T.C.
GÜMÜŞHANE ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ VE AKILLI SİSTEMLER ANA BİLİM DALI

OTOMATİK AVUÇ İÇİ BÖLÜTLENMESİNİN GERÇEKLEŞTİRİLMESİNDE FARKLI DERİN ÖĞRENME MODELLERİNİN PERFORMANSLARININ İNCELENMESİ

YÜKSEK LİSANS

Kadir YALÇIN

OCAK-2025
GÜMÜŞHANE

T.C.
GÜMÜŞHANE ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ VE AKILLI SİSTEMLER ANA BİLİM DALI

OTOMATİK AVUÇ İÇİ BÖLÜTLENMESİNİN GERÇEKLEŞTİRİLMESİNDE FARKLI DERİN ÖĞRENME MODELLERİNİN PERFORMANSLARININ İNCELENMESİ

INVESTIGATION OF THE PERFORMANCES OF DIFFERENT DEEP LEARNING MODELS IN THE IMPLEMENTATION OF AUTOMATIC PALM PRINT SEGMENTATION

YÜKSEK LİSANS

Kadir YALÇIN

OCAK-2025
GÜMÜŞHANE

T.C.
GÜMÜŞHANE ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ VE AKILLI SİSTEMLER ANA BİLİM DALI

OTOMATİK AVUÇ İÇİ BÖLÜTLENMESİNİN GERÇEKLEŞTİRİLMESİNDE FARKLI DERİN ÖĞRENME MODELLERİNİN PERFORMANSLARININ İNCELENMESİ

INVESTIGATION OF THE PERFORMANCES OF DIFFERENT DEEP LEARNING MODELS IN THE IMPLEMENTATION OF AUTOMATIC PALM PRINT SEGMENTATION

YÜKSEK LİSANS

Kadir YALÇIN

Danışman: Dr.Öğr.Üyesi Ramazan Özgür DOĞAN

OCAK-2025
GÜMÜŞHANE

KABUL VE ONAY

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Yüksek Lisans Tezi olarak hazırlamış olduğum "**Otomatik avuç içi bölütlenmesinin gerçekleştirilmesinde farklı derin öğrenme modellerinin performanslarının incelenmesi**" isimli bu tezimin, tamamen kendi çalışmam olduğunu, her alıntıya kaynak gösterdiğimi, alıntı yaptığım tüm çalışmaları kaynakçada belirtiğimi ve Gümüşhane Üniversitesi'nin lisanslı kullanıcısı olduğu intihal yazılım programı ile Lisansüstü Eğitim Enstitüsü'nün belirlediği kıstaslara uygun olarak raporladığımı taahhüt ederim. Tezimin kâğıt ve elektronik kopyalarının Gümüşhane Üniversitesi Lisansüstü Eğitim Enstitüsü arşivinde saklanmasına izin verdiğimi onaylarım.

Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

22/01/2025

Kadir YALÇIN

TEŞEKKÜR

Çalışmalarım sırasında bilgi ve tecrübelerinden faydalandığım tez danışmanım Trabzon Üniversitesinden Sayın Dr.Öğr.Üyesi Ramazan Özgür DOĞAN'a, Gümüşhane Üniversitesi Yazılım mühendisliği bölümünden Sayın Dr.Öğr.Üyesi Özkan BİNGÖL'e, Gümüşhane Üniversitesi Elektrik Elektronik mühendisliği bölümünden Sayın Dr.Öğr.Üyesi Gökhan ÇETİN'e, Ağrı İbrahim Çeçen üniversitesinden Sayın Prof. Dr. Mehmet YALÇIN'a teşekkür ederim.

Bu tez çalışması TÜBİTAK 1001 - Bilimsel ve Teknolojik Araştırma Projelerini Destekleme Programı tarafından desteklenen 122E402 nolu "Kısıtlamasız Ortamda Alınan El Görüntülerinden Biyometrik Doğrulama İçin Güvenilir Avuç İzi Bölgesinin Tespiti" isimli proje kapsamında gerçekleştirilmiştir. Bu desteklerinden dolayı TÜBİTAK'a teşekkürlerimi sunarım.

Kadir YALÇIN

GÜMÜŞHANE – 2025

ÖZET

Tıbbi görüntüleme teknolojilerindeki ilerlemeler, özellikle derin öğrenme tabanlı yöntemlerin kullanımıyla, el bölütleme gibi spesifik görevlerde dikkat çekici başarılar elde etmiştir. El bölütleme, biyometrik doğrulama sistemlerinde kimlik tespiti için kritik bir öneme sahiptir. Avuç içi damar yapıları, parmak uzunlukları ve deri dokusu gibi benzersiz biyometrik özelliklerin doğru bir şekilde çıkarılması, el bölütleme algoritmalarının doğruluğuna bağlıdır. Bu tez çalışmasında, el bölütleme probleminin çözümü için oto kodlayıcı tabanlı U-Net, MA-Net, FPN ve PSP-Net modelleri, Vision Transformer, EfficientNet-B5 ve ResNet-34 omurgaları ile değerlendirilmiştir. Modeller, Doğruluk (ACC), Intersection over Union (IoU), Dice katsayısı ve F1 skor gibi metriklerle analiz edilmiştir. Sonuçlar, Vision Transformer ile entegre edilen U-Net modelinin bölütleme de en yüksek başarı sağladığını göstermiştir (IoU: 0.8917, Dice: 0.9892). MA-Net, ResNet-34 omurgasıyla kullanıldığında en yüksek doğruluk değerine (0.9921) ulaşmıştır. Ayrıca bu kombinasyon, en yüksek F1 skorunu (0.9584) da elde etmiştir. EfficientNet-B5, MA-Net ve FPN ile etkili sonuçlar sunmuştur. Çalışmanın sonuçları, derin öğrenme modelleriyle elde edilen bölütleme başarısının biyometrik doğrulama sistemlerinin güvenilirliğini artırabileceğini ortaya koymaktadır. Özellikle U-Net - Vision Transformer ve MA-Net – ResNet-34 kombinasyonları, elin bölütlenmesinde yüksek performans göstermiştir. Bu bulgular, el bölütleme uygulamalarında model ve omurga seçimlerinin kritik olduğunu vurgulamaktadır.

Anahtar Kelimeler: El bölütleme, FPN, MA-Net, PSP-Net, U-Net

SUMMARY

Advances in medical imaging technologies, particularly with the use of deep learning based methods, have achieved remarkable success in specific tasks such as hand segmentation. Hand segmentation is of critical importance for identification in biometric verification systems. Accurate identification of unique biometric features such as palm vein structures, finger lengths and skin texture depends on the accuracy of hand segmentation algorithms. In this study, autoencoder based U-Net, MA-Net, FPN and PSP-Net models have been evaluated with Vision Transformer, EfficientNet-B5 and ResNet-34 backbones for solving the hand segmentation problem. The models were analyzed with metrics such as Accuracy (ACC), Intersection over Union (IoU), Dice coefficient and F1 score. The results showed that the U-Net model integrated with Vision Transformer provided the highest success in segmentation (IoU: 0.8917, Dice: 0.9892). MA-Net achieved the highest accuracy value (0.9921) when used with ResNet-34 backbone. In addition, this combination achieved the highest F1 score (0.9584). EfficientNet-B5 provided effective results with MA-Net and FPN. The results of the study revealed that the segmentation success obtained with deep learning models could increase the reliability of biometric verification systems. In particular, U-Net - Vision Transformer and MA-Net – ResNet-34 combinations showed high performance in hand segmentation. The findings emphasize that model and backbone selections are critical in hand segmentation applications.

Keywords: Hand segmentation, FPN, MA-Net, PSP-Net, U-Net

İÇİNDEKİLER

TABLOLAR DİZİNİ

ŞEKİLLER DİZİNİ

SİMGELER VE KISALTMALAR DİZİNİ

3DFCN	: 3 Boyutlu Tam Evrişim Ağı
AR	: Artırılmış Gerçeklik
BERT	: Transformatörlerin Çift Yönlü Kodlayıcı Gösterimleri
CNN	: Evrişimli Sinir Ağı
CT	: Bilgisayarlı Tomografi
FPN	: Özellik Piramit Ağı
GAN	: Üretken Çatışmacı Ağlar
GPT	: Üretici Önceden Eğitilmiş Dönüştürücü
GPU	: Grafik İşlemci Ünitesi
MA-Net	: Çok Ölçekli Dikkat Ağı
NLP	: Doğal Dil İşleme
PSP-Net	: Piramit Sahne Ayrıştırma Ağı
Resnet	: Derin Artık Ağlar
R-CNN	: Bölge Tabanlı Evrişimli Sinir Ağı
VAE	: Varyasyonel Otokodlayıcı
VGG	: Çok Derin Evrişimli Ağlar
ViT	: Vision Transform

1. GİRİŞ

İnsan vücudunun kendine özgü biyolojik veya davranışsal özelliklerini kullanarak güvenilir ve etkili kimlik doğrulama çözümleri sunan bir teknoloji olan biyometri, günümüzde oldukça önemli bir yere sahiptir (Yang vd., 2019). Yapılan çalışmalar, biyometrik sistemlerin farklı alanlarda kullanımının yaygınlaştığını göstermektedir (Alsaadi vd., 2021). Biyometrinin temel amacının kimlik doğrulama olduğu ve bu doğrulamanın kişiye özgü biyometrik özellikler sayesinde yüksek doğruluk oranlarıyla gerçekleştirildiği literatürde vurgulanmaktadır. Güvenlik sistemlerinden kişisel cihazlara kadar birçok ticari üründe biyometrik teknolojiler kullanılmaktadır (Dargan ve Kumar, 2020). Yüz, parmak izi ve iris tanıma gibi teknolojiler, biyometrik sistemler arasında en yaygın kullanılan yöntemler olarak öne çıkmaktadır. Bu sistemler, güvenilirlikleri ve kullanım kolaylıkları sayesinde çeşitli alanlarda yaygın olarak benimsenmiştir. Son yıllarda ise avuç içi biyometrisinin, kullanım kolaylığı, geniş yüzey alanı ve diğer yöntemlere göre daha fazla özellik barındırması gibi avantajlarla dikkat çektiği gözlenmektedir (Alausa vd., 2022).

Avuç içi biyometrisinin biyometrik tanıma sistemlerinde öne çıkmasının başlıca sebeplerinden biri, avuç içlerinde bulunan zengin özellik çeşitliliğidir. Yaşam çizgisi ve kalp çizgisi gibi el yapısındaki ana çizgiler, bu özelliklerin en bilinenleridir. Ayrıca, parmak izlerine özgü avuç içindeki küçük detaylar da tanıma işlemlerinde kullanılabilir. Avuç içlerindeki bu çok katmanlı özellikler, biyometrik tanıma ve doğrulama sistemlerinin doğruluk oranlarını ve güvenilirliğini artıran önemli faktörler olarak kabul edilmektedir (Alsaadi vd., 2021).

Avuç içi biyometrisinin yaygınlaşmasındaki bir diğer etken ise kullanıcı dostu olmasıdır. Parmak izi biyometrisine benzer şekilde, avuç içi tanıma sistemleri de kullanıcıların kolayca etkileşim kurabileceği bir yapı sunmaktadır. Avuç içleri, daha geniş bir tarama alanı sağladığı için tarama cihazlarının doğruluğunu artırma potansiyeline de sahiptir. Bu özellik, fiziksel temasın azaltıldığı ve tarama sürecinin daha verimli hale getirildiği uygulamalarda kullanım kolaylığı sağlamaktadır (Alausa vd., 2022). Biyometrik özelliklerin tanınmasında kullanılan çeşitli algoritmalar ve derin öğrenme modelleri, bu sistemlerin performansını daha da iyileştirmektedir. Özellikle son yıllarda yapılan çalışmalarda, makine öğrenimi ve örüntü işleme teknikleri kullanılarak tanıma süreçlerinin daha otomatik, hızlı ve güvenilir hale getirildiği görülmektedir. Bu

gelişmeler, avuç içi biyometrisinin kimlik doğrulama sistemlerindeki önemini artırmaktadır (Minaee vd., 2023).

Avuç içi okuma ve el algılama sistemlerinin herhangi bir kural veya sınırlama olmaksızın işlediği durumlara "kısıtlamasız ortam" denir. Alışılagelmiş avuç içi tanıma metotlarında, sistemler genellikle kapalı bir kutu içinde veya elin özel bir aparata yerleştirildiği kontrollü şartlar altında çalışır. Bu tür sistemler, elin sabit bir duruşta olmasını ve çevresel etkenlerin kontrol edilmesini sağlayarak yüksek doğruluk oranları sunar. Ancak bu çalışmada benimsenen yaklaşım bu tarz kısıtlamaları içermez. Kısıtlamasız ortam, gerçek yaşamın değişken koşullarında, elin herhangi bir pozisyonda veya doğal şekilde durduğu hallerde dahi doğru sonuçlar üretebilen bir sistemi tanımlar (Han vd,. 2007). Bu yaklaşımın en büyük zorluğu, elin karmaşık zeminlerden veya farklı ışıklandırma koşullarından etkilenmeden doğru biçimde tespit edilmesidir. Çalışmada benimsenen, kısıtlamasız bir ortamda elin konumunu ve sınırlarını saptamaktır. Buradaki asıl nokta, avuç içi tanımanın ötesinde, elin kendisinin bu doğal ortamlarda güvenilir bir şekilde algılanabilmesidir. Esas yaklaşım, herhangi bir kısıtlama olmaksızın elin bulunmasıdır. Bu hedef doğrultusunda, kısıtlamasız ortamda el bölütleme teknolojisinin ilerlemesi, yalnızca biyometrik sistemlerde değil, aynı zamanda insan-bilgisayar etkileşimleri ve artırılmış gerçeklik uygulamaları gibi birçok alanda yeni imkanlar sunabilmesidir. Kısıtlamasız ortamların sağladığı adaptasyon yeteneği, sistemlerin daha geniş bir kullanıcı kitlesine ulaşmasını mümkün kılarak, hem kullanım kolaylığı hem de işlevsellik açısından önemli bir gelişme olacaktır.

