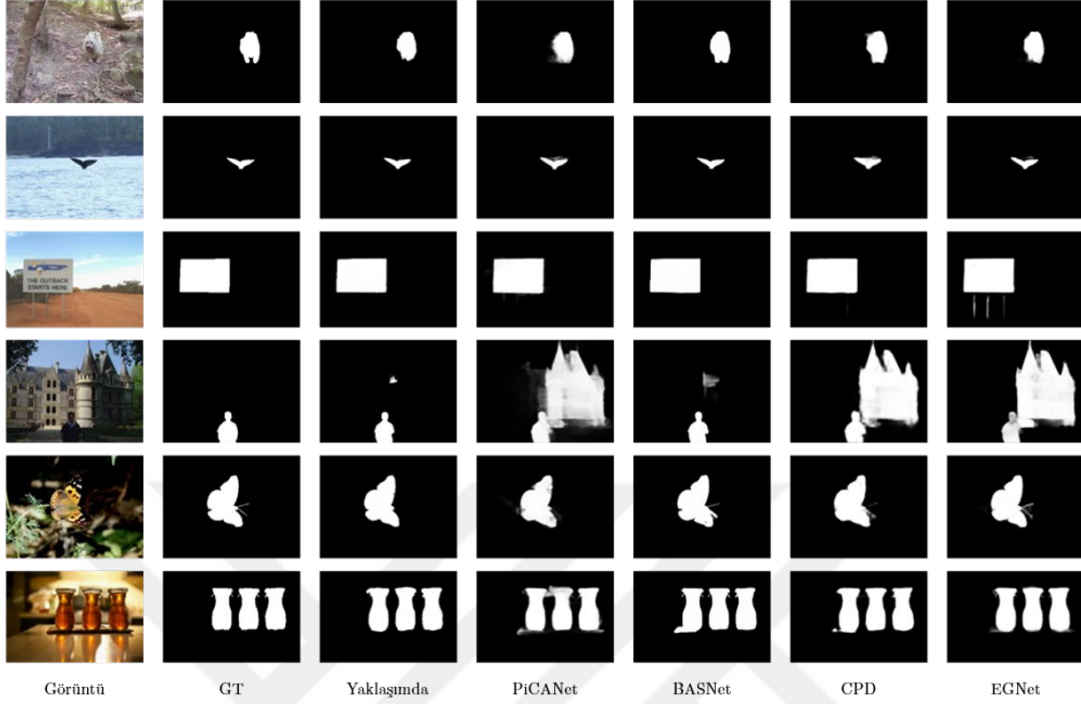
Şekil 3.12, yöntemin diğer güncel yöntemlerle görsel karşılaştırma sonuçlarını göstermektedir. Önerilen yöntemin, çeşitli karmaşık sahneler altında kamuflaj nesnelerinin belirli konumlarını doğru konturla başarılı bir şekilde böldüğü gözlemlenebilir. Ayrıca kamuflaj edilmiş nesneyi benzer arka plandan doğru bir şekilde ayırmaktadır.

Çoklu bölütleme görevi için Şekil 3.13, tıbbi derin tabanlı modellerle karşılaştırmalı olarak algılama sonuçlarını göstermektedir. Görülebileceği gibi önerilen model genel olarak makul sonuçlar sunmaktadır. Hatta bazı durumlarda, hastalıklı bölgeyi diğer yöntemlerden daha doğru tahmin etmektedir (5.satır ve 6.satır).



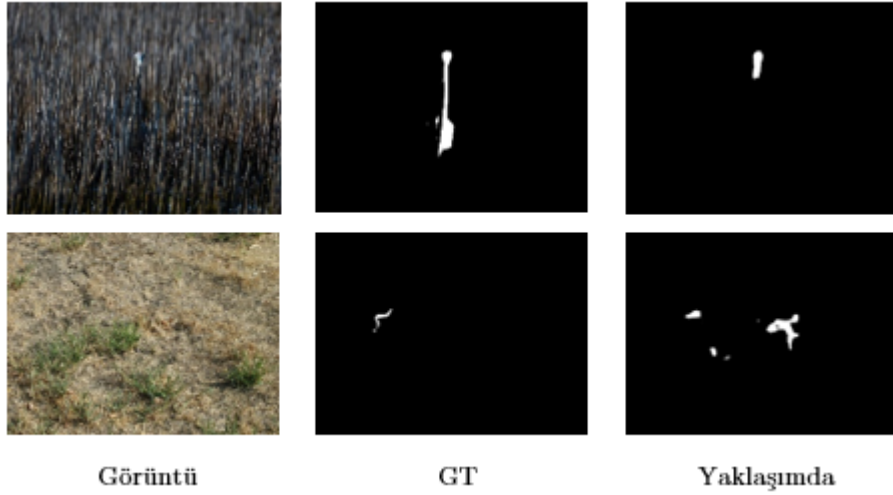
Şekil 3.13. Polip segmentasyonu için görsel karşılaştırma

Şekil 3.14, kamuflajsız nesneler için tespit sonuçlarını sunmaktadır. Görülebileceği gibi, önerilen yöntem diğer yöntemlerle benzer performansa sahiptir, ancak 4. satırda gösterildiği gibi bazı durumlarda daha iyi bir sonuç sunabilir. Önerilen yöntemin tespit sonucu GT'ye en yakın olanıdır.



Şekil 3.14. Kamuflajsız nesne tespiti için görsel karşılaştırma

Yanlış tespit durumları: Önerilen model yüksek performansa ulaşsa da deneylerde bazı başarısızlık durumları vardır. Şekil 3.15'te gösterildiği gibi, başarısız durumlara ilişkin örneklerin tümü aşırı durumlarda, yani ön plan ve arka planın oldukça benzer olduğu karmaşık ortamlarda meydana gelmektedir. Bu koşullarda, insan görsel mekanizması da kamufla edilmiş nesneleri doğru bir şekilde ayırt edemez ve tespit edemez. Bu başarısızlıklardan, önerilen modelin dokuları arka planlara benzeyen bazı nesneleri kamufla edilmiş nesneler veya kamufla edilmiş nesnelerin parçaları olarak yanlış tespit ettiği belirtilebilir. Bunun nedeni, modelin ön plan ile arka planı karıştırması ve özellik ipuçlarını çıkarırken gereksiz bilgiye yol açması olabilir. Gelecekteki çalışmalarda, daha iyi bir tespit için gereksiz bilgiden kaçınmak amacıyla ön plan ve arka plan arasında tam olarak nasıl ayırım yapılacağına odaklanılması faydalı olacaktır.



Şekil 3.15. Yanlış tespit durumlarına örnekler

3.5. Önerilen Yöntemin En Yeni Yöntemlerle Zaman Karşılaştırması

Bir yöntemin üstünlüğünü değerlendirmede hesaplama hızı önemli bir ölçüdür. Bir yöntem gerçek zamanlı bir sisteme uygulanabiliyorsa, hesaplama hızı uygulanabilirliğini belirler. Yöntemin doğruluğunun beklenen gereksinimleri karşıladığını varsayarsak, hesaplama hızı ne kadar yükseltirse, genel performansı o kadar iyi olur.

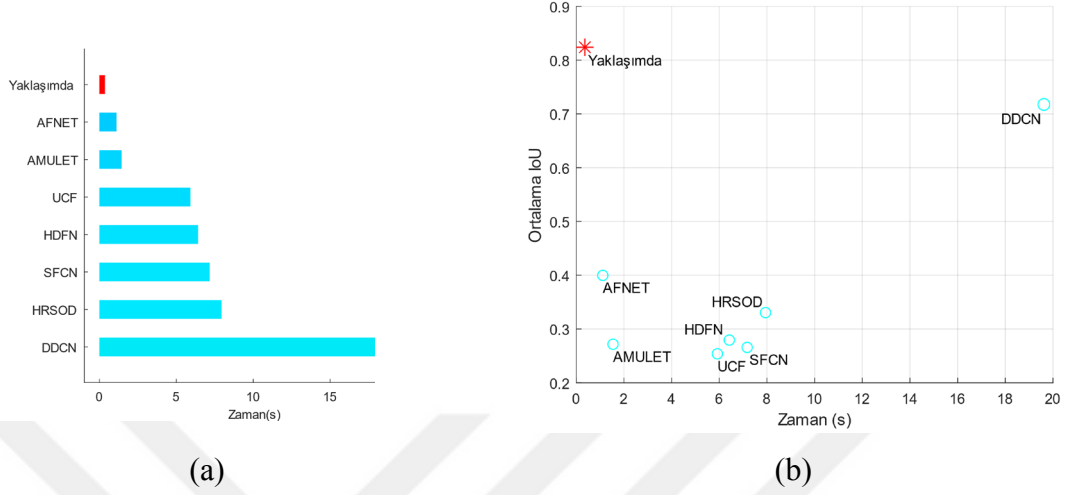
Tablo 3.11, ortalama çalışma sürelerinin bir karşılaştırmasını göstermektedir. Bu değerlendirme, 2,50 GHz hızında çalışan Intel Core i7 işlemcili ve 8G bellekli makinede gerçekleştirilmiştir. Görüldüğü gibi, önerilen yöntem diğer yöntemlerden önemli ölçüde daha hızlıdır.

Tablo 3.11. Önerilen yöntemin ortalama çalışma süresi ve CPD1 veri setinde karşılaştırılan modeller

	Önerilen Yöntem	DDCN	AMULET	UCF	SFCN	HDFN	HRSOD	AFNET
Zaman (s)	0.36	19.62	1.54	5.92	7.17	6.42	7.94	1.11

Şekil 3.16'da gösterildiği gibi, önerilen model diğer yöntemlerle karşılaştırıldığında çok hızlı ve etkili bir şekilde sonuç verirken, makul bir 0.823 ortalama IoU'ya ulaşır. DDCN

[48] modelinin %10,6 daha düşük doğrulukla en yavaş model olduğunu belirtmekte fayda vardır.

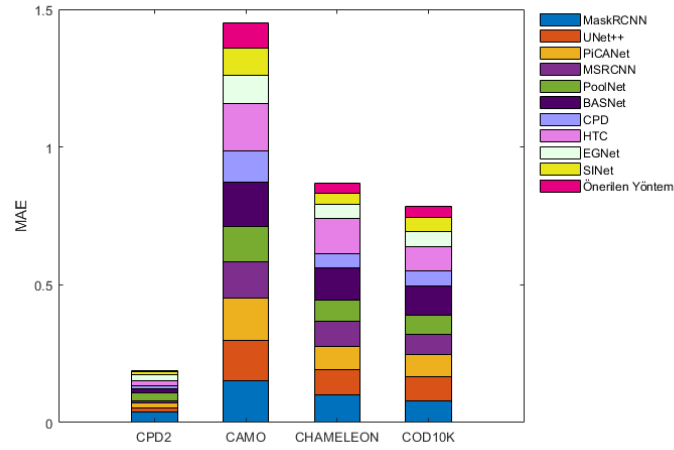


Şekil 3.16. (a) Ortalama çalışma süresinin karşılaştırılması, (b) CPD1 veri seti üzerinde ortalama çalışma süresiyle IoU ilişkisi

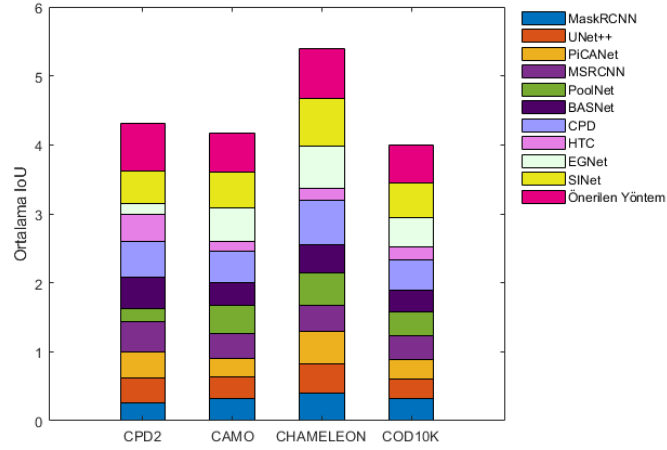
3.6. COD Veri Setleri Genel Performans Değerlendirmesi

Şekil 3.17 bir yöntemin çeşitli veri setleri üzerindeki performansının genel görünümünü yakalamak için MAE metriğindeki yöntemlerin 4 veri seti üzerinde karşılaştırmalı sonuçlarını göstermektedir. Şekilden görülebileceği gibi, yöntemlerin CAMO veri setindeki performansı daha kötüdür. Bunun nedeni, CAMO veri setinin birden çok nesne içermesi, ayrık tek nesneler ve uzamsal olarak bağlı/örtüşen nesneler içermesidir. Bu da veri setini kamufle edilmiş nesne segmentasyonu için daha zor hale getirir.

Şekil 3.18, farklı veri setlerinde ortalama IoU değerleri için yöntemlerin karşılaştırmalı sonuçlarını göstermektedir. Puan ne kadar yüksek olursa performans o kadar iyidir. Yöntemlerin CHAMELEON veri setinde daha iyi performans gösterdiği gözlemlenmektedir. Bunun nedeni sadece 76 görüntü içerdiği için veri setinin küçük boyutlu olmasıdır. Buna karşılık COD10K veri seti, daha zorlu senaryolara sahip 2026 test görüntüleri ile en büyük veri seti olduğu için en düşük performansa sahiptir.



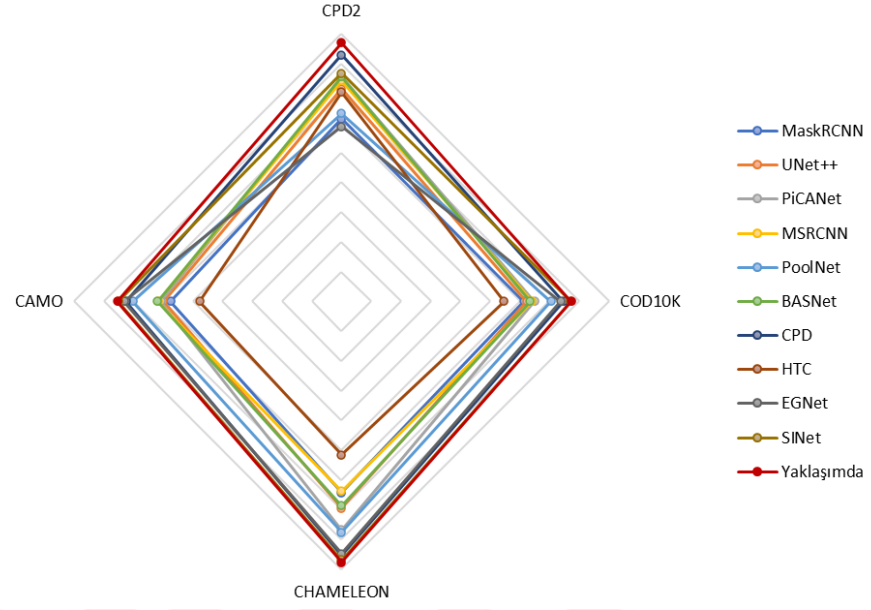
Şekil 3.17. MAE performansı



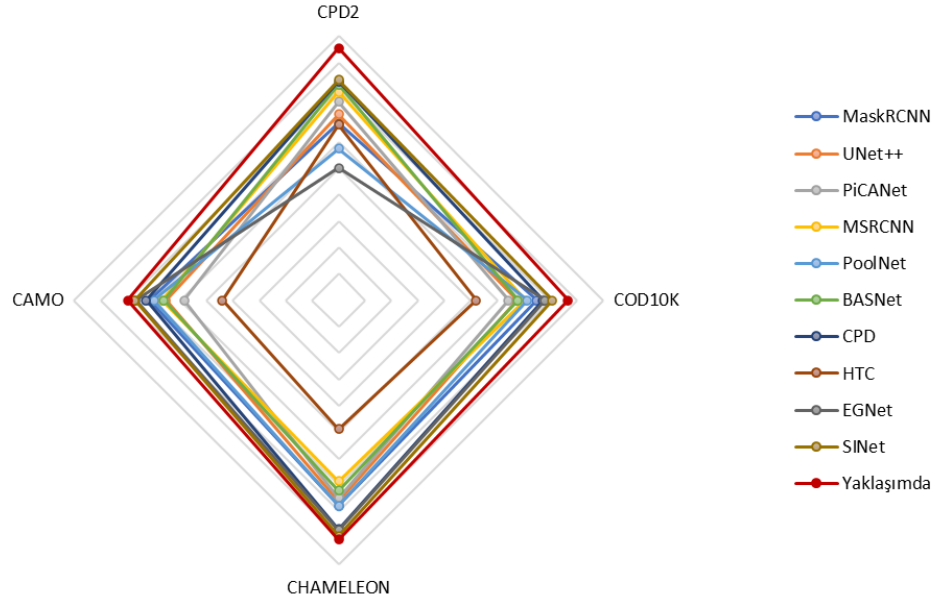
Şekil 3.18. Ortalama IoU performansı

Şekil 3.19 ve Şekil 3.20, farklı veri setlerinde S-ölçümü ve E-ölçümü için yöntemlerin karşılaştırmalı sonuçlarını göstermektedir. Her şekil, radar grafiği, her bir eğrinin veri setleri üzerindeki bir yöntemin performansını gösterdiği 11 kapalı eğri içerir. Radar grafiği, bir veri seti içindeki değişkenlerin yüksek veya alçak değerlerini görmek ve bunların performanslarının karşılaştırılmasına da olanak tanır. Kapalı eğrinin alanı, yöntemin tüm veri setleri üzerindeki performansını yansıtmaktadır. Alan ne kadar büyük olursa performans o kadar iyi olur. Burada her bir yöntemin alanı, tüm veri setleri için karşılık gelen tüm ölçüm puanlarının toplamı olarak hesaplanır.