Şekil 1. Kısıtlamalı ortam örnek bir görüntü

Şekil 2. Kısıtlamasız ortam örnek bir görüntü

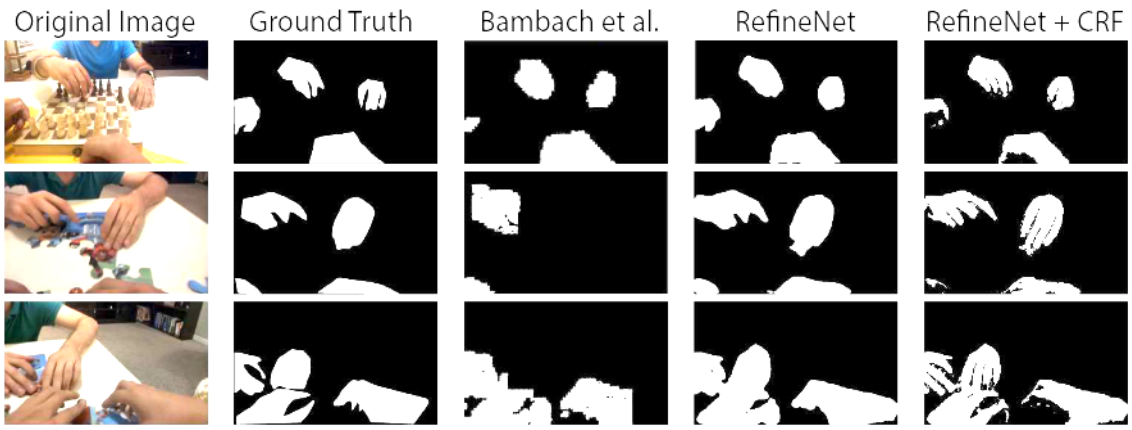
2. LİTERATÜR ARAŞTIRMASI

El bölütleme için derin öğrenme tabanlı bir model sunulan çalışmada, derin öğrenme modeli, iki ana öğrenme yolu içerir: Direct Learner (Doğrudan Öğrenici) ve Context Learner (Bağlamsal Öğrenici) (Neverova vd., 2014). Her iki yol da derin evrişimsel sinir ağlarına (CNN) dayalıdır ve aynı mimariye sahiptir. Doğrudan Öğrenici, pikselleri bağımsız olarak sınıflandırırken; Bağlamsal Öğrenici, çevre piksellerin ilişkisini analiz ederek bağlamsal bilgilerden faydalanır. Modelin mimarisi, üç evrişim katmanı içerir. Her katmanda filtreler ile özellik çıkarımı yapılır. Filtreleme sonrasında azalma olmadan çözünürlüğü korumak için özel bir alt örnekleme yöntemi kullanılır. Max pooling işlemi, görüntü boyutlarını küçültmeden bilgi yoğunluğunu artırmak için optimize edilmiştir. Modelin öğrenme süreci, sentetik verilerle başlatılan gözetimli eğitim ve etiketsiz gerçek verilerin kullanıldığı yarı-gözetimli bir yaklaşımla geliştirilmiştir. Ayrıca, yerel ve küresel yapı bilgileri, modelin öğrenme sürecine dahil edilerek bölütleme doğruluğu artırılmıştır. Yerel ve küresel yapı bilgilerini içeren bu yöntem, düşük çözünürlüklü görüntülerde bölütleme performansını artırmıştır.

Birinci şahıs görüşlerinde ellerin çok doğru maskelerini çıkarmak için bir yöntem sunulan bu çalışmada (Vodopivec vd., 2016), mevcut derin Öğrenme yöntemlerinin aksine, giriş görüntüsünün düşük boyutlu bir temsiline uygulanan ölçekleme katmanları kullanmayan yeni bir derin öğrenme mimarisine dayanmaktadır. Bunun yerine, özellikler evrişimli katmanlarla çıkarılır ve tam bağlantılı bir katmanla doğrudan bir bölütleme maskesine eşlenir. Ağ birinci bölümde düşük çözünürlüklü bir görüntü ile kaba bölütleme yaparken, ikinci bölümde orijinal görüntüyü ve ilk çıktıyı birleştirerek tam çözünürlüklü sonuçlar sağlar. Ayrıca Havuzlama katmanlarının kullanılmaması, daha ince detayların öğrenilmesine olanak tanır. Bu yaklaşımın çok ölçekli bir şekilde uygulandığında hem doğru hem de gerçek zamanlı olarak yeterince verimli olduğu gösterilmiştir. Bu, dış mekanlardan ofislere kadar çeşitli ortamlarda çekilen görüntülerden oluşan yeni bir veri kümesi üzerinde gösterilmektedir.

(Urooj vd., 2018) gerçek ortam koşullarında el bölütleme problemini çözmek için RefineNet modeli kullanılmıştır. Çalışmada hem mevcut veri setleri (EgoHands, GTEA) hem de yeni oluşturulan veri setleri (EgoYouTubeHands, HandOverFace, EgoHands+) kullanılmıştır. Bu veri setleri, günlük yaşam aktivitelerinde farklı el tiplerini ve koşullarını temsil eden görüntülerden oluşmaktadır. bölütleme doğruluğunu artırmak için Koşullu Rastgele Alanlar (CRF) kullanılmıştır. RefineNet modeli, bölütleme başarısını

mevcut yöntemlere göre önemli ölçüde artırmıştır. Çalışmanın sonuçları ile ilgili birkaç görüntü Şekil 3'de görülmektedir.



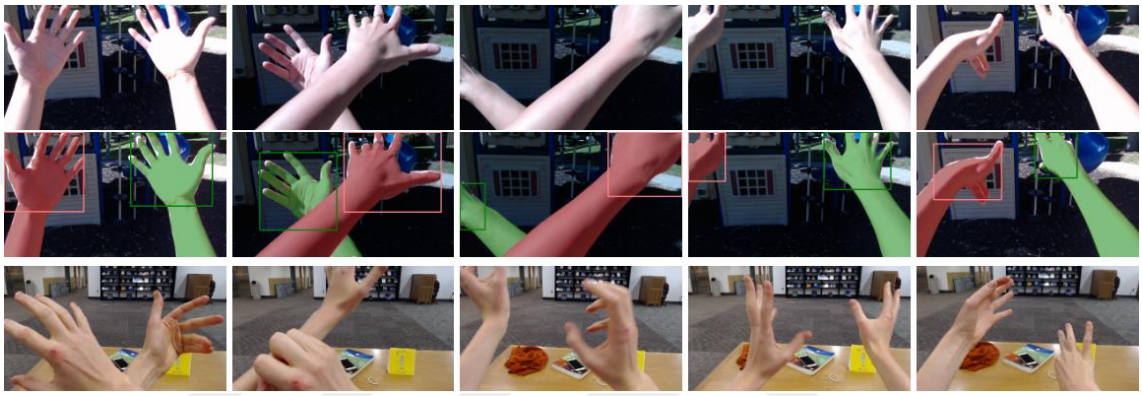
Şekil 3. RefineNet modeli kullanılarak yapılan çalışmanın görsel sonuçları

(Bojja vd., 2019) tarafından derinlik görüntülerinden el bölütlemesi için otomatik olarak etiketlenmiş bir veri kümesi olan HandSeg'i tanıtılmaktadır. Çalışmada, RGBD kameralar ve renkli eldivenler kullanılarak el hareketleri kaydedilmiş ve bu verilerle piksel düzeyinde yüksek kaliteli etiketler oluşturulmuştur. Verilerin etiketlenmesi, HSV eşikleme, GrabCut ve doğrusal SVM gibi üç aşamalı bir otomatik bölütleme yöntemiyle gerçekleştirilmiştir. Bu yöntem, manuel müdahaleyi en aza indirerek büyük ölçekli bir veri kümesinin kolayca oluşturulmasını sağlamıştır. Önerilen veri kümesi, toplam 210.000 çerçeve ve 13 katılımcıyı içerirken, sol ve sağ el ayrımını yapan mevcut ilk büyük ölçekli veri kümesi olarak öne çıkmaktadır.

(Cai vd., 2020) tarafından birinci şahıs bakış açısından videolarda el bölütlemede yeni ve görülmemiş alanlarda daha etkili uygulanabilmesi için belirsizlik tabanlı bir model adaptasyonu yöntemi sunulmuştur. Bu yöntem ile hedef alanda bölütleme etiketlerine ihtiyaç duyulmaz. Bunun için, bir Bayesian CNN kullanılarak modelin belirsizlik tahminleri yapılır ve güvenilir tahminler sahte etiketler olarak kullanılarak model iteratif bir şekilde adapte edilir. Ayrıca, farklı ortam koşullarında el şeklinin genel olarak tutarlı olması avantajından yararlanılarak el şekil önceliği kullanılmıştır. Bu iki önemli faktör, modelin belirsizlik tabanlı kendi kendine eğitim sürecinde yönlendirme sağlar. Yöntem, RefineNet tabanlı standart evrişimsel sinir ağlarına kıyasla daha iyi performans sergilemiştir. Sonuçlara göre, önerilen yöntem tüm veri setlerinde el bölütleme için daha yüksek doğruluk sağlamıştır.

(Lin vd., 2020) tarafından Ego2Hands adında gerçek zamanlı el bölütleme ve tespiti için geniş çaplı bir RGB tabanlı veri seti sunulmuştur. Mevcut veri setlerinin sınırlı

çeşitlilik ve doğruluk sağlaması nedeniyle bu çalışma, otomatik etiketleme ve kompozisyon tabanlı veri artırma yöntemlerini içeren yenilikçi bir yaklaşım geliştirmiştir. Ego2Hands, 188,000'den fazla sağ el görüntüsü içermekte ve rastgele seçilen sağ ellerin yatay çevrilmesiyle sol el verisi oluşturulmaktadır. Eğitimde gri tonlama görüntüler ve kenar haritaları kullanılarak renk değişkenliğine karşı dayanıklılık sağlanmıştır. Ayrıca, eğitim sırasında veri çeşitliliğini artırmak için farklı aydınlatma ve gölge koşullarında çekimler yapılmış ve çeşitli arka planlarla birleştirilmiştir. Çalışmanın sonuçlarına ait örnek görüntüler Şekil 4'de görülmektedir.



Şekil 4. Ego2Hand dataset kullanılarak elde elilen sonuçlar

Uzayda insan ve robot arasındaki etkileşimde kritik bir rol oynayan el algılama ve konumlandırma konusuna odaklanan bu çalışma (Gao vd., 2020), Feature-Map-Fused Single Shot Detector (FF-SSD) adı verilen yeni bir sinir ağı modeli geliştirilmiştir. FF-SSD, küçük ellerin ve karmaşık el hareketlerinin hassas bir şekilde algılanmasını sağlamak için tasarlanmıştır. Bu model, derin ve yüzeysel katmanlardaki bilgileri birleştirerek küçük nesnelerin tespitinde yüksek doğruluk elde ederken, aynı zamanda gerçek zamanlı uygulamalar için gerekli olan hızı da korur. Çalışmada, Egohands ve Oxford Hand gibi yaygın olarak kullanılan veri setlerinin yanı sıra, özellikle bu araştırma için oluşturulmuş olan SRSSL veri seti kullanılmıştır. FF-SSD modeli, farklı el boyutlarını ve çeşitli el hareketlerini etkili bir şekilde algılayabilmesi için bu veri setleri üzerinde eğitilmiştir.

Hijyenik ve kullanıcı dostu bir biyometrik tanıma yöntemi olan temassız parmak izi tanıma sistemleri için el bölütleme üzerine odaklanılan çalışmada, derin öğrenme tabanlı bir semantik segmentasyon modeli olan DeepLabv3+ kullanılmış ve elin arka plandan hassas bir şekilde ayrıştırılması hedeflenmiştir (Priesnitz vd., 2021). DeepLabv3+ modeli, Hand Gesture Recognition (HGR) veri tabanı kullanılarak eğitilmiş ve performansı, yaygın olarak kullanılan bir renk tabanlı yöntemle karşılaştırılmıştır.

Sonuçlar, DeepLabv3+ modelinin hem homojen hem de karmaşık arka planlarda el ve arka plan sınırlarını daha doğru ve tutarlı bir şekilde belirlediğini göstermiştir.

MASK-RCNN ile Çekirge Optimizasyonunu birleştirerek 3 boyutlu el hareketleri bölütleme ve sınıflandırması için bir yöntem sunulmaktadır (Khan vd., 2021). Intel Kinetic ve Intel Real sense d435i kamera kullanılarak özel bir 3 boyutlu ve RGB el hareketleri veri seti oluşturulmuş, ardından insan kinematiğini kullanarak RGB el hareketleri için temel noktaları tahmin etmek için bir model önerilmiş, daha sonra temel noktalar en iyi serbestlik derecesini (DoF) elde etmek için kullanılmıştır. Minimum mesafe fonksiyonunun yanı sıra çekirge optimizasyonu, 3 boyutlu el hareketleri veri setinden en iyi derin özellikleri elde etmek için uygulanmıştır. ResNet50 ağı, bölütleme için Örtüşme Katsayısını (OC) hesaplamak için omurga olarak kullanılır ve 3 boyutlu el hareketleri için sınıflandırmayı hesaplamak için ResNet50, ResNet101 ağları kullanılır.

EgoGAN, benmerkezci videolarda gelecekteki el bölütleme tahmin etmek için geliştirilmiş bir derin öğrenme yöntemidir (Jia vd., 2022). Bu model, 3 boyutlu tam evrişimli ağlar (3DFCN) ve Generative Adversarial Network (GAN) birleşimini kullanır. 3DFCN, videodan mekansal-zamansal (spatio-temporal) özellikleri çıkarırken, GAN gelecekteki hareketlerini modelleyerek el maskesi tahminini destekler. GAN'ın ana katkısı, gelecekteki hareketlerin olası dağılımlarını tahmin ederek el bölütleme için daha kesin ve tutarlı bilgiler sunmasıdır. Model, EPIC-Kitchens ve EGTEA Gaze+ veri setlerinde test edilmiş ve geçmiş yöntemlerle kıyaslandığında belirgin iyileşmeler göstermiştir. Modelin görsel sonuçları Şekil 5'de görülmektedir.