Tablo 3.12 ve Tablo 3.13, sırasıyla Şekil 3.19 ve Şekil 3.20'deki 11 eğrinin (11 yönteme karşılık gelen) ayrıntılı alanlarını göstermektedir.



Şekil 3.19. S-ölçü performansı



Şekil 3.20. E-ölçü performansı

Tablo 3.12. Şekil 3.19'daki her yöntemin alanı

Yöntem	Alan
MaskRCNN	2.444
UNet++	2.631
PiCANet	2.779
MSRCNN	2.638
PoolNet	2.815
BASNet	2.695
CPD	3.152
HTC	2.243
EGNet	2.904
SINet	3.157
Önerilen Yöntem	3.271

Tablo 3.13. Şekil 3.20'deki her yöntemin alanı

Yöntem	Alan
MaskRCNN	2.911
UNet++	2.793
PiCANet	2.725
MSRCNN	2.848
PoolNet	2.764
BASNet	2.877
CPD	3.196
HTC	2.112
EGNet	2.918
SINet	3.308
Önerilen Yöntem	3.526

Şekil 3.19 ve Şekil 3.20, yöntemlerin CHAMELEON veri setinde en yüksek S-ölçü ve E-ölçü puanlarına ulaştığını göstermektedir. CAMO ve COD10K veri seti daha küçük

boyutlu hedef nesneler içerdüğinden ve tespit edilmesi daha zor olduğundan bu yöntemler diğer veri setlerine kıyasla daha az performansta kalır.

Tablo 3.13 ve Tablo 3.14'ten, önerilen modelin en yüksek S-ölçü ve E-ölçü puanlarına sahip olduğu, SINet ve CPD'nin hepsi en iyi 3 puan arasında yer alan kabul edilir sonuçlar aldığı, HTC'nin pek iyi performans göstermediği gözlemlenmektedir.



4. SONUÇLAR

Bu tez, karmaşık arka plandan kamufle edilmiş nesneyi bulmayı amaçlayan kamufle edilmiş nesne tespiti problemlerine odaklanmaktadır. Genellikle kamufle edilmiş nesnelerin tespiti zor bir problem olduğundan geleneksel yöntemler yanlış değerlendirmelere yol açmaktadır. Bu nedenle kamufleli nesne tespiti, geleneksel nesne segmentasyonuna göre ciddi yanlış değerlendirmelere daha yatkındır. Nesne tespiti görüntü düzenleme, görüntü alma, fotoğraf kompozisyonu ve hedef izleme gibi birçok görüntü anlama ve işleme görevinin ilk adımları olarak kullanılabilir. Ayrıca, kamufleli nesne tespiti araştırması tıbbi görüntü, askeri, tarımsal ve vahşi yaşamı koruma alanlarının gelişimine katkıda bulunacaktır.

Bu tezde, kamufleli nesne tespiti için kapsamlı araştırma yapılmış ve derin sinir ağına dayalı yeni bir model önerilmiştir.

Önerilen yöntemde özellik çıkarma işlemi ResNet-101 omurgasına dayanmaktadır. Bir ağ derinleştikçe, alıcı alan giderek genişler, anlamsal bilgi ise daha gelişmiş hale gelir; özellik haritası küçülür ve uzamsal bilgi sürekli olarak azalır. Bu sorunu çözmek için, özellik haritasının uzamsal boyutunu korurken alıcı alanların boyutunu artırabilen yüksek seviyeli özelliklere genişletilmiş bir konvolüsyon eklenmiştir. Özellik haritalarının daha iyi bütünleşmesini sağlamak için farklı konvolüsyonların kullanıldığı bir başlangıç modülü tasarlanmıştır. Genişletilmiş konvolüsyon ve deforme edilebilir konvolüsyon kombinasyonları, orijinal IM'ye kıyasla performansı iyileştirmiştir ve bu konvolüsyonların modelin performansını iyileştirmedeki etkinliği kanıtlanmıştır. Daha sonra, zengin semantik bilgi ve kamufle edilmiş hedef konum bilgisi içeren bir özellik haritası elde etmeyi amaçlayan etkili bir yukarıdan aşağıya yoğun özellik piramidi sunulmuştur.

Ayrıca, görüntüdeki kamufle edilmiş bölgeyi daha da genişletmek ve iyileştirmek için, farklı ölçeklerdeki hedef bilgilerinden ve bağlam bilgilerinden tam olarak yararlanabilen bir çoklu dikkat modülü önerilmiştir. MAM üç farklı bloktan oluşmaktadır: dönüşüm dikkat modülü, kanal dikkat modülü ve uzamsal dikkat modülü. Bu modeller, kamufle edilen bölgeyi arka plandan daha da uzaklaştırmak amacıyla tasarlanmıştır. Tanıtılan modüller

üzerinde yapılan deneylerden sonra, bu kombinasyonların performansı artırdığı ve her bir bileşenin önemi ortaya konulmuştur.

Modelin etkinliğini kanıtlamak için sistem farklı koşullarda test edilmiş ve F-ölçü, S-ölçü, E-ölçü, MAE, IoU, AUC, kesinlik ve duyarlılık skorlarının kullanıldığı çeşitli metrikler değerlendirilmiştir.

İlk olarak, önerilen model kamufle edilmiş nesne tespiti için 5 veri seti kullanılmış ve farklı 11 güncel yöntemle karşılaştırma yapılmıştır. Model, test edilen tüm veri setlerinde niceliksel ve niteliksel olarak karşılaştırılan yöntemlerden daha iyi performans verdiği kanıtlanmıştır.

İkinci olarak, önerilen yöntem kamuflajsız nesneler için 3 büyük veri seti ile değerlendirilmiştir. Güncel yöntemlerle karşılaştırıldığında, önerilen yöntemin, iyi genelleme yeteneklerine sahip olduğunu doğrulayan deneysel sonuçlar elde etmiştir.

Çalışmanın genelliğini göstermek amacıyla tezde, polip içeren tıbbi görüntülerin segmentasyonu da irdelenmiştir. Bilindiği üzere, tıbbi görüntüleme yoluyla erken teşhis, hastalıkların tedavisinde kilit bir rol oynamaktadır. Erken hastalık bölgesi genellikle çevre dokularla yüksek derecede homojenliğe sahiptir ve doktorlar için lezyon bölgesini erken aşamada tıbbi görüntüden tanımlamak zordur. Tipik bir örnek, kolorektal kanser tanısında yaklaşık %30'luk bir düşüşe sebep olan polipleri segmentlere ayırmak için erken kolonoskopi görüntülerinin yorumlamasıdır [132]. Gizli nesne tespitine benzer şekilde, polip bölütleme de görünümdeki değişkenlik ve bulanık sınırlar gibi çeşitli zorluklarla karşı karşıyadır.

Bu bakış açısıyla, önerilen model çoklu bölütleme durumu için tıbbi görüntü bölütleme için değerlendirilmiştir. Önerilen model, 4 farklı veri setinde 4 derin öğrenme tabanlı tıbbi bölütleme ağı ile karşılaştırılmıştır. Deneyler, önerilen modelin tüm veri setlerinde özellikle MAE ve ortalama IoU skorları açısından son yaklaşımlara göre önemli bir iyileşme elde ettiğini göstermektedir.