Şekil 5. EgoGAN modelinin görsel sonuçları

Derin öğrenme, biyometrik sistemler için el bölütleme çalışmalarında önemli bir yer edinmiş ve farklı modeller kullanılarak yüksek performans elde edilmeye çalışılmıştır. Özellikle Freihand gibi dinamik veri setlerinde, farklı omurga mimarilerine dayalı gelişmiş bölütleme modellerinin etkisi henüz tam olarak araştırılmamıştır. Bu çalışmada autoencoder tabanlı, U-Net, FPN, PSP-Net ve MA-Net gibi modern bölütleme modelleri kullanılarak el bölütleme performansı değerlendirilecektir. Omurga mimarileri olarak Vision Transformer, EfficientNet-B5 ve ResNet-34 kullanılacaktır. Önceden eğitilmiş modeller, transfer öğrenme yöntemiyle kullanılarak eğitim süresi kısaltılacak ve sınırlı veri setinde yüksek doğruluk oranları hedeflenmektedir. Bu çalışma, Freihand veri seti üzerindeki el bölütleme performansını iyileştirerek el bölütleme alanına katkı sağlamayı amaçlamaktadır.

3. MATERYAL VE YÖNTEM

3.1. Bölütleme

Bölütleme, bir metin, veri ya da görüntü gibi bir bütünü, anlamlı ve işlevsel alt birimlere ayırma sürecidir (Guo vd., 2018). Bu işlem, verilerin daha etkili bir şekilde analiz edilmesini ve kullanılmasını sağlar. Örneğin, metin bölütleme, bir metni cümlelere veya kelimelere ayırırken, görüntü bölütleme, bir görüntüyü nesneler veya bölgeler şeklinde ayırmayı hedefler. Bu teknik, doğal dil işleme, görüntü işleme ve veri madenciliği gibi birçok alanda temel bir rol oynar. Bölütleme türleri arasında örnek bölütleme ve anlamsal bölütleme ön plana çıkar. Örnek bölütleme, bir verideki her bir örneği veya nesneyi birbirinden ayırt etmeye odaklanır. Örneğin, bir görüntüde aynı sınıfa ait nesneler birbirinden ayrı olarak algılanır ve işlenir. Buna karşın, anlamsal bölütleme, tüm veri kümesi içinde aynı sınıfa ait olan tüm öğeleri bir bütün olarak ele alır. Örneğin, bir görüntüdeki tüm ağaçlar veya yollar, aynı kategoriye ait tek bir "anlamlı bölge" olarak işlenir. Her iki yöntemin de uygulama alanları farklıdır; örnek bölütleme genelde otonom araçlar ve robotik gibi nesne düzeyinde işlem gerektiren sistemlerde tercih edilirken, anlamsal bölütleme tıbbi görüntü analizi ve uydu görüntü işleme gibi alanlarda kullanılır (Garcia-Garcia vd., 2017). Diğer bazı derin öğrenme, Mask R-CNN (He vd., 2017) ve YOLO (Redmon vd., 2016) gibi yöntemler, nesne tespiti ve bölütleme işlemlerini etkili bir şekilde gerçekleştirebilen güçlü algoritmalardır. Özellikle YOLO, yüksek hızda nesne tespiti için optimize edilmiştir, ancak bölütleme gibi daha karmaşık işlemleri doğrudan çözmek yerine, bu yetenekler ek genişletmelerle sağlanabilir. Bu tür algoritmalar genellikle genel amaçlıdır ve birden fazla nesne türü için tasarlanmıştır. Ancak, bu geniş kapsamlı tasarım, yüksek hesaplama ve eğitim maliyetlerini de beraberinde getirmektedir. Bizim çalışmamızda, yalnızca bir el nesnesinin bölütlenmesine odaklanıldığından, bu tür genel ve karmaşık algoritmalar gereksiz derecede maliyetli ve hesaplama açısından ağır kalmaktadır. Daha spesifik ve optimize edilmiş yaklaşımların kullanılması, hem kaynakların daha verimli bir şekilde kullanılması hem de hedeflenen doğruluğa ulaşılması açısından daha uygun bir seçenek olacaktır. Bu sebeple çalışmada literatürdeki birçok çalışmada kullanılan modellerin ve omurga mimarilerinin özellikleri dikkate alınarak model ve omurga seçimi yapılmıştır. Otokodlaycıların özellik çıkarma yeteneği, kolayca derin öğrenme modellerinin yapısına entegre olması, basit ve sade yapısı nedeniyle çalışmamızda otokodlayıcı temelli modeller tercih edilmiştir. U-Net mimarisin seçiminde bu modelin otokodlayıcı yapısı dolayısı ile çok iyi özellik çıkarımı,

modelin ince detayları yakalama yeteneği, az kaynak harcaması, az verilerde bile iyi performans göstermesi,basit ve sade yapısı modeli tercih sebeplerimizdendir. MA-Net'in çok ölçekli özellik çıkarımı ve sahip olduğu dikkat mekanizması bu modeli daha yüksek bölütleme kapasitesi sunmasından dolayı tercih edilmiştir. PSP-Net modelinin tercih sebebi, sahip olduğu piramit modülü sayesinde farkli havuzlama katmanları kullanarak özellik çıkarabilme yeteneği modeli diğer modellerden ayırmaktadır. FPN modelinin diğer modellerden farklı olarak görüntüleri farklı ölçeklerde işleme kapasitesi sayesinde daha fazla özellik çıkarma kapasitesi tercih nedeni olmuştur. Omurga seçiminde ise ResNet-34 mimarisinin çok katmanlı derin yapısı ve gradyan kaybolması gibi sorunları önleme yeteneği sayesinde diğer omurgalardan ayrılmaktadır. EfficientNet-B5 omurgası literatürde de bahsedilen dengeli bir yapıda olması, eğitime göre modelin katman ve derinliğinin kolay bir şekilde değiştirilebilmesi, eğitim süresinin kısa olması onun diğer omurgalardan ayrılmasını sağlamaktadır. Vision transformer omurgası, günümüzde doğal dil işleme modellerinin başarısındaki ana etken olan dikkat mekanizmasını etkili bir şekilde kullanması tercih edilmesini sağlamıştır. Ayrıca modellerin eğitiminde trasnfer öğrenme kullanılarak hem daha hızlı öğrenme ve daha az kaynak kullanarak modelimizin kullandığı omurga sayesinde daha fazla öznitelik çıkarması ve dolayısı ile daha iyi performans, bizim çalışmamızda daha fazla bölütleme başarısı göstermesi amaçlanmıştır.

3.1.1. Oto Kodlayıcı

Oto kodlayıcı, denetimsiz öğrenme yöntemlerinden biri olarak sinir ağları aracılığıyla veri temsili oluşturmak için kullanılan bir modeldir. Amacı, veri boyutunu düşürerek, verinin önemli özelliklerini koruyup yeniden yapılandırmaktır. Veri ön işleme, boyut indirgeme, gürültü temizleme ve özelliklerin öğrenimi gibi alanlarda yaygın olarak kullanılır (Goodfellow vd., 2016).

Oto kodlayıcının, iki bileşeni vardır: kodlayıcı ve çözücü, kodlayıcının görevi, giriş verisini sıkıştırıp daha küçük boyutlu bir temsile dönüştürmektir. Bu aşamada, orijinal verinin sadece temel nitelikleri korunur. Çözücü ise, bu sıkıştırılan temsili kullanarak, veriyi tekrar oluşturmaya çalışır (Kingma ve Welling, 2019). Böylece, kodlayıcıda sıkıştırılan verinin, özelliklerinin kaybolmaması sağlanır.

Verinin türüne ve uygulama alanına göre çeşitli oto kodlayıcı türleri vardır:

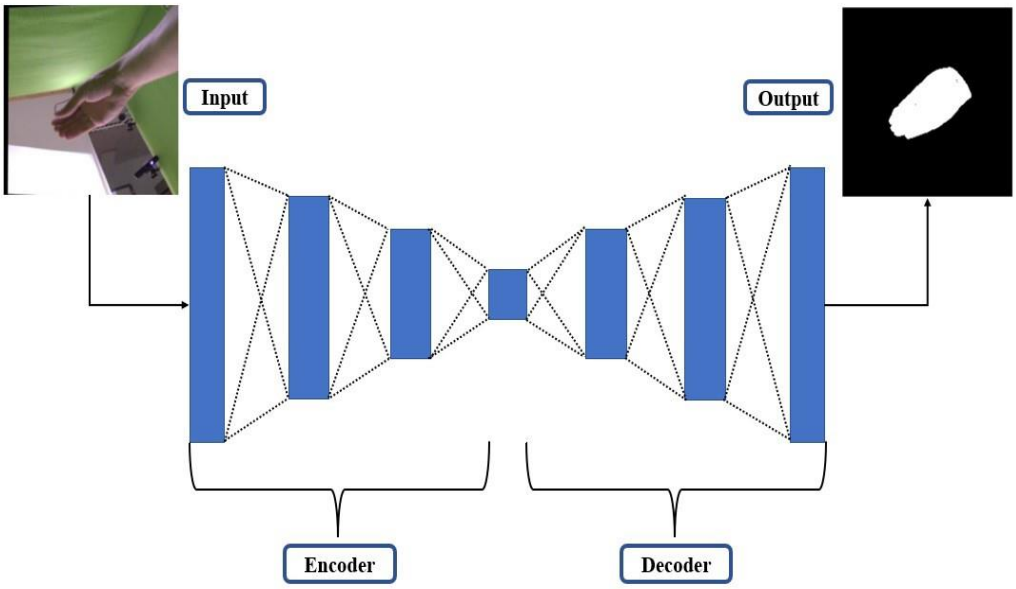
- Vanilla Oto kodlayıcı**:** En basit türdeki Oto kodlayıcı dır. Temel birkaç gizli katman dan oluşur. Temel veri sıkıştırma ve yeniden yapılandırma görevini yapar (Goodfellow vd., 2016).

- Sparse Oto kodlayıcı**:** Bu türde, gizli katmanlarda seyrek bir yapı bulunur. Böylece sadece az sayıdaki nöronun aktif olması ile verideki temel özellikleri öğrenilir. Bu model, en çok özellik çıkarımında etkilidir (Goodfellow vd., 2016).
- Denoising Oto kodlayıcı**:** Denoising Oto kodlayıcı modeli, giriş verisine gürültü ekleyerek modelin sağlamlığını arttırmayı amaçlar. Gürültülü veriyi orjinal haline en yakın şekilde yeniden yapılandırarak, modelin gürültüye karşı daha dirençli olmasını sağlar (Vincent vd., 2008).
- Variational Oto kodlayıcı**:** Variational Oto kodlayıcı (VAE), verilerin olasılıksal bir temsilini öğrenerek yeni veri örnekleri üretebilir. VAE, özellikle generatif modelleme için uygun bir yapıdır. Genelde görüntü veya ses gibi verilerin üretiminde tercih edilir (Doersch, 2016; Kingma ve Welling, 2019).
- Convolutional Oto kodlayıcı**:** Özellikle görüntü verisi üzerinde çalışan bu model, evrişimsel katmanlarının kullanımıyla verinin mekansal özelliklerini korur. Görüntü sıkıştırma ve gürültü giderme gibi görevlerde oldukça başarılıdır (Goodfellow vd., 2016).

Oto kodlayıcı modelleri, çeşitli alanlarda kullanılırlar:

- Boyut İndirgeme**:** Büyük veri setlerinin daha küçük bir boyuta indirgenmesi ve geleneksel boyut azaltma yöntemlerine kıyasla daha karmaşık veri temsillerini daha esnek biçimde sağlar (Goodfellow vd., 2016).
- Gürültü Giderme: Denoising Oto kodlayıcı, özellikle görüntü ve sinyal işleme gibi alanlarda gürültü temizleme işlevini başarılı bir şekilde yerine getirir (Vincent vd., 2008).
- Anomali Tespiti: Oto kodlayıcı, normal veri örüntülerini öğrenerek anormal durumları tespit etmek için kullanılır. Bu yaklaşım, siber güvenlik ve sahtekarlık tespiti gibi alanlarda yaygındır (Zong vd., 2018).
- Generatif Modelleme: VAE ve GAN gibi modellerle birlikte kullanılan Oto kodlayıcılar, yeni örneklerin üretiminde sıklıkla kullanılır (Kingma ve Welling, 2019; Doersch, 2016).
- Veri Görselleştirme: Yüksek boyutlu verilerin daha düşük boyutlu temsillerini oluşturarak, verilerin daha anlaşılır ve görsel olarak daha kullanılabilir hale gelmesini sağlar (Goodfellow vd., 2016).

Oto kodlayıcı modelleri, sinir ağları ve denetimsiz öğrenmenin güçlü bir birleşimidir ve veri işleme, sıkıştırma ve özellik çıkarımı gibi görevlerde oldukça etkilidir. Bu modeller, verilerin temel yapısını öğrenerek, verinin daha anlaşılır hale gelmesini ve ileriye yönelik geniş bir uygulama yelpazesine sahip olmayı sürdürmektedir.