5. ÖNERİLER

Nesne tespiti alanında son yıllarda kaydedilen ilerlemeye rağmen, kamufle edilmiş nesne tespiti için derin öğrenme çağının önde gelen algoritmaları sınırlı kalmakta ve gerçek dünyadaki zorlukları henüz etkili bir şekilde çözmemektedir. Gelecekteki araştırmalar için bir dizi öngörülebilir yön vardır.

Gizli Örnek Segmentasyonu, sınıf başına bir segment yerine her farklı nesnenin segmente edildiği Görüntü Segmentasyonunun bir çeşididir. Gizli örnek segmentasyonunda, nesne segmentasyon maskeleriyle birlikte mevcut birden fazla kategorinin her örneği için sınırlayıcı çerçeveler oluşturulur ve aynı sınıftaki birden fazla nesne farklı örnekler olarak ele alınır. Genel örnek segmentasyonu [122] ile karşılaştırıldığında, kamufle edilmiş örnek segmentasyonunun yüksek özellik benzerliğine sahip daha karmaşık senaryolarda gerçekleştirilmesi gerekir. Ayrıca, çeşitli örnekler farklı kamuflej görüntüleri sergileyebilir. Türetilen bu kamuflejler, örnek segmentasyonunu daha da karmaşık hale getirir. İnsanlar yoğun şekilde kamufle edilmiş bir sahneye baktıklarında, görsel sistem içgüdüsel olarak değerli ipuçları aramak için tüm sahne boyunca bir dizi yerel kapsamı tarayacaktır [133]. Küresel bir perspektifte tüm konumlardaki önemli bilgileri titizlikle yakalayan ve doğrudan kamufle edilmiş örnek maskeleri oluşturan kamufle edilmiş örnek segmentasyon modelleri önerilebilir.

Günümüzde, COD alanındaki araştırmalar RGB görüntülerine dayanmaktadır. Buraya derinlik bilgisi de eklenebilir. Araştırmalar, derinliğin gözlemciyi gerçek sınıra daha duyarlı hale getiren ve böylece kamuflejin daha az etkili olmasını sağlayan 3B geometrik bilgi sağladığını göstermektedir [134, 135]. Buna göre, derinlik bilgisinin RGB görüntülerle birleştirilmesi, kamufle edilmiş nesne tespitindeki zorlukları çözülmesinde yeni bir yol olarak düşünülebilir. Ayrıca, birçok türün görsel sistemi gerçek 3B sahnelerde çalışır ve nesne segmentasyonu için yapılan birçok yeni çalışma [136-139] sahnenin 3B düzeni ve nesne şekli bilgisi hakkındaki özellikleri çıkarmak için RGB görüntülerin üzerine başka bir modal olarak derinlik haritalarını entegre etmiştir. Hem RGB dalından hem de RGB-D dalından gelen model tahminlerini uyarlamalı olarak birleştirmek için derinlik katkısı tahmin çözümlerine dayanan yeni bir yaklaşım önerilebilir.

Kameralarla donatılmış droneler, hava fotoğrafçılığı, keşif, gözetleme vb. dahil olmak üzere geniş bir uygulama yelpazesine hızla yayılmıştır. Nesne tespiti için ideal bir platform olan drone, rahat hareket kabiliyeti ve havadan görüntüleme sistemi sayesinde sahneyi yukarıdan görüntüleyebilir [140]. Bununla birlikte, sıradan görüntülerde nesne tespiti ile karşılaştırıldığında, dronelerde dikey olarak çekilen görüntülerden hedefleri tespit etmek için önemli bir fark vardır: birincisi, hava görüntülerindeki hedeflerin açıları çoğunlukla üstten ve eğik olarak görüntülenir ve hava görüntülerindeki hedef özelliklerinin tespit edilmesi daha zordur; ikincisi, hedefin boyutu çekim açısıyla değişir ve uçuş yüksekliği arttıkça hedef bir nokta veya düzlem olarak görünme eğilimindedir [141,142]. Özellikle, kamufle olmuş insanları tespit etme görevi için hedef, çevredeki ortamla kolayca karışır. Drone belirli bir yükseklikte hava fotoğrafı çektiğinde, kamufle edilmiş kişilerin özellikleri daha az belirgin hale gelir. Bu nedenle, kamufle edilmiş hedefi hızlı bir şekilde tespit eden drone tabanlı gerçek zamanlı bir kamufle edilmiş insan tespit sistemi geliştirilebilir. Küçük ölçekli özel operasyonlar kapsamında dronelerin uygulama senaryolarını zenginleştirebilir ve bu uygulamalar geniş bir askeri uygulama beklentisine sahiptir.

Ayrıca konunun önemi dikkate alınarak çeşitli iklim koşullarına bağlı şekilde veri setlerinin sınıflandırılması kamufle görüntülerden nesne tespitinde performans değerlendirmesini yükseltir.

Bu tezde, gizli nesne tespiti için nesnenin segmentasyonu üzerinde durulmaktadır. Bu ise sınırlayıcı çerçeveleri üreten geleneksel nesne tespitinden farklıdır. Bu nedenle, gizli nesne önerisi ve takibi gelecekteki çalışmalar için yeni ve ilginç bir yön olabilir.

6. KAYNAKLAR

1. Copeland, A.C. ve Trivedi, M.M., Models and metrics for signature strength evaluation of camouflaged targets, Proceedings of SPIE on. Algorithms for Synthetic Aperture Radar Imagery IV (1997) 194–199.
2. Le, T.N., Nguyen, T.V., Nie, Z., Tran, M.T. ve Sugimoto, A., Anabranh network for camouflaged object segmentation, Computer Vision and Image Understanding, 184 (2019) 45–56.
3. Du, H. ve Shu, L., Camouflage images based on mean value interpolation, International Conference on. Information Technology and Software Engineering. Springer, 2013, 775–782.
4. Friškovec, M. ve Gabrijelčič, H., Development of a procedure for camouflage pattern design. Fibres & Textiles in Eastern Europe, 4.81 (2010) 68-76.
5. Xue, F., Xu, S., Luo, Y.T. ve Jia, W., Design of digital camouflage by recursive overlapping of pattern templates, Neurocomputing, 172 (2016) 262–270.
6. Yang, H. ve Yin, J., A digital camouflage generation algorithm using color similarity, International Journal of Multimedia and Ubiquitous Engineering, 10, 6 (2015) 159–164.
7. Culpepper, J.B., Texture metric that predicts target detection performance, Optical Engineering, 54,12 (2015) 123101.
8. Nyberg, S. ve Bohman, L., Assessing camouflage using textural features, Targets and Backgrounds VII: Characterization and Representation, 4370 (2001) 60–71.
9. Song, L. ve Geng, W., A new camouflage texture evaluation method based on WSSIM and nature image features, IEEE International Conference on Multimedia Technology, 2010, 1–4.
10. Toet, A., Structural similarity determines search time and detection probability, Infrared Physics and Technology, 53, 6 (2010) 464–468.
11. Rong, X., Jia, Q., Xu, W., Lv, X. ve Hu, J., Camouflage effect evaluation of pattern painting based on moving object detection, International Conference on. Energy, Power and Electrical Engineering, 2016, 244-247.
12. Singh, S.K., Dhawale, C.A. ve Misra, S., Survey of object detection methods in camouflaged image, IERI Procedia, 4 (2013) 351–357.
13. Thayer, G. H., Concealing-coloration in the animal kingdom: an exposition of the laws of disguise through color and pattern: being a summary of Abbott H. Thayer's discoveries, Macmillan Company, 1918.

14. Krizhevsky, A., Sutskever, I., Hinton, G. E., Imagenet classification with deep convolutional neural networks, In *Advances in neural information processing systems*, 2012, 1097-1105.
15. Park, J.H, Choi, S.C, Ahn, I.Y ve Kim, J., Multiple UAVs-based surveillance and reconnaissance system utilizing IoT platform, *IEEE International Conference on Electronics, Information, and Communication (ICEIC)*, 2019.
16. Yang, X., Xu, W., Jia, Q. ve Liu, J., MF-CFI. A fused evaluation index for camouflage patterns based on human visual perception. Defence Technology, 17, 5 (2021) 1602-1608.
17. Itti, L., Koch, C. ve Niebur, E., A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on. Pattern Analysis and Machine Intelligence*, 20 (1998) 1254–1259.
18. Hou, X. ve Zhang, L., Saliency detection: A spectral residual approach, In *Proceedings of the 2007 IEEE Conference on. Computer Vision and Pattern Recognition*, 2007, 1–8.
19. Jian, M., Zhang, W., Yu, H., Cui, C., Nie, X., Zhang, H. ve Yin, Y., Saliency detection based on directional patches extraction and principal local color contrast. Journal of Visual Communication Image Representation, 57 (2018) 1–11.
20. Jian, M., Zhou, Q., Cui, C., Nie, X., Luo, H., Zhao, J. ve Yin, Y., Assessment of feature fusion strategies in visual attention mechanism for saliency detection, Pattern Recognition Letter, 127 (2019) 37–47.
21. Jian, M., Zhao, R., Sun, X., Luo, H., Zhang, W., Zhang, H., Dong, J., Yin, Y. ve Lam K., M., Saliency detection based on background seeds by object proposals and extended random walk. Journal of Visual Communication Image Representation 57 (2018) 202–211.
22. Jian, M., Qi, Q., Dong, J., Sun, X., Sun, Y. ve Lam K., M., Saliency detection using quaternionic distance based weber local descriptor and level priors. Multimedia tools applications 77, 11 (2018) 14343–14360.
23. Wang, L., Dong, S., L., Li, H., S. ve Zhu, X., B., A brief survey of low-level saliency detection, *Information System and Artificial Intelligence (ISAI), International Conference on. IEEE*, 2016, 590–593.
24. Judd, T., Ehinger, K., Durand, F. ve Torralba, A., Learning to predict where humans look, *Computer Vision, IEEE 12th international conference on. IEEE*, 2009, 2106–2113.
25. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X. ve Shum, H., Y., Learning to Detect a Salient Object, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33, 2 (2011) 353–367.

26. Yang, J. ve Yang, M., H., Top-down visual saliency via joint CRF and dictionary learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 3 (2017) 576–588.
27. Borji, A., Boosting bottom-up and top-down visual features for saliency estimation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, 438–445.
28. Wang., Yuan, Y., Yan, P. ve Li, X., Saliency detection by multiple-instance learning, *IEEE transactions on cybernetics* 43, 2 (2013) 660–672.
29. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N. ve Li, S., Salient object detection: a discriminative regional feature integration approach. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, 2083–2090.
30. Wang, J., Jiang, H., Yuan, Z., Cheng, M-M, Hu, X. ve Zheng, N., Salient Object Detection: A Discriminative Regional Feature Integration Approach, *International Journal of Computer Vision*, 123 (2017) 251–268.
31. Tong, N., Lu, H., Ruan, X. ve Yang, M., H., Salient object detection via bootstrap learning, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
32. Long, J., Shelhamer, E. ve Darrell, T., Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, 3431–3440.
33. Wang, L., Wang, L., Lu, H., Zhang, P. ve Ruan, X., Salient object detection with recurrent fully convolutional networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2018) 1734–1746.
34. Chen, S., Tan, X., Wang, B., Lu, H., Hu, X. ve Fu, Y., Reverse Attention-Based Residual Network for Salient Object Detection, *IEEE Transactions on Image Processing*, 29(2020) 3763–3776.
35. Zhang, J., Zhang, T., Dai, Y., Harandi, M. ve Hartley, R., Deep unsupervised saliency detection: A multiple noisy labeling perspective, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, 9029–9038.
36. Achanta, R., Hemami, S., Estrada, F. ve Susstrunk, S., Frequency-tuned salient region detection. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, 1597–1604.
37. Luo, Z., Mishra, A., Achkar, A., Eichel, J., Li, S. ve Jodoin, P., M., Non-local deep features for salient object detection, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 6609–6617.

38. Feng, G., Bo, H., Sun, J., Zhang, L. ve Lu, H., CACNet: Salient Object Detection via Context Aggregation and Contrast Embedding, Neurocomputing, 403 (2020) 33–44.
39. Galun, M., Sharon, E., Basri, R. ve Brandt, A., Texture segmentation by multiscale aggregation of filter responses and shape elements, In: Proceedings of 9th IEEE International Conference on Computer Vision, 2003, 716–723.
40. Bhajantri, N., U. ve Nagabhushan, P., Camouflage defect identification: a novel approach, In: 9th IEEE International Conference on Information Technology, 2006, 145–148.
41. Robert, M., H., Shanmugam, K. ve Dinstein, I., Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics, 3, 6 (1973) 610–621.
42. Sengottuvelan, P., Wahi, A. ve Shanmugam, A., Performance of decamouflaging through exploratory image analysis, In: 1st IEEE International Conference on Emerging Trends in Engineering and Technology, ICETET'08, 2008, 6–10.
43. Pulla Rao C., Guruva Reddy A., Rama Rao C., B., Camouflaged object detection for machine vision applications, International Journal of Speech Technology, 23 (2020) 327–335.
44. Tankus, A. ve Yeshurun, Y., Computer vision, camouflage breaking and countershading, Philosophical Transactions of the Royal Society B: Biological Sciences, 364, 1516 (2009) 529–536.
45. Tankus, A. ve Yeshurun, Y., Convexity-based camouflage breaking, In: Proceedings of 15th International Conference on Pattern Recognition, 2000, 454–457.
46. Pan, Y., Chen, Y., Fu, Q., Zhang, P. ve Xu, X., Study on the camouflaged target detection method based on 3D convexity, Modern Applied Science, 5, 4 (2011) 152–156.
47. Li, C., Zhao, X. ve Wan, Y., The camouflage color target detection with deep networks, In: International Conference on Neural Information Processing. Springer, 2017, 476–485.
48. Zheng, Y., Zhang, X., Wang, F., Cao, T., Sun, M. ve Wang, X., Detection of people with camouflage pattern via dense deconvolution network, IEEE Signal Processing Letters, 26, 1 (2018) 29–33.
49. Fang, Z., Zhang, X., Deng, X., Cao, T. ve Zheng, C., Camouflage people detection via strong semantic dilation network, In: Proceedings of the ACM Turing Celebration Conference, 2019, 1–7.
50. Yan, J., Le, T.N., Nguyen, K.D., Tran, M.T., Do, T.T. ve Nguyen, T.V., Mirrornet: Bio-inspired camouflaged object segmentation, IEEE Access, 9 (2021) 43290 – 43300.

51. Fan, D.P., Ji, G.P., Sun, G., Cheng, M.M., Shen, J. ve Shao, L., Camouflaged object detection, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, 2777–2787.
52. Wang, Y., Li, L., Yang, X., Wang, X. ve Liu, H., A camouflaged object detection model based on deep learning, In: IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS), 2020, 150–153.
53. Hinton, G. E., Osindero, S. ve Teh, Y. W., A fast learning algorithm for deep belief nets, Neural Computation, 18, 7(2006) 1527–1554.
54. Bengio, Y., Lamblin, P., Popovici, D. ve Larochelle, H. , Greedy layer-wise training of deep networks, in Advances in Neural Information Processing Systems, 2007, 153–160.
55. Rumelhart, D. E. ve McClelland, J. L., Information Processing in Dynamical Systems: Foundations of Harmony Theory, in Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations, MIT Press, 1987, 194–281.
56. Fischer, A., ve Igel, C., An introduction to restricted Boltzmann machines, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2012, 14–36.
57. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., ve Manzagol, P. A., Stacked denoising autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, Journal of Machine Learning Research, 11, 3 (2010) 3371–3408.
58. Lee, H., Grosse, R., Ranganath, R., ve Ng, A. Y., Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in Proceedings of the 26th International Conference on Machine Learning, 2009, 609–616.
59. Rumelhart, D., Hinton, G. ve Williams, R., Learning representations by back-propagating errors, Nature, 323 (1986) 533–536.
60. Hochreiter, S. ve Schmidhuber, J., Long Shortterm Memory, Neural Computation, 9, 8 (1997) 1735–1780.
61. Haykin, S., ve Kosko, B., GradientBased Learning Applied to Document Recognition, in Intelligent Signal Processing, 2001, 306 – 351.
62. Deng, L. ve Yu, D., Deep convex net: A scalable architecture for speech pattern classification, in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2011, 2285–2288.
63. Bengio, Y., Learning deep architectures for AI, Now Publishers ,2009.