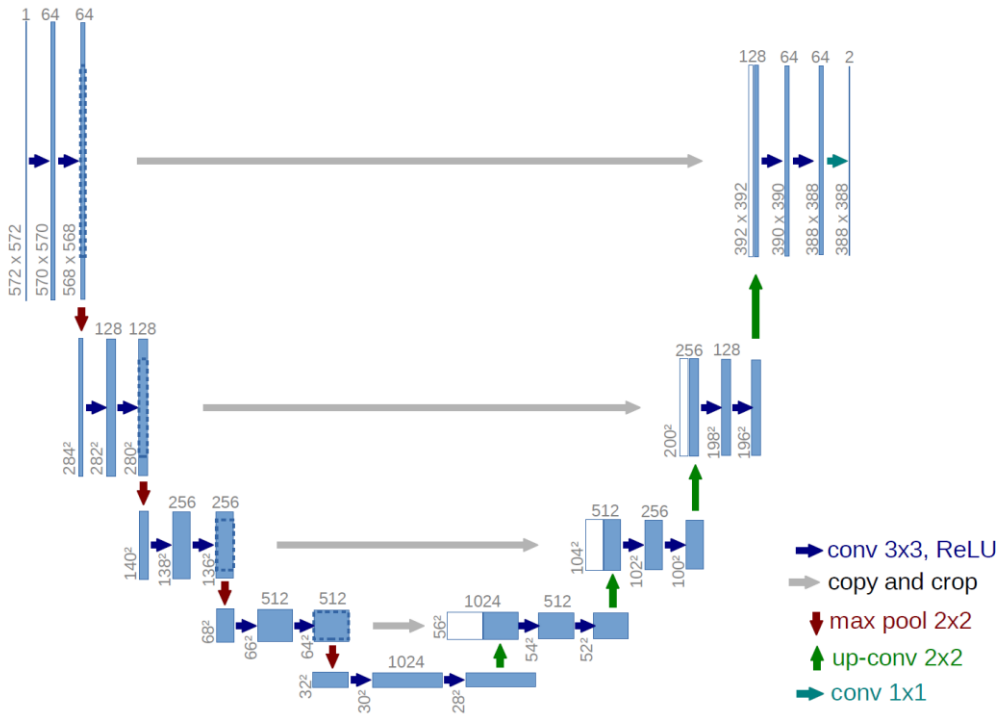
Şekil 6. El bölütleme için temel oto kodlayıcı yapısı

3.1.2. U-Net

U-Net, özellikle tıbbi görüntü analizinde önemli bir derin öğrenme modelidir (Ronneberger vd., 2015). Görüntü bölütleme alanında uzmanlaşmış olan U-Net, az sayıda örnekten bile etkili bir şekilde öğrenebilir. Bu özelliği, tıbbi görüntülerde etiketleme işleminin zor ve zaman alıcı olması sebebiyle büyük önem taşır. U-Net, temel olarak bir kodlayıcı ve bir çözücüden oluşan simetrik bir yapıya sahiptir. Kodlayıcı, girdi görüntüsünü işleyerek önemli özelliklerini çıkarır ve boyutunu kademeli olarak küçültür. Bu sırada, görüntüdeki detaylar daha soyut bir temsile dönüştürülür. Çözücü ise kodlayıcının çıkardığı bu özeti alır ve tekrar orijinal görüntü boyutuna genişletirken, aynı zamanda piksel bazında sınıflandırma yaparak bölütleme işlemini gerçekleştirir. U-Net'in en önemli özelliklerinden biri, kodlayıcı ve çözücü arasındaki "atlamalı bağlantılar"dır (Drozdzal vd., 2016). Bu bağlantılar, kodlayıcıda öğrenilen detayların çözücüye aktarılmasını sağlayarak bilgi kaybını en aza indirir. Böylece, hem genel yapısal bilgiler hem de ince detaylar korunarak daha hassas bölütleme sonuçları elde edilir. Bu bağlantılar ayrıca gradyan kaybolmasını engelleyerek özellikle derin ağlarda gradyan kaybolmasını, erken katmanlarda daha etkili bir şekilde geri yayılır (He vd., 2016). Ağın parametkerini daha verimli bir şekilde optimize etmeye yardımcı olur. Bu özellikle derin modellerin öğrenme hızını artırır ve modelin daha iyi minumuna oluşmasını sağlar (Sribastava vd., 2015).

U-Net, başlangıçta biyomedikal görüntü analizi için geliştirilmiş olsa başarısı diğer alanlarada yayılmıştır (Çiçek vd., 2016). Uydu görüntülerinin analizi (Garcia-Garcia vd.,

2017), tarımsal alanların sınıflandırılması (Lu vd., 2017) ve otonom sürüş sistemleri (Neven vd., 2018) gibi farklı görevlerde de U-Net'in etkili olduğu görülmüştür. U-Net'in başarısının ardında yatan temel faktörler, etkili mimarisi, atlamalı bağlantılar ve az sayıda veriyle öğrenme kapasitesidir. Bu özellikler, U-Net'i görüntü bölütleme alanında güçlü bir araç haline getirmiştir. Modelin temek mimarisi Şekil 3'de görünmektedir.



Şekil 7. Temel U-Net mimarisi (Ronneberger vd., 2015).

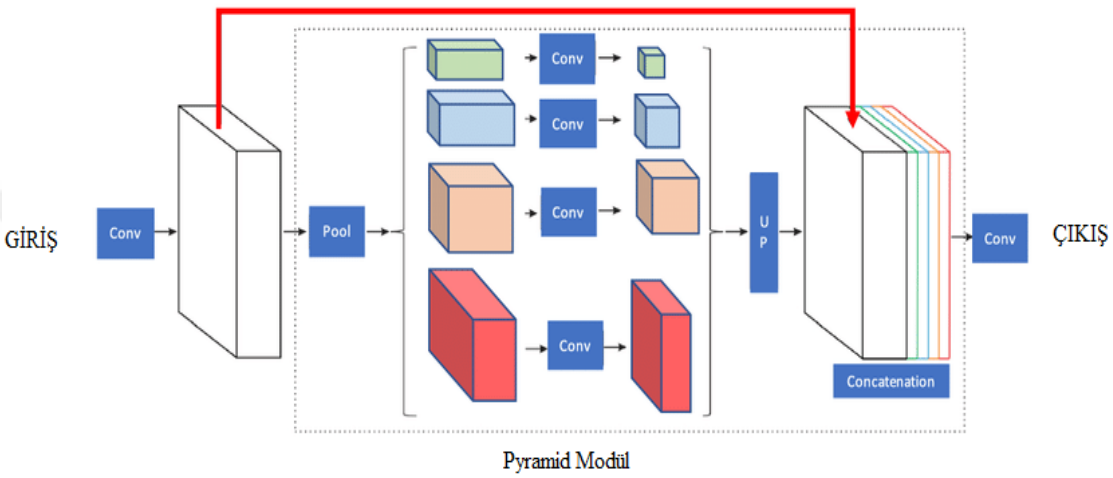
3.1.3. PSP-Net

Zhao ve arkadaşları tarafından 2017 yılında Microsoft Research Asia'da geliştirilen PSP-Net, "piramit havuzlama" adı verilen yeni bir modül kullanarak görüntülerdeki çok ölçekli bağlamsal bilgileri etkili bir şekilde yakalamışlardır (Zhao vd., 2017). Bu modül, farklı boyutlardaki bölgelerden gelen bilgileri birleştirerek hem küçük hem de büyük nesnelerin daha doğru bir şekilde sınıflandırılmasını sağlar. PSP-Net'in temel gücü, piramit havuzlama modülü aracılığıyla farklı ölçeklerdeki özellikleri tek bir bağlamsal özellik haritasında bir araya getirmesinden kaynaklanır. Bu yaklaşım, modelin hem yerel hem de global bağlamsal bilgileri kullanarak daha doğru bölütleme sonuçları üretmesini sağlar. Zhao vd., (2017), PSP-Net'in ADE20K ve Cityscapes gibi zorlu veri setlerinde elde ettiği yüksek doğruluk oranlarıyla modelin etkinliğini kanıtlamıştır.

PSP-Net'in mimari yapısı, üç aşamadan oluşur:

- Özellik Çıkarma: İlk aşamada, ResNet gibi bir evrişimsel sinir ağı kullanılarak girdi görüntüsünden özellikler çıkarılır.

- Piramit Havuzlama: Bu aşamada, özellik haritası farklı boyutlardaki (örneğin, 1x1, 2x2, 3x3 ve 6x6) havuzlama katmanlarından geçirilir. Her havuzlama katmanı, farklı ölçeklerde bağlamsal bilgi yakalar ve sonra bu bilgiler birleştirilerek kapsamlı bir özellik temsili oluşturulur.
- Nihai Özellik Haritasının Üretilmesi: Son aşamada, piramit havuzlama modülünden elde edilen zenginleştirilmiş özellikler, piksel düzeyinde sınıflandırma için kullanılarak son bölütleme haritası üretilir.



Şekil 8. PSP-Net mimarisi (Zhao vd., 2017).

3.1.4. FPN

Bilgisayarla görü alanında nesne algılama ve bölütleme gibi görevler, görüntülerdeki nesnelerin hem konumlarını hem de boyutlarını doğru bir şekilde belirlemeyi gerektirir. Bu tür görevlerde karşılaşılan temel zorluklardan biri, görüntülerde farklı ölçeklerde nesnelerin bulunabilmesidir. Feature Pyramid Network (FPN), bu soruna etkili bir çözüm sunan ve nesne algılama performansını önemli ölçüde artıran bir sinir ağı mimarisidir (Lin vd., 2017). FPN, görüntünün farklı çözünürlüklerdeki bilgi katmanlarını, yani çok ölçekli özellik haritalarını, bir araya getirerek hem küçük hem de büyük nesneleri aynı anda algılayabilme yeteneği kazanır. Bu, özellikle küçük nesnelerin algılanmasında büyük bir avantaj sağlar, çünkü düşük çözünürlüklü katmanlardaki genel bilgiler, yüksek çözünürlüklü katmanlardaki detaylarla birleştirilerek daha güçlü bir nesne temsili elde edilir (Lin vd., 2017).

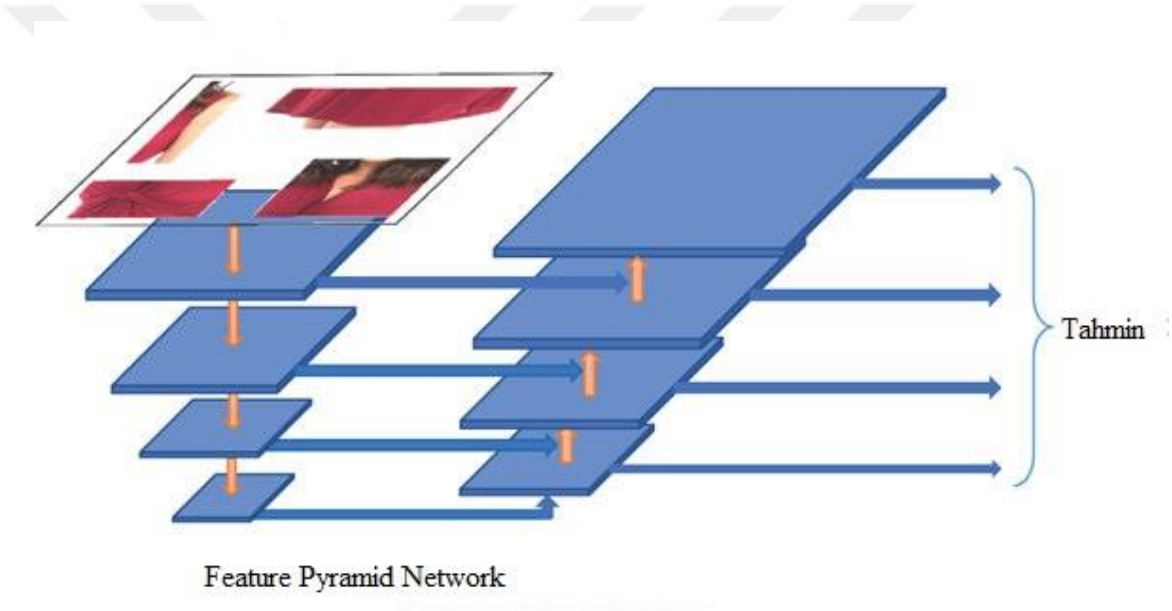
FPN, iki yönlü bir ağ yapısı kullanır:

- Bottom-up Pathway: Bu yol, geleneksel bir evrişimsel sinir ağı mimarisindeki gibi, girdi görüntüsünden hiyerarşik olarak özellikler çıkarır. Her katmanda,

daha yüksek seviyeli ve daha soyut özellikler elde edilirken, uzamsal çözünürlük azalır.

- Top-down Pathway: Bu yol, daha yüksek çözünürlüklü özelliklerin, daha düşük çözünürlüklü katmanlardaki bilgilerle birleştirilmesini sağlar. Bu, "skip connection" adı verilen bağlantılar aracılığıyla gerçekleştirilir. Skip connectionlar, farklı katmanlardaki bilgilerin doğrudan aktarılmasını sağlayarak, özelliklerin daha etkin bir şekilde entegre edilmesine olanak tanır (He vd., 2017).

FPN'in bu iki yönlü yapısı, farklı ölçeklerdeki özellik haritalarının aynı anda işlenmesini ve böylece hem büyük hem de küçük nesnelerin eşzamanlı olarak algılanmasını mümkün kılar (Lin vd., 2017).



Şekil 9. FPN mimarisi (Lin vd., 2017).

3.1.5. MA-NET

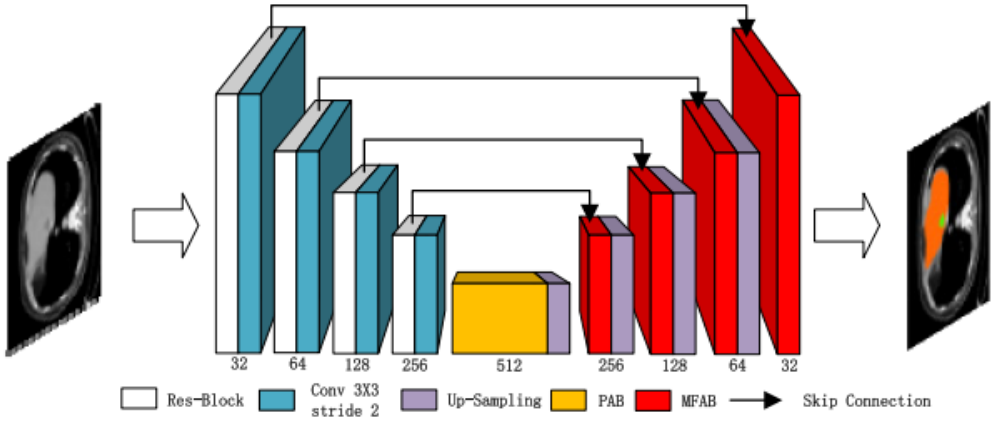
Farklı ölçeklerdeki bilgileri etkili bir şekilde kullanabilen modeller, özellikle görüntü bölütleme, süper çözünürlük ve sınıflandırma gibi karmaşık görevlerde öne çıkmaktadır. Çok Ölçekli Dikkat Ağı (MA-Net), bu tür görevlerde yüksek performans gösteren ve dikkat mekanizmalarını kullanarak çok ölçekli özellikleri birleştiren bir mimaridir. MA-Net'in başarısının altında yatan temel prensip, görüntüdeki yerel detayları küresel bağlamla bütünleştirerek daha zengin bir anlayış elde etmektir. Bu sayede, model karmaşık desenleri daha iyi tanıyabilir ve daha doğru tahminler yapabilir. MA-Net, tıbbi görüntü analizi, uzaktan algılama ve robotik gibi çeşitli alanlarda büyük potansiyel

sunmaktadır.MA-Net'i diğer derin öğrenme modellerinden ayıran bazı önemli özellikler şunlardır:

- Çok Çözünürlüklü Bilgi İşleme: MA-Net, farklı çözünürlüklerdeki özellik haritalarını birleştirerek görüntüden daha kapsamlı ve ayrıntılı bilgi çıkarır. Bu, hem küçük hem de büyük ölçekli nesnelerin etkili bir şekilde algılanmasını sağlar (Fan vd., 2020).
- Bağlamsal İlişkileri Öğrenme: Self-attention mekanizması sayesinde, MA-Net görüntünün farklı bölgeleri arasındaki bağımlılıkları öğrenir ve bu bilgileri kullanarak daha anlamlı özellik haritaları oluşturur. Bu, dikkat mekanizmasının gücünü çok ölçekli bağlamlarda kullanmayı sağlar (Wang vd., 2024).
- Esnek ve Genişletilebilir Tasarım: MA-Net, modüler bir yapıya sahiptir, bu da modelin kolayca genişletilmesini ve diğer ağlarla entegre edilmesini sağlar. Bu esneklik, MA-Net'in farklı uygulama senaryolarına uyarlanabilmesini kolaylaştırır (Piao vd., 2019).