64. Arel, I., Rose, D. ve Karnowski, T., Deep machine learning-A new frontier in artificial intelligence research, IEEE Computational Intelligence Magazine, 5, 4 (2010) 13–18.
65. Cho, K., Van Merriënboer, B., Bahdanau, D., ve Bengio, Y., On the Properties of Neural Machine Translation: Encoder–Decoder Approaches, CoRR, (2015) 103–111.
66. Shen, D., Wu, G. ve Suk, H., Deep learning in medical image analysis, Annual review of biomedical engineering, 19 (2017) 221–248.
67. Nair, V. ve Hinton, G. E., Rectified linear units improve restricted boltzmann machines, In Proceedings of the 27th international conference on machine learning, 2010, 807–814.
68. Michalski, R.S., Carbonell, J.G. ve Mitchell, T.M, Machine learning: An artificial intelligence approach. Springer Science & Business Media, 2013.
69. Hubel, D. ve Wiesel, T., Receptive fields and functional architecture of monkey striate cortex, Journal of Physiology, 195, 1(1968) 215-243.
70. Collobert, R. ve Weston, J., A unified architecture for natural language processing: Deep neural networks with multitask learning, In: Proceedings of the 25th international conference on Machine learning. ACM, 2008, 160–167.
71. Clark, C. ve Storkey, A., Teaching Deep Convolutional Neural Networks to Play Go. arXiv preprint arXiv:1412.3409, 2014.
72. LeCun, Y., Bottou, L., Bengio, Y. ve Haffner, P., Gradient-based learning applied to document recognition, In: Proceedings of the IEEE, 86, 11(1998), 499–514.
73. Ronneberger, O., Fischer, P. ve Brox, T., U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer Assisted Intervention. Springer, 2015, 234–241.
74. Yu, F. ve Koltun, V., Multi-scale context aggregation by dilated convolutions, International Conference on Learning Representations (ICLR), 2016.
75. Lowe, D.G., Object recognition from local scale-invariant features, Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.
76. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G. ve ark., Deformable convolutional networks, IEEE International Conference on Computer Vision, 2017.
77. Shlens, J., Notes on Kullback-Leibler Divergence and Likelihood, CoRR, abs/1404, 2000 (2014).
78. 78. Ruder, S., An overview of gradient descent optimization algorithms, CoRR, arXiv:1609.04747 (2016).

79. Kingma, D.P., ve Ba, J., Adam: A method for stochastic optimization, The 3rd International Conference for Learning Representations, 2015.
80. Duchi, J., Hazan, E., ve Singer, Y., Adaptive subgradient methods for online learning and stochastic optimization, Journal of Machine Learning Research, 12 (2011) 2121–2159.
81. Tieleman, T., ve Hinton, G., RmsProp: Divide the gradient by a running average of its recent magnitude, COURSE: Neural Networks for Machine Learning, 2012.
82. Wen, W., Wu, C., Wang, Y., Chen, Y. ve Li, H., Learning structured sparsity in deep neural networks, In Advances in neural information processing systems, 2016, 2074–2082.
83. Deng, J., Dong, W., Socher, R., Li, L., Li, K., ve Fei-Fei, L., ImageNet: A large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, 248–255.
84. Elman, J.L, Finding Structure in Time, Cognitive Science, 14, 2 (1990) 179-211.
85. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., ve ark., Attention Is All You Need, Advances in Neural Information Processing Systems 30, 2017, 16-27.
86. Kipf, T.N., Welling, M., Semi-Supervised Classification with Graph Convolutional Networks, International Conference on Learning Representations, 2017.
87. Zeiler, M.D. ve Fergus, R., Visualizing and Understanding Convolutional Networks, European Conference on Computer Vision (ECCV), 2014, 818–833.
88. Simonyan, K. ve Zisserman, A., Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representations (ICLR'15), 2015.
89. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. ve Rabinovich, A., Going Deeper with Convolutions IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
90. He, K., Zhang, X., Ren, S. ve Sun J., Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770–778.
91. Adelson, E., Anderson, C., Bergen, J., Burt, P. ve Ogden, J., Pyramid Methods in Image Processing, In: RCA, 1983, 33-41.
92. He, K., Zhang, X., Ren, S., Sun, J., Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, European Conference on Computer Vision, 2014, 346–361.

93. Cai, Z., Fan, Q., Feris, R.S. ve Vasconcelos, N., A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection, In European Conference on Computer Vision, 2016
94. Girshick, R., Donahue, J., Darrell, T. ve Malik, J., Rich feature hierarchies for accurate object detection and semantic segmentation, The IEEE Computer Vision and Pattern Recognition Conference, 2014.
95. Girshick, R., Fast r-cnn, Proceedings of the IEEE international conference on computer vision, Aralık 2015, Santiago, Bildiriler Kitabı: 1440–1448.
96. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. ve Berg, A.C, SSD: Single Shot MultiBox Detector, European Conference on Computer Vision, 2016, 21–37.
97. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. ve Belongie, S., Feature pyramid networks for object detection, The IEEE Computer Vision and Pattern Recognition Conference, 2017.
98. Liu, S., Qi, L., Qin, H., Shi, J. ve Jia, J., Path Aggregation Network for Instance Segmentation, The IEEE Computer Vision and Pattern Recognition Conference, 2018, 8759-8768.
99. Ghiasi, G., Lin, T., Pang, R., ve Le, Q., NASFPN: Learning scalable feature pyramid architecture for object detection, arXiv:1904.07392, 2019.
100. Skurowski, P., Abdulameer, H., Błaszczuk, J., Depta, T., Kornacki, A. ve Koziel, P., Animal camouflage analysis: Chameleon database, unpublished Manuscript, 2018.
101. Yan, Q., Xu, L., Shi, J. ve Jia, J., Hierarchical saliency detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, 1155–1162.
102. Yang, C., Zhang L., Lu H., Ruan X. ve Yang, M., Saliency detection via graph-based manifold ranking, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, 3166–3173.
103. Wang, L., Lu, H., Wang Y., Feng, M., Wang D., Yin B. ve Ruan, X., Learning to detect salient objects with image-level supervision, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 136–145.
104. Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C. ve Vilariño, F., WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, Computerized Medical Imaging and Graphics, 43 (2015) 99–111.
105. Jha, D., Smedsrud, P. H. , Riegler, M. A. , Halvorsen, P., de Lange, T. , Johansen D. ve Johansen, H. D., Kvasir-seg: A segmented polyp dataset, International Conference on Multimedia Modeling. Springer, 2020, 451–462.

106. Silva, J., Histace, A., Romain, O., Dray, X. ve Granado, B., Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer, International Journal of Computer Assisted Radiology and Surgery, 9, 2 (2014) 283–293.
107. Vázquez, D., Bernal, J., Sánchez, J., Fernández-Esparrach, G., López, A.M, Romero, A., Drozdal, M. ve Courville, A., A benchmark for endoluminal scene segmentation of colonoscopy images, Journal of Healthcare Engineering, 2017.
108. Everingham, M., Eslami, S.A., Gool, L.V., Williams, C.K., Winn, J. ve Zisserman, A., The pascal visual object classes challenge: A retrospective, International Journal of Computer Vision, 111, 1(2014) 98–136.
109. Arbelaez, P., Maire, M., Fowlkes, C. ve Malik, J, Contour detection and hierarchical image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33 (2010) 898–916.
110. Fan, D.P., Cheng, M.M., Liu, Y., Li, T. ve Borji, A., Structure-measure: A new way to evaluate foreground maps, In Proceedings of the IEEE International Conference on Computer Vision, 2017, 4548–4557.
111. Perazzi, F., Krähenbühl, P., Pritch, Y. ve Hornung, A., Saliency filters: Contrast based filtering for salient region detection, In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, 733–740.
112. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M. ve Borji, A., Enhanced-alignment measure for binary foreground map evaluation, In Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018, 698–704.
113. Szegedy, C., Vanhoucke, V., Loffe, S., Shlens, J. ve Wojna, Z., Rethinking the inception architecture for computer vision, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 2818–2826.
114. Zhao, H., Shi, J., Qi, X., Wang, X., ve Jia, J., Pyramid scene parsing network, In: The IEEE conference on computer vision and pattern recognition, 2017, 2881–2890.
115. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. ve Darrell, T., Caffe: Convolutional architecture for fast feature embedding, ACM international conference on Multimedia, 2014, 675–678.
116. Zhang, P.P., Wang, D., Lu, H.C. ve Ruan, X., Amulet: Aggregating multi-level convolutional features for salient object detection, In Proceedings of the 2017 IEEE International Conference on Computer Vision, 2017, 202–211.
117. Zhang, P.P., Wang, D., Lu, H.C., Wang, H.Y. ve Yin, B.C., Learning uncertain convolutional features for accurate saliency detection, In Proceedings of the 2017 IEEE International Conference on Computer Vision, 2017, 212–221.

118. Zhang, P.P., Liu, W., Lu, H.C. ve Shen, C.H., Salient object detection by lossless feature reflection, In Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018, 1149–1155.
119. Zhang, P., Liu, W., Lei, Y. ve Lu, H., Hyperfusion-Net: Hyper-densely reflective feature fusion for salient object detection, Pattern Recognition, 93 (2019) 521-533.
120. Zeng, Y., Zhang, P., Zhang, J., Lin, Z. ve Lu, H., Towards High-Resolution Salient Object Detection, IEEE International Conference on Computer Vision (ICCV), 2019, 7234–7243.
121. Feng, M., Lu, H. ve Ding, E., Attentive feedback network for boundary-aware salient object detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, 1623–1632.
122. He, K., Gkioxari, G., Dollar, P., ve Girshick, R., Mask R-CNN, IEEE International Conference on Computer Vision, 2017, 2980–2988.
123. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N. ve Liang, J., Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, Journal of IEEE Transactions on Medical Imaging, 39, 6 (2020) 1856 – 1867.
124. Liu, N., Han, J. ve Yang, M.-H., Picanet: Learning pixel-wise contextual attention for saliency detection, IEEE Conference on Computer Vision and Pattern Recognition, 2018, 3089–3098.
125. Huang, Z., Huang, L., Gong, Y., Huang, C. ve Wang, X., Mask scoring R-CNN, IEEE Conference on Computer Vision and Pattern Recognition, 2019, 6409–6418.
126. Liu, J.J., Hou, Q., Cheng, M.M., Feng, J. ve Jiang, J., A Simple Pooling-Based Design for Real-Time Salient Object Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
127. Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M. ve Jagersand, M., Basnet: Boundary-aware salient object detection, IEEE Conference on Computer Vision and Pattern Recognition, 2019, 7479–7489.
128. Wu, Z., Su, L. ve Huang, Q., Cascaded partial decoder for fast and accurate salient object detection, IEEE Conference on Computer Vision and Pattern Recognition, 2019, 3907–3916.
129. Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., ve ark., Hybrid task cascade for instance segmentation, IEEE Conference on Computer Vision and Pattern Recognition 2019, 4974–4983.
130. Zhao, J.-X., Liu, J.-J., Fan, D.-P., Cao, Y., Yang, J. ve Cheng, M.-M., Egnet: Edge guidance network for salient object detection, Proceedings of the IEEE International Conference on Computer Vision, 2019, 8779–8788.

131. Fang, Y., Chen, C., Yuan, Y. ve Tong, K., Selective feature aggregation network with area boundary constraints for polyp segmentation, *International Conference on Medical Image Computing and Computer Assisted Intervention*, Springer, 2019, 302–310.
132. Fan, D.-P., Ji, G.-P. Zhou, T., Chen, G., Fu, H., Shen, J. ve Shao, L., Pranet: Parallel reverse attention network for polyp segmentation, *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2020, 263–273.
133. Matthews, O., Liggins, E., Volonakis, T., Scott-Samuel, N., Baddeley, R. ve Cuthill, I., Human visual search performance for camouflaged targets. *Journal of Vision*, 15,12 (2015) 1164–1164.
134. Adams, W-J., Graf, E-W. ve Anderson, M., Disruptive coloration and binocular disparity: breaking camouflage, *Proceedings of the Royal Society B*, 286, 1896 (2019) 2018-2045.
135. Kelman, E-J., Osorio, D. ve Baddeley., R-J., A review of cuttlefish camouflage and object recognition and evidence for depth perception. *Journal of Experimental Biology*, 211, 11 (2008)1757–1763.
136. 136. Zhang, M., Yao, S., Hu, B., Piao, Y. ve Ji, W., C2DFNet: Criss-Cross Dynamic Filter Network for RGB-D Salient Object Detection, in *IEEE Transactions on Multimedia*, (2022) 1 – 13.
137. Lee, M., Park, C., Cho, S., Lee, S., SPSN: Superpixel Prototype Sampling Network for RGB-D Salient Object Detection, *European Conference on Computer Vision*, 2022, 630–647.
138. Ji, W., Li, J., Yu, S., Zhang, M., Piao, Y., Yao, S., Bi, Q., Ma, K., Zheng, Y., Lu, H., ve ark, Calibrated rgb-d salient object detection, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 9471–9481.
139. Zhang, J., Fan, D.P., Dai, Y., Anwar, S., Saleh, F.S., Zhang, T., Barnes, N., Ucnnet: Uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, 8582–8591.
140. Zhu, P., Du, D., Wen, L., Bian, X., Ling, H., Hu, Q., ve ark, VisDrone-VID2019: the vision meets drone object detection in video challenge results, *IEEE/CVF International Conference on Computer Vision Workshop*, 2019.
141. Tan, L., Lv, X., Lian, X., Wang, G., YOLOv4_Drone: UAV image target detection based on an improved YOLOv4 algorithm, *Computers and Electrical Engineering*, 93 (2021).
142. Fu, L., Gu, W-B, Li W, Chen, L., Ai Y-B, Wang H-L, Bidirectional parallel multibranch convolution feature pyramid network for target detection in aerial images of swarm UAVs, *Defence Technology*, 17, 4 (2021) 1531-1541.

143. Stone, M., Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society*, 36, 2 (1974) 111–147.
144. Wei, J., Wang, S. ve Huang, Q., F3Net: Fusion, Feedback and Focus for Salient Object Detection, In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.
145. Zhao, X., Pang, Y., Zhang, L., Lu, H. ve Zhang, L., Suppress and Balance: A Simple Gated Network for Salient Object Detection, *European Conference on Computer Vision*, 2020.
146. Chen, Z., Xu, Q., Cong, R. ve Huang, Q., Global Context-Aware Progressive Aggregation Network for Salient Object Detection. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.
147. Pang, Y., Zhao, X., Zhang, L., ve Lu, H., Multi-Scale Interactive Network for Salient Object Detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 9410–9419.
148. Qin, X., Zhang, Z. V., Huang, C., Dehghan, M., Zaiane, O. R. ve Jagersand, M., U2-Net: Going deeper with nested U-structure for salient object detection, *Pattern Recognition*, 106 (2020).
149. Wei, J., Wang, S.; Wu, Z., Su, C., Huang, Q. ve Tian, Q., Label Decoupling Framework for Salient Object Detection. In *IEEE/CVF Conference on Computer Vision*, 2020.
150. Liu, N., Zhang, N., Wan, K., Shao, L. ve Han, J., Visual Saliency Transformer. In *IEEE/CVF International Conference on Computer Vision*, 2021, 4702–4712.
151. Yun, Y.K ve Lin, W., SelfReformer: Self-Refined Network with Transformer for Salient Object Detection, *arXiv preprint arXiv:2205.11283*, (2022).
152. Zhu, J., Zhang, X., Zhang, S. ve Liu, J., Inferring camouflaged objects by texture-aware interactive guidance network, in *AAAI Conference on Artificial Intelligence*, 2021, 3599–3607.
153. Mei, H., Ji, G.-P., Wei, Z., Yang, X., Wei, X., ve Fan, D.-P., Camouflaged object segmentation with distraction mining, in *Conference on computer vision and pattern recognition*, 2021, 8772–8781.
154. Zhai, Q., Li, X., Yang, F., Chen, C., Cheng, H. ve Fan, D.-P., Mutual graph learning for camouflaged object detection, in *Conference on computer vision and pattern recognition*, 2021, 12997–13007.
155. Zhai, W., Cao, Y., Xie, H. ve Zha, Z. -J., Deep Texton-Coherence Network for Camouflaged Object Detection, in *IEEE Transactions on Multimedia*, 2022.