MA-Net, çeşitli görüntü işleme görevlerinde başarıyla kullanılmıştır:

- Tıbbi Görüntülerde Hassas Bölütleme: Karaciğer ve tümör bölütleme gibi tıbbi görüntü analizinde kritik öneme sahip görevlerde, MA-Net yüksek doğruluk oranları elde etmiştir (Fan vd., 2020).
- Görüntü Kalitesini Artırma: Tek görüntü süper çözünürlük uygulamalarında, MA-Net farklı çözünürlüklerdeki bilgileri birleştirerek görüntü kalitesini önemli ölçüde artırabilir (Wang vd., 2024).
- Görüntü Onarma: MA-Net'in çok ölçekli özellik çıkarımı, eksik veya bozuk görüntü bölgelerini doldurma (inpainting) işleminde de etkili bir şekilde kullanılabilir (Bai vd., 2021).
- Hastalık Teşhisi: Retina görüntülerindeki hasarı tespit ederek diyabetik retinopati gibi hastalıkların teşhisinde MA-Net başarılı bir şekilde uygulanmıştır (Al-Antary ve Arafa, 2021).



Şekil 10. Örnek bir MA-Net mimarisi (Fan vd., 2020).

3.2. Transfer Öğrenme

Transfer öğrenme, özellikle sınırlı veri veya sınırlı hesaplama kaynaklarına sahip olduğumuz durumlarda, bir alan veya görevde öğrenilmiş bilgilerin başka bir alan veya göreve aktarılmasına dayanan bir makine öğrenimi ve derin öğrenme tekniğidir. Bu teknik, genellikle derin öğrenme modellerinin eğitimi için büyük veri setlerinin gerektiği durumlarda, oto kodlayıcı temelli modellerde, önceden eğitilmiş modellerin encoder kısmının, yani girdi verilerini işleyip öznitelik çıkaran ve omurga olarak da adlandırılan kısmının, yeniden kullanımı ile yeni bir görev için başarılı sonuçlar elde etmek amacıyla tercih edilmektedir (Yosinski vd., 2014). Transfer öğrenme, modelin önceden bir görev için eğitildiği bilgiyi, omurga aracılığıyla, başka bir göreve uyarlamasını sağlar. Bu yaklaşım, özellikle doğal dil işleme ve bilgisayarla görme gibi alanlarda yaygın olarak kullanılır. Örneğin, bir model geniş bir görüntü veri seti üzerinde eğitildikten sonra, farklı ancak benzer bir görev olan tıbbi görüntü sınıflandırmasına transfer edilebilir. Böylece, modelin daha önce omurganın öğrendiği temel özellik çıkarma bilgileri, yeni görev için yeniden kullanılır ve bu sayede eğitim süresi önemli ölçüde azalırken doğruluk oranları artar (Yosinski vd., 2014).

Transfer öğrenme genelde şu adımları içerir:

- Önceden Eğitilmiş Modelin Seçilmesi**:** Transfer öğrenme için yaygın olarak kullanılan modeller genellikle büyük veri setlerinde (örneğin, ImageNet) eğitilmiş ResNet, VGG gibi derin öğrenme modelleridir. Bu modellerin omurga yapıları, farklı görevler için güçlü öznitelik çıkarıcılar olarak kanıtlanmıştır (Devlin vd., 2018).
- İnce Ayar (Fine-Tuning): Yeni görev için modelin bazı katmanlarının eğitimi yeniden yapılabilir. Bu aşamada, özellikle son katmanların eğitimi, yeni görevin

gerekliliklerine uygun olacak şekilde yapılırken, omurganın bazı katmanları da ince ayar için açılabilir (Tan vd., 2018).

- Özellik Çıkarımı**:** Modelin alt katmanlarında, yani omurgada öğrenilen özellikler, çoğu zaman genel geçer özellikler olduğu için, bu katmanlar sabit tutularak yeni veri setinde yalnızca üst katmanlar yeniden eğitilir (Tan vd., 2018).

Transfer öğrenmenin temel avantajı, çok büyük veri setleri ve hesaplama kaynakları gerektiren derin öğrenme modellerinin, bu gereksinimleri daha az karşılayan projelerde kullanılabilmesini sağlamaktır. Omurganın yeniden kullanımı, özellikle sınırlı veri olan alanlarda transfer öğrenmenin etkili bir seçenek olmasına yol açmıştır.

Transfer öğrenme, birçok farklı alanda etkili bir şekilde kullanılmaktadır:

- Doğal Dil İşleme (NLP): BERT, GPT gibi modeller, geniş dil veri setlerinde önceden eğitilmiş olup, farklı metin sınıflandırma veya dil modeli görevlerinde yeniden kullanılabilmektedir. Bu modellerin güçlü omurganın yapıları, metin verilerinden anlamlı öznitelikler çıkararak farklı NLP görevlerinde başarı sağlar (Devlin vd., 2018).
- Görüntü İşleme ve Bilgisayarla Görme: Görüntü sınıflandırma, nesne tespiti ve bölütleme gibi görevlerde ImageNet üzerinde eğitilmiş modellerin tekrar kullanımı yaygındır. Örneğin, bir otonom araç sisteminde yol işaretlerini tanımlamak için transfer öğrenme sıkça kullanılır. Bu modellerin omurgaları, genel görüntü özelliklerini etkili bir şekilde öğrenir (Huh vd., 2016).
- Tıbbi Görüntüleme: Özellikle sınırlı sayıda etiketli veri bulunan tıbbi görüntü sınıflandırma projelerinde transfer öğrenme, önceden eğitilmiş bir görüntü sınıflandırma modelini ince ayar yaparak daha hızlı ve başarılı sonuçlar sağlar. Omurganın genel görüntü özelliklerinden yararlanılması, tıbbi görüntülerdeki spesifik özelliklerin öğrenilmesini kolaylaştırır (Raghu vd., 2019).

Transfer öğrenme, büyük veri setlerine erişimin sınırlı olduğu veya yeni bir görevin daha önce eğitilmiş benzer bir modelle çözülebileceği durumlarda, öğrenme sürecini hızlandırarak yüksek başarı sağlayan etkili bir tekniktir. Bu yöntemin günümüzde geniş kullanım alanı bulması, onu derin öğrenme ve makine öğrenimi alanında önemli bir araç haline getirmiştir. U-Net özelinde örnek bir transfer öğrenme ve omurga kullanımı Şekil 7'de gösterilmiştir. Burda renkli çerçeve ile gösterilenler modellerin kodlayıcı kısımlarıdır. Kodlayıcı kısımları modellerin özellik çıkarımı yaptıkları yerdir. Transfer öğrenme için kullanacağımız omurgalar bu kodlayıcı kısıma eklenmektedir. Amaç milyonlarda resimden düşük seviyede özellik çıkarmak için eğitilen modelin ki biz burda ImageNet modeli üzerinden eğitilen modeller kullanıyoruz, kodlayıcı kısmında

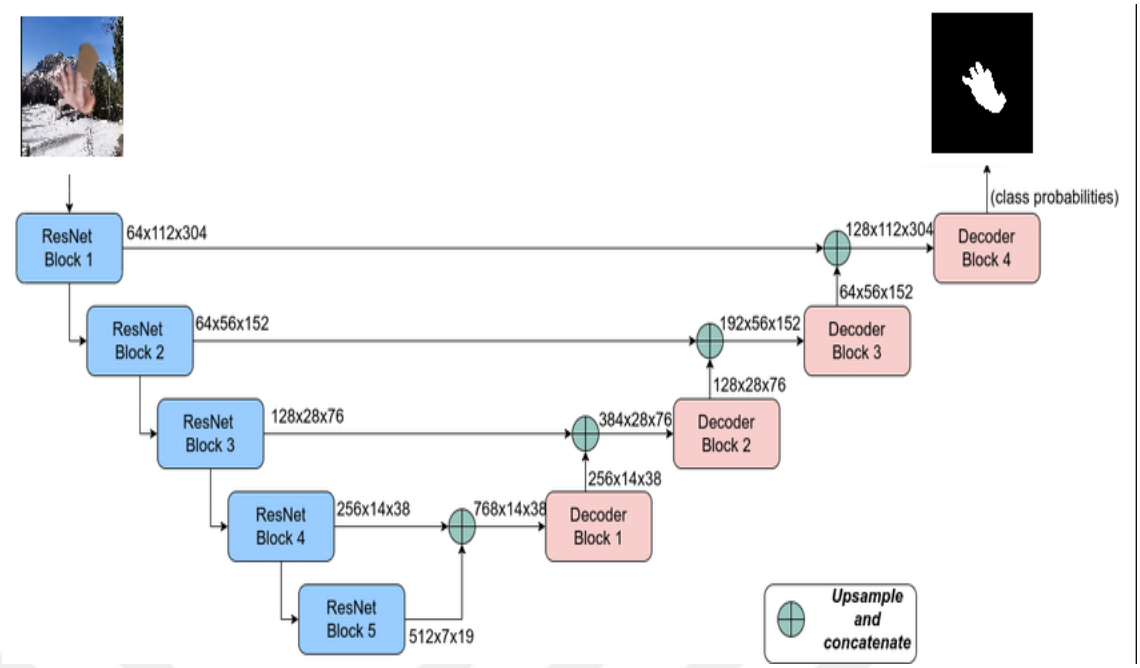
omurgalar düşük seviyedeki bilgiler; kenarlar, dokular gibi temel özellikleri daha yüksek bir seviyede çıkararak modelin başarısını artırmaktadırlar. Omurga eklemek için, omurga kısmının son katmanları olan sınıflandırma katmanları çıkarılarak modelin kodlayıcı kısmına eklenir. Daha sonra omurganın elde ettiği özellik haritaları latent katmanına (gizli temsil) sıkıştırılır. Decoder kısmı bu sıkıştırılan bu verileri alıp orijinal verinin yeniden oluşturulmasını sağlar.

3.2.1. ResNet-34

ResNet-34, bilgisayarla görme ve derin öğrenme projelerinde özellik çıkarımı için yaygın olarak kullanılan bir omurga modelidir. ResNet mimarisi, 2015 yılında He vd.,tarafından önerilmiş olup, derin sinir ağlarının eğitim sürecinde karşılaşılan gradyan kaybolması sorununa yönelik getirdiği yenilikçi çözüm ile öne çıkmıştır . Özellikle residual blok adı verilen bir yapı sayesinde, ağın daha derin katmanlarının daha verimli ve hızlı bir şekilde eğitilmesine olanak tanır. Model, 34 katmanlı bir yapıya sahip olup, güçlü bir omurga seçeneği olarak kullanılabillir (He vd., 2016).

ResNet-34, derin ağlarda gradyan kaybı problemine çözüm sunan atlama bağlantısı (skip connection) olarak bilinen bir mimari ile tasarlanmıştır. Bu bağlantılar, ağın daha derin katmanlarında girdi bilgilerini doğrudan atlayarak, bu katmanların yalnızca farkı öğrenmesini sağlar. Bu yapı sayesinde ResNet-34, klasik evrişimsel ağların aksine, daha derin katmanlarda bile gradyanın etkili bir şekilde iletilmesine olanak tanır. Böylece, eğitim süreci daha stabil hale gelir ve daha doğru sonuçlar elde edilir (He vd., 2016).

ResNet-34, her biri 3x3 boyutunda filtreler içeren toplamda 34 katmanlı bir mimariden oluşur. Bu katmanlar dört ana blokta gruplanmıştır ve her bir blokta farklı sayıda residual bloğu bulunmaktadır. Özellikle 34 katmanlı yapısı, temel özellikleri başarılı bir şekilde çıkarmaya yönelik olarak orta derinlikte bir yapı sunar ve görüntü sınıflandırma, nesne tespiti ve bölütleme gibi görevlerde verimli bir omurga olarak kullanılır. ResNet-34, ImageNet gibi büyük ölçekli veri setlerinde eğitim aldıktan sonra birçok göreve transfer öğrenme yöntemiyle kolayca adapte edilebilmektedir. ResNet-34, daha az parametre ve daha düşük hesaplama gereksinimi sayesinde daha derin ResNet varyantlarına göre daha hafiftir ve bu nedenle özellikle hesaplama kaynaklarının sınırlı olduğu durumlarda tercih edilir.



Şekil 11. Transfer öğrenmede omurga kullanımı U-Net-ResNet (Neven ve Goedemé, 2021).

Bu model, derin özellik çıkarım gerektiren birçok görevde yaygın olarak kullanılır:

- Görüntü Sınıflandırma: ResNet-34, ImageNet gibi geniş veri setlerinde eğitilerek yüksek doğruluk oranları sunan bir sınıflandırma modeli olarak kullanılır. Özellikle sınırlı veri bulunan alanlarda transfer öğrenme yöntemiyle farklı sınıflandırma görevlerine uyarlanabilir (He vd., 2016).
- Nesne Tespiti ve anlamsal bölütleme: Mask R-CNN ve Faster R-CNN gibi daha karmaşık nesne tespiti modellerinde omurga olarak ResNet-34, özellik çıkarımı için kullanılır. Bu sayede model, geniş sahnelerdeki nesnelerin tespitini ve bölütlemesini doğru bir şekilde yapabilir (Ren vd., 2015; He vd., 2017).
- Medikal Görüntüleme: Tıbbi görüntülerin analizinde, hücresel detayların çıkarımı gibi hassas görevlerde ResNet-34, başarılı bir şekilde özellik çıkarımı yaparak sınırlı veri setlerinde bile yüksek doğruluk elde edilmesini sağlar (Raghu vd., 2019).

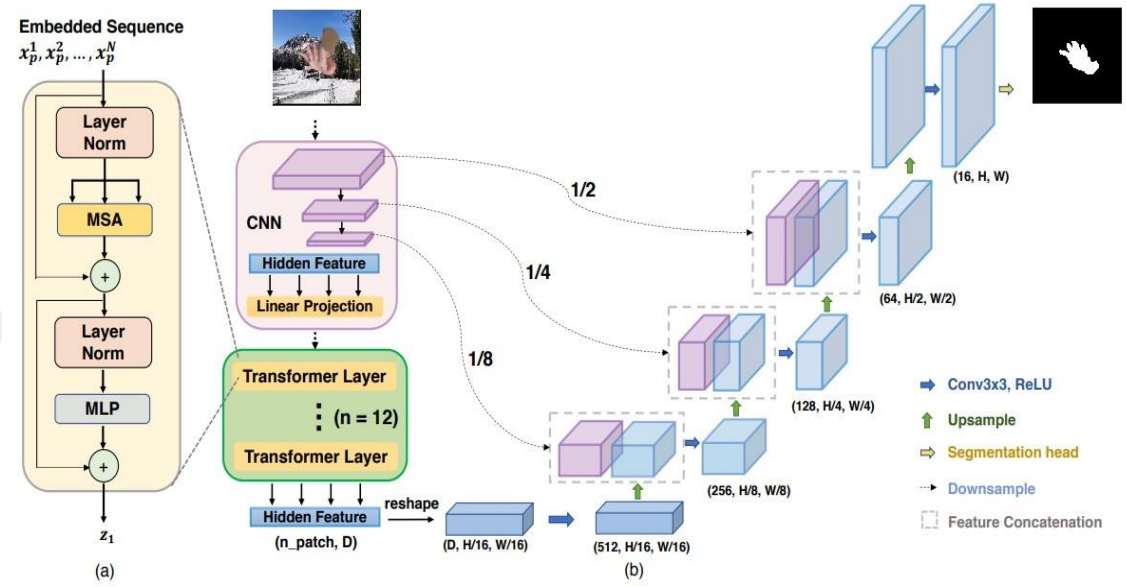
3.2.2. Vision Transformer

Vision Transformer (ViT) modelleri, öncelikle doğal dil işleme alanında yaygın olan transformer yapısını görsel verilerde kullanmak üzere tasarlanmıştır. Vision Transformer'ınn ana yapısı, görüntüleri küçük parçalara ayırıp her parçayı bir "token" olarak işlemeye dayanır. Bu yapı, görüntü işleme modellerine getirilen yeniliklerden biri olup, görsel bilgilerin daha verimli ve etkili işlenmesini sağlamaktadır.

Vision transformer yapısı;

- Girdi ve Parçalama: Vision Transformer, bir görüntüyü sabit boyutlu karelere bölerek her bir kareyi bir vektör olarak temsil eder. Bu işlemde, her görüntü parçası, bir "embedding" tabakasına girer ve her bir parçanın boyutu sabitlenmiş bir vektör olarak çıkar. Bu yöntem, her parçanın diğer parçalarla bağımsız olarak işlenmesini sağlayarak klasik evrişimsel sinir ağlarının sınırlamalarını aşar (Han vd., 2022).
- Pozisyon Kodlaması: Transformer yapısında olduğu gibi, Vision Transformer modellerinde de pozisyon kodlaması önemlidir. Görüntü parçalarının ardışıklığı olmadığı için modelin bu parçaların konumunu öğrenmesi gerekmektedir. Pozisyon kodlama, her parçanın konum bilgilerini ekleyerek görüntüdeki yer bilgilerini modelin dikkate almasını sağlar (Arnap vd., 2021).
- Çoklu Başlıklı Kendine Dikkat (Self-Attention) Mekanizması: Vision Transformer'ının özünde yer alan çok başlıklı kendine dikkat mekanizması, her görüntü parçasının diğer tüm parçalarla ilişkisini öğrenmesini sağlar. Böylece model, her pikselin diğer piksellerle olan ilişkisini öğrenerek daha iyi bir özellik çıkarımı sağlar. Bu mekanizma sayesinde model, görüntüdeki uzun mesafeli bağıntıları daha iyi öğrenir ve karmaşık görsel yapıları daha etkili işler (Vaswani vd., 2017).
- Transformer Blokları: Vision Transformer modelleri, ardışık transformer bloklarından oluşur. Her blok, bir "multi-head attention" katmanı ve ardından bir "feed-forward" katmandan oluşur. Bu yapıda, her bir bloğun çıktısı bir sonraki bloğa girdi olarak geçer ve modelin derin öğrenme kabiliyeti artırılır. Bu çok katmanlı yapı, modelin derin özellikleri daha iyi öğrenmesine imkan tanır (Liu vd., 2021).
- Dönüşümler ve Ağır Öğrenme: Vision Transformer modelleri, eğitim sırasında oldukça büyük veri kümeleri gerektirir. Büyük veri kümeleri üzerinde eğitimle, modelin daha karmaşık yapıları öğrenmesi sağlanır. Bu yapı, özellikle büyük veri kümelerinde evrişimsel sinir ağlarının performansını aşmaktadır. Modelin eğitim sürecinde verimliliği artırmak için çeşitli optimizasyon yöntemleri kullanılmaktadır. Örneğin, Swin Transformer modeli, kayan pencereler ile işlem yaparak belleği daha verimli kullanır ve geniş veri kümelerinde dahi performansını korur (Tang vd., 2021).
- Çıkış ve Sınıflandırma Katmanı: Modelin son aşaması sınıflandırma katmanıdır. Transformer blokları tarafından işlenen görüntü parçaları, son aşamada

birleştirilerek bir "global" özellik vektörü oluşturulur ve bu vektör sınıflandırma katmanına gönderilir. Sınıflandırma katmanı, gelen bilgiyi analiz ederek görüntünün sınıf etiketini belirler. Vision Transformer modelleri, bu yapısıyla görsel sınıflandırma görevlerinde oldukça yüksek başarı oranına ulaşmaktadır (Chen vd., 2021).



Şekil 12. Vision Transformer oto kodlayıcı kullanımı (Chen vd., 2021).

Vision Transformer, bilgisayarla görme alanında görüntü işlemeye yenilikçi bir yaklaşım getirmiştir. Çeşitli varyantları ve optimizasyon teknikleri ile Vision Transformer modelleri, farklı uygulama alanlarında etkili sonuçlar vermekte ve araştırmacılar için geniş bir çalışma alanı sunmaktadır.

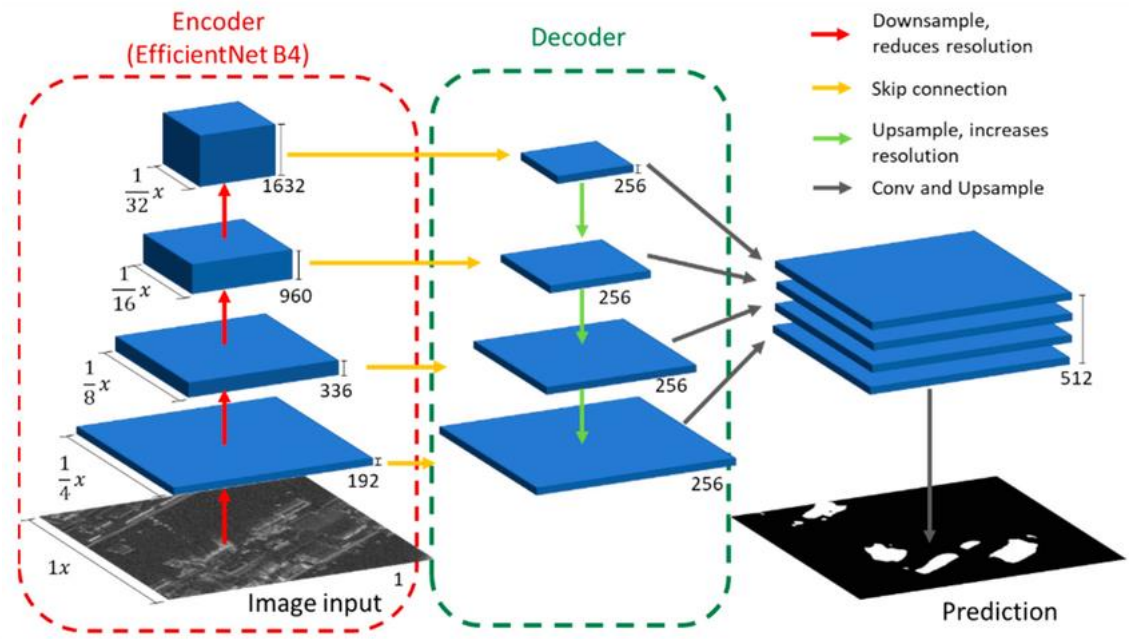
3.2.3. EfficientNet-B5

EfficientNet-B5 modeli, makine öğreniminde popüler hale gelen ve yüksek doğruluk oranlarını düşük hesaplama maliyetiyle birleştiren bir evrişimli sinir ağıdır. EfficientNet-B5, modelin genişlik, derinlik ve çözünürlük boyutlarını verimli bir şekilde ölçekleyerek daha küçük bir model yapısıyla yüksek performans sağlamayı hedefler. Bu ölçekleme stratejisi, modelin daha yüksek doğruluk seviyelerine ulaşmasına imkan tanır (Tan ve Le, 2021). Örnek bir kullanım Şekil 14'de görülmektedir.

EfficientNet-B5'in Yapısı:

- Verimli Model Ölçekleme: EfficientNet, model genişliğini yani katman sayısını, derinliğini ve çözünürlüğünü orantılı bir şekilde artırarak verimliliği optimize eder. Böylece, özellikle yüksek doğruluk gerektiren görevlerde modelin hesaplama maliyetini düşürür (Tan ve Le, 2021).
- MBConv Blokları: EfficientNet-B5, mobil cihazlarda daha verimli çalışabilmesi için MobilNetV2'nin kullandığı MBConv bloklarını içerir. Bu bloklar, dar alanlarda bile bilgi kaybını azaltarak etkin bir özellik çıkarımı sağlar. MBConv bloklarının kullanımı, modelin hesaplama maliyetini düşürür (Vijayan vd., 2023).
- Sıkıştırma ve Uyarılma (Squeeze-and-Excitation) Katmanları: Model, her katmandan çıkan özellik haritalarını güçlendirmek için sıkıştırma ve uyarılma işlemleri kullanır. Bu, modelin görsel verilerde belirli alanlara daha fazla odaklanmasını sağlayarak verimliliğini artırır (Mahasin ve Dewi. 2022).

  EfficientNet-B5'in omurga olarak kullanımı, genellikle sınıflandırma veya bölütleme gibi çeşitli görevlerde modelin daha geniş yapılarla birleştirilmesine olanak tanır.
- Tıbbi Görüntü bölütleme: EfficientNet-B5, çeşitli bölütleme modellerinde omurga olarak kullanılarak görüntüdeki hastalık alanlarını doğru bir şekilde belirler.Örneğin, göğüs röntgenlerinde EfficientNet-B3 omurga olarak kullanılarak pnömotoraks tespiti yapılmıştır (Abedalla vd., 2021).
- U-Net++ ile Birleştirilmesi: EfficientNet-B5, U-Net++ gibi bölütleme modelleriyle birleştirilerek tıbbi görüntülerde detaylı bölgesel analiz yapmayı sağlar. EfficientNet-B0 ile başlatılan bölütleme modeli, yüksek çözünürlüklü uydu görüntülerinde bina gibi karmaşık nesneleri tespit etmede başarılı sonuçlar verir (Ovi vd., 2023).
- COVID-19 CT Görüntülerinde Kullanımı: EfficientNet-B5 modeli, COVID-19'a özgü CT görüntülerinde sınıflandırma ve teşhis amaçlı omurga olarak kullanılmıştır. Özellikle CLAHE gibi görüntü iyileştirme yöntemleriyle entegre edildiğinde model, daha yüksek doğruluk sağlar (Ebenezer vd., 2022).



Şekil 13. FPN-EfficientNet-B5 oto kodlayıcı kullanımı (Wangiyana vd., 2022).

EfficientNet-B5'in farklı varyantları (B0-B7) bu tip görevlerde etkin bir şekilde kullanılmakta olup, modelin ölçeklenebilirliği sayesinde geniş veri kümeleri üzerinde kolaylıkla ayar yapılabilir. EfficientNet'in omurga olarak kullanımı, özellik çıkarımı açısından hesaplamada optimizasyon sağlamakta ve yüksek doğruluk gerektiren birçok alanda tercih edilmektedir.

3.3. Değerlendirme Metrikleri

3.3.1. Karmaşıklık Matrisi

Makine öğrenmesinde sınıflandırma problemleriyle çalışırken, modelin performansını detaylı bir şekilde değerlendirmek önemlidir. Bu noktada Karmaşıklık Matrisi, modelin tahminlerinin doğruluğunu farklı açılardan analiz etmek için güçlü bir araç olarak karşımıza çıkar (Powers, 2020). Modelin gerçekte pozitif olanları ne kadar doğru tahmin ettiğini, gerçekte negatif olanları ne kadar doğru tahmin ettiğini ve hangi tür hataları yaptığını gösterir.

Tablo 1. Karmaşıklık matrisi

	Tahmin Edilen: Pozitif	Tahmin Edilen: Negatif
Gerçek: Pozitif	True Positive (TP)	False Negative (FN)
Gerçek: Negatif	False Positive (FP)	True Negative (TN)

Karmaşıklık Matrisi, temel olarak dört farklı kategoriye odaklanır:

- Doğru Pozitif (TP): Modelin doğru bir şekilde pozitif olarak tahmin ettiği örnekler.
- Doğru Negatif (TN): Modelin doğru bir şekilde negatif olarak tahmin ettiği örnekler.
- Yanlış Pozitif (FP) (Tip I Hata): Modelin yanlış bir şekilde pozitif olarak tahmin ettiği örnekler.
- Yanlış Negatif (FN) (Tip II Hata): Modelin yanlış bir şekilde negatif olarak tahmin ettiği örnekler.

Bu kategoriler, modelin hangi durumlarda başarılı olduğunu ve hangi durumlarda hata yaptığını anlamamıza yardımcı olur (Fawcett, 2006). Karmaşıklık Matrisi'nden elde edilen bilgiler, modelin performansını daha derinlemesine anlamak ve iyileştirmek için kullanılabilir. Ayrıca, Karmaşıklık Matrisi'ni kullanarak Kesinlik, Duyarlılık ve Özgüllük gibi önemli metrikleri hesaplayabiliriz (Sokolova ve Lapalme, 2009):

Kesinlik: Modelin pozitif olarak tahmin ettiği örnekler arasında gerçekte kaçının pozitif olduğunu gösterir. Yanlış pozitiflerin maliyetli olduğu durumlarda önem kazanır (Kitrungrotsakul ve Jirapatnakul, 2023).

$$\text{Kesinlik} = \frac{TP}{TP+FP} \qquad \text{(Eşitlik-1)}$$

Duyarlılık: Gerçekte pozitif olan örnekler arasında modelin kaçını doğru bir şekilde pozitif olarak tahmin ettiğini gösterir. Yanlış negatiflerin maliyetli olduğu durumlarda kullanılır (Powers, 2020).

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \qquad \text{(Eşitlik-2)}$$

Özgüllük: Gerçekte negatif olan örnekler arasında modelin kaçını doğru bir şekilde negatif olarak tahmin ettiğini gösterir. Yanlış pozitiflerin maliyetli olduğu durumlarda önemlidir (Powers, 2020).

$$\text{Özgüllük} = \frac{TN}{TN+FP} \quad \text{(Eşitlik-3)}$$

Bu metrikler, modelin performansını sadece genel doğruluk üzerinden değerlendirmek yerine, farklı hata türlerini ve performans özelliklerini ayrıntılı bir şekilde incelememizi sağlar. Özellikle veri setinin dengesiz olduğu durumlarda, yani bir sınıfın diğerlerinden çok daha fazla temsil edildiği durumlarda, Karmaşıklık Matrisi ve bu matristen türetilen metrikler modelin performansını değerlendirmek için daha güvenilir bir yol sunar. Modelin farklı sınıflar üzerindeki performansını ve yaptığı hata türlerini analiz ederek, modelin optimizasyonu için gerekli yönlendirmeleri sağlar (Kitrungrotsakul ve Jirapatnakul, 2023).

3.3.2. Accuracy

Makine öğrenmesi modellerinin performansını değerlendirmek, özellikle sınıflandırma problemlerinde büyük önem taşır. Bu amaçla kullanılan birçok metrik arasında, doğruluk (accuracy) en yaygın ve anlaşılır olanlardan biridir (Powers, 2020). Doğruluk, modelin yaptığı tüm tahminler içerisinde doğru olanların oranını ifade eder ve modelin genel performansı hakkında bilgi verir (Kitrungrotsakul ve Jirapatnakul, 2023). Yüksek doğruluk, modelin verileri ne kadar iyi öğrendiğini ve yeni verilere ne kadar başarılı bir şekilde genelleme yapabildiğini gösterir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(Eşitlik-4)}$$

3.3.3. Intersection Over Union- IoU

IoU (Keşisimlerin birleşimi), nesne tespiti ve görüntü bölütleme gibi bilgisayarlı görü görevlerinde modelin başarımını değerlendirmek için kullanılan önemli bir metriktir (Rezatofighi vd., 2019). Temel olarak, modelin tahmin ettiği nesnenin sınırlayıcı kutusu veya bölütlenmiş alanının gerçek sınırlayıcı kutu veya alanla ne kadar örtüştüğünü ölçer (Rahman ve Wang, 2018). IoU, modelin nesneleri arka plandan ne kadar iyi ayırt ettiğini nicel olarak belirleyerek, farklı modelleri karşılaştırmak ve bir modelin performansını

iyileştirmek için hangi alanlara odaklanılması gerektiğini anlamak için kullanışlı bir araçtır.

$$\text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}} \qquad \text{(Eşitlik-5)}$$

3.3.4. Dice Coefficient

Görüntü bölütlemesinde model başarımını değerlendirmek için kullanılan metriklerden biri de Zar Katsayısıdır (Dice coefficient). Tıbbi görüntü bölütlemede sıklıkla tercih edilen bu metrik, iki alan arasındaki benzerliği ölçer. IoU ile benzerlik göstersede, Dice coefficient formülünde kesişim alanının iki kat ağırlıklandırılmasıyla küçük nesnelerin bölütlenmesinde daha hassas sonuçlar elde edilmesini sağlar. Bu sayede, özellikle küçük detayların önemli olduğu tıbbi görüntülerde daha doğru bir değerlendirme yapılabilmektedir (Padilla vd., 2020).

$$\text{Dice} = \frac{\text{2xTP}}{\text{2xTP+FP+FN}} \qquad \text{(Eşitlik-6)}$$

3.3.5. F1 Score

Özellikle eşit olmayan verilerde kullanılan, modelin performansını ölçmek için kullanılan bir metriktir. Kesinlik (Precision) ve Duyarlılık (Recall) arasındaki dengeyi ölçer. Bu dengeyi sağlamak için duyarlılık ve kesinliğin harmonik ortalamasını alır ve böylece dengeli bir ölçüm gerçekleştirmiş olur. Ayrıca harmonik ortalama ile uç değerlerin etkisini azaltarak dengeli bir ölçüm sağlar. f1 score"un yüksek olması ile modelin,hem yüksek duyarlılık hemde yüksek hassasiyete sahip olduğu anlaşılır.(Padilla vd., 2020)

$$\text{F1} = 2x\ \frac{\text{TP}}{\text{2xTP+FP+FN}} \qquad \text{(Eşitlik-7)}$$

3.4. Loss Functions

Derin öğrenme modellerinin başarısı, tahminlerini gerçek değerlerle karşılaştıran ve aradaki farkı minimize eden kayıp fonksiyonlarına bağlıdır. Bu fonksiyonlar, modelin öğrenme sürecini yönlendirir ve ağırlıkların optimize edilmesini sağlar. Sınıflandırma ve bölütleme gibi görevlerde, kullanılan kayıp fonksiyonu uygulamaya ve veri setine uygun olmalıdır. Örneğin, dengesiz veri setlerinde Dice Loss ve Jaccard Loss tercih edilirken, genel sınıflandırma problemlerinde Binary Cross-Entropy yaygın olarak kullanılır. Doğru

kayıp fonksiyonu seçimi, modelin performansını ve genelleme yeteneğini önemli ölçüde etkiler (Abraham ve Khan, 2019). Bu çalışmada Binary Cross-Entropy nin genelleştirilmiş bir versiyonu olan Tversky loss kayıp fonksiyonu kullanılmıştır.

3.4.1. Binary Cross-Entropy

Binary Cross-Entropy (İkili Çapraz Entropi), makine öğrenmesinde, özellikle ikili sınıflandırma problemlerinde kullanılan bir kayıp fonksiyonudur. Modelin, bir girdiyi iki sınıftan birine ait olma olasılığı olarak tahmin ettiği değer ile gerçek etiket arasındaki farkı ölçer (Ho vd., 2019). Örneğin, bir e-postanın spam olup olmadığını tahmin eden bir model düşünelim. Model, bir e-posta için spam olma olasılığını 0.7 olarak tahmin ederse ve e-posta gerçekten spam ise, Binary Cross-Entropy bu tahminin doğruluğunu ölçer ve modelin öğrenme sürecini yönlendirir. Bu fonksiyon, modelin tahminlerini gerçek etiketlere yaklaştırmak için kullanılır. Amaç, modelin yaptığı tahminler ile gerçek etiketler arasındaki farkı minimize etmektir. Binary Cross-Entropy değeri ne kadar düşükse, modelin performansı o kadar iyi demektir. Bu nedenle, model eğitimi sırasında, modelin parametreleri, Binary Cross-Entropy değerini minimize edecek şekilde ayarlanır (Ruby vd., 2020).

$$BCE = \frac{1}{N}(TP \cdot \log(p) + FN \cdot \log(1-p) + FP \cdot \log(1-p) + TN \cdot \log(1-p)) \qquad \text{(Eşitlik-8)}$$

N: Toplam örnek sayısı
p: Modelin pozitif sınıf için tahmin ettiği olasılık
log: Logaritma

3.4.2. Tversky Loss

Görüntü bölütlemede, özellikle de sınıf dengesizliği söz konusu olduğunda, Tversky Loss etkili bir kayıp fonksiyonu olarak kullanılmaktadır. Jaccard Index'in genelleştirilmiş bir hali olup, yanlış pozitif ve yanlış negatiflerin oranını ayarlamak için α ve β hiperparametrelerini kullanır (Salehi vd., 2017). Jaccard Index; IoU ile hesaplama olarak aynı olsada daha genel bir kavram olmakla birlikte, daha genel bir benzerlik ölçer ve küme teorisi gibi daha geniş bir matematiksel bağlamlarda kullanılır. Tversky Loss'un bu esnekliği, az temsil edilen sınıfların ve küçük nesnelerin bölütlenmesinde modelin performansını iyileştirir. Örneğin, tıbbi görüntülerde küçük tümörlerin tespiti kritik öneme sahiptir ve Tversky Loss bu gibi durumlarda başarılı bir çözüm sunar (Abraham ve Khan, 2019). Bölütleme performansını daha da optimize etmek amacıyla Tversky

Loss'un farklı varyasyonları geliştirilmiştir. Focal Tversky Loss, zorlu bölgelerde sınıflandırma doğruluğunu artırırken, Accelerated Tversky Loss ise eğitim sürecini hızlandırarak büyük veri setlerinde verimliliği yükseltir (Nasalwai vd., 2021). Ayrıca, Tversky Loss'un parametreleri ayarlanarak sınıf dengesizliği gibi sorunlara çözüm getirilebilir, bu da onu Dice Loss ve Cross-Entropy Loss gibi klasik yöntemlere göre daha esnek bir alternatif yapar (Jadon, 2020).

$$\text{Tversky Loss} = 1 - \frac{TP}{TP + \alpha * FP + \beta * FN} \qquad \text{(Eşitlik-9)}$$

3.5. FreiHAND Veri Seti

El hareketlerini ve şekillerini 3 boyutlu olarak algılamak, özellikle robotik, artırılmış gerçeklik ve insan-bilgisayar etkileşimi gibi alanlarda oldukça önemlidir. Ancak bu karmaşık görevi RGB görüntülerden çözmek, mevcut derin öğrenme algoritmaları için hala büyük bir zorluk teşkil etmektedir. Çoğu algoritma, eğitildikleri veri setine aşırı derecede uyum sağlamakta ve bu da gerçek dünya senaryolarında veya farklı veri setlerinde performans düşüşlerine yol açmaktadır. Bu sorunu ele almak için Zimmermann vd., (2019), FreiHAND adlı yeni bir veri setini tanıttı. Bu veri seti, 8 kamera kullanılarak 32 katılımcıdan toplanan 130.240 RGB görüntü den oluşmaktadır. FreiHAND, el pozlarının ve şekillerinin geniş bir çeşitliliğini kapsamakta ve her görüntü için 2 ve 3 boyutlu eklem konumları, bölütleme maskeleri ve 3 boyutlu yüzey modelleri gibi zengin işaretlemeler sağlamaktadır. Ayrıca, bazı görüntüler yeşil ekran önünde çekilerek, farklı arka planlara entegre edilebilme ve modellerin gerçek dünya koşullarına daha iyi genelleme yapabilme yeteneğini geliştirme olanağı sunmaktadır. FreiHAND, el pozu tahmini ve el bölütleme alanındaki araştırmalar için önemli bir kaynaktır. Bu veri seti, derin öğrenme modellerinin eğitilmesi ve değerlendirilmesi için daha önce elde edilemeyen bir ölçekte ve çeşitlilikte veri sunmaktadır. FreiHAND'in sunduğu zengin anotasyonlar ve gerçekçi görüntüler, daha doğru ve güvenilir 3 boyutlu el modeli tahminlerine yol açabilir ve böylece robotik, AR ve insan-bilgisayar etkileşimi gibi alanlarda yeni uygulamaların önünü açabilir. Veri setinden örnek resimler Şekil-1 de gösterilmiştir. Araştırmacıların kullanımına sunulan diğer verisetleri (EgoHands, HandOverFace, RHD, GTEA Gaze+, BigHand, Oxford Hand) gibi verisetlerine göre FreiHAND birçok avantaja sahiptir.

FreiHAND'ı diğer veri setlerinden ayıran temel özellikler şunlardır:

- Büyüklük ve çeşitlilik**:** FreiHAND, 32 farklı kişiden alınan 130 binden fazla el görüntüsü içerir ve bu da onu mevcut veri setlerinden önemli ölçüde daha büyük ve çeşitli kılar.
- 3 boyutlu el şekli ek açıklamaları**:** Diğer veri setlerinin çoğu yalnızca 3 boyutlu el poz ek açıklamaları sağlarken, FreiHAND hem 3 boyutlu el poz hem de şekil ek açıklamaları içerir. Bu, araştırmacıların ek açıklamalı veriler üzerinde tam olarak denetlenen bir şekilde şekil tahmin modelleri eğitmelerine olanak tanır.
- Gerçek dünya koşulları**:** FreiHAND'deki görüntüler, kontrollü bir ortamda çekilmiştir, ancak yine de gerçek dünyadaki senaryolarda bulunan aydınlatma, arka plan ve el duruşlarında önemli farklılıklar gösterirler. Bu, FreiHAND üzerinde eğitilmiş modellerin gerçek dünya ortamlarına daha iyi genelleme yapmasına yardımcı olur.
- Yarı otomatik etiketleme süreci**:** FreiHAND'ı oluşturmak için yazarlar, büyük ölçekli 3 boyutlu el şekli ve poz verilerini verimli bir şekilde yakalamak için seyrek 2 boyutlu ek açıklamalarını kullanan yeni bir yinelemeli, yarı otomatik "döngüde insan" yaklaşımı önerilmiştir. Yarı otomatik kavramı; etiketleme ve analiz süreçleri kısmen otomatik olarak yapılır. Ancak hala insan müdahelesine ihtiyaç vardır. Algoritmaların ürettiği tahminler, insanlar tarafından kontrol edilir ve gerekli ise düzeltilir. Döngüde insan kavramı; insanın aktif olarak rol aldığı durumdur. Modelin çıktıları üzerinde geri bildirim verir, hataları düzeltir veya belirli durumlarda el ile etiketleme yapar. Bu modelin doğruluğunu artırmak için insan makine işbirliği mekanizmasıdır.

(en soldan başlayarak; yeşil arka plan görüntü, arka plan değiştirilmiş görüntü, maskeler)

Şekil 14. Freihand verisetinden örnek görüntüler (Zimmermann vd., 2019).

Bu verisetini seçmemizdeki temel amaç veri setinin gerçek dünya şartlarında elin herhangi bir açıda, herhangi bir pozisyonda olabileceği, elin farklı ışık altında olabileceği gibi durumlar için oluşturulan görüntülerden oluşmasıdır. Böylece oluşturulan modellerin gerçek dünya koşullarında genelleme yeteneği artırılmak istenmiştir. Görüntülerin arkaplanı değiştirilerek farklı arkaplanlar ile modelin daha farklı arkaplan görüntülerde el bölgesi tesbit edip, modelin başarısı artırılmak istenmiştir. Ayrıca verisetindeki görüntülere farklı arkaplanlar eklenerek aşırı öğrenme sorununun önüne geçmeye çalışmıştır. Diğer birçok verisetindeki temel sorun belli ışık altında, belli açılardaki eller veya belli bir arkaplan el görüntülerinden oluşmasıdır. Ayrıca verisetlerindeki görüntü sayılarının az olması, veri setlerindeki düşük görüntü çözünürlüğü diğer birkaç etkendir. Bu sebeplerden dolayı bu verisetleri modellerin genelleme yeteneğini oldukça azaltmaktır. Oysa bizim istediğimiz herhangi bir kısıtlı ortamının olmadığı, model için elin hangi açı ya da durumda olduğunun önemli olmadığı, ortamdaki ışık etkisinin sorun olmadığı bir yapı geliştirmektir. Freihand veriseti bize diğer verisetlerine göre daha falza genelleme yeteneği, aşırı öğrenmenin önüne geçebilme, kısıtlamasız ortamlarda daha yüksek başarım sergileme olanağı sunmaktadır.

4. BULGULAR

Bu çalışmada, el bölütleme alanında başarılı sonuçlar elde etmek için Zimmermann ve arkadaşları tarafından 2019'da yayınlanan FreiHAND veri seti temel alınmıştır. Bu veri seti, el maskeleri, 2 boyutlu ve 3 boyutlu eklem noktaları ve 3 boyutlu el yüzeyi modelleri gibi zengin bilgiler içermesiyle öne çıkmaktadır. FreiHAND, el bölütleme gibi karmaşık görevler için sağlam bir zemin oluşturmaktadır. Modelin daha iyi genelleme yapabilmesi ve farklı durumlara uyum sağlayabilmesi için veri seti üzerinde çeşitli ön işlemler gerçekleştirilmiştir. Örneğin, FreiHAND veri setindeki görüntülerin çoğunluğu yeşil perde önünde çekildiğinden, arka planlar değiştirilerek modelin farklı ortamlara maruz kalması sağlanmıştır. Bu sayede, modelin gerçek dünya koşullarında daha başarılı olması hedeflenmiştir. Hazırlanan veri seti, eğitim, doğrulama ve test olmak üzere üç bölüme ayrılmıştır. Modelin öğrenme aşamasında kullanılmak üzere veri setinin %70'i eğitim setine ayrılmıştır. %20'lik kısım olan doğrulama seti, modelin öğrenme sürecindeki performansını izlemek ve ayarlamalar yapmak için kullanılmıştır. Son olarak, veri setinin %10'u test seti olarak ayrılmış ve bu set, modelin eğitildikten sonraki başarısını ölçmek için kullanılmıştır. Bu üçlü ayrım, modelin hem öğrenme aşamasında başarılı olmasını hem de gerçek dünya verilerine genelleme yapabilmesini sağlamaktadır.

Modelin eğitimi, Google Colaboratory (Colab) platformunda gerçekleştirilmiştir. Colab, ücretsiz GPU erişimi sunan ve herhangi bir kurulum gerektirmeyen bir Jupyter Notebook hizmetidir. Bu çalışmada Colab Pro platformu tercih edilmiş ve bu sayede daha yüksek işlem gücünden yararlanılarak büyük veri seti üzerinde hızlı ve verimli bir eğitim süreci gerçekleştirilmiştir. Eğitim sürecinde Python programlama dili ve PyTorch kütüphanesi kullanılmıştır. Çalışmada kullanılan veri seti, modeller, omurga mimarileri ve hiperparametreler Tablo 2'de detaylı bir şekilde sunulmuştur.

Tablo 2. Model eğitim parametreleri

Parametre	Değer
Eğitim verisi (Train)	23443
Doğrulama (Validation)	6512
Test	2605
Görüntü çözünürlük	256 x 256
Kodlayıcı (Encoder)	Vision Transformer, ResNet-34, Efficientnet-B5
Kodlayıcı ağırlıkları (Encoder weights)	Imagenet
Adım sayısı (Epochs)	20
Kayıp fonksiyonu (Loss fonction)	Tversky Loss
Öğrenme oranı (Learning Rate)	0.0001
Parça boyutu (batch Size)	64
Optimizer	Adam
Metrikler	İou, Dice coefficient, F1 score, Accuracy

4.1. Çalışmanın Sayısal Sonuçları

Bu çalışma, bölütleme alanında kullanılan dört farklı oto kodlayıcı tabanlı bölütleme modelinin MA-Net, FPN, U-Net ve PSP-Net performansını, üç farklı omurga mimarisi EfficientNet-B5, Vision Transformer ve ResNet-34 ile performansları analiz edilmiştir. Bu modellerin el bölütlemedeki etkinliğini değerlendirmek için doğruluk, Intersection over Union (IoU), F1 skoru ve Dice katsayısı gibi yaygın olarak kullanılan metrikler kullanılmıştır. Elde edilen sonuçlar, her model ve omurga kombinasyonunun güçlü ve zayıf yönlerini ayrıntılı bir şekilde ortaya koymaktadır. Elde edilen sonuçlar Tablo 3'te gösterilmiştir.

Tablo 3. Model eğitim sonuçları

Model	**Omurga**	**Acc**	**IoU**	**F1 skor**	**Dice**
MA-Net	EfficientNet-B5	0.9876	0.8776	0.9349	0.9876
	Vision Transform	0.9889	0.8912	0.9425	0.9889
	Resnet-34	**0.9921**	0.8679	**0.9584**	0.9863
FPN	EfficientNet-B5	0.9879	0.8801	0.9362	0.9879
	Vision Transform	0.9886	0.8867	0.9399	0.9886
	Resnet-34	0.9861	0.8689	0.9298	0.9865
U-Net	EfficientNet-B5	0.9881	0.8839	0.9384	0.9881
	Vision Transform	0.9892	**0.8917**	0.9427	**0.9892**
	Resnet-34	0.9874	0.8760	0.9339	0.9873
PSP-Net	EfficientNet-B5	0.9853	0.8579	0.9235	0.9853
	Vision Transform	0.9855	0.8689	0.9241	0.9855
	Resnet-34	0.9848	0.8515	0.9118	0.9848

4.2. Modellerin Genel Performansı

U-Net: U-Net modeli, tüm omurga mimarileriyle birlikte kullanıldığında tutarlı ve güvenilir bir performans göstermiştir. Özellikle Vision Transformer omurgasıyla birleştirildiğinde en yüksek Dice ve IoU değerlerine ulaşarak el bölütleme görevinde yüksek bir başarı elde etmiştir. Bu kombinasyon, IoU metriğinde 0.8917 ve Dice metriğinde 0.9892 gibi iyi sonuçlar üretmiştir. Buradan U-net modelinin omurga kısmında Vision transformer kullanımı, dikkat mekanizmanın çok daha iyi özellik çıkarması ve bu çıkarılan özellikler ile modelin atlamalı bağlantıları kullanarak daha yüksek bir bölütleme performansı göstermiş olduğu görülmektedir.

MA-Net: Çok ölçekli dikkat mekanizmalarını kullanan MA-Net, ResNet-34 omurgasıyla kullanıldığında en yüksek doğruluk değerine (0.9921) ulaşmış. Ayrıca bu kombinasyon, en yüksek F1 skorunu (0.9584) da elde etmiştir. MA-Net'in ResNet-34 ile birlikte kullanıldığında diğer modellere ve omurgalara göre daha iyi performans gösterdiği görülmektedir. Model sahip olduğu dikkat mekanizması sayesinde, el bölgesine daha fazla odaklandığı, Omurga olarak ResNet-34 kullanarak, omurganın derin yapısı sayesinde daha fazla özellik çıkarımı yapılmış, model kendi dikkat mekanizması ve çok ölçekli mimarisi ile yüksek değerler elde etmiştir.

FPN: Özellik piramit ağı (FPN) modeli, EfficientNet-B5 ve Vision Transformer omurga yapılarıyla iyi sonuçlar elde etmiş olsa da, genel performansı U-Net ve MA-Net'in gerisinde kalmıştır. FPN, IoU ve Dice skorlarında belirli bir tutarlılık sağlamış, ResNet-34 omurgasıyla entegre edildiğinde performansında düşüş gözlemlenmiştir. FPN, farklı çözünürlükteki öznitelik haritalarını birleştirerek, hem düşük çözünürlükteki bilgileri hemde yüksek çözünürlükteki detayları harmanlar. Özellikle çok ölçekli öznitelikleri birleştirme konusunda etkili bir yapıdadır. EfficientNet-B5'in her blok katmanı farklı boyutlarda öznitelik haritaları üretir. Bu çok ölçekli bilgiye erişim sağlar. Modelin, omurganın bu farklı seviyelerdeki çıkışlarına entegre olarak iyi bir uyum yakaladığı görülmektedir.

PSP-Net: Piramit sahne ayrıştırma ağı modeli, performans açısından diğer modellere göre daha zayıf sonuçlar ortaya koymuştur. IoU ve Dice skorları, diğer modellerin elde ettiği değerlerin altındadır. Bu durum, PSP-Net'in el bölütleme gibi daha özel ve zorlu görevlerde diğer modellere göre daha az etkili olduğunu göstermektedir. Örnek olarak en düşük IoU ve Dice skorları PSP-Net ve ResNet-34 ile alınmış. PSP-Net modeline ResNet-34'ün eklenmesi, sınırlı derinlik nedeniyle beklenen performansı sağlamayabilir. ResNet-34, daha derin ağlara (örneğin ResNet-50 veya ResNet-101) kıyasla daha az soyutlama yapma kapasitesine sahip olup, bu da özellikle karmaşık

bölütleme görevlerinde yetersiz kalabilir. Derin ağlar daha zengin ve detaylı özellikler çıkarabildiğinden, PSP-Net'in piramidal çözümleme yapıları bu güçlü özelliklerle daha etkili çalışır. Ayrıca, ResNet-34 ile transfer öğrenimi yapılırken öğrenilen özellikler sınırlı olabilir, bu da bölütleme performansını düşürür. Modelin başarısı, eğitim süreci ve hiperparametre ayarlamalarına da bağlıdır; yanlış hiperparametreler veya derin ağ eksikliği, modelin overfitting veya underfitting yapmasına yol açabilir. Bu nedenle, daha derin bir omurga ağı ve dikkatli bir ince ayar, daha iyi sonuçlar elde edilmesini sağlayabilir.

4.3. Omurga Mimarilerinin Etkisi

EfficientNet-B5: EfficientNet-B5, düşük hesaplama maliyetiyle birlikte yüksek doğruluk oranları sunarak iyi bir performans sergilemiştir. IoU ve Dice metriklerinde istikrarlı bir performans sunan EfficientNet-B5, U-Net ve MA-Net modelleriyle etkili bir uyum sağlamıştır. Ancak, Vision Transformer omurgasına kıyasla IoU ve Dice skorlarında biraz geride kalmıştır.

Vision Transformer: Genel olarak en yüksek performansı sağlayan Vision Transformer, U-Net ve MA-Net modelleri ile entegre edildiğinde el bölütlemede dikkate değer bir başarı elde etmiştir. IoU metriğinde 0.8917'ye kadar ulaşan Vision Transformer, dikkat mekanizması sayesinde güçlü özellik çıkarımı yeteneğiyle bu yüksek performansı desteklemiştir. Vision Transformer (ViT) modelinin, U-Net ve MA-Net gibi yapılarla birleşerek iyi performans göstermesinin nedeni, her iki modelin güçlü özellik çıkarımı ve farklı ölçeklerdeki bilgileri başarılı şekilde birleştirme yetenekleridir. U-Net, simetrik kodlayıcı-kod çözücü yapısıyla görüntü bölütlemede yüksek çözünürlükteki detayları koruyarak başarılı sonuçlar verirken, MA-Net'in çok ölçekli dikkat mekanizmaları, düşük ve yüksek seviyedeki özelliklerin etkili bir şekilde entegrasyonunu sağlar. ViT, görsel veriyi ardışık dizi olarak işleyerek uzun menzilli bağıntıları öğrenme