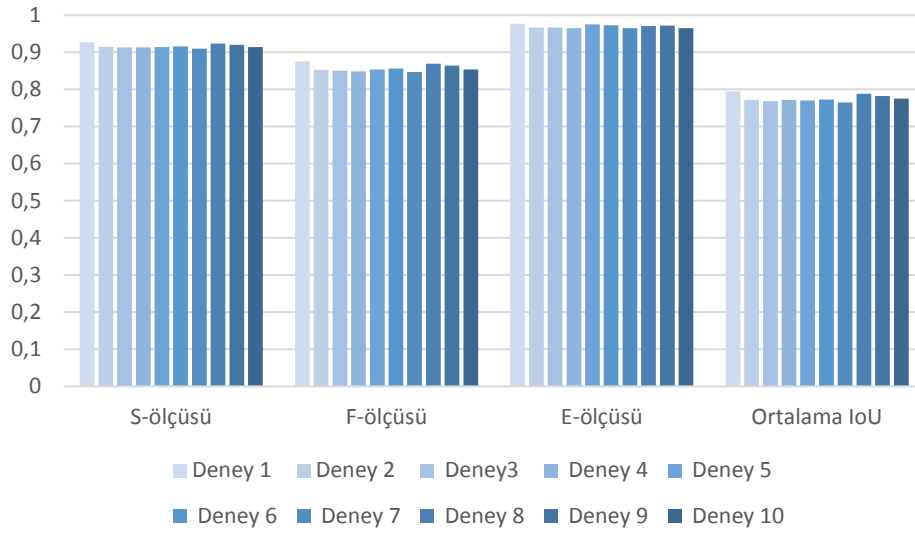
7. EKLER

Ek 1. N-Kat apraz Doğrulama

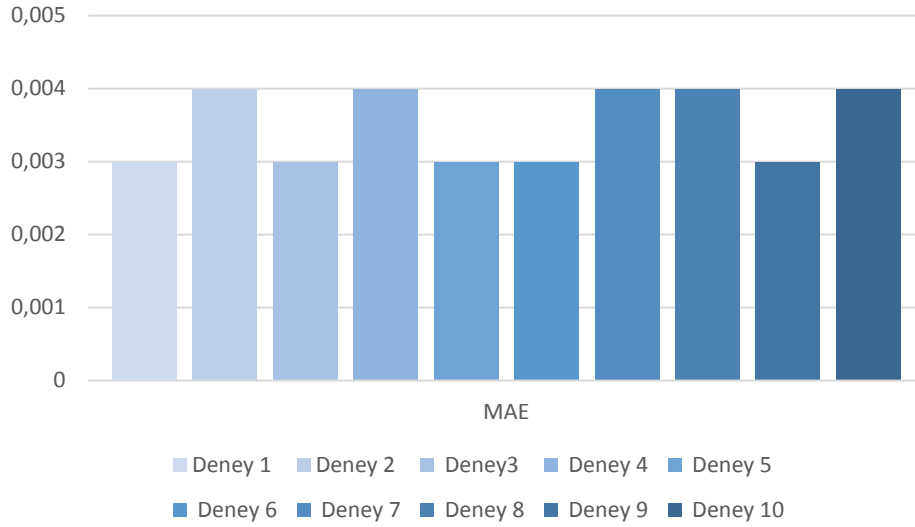
N-kat apraz doğrulama [143] makine öğrenmesi deneylerinde sonuçların geçerliliğini artırmak için kullanılan bir tekniktir. Her sınıfın S- ölçümü, F- ölçümü, E- ölçümü, MAE, IoU puanı her deney için hesaplanır ve bu değerlerin ortalama değerleri CPD1 veri seti için yöntemin performansını ölçmek için alınır.

Tablo 7.1. CPD1 veri seti için N-kat apraz doğrulama sonuçları

Deney No	$S_{\alpha} \uparrow$	$F_{\beta}^{ort} \uparrow$	$E_{\phi}^{ort} \uparrow$	$M \downarrow$	$IoU^{ort} \uparrow$
1	0.927	0.875	0.977	0.003	0.795
2	0.915	0.853	0.967	0.004	0.772
3	0.913	0.850	0.967	0.003	0.768
4	0.913	0.849	0.965	0.004	0.772
5	0.914	0.854	0.975	0.003	0.770
6	0.916	0.856	0.973	0.003	0.773
7	0.910	0.847	0.965	0.004	0.765
8	0.924	0.869	0.971	0.004	0.788
9	0.920	0.864	0.972	0.003	0.782
10	0.914	0.854	0.965	0.004	0.775
Ortalama	0.916	0.857	0.969	0.003	0.776



Şekil 7.1. Tüm testler için S- ölçümü, F- ölçümü, E- ölçümü ve IoU ölçümü puanları



Şekil 7.2. Tüm testler için MAE puanları

Ek 2. Önerilen Yöntemin Yeni Kamuflajsız Nesne Algılama Yöntemlerle Nicel Karşılaştırması

Tablo 7.2. Yeni SOD modellerinin önerilen yöntemle değerlendirme sonuçları

	Yöntem	$S_\alpha \uparrow$	$F_\beta^{ort} \uparrow$	$E_\phi^{ort} \uparrow$	$M \downarrow$
DUTS-TE	F3Net [144]	0.888	0.852	0.920	0.035
	GateNet [145]	0.885	0.837	0.906	0.040
	GCPA [146]	0.891	0.841	0.912	0.038
	MINet [147]	0.884	0.844	0.917	0.037
	U2Net [148]	0.874	0.826	0.897	0.045
	LDF [149]	0.892	0.861	0.925	0.034
	VST [150]	0.896	0.845	0.920	0.037
	SELFREFORMER [151]	0.911	-	-	0.026
	Önerilen Yöntem	0.889	0.873	0.926	0.032
DUT-OMRON	F3Net [144]	0.838	0.766	0.864	0.053
	GateNet [145]	0.838	0.757	0.856	0.055
	GCPA [146]	0.839	0.756	0.853	0.056
	MINet [147]	0.833	0.757	0.860	0.056
	U2Net [148]	0.847	0.775	0.867	0.054
	LDF [149]	0.839	0.770	0.865	0.052
	VST [150]	0.850	0.774	0.871	0.058
	SELFREFORMER [151]	0.856	-	-	0.041
	Önerilen Yöntem	0.832	0.775	0.865	0.050
ECSSD	F3Net [144]	0.924	0.925	0.948	0.033
	GateNet [145]	0.920	0.913	0.936	0.040
	GCPA [146]	0.927	0.916	0.944	0.035
	MINet [147]	0.925	0.923	0.950	0.033
	U2Net [148]	0.928	0.923	0.947	0.033
	LDF [149]	0.924	0.927	0.948	0.034
	VST [150]	0.932	0.920	0.951	0.033
	SELFREFORMER [151]	0.935	-	-	0.027
	Önerilen Yöntem	0.916	0.930	0.945	0.035

Ek 3. Önerilen Yöntemin Yeni Kamuflej Nesne Algılama Yöntemlerle Nicel Karşılaştırması

Önerilen model, yakın zamanda önerilen COD ağları ile test edilmiştir. Adil bir karşılaştırma için, önerilen yöntem literatürde kullanılan aynı veri setleri ile 4.040 örnek içeren hibrit veri seti (COD10K-Tr: 3.040 görüntü + CAMO-Tr:1.000 görüntü üzerinde yeniden eğitilmiş ve önerilen yöntem yukarıdaki ölçütler (COD10K-TE: 2.026 görüntü ve CAMO-TE:250 görüntü) üzerinde değerlendirilmiştir.

Tablo 7.3. COD10K veri seti için önerilen yöntem ile son COD modellerinin değerlendirme sonuçları

Yöntem	Parametre sayısı (m)	$S_{\alpha} \uparrow$	$F_{\beta}^{ort} \uparrow$	$E_{\phi}^{ort} \uparrow$	$M \downarrow$
SINet [51]	48.95	0.771	0.635	0.808	0.051
TINet [152]	28.56	0.793	0.679	0.861	0.042
PFNet [153]	46.5	0.800	0.701	0.877	0.040
S-MGL [154]	63.6	0.811	0.702	0.845	0.037
DTC-Net [155]	-	0.790	-	-	0.041
Önerilen Yöntem	62.79	0.783	0.689	0.879	0.036

ÖZGEÇMİŞ

İlkokul, ortaokul ve lise öğrenimini Tunus'ta tamamladı. 2008 yılında Uygulamalı Matematik ve Bilgisayar Bilimleri Üniversitesi'nde ses ve görüntülerde dijital teknolojiler bölümünü başladı. 2011 yılında Enformatik ve Multimedya Üniversitesi, bilgisayar bilimler ve multimedya 'da başladığı yüksek lisansını tamamladı. 2016 yılında Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalında doktora çalışmalarına başladı.

Doktora çalışması sırasında, 2 SCI indeksli dergide makale yayınlanmış 1 makale ise TR indeksli dergiye gönderilmiştir.

Uluslararası hakemli dergilerde yayınlanan makaleler (SCI/SCI-E)

1. Hendaoui,R., NabiyeV, V. (2022). An end-to-end neural network for detecting hidden people in images based on multiple attention network, Multimedia Tools and Applications, 81(13): 18531-18542.

2. Hendaoui,R., NabiyeV, V. (2022). A novel deep neural network for hidden target detection in images. Kuwait Journal of Science, 49 (4):1-9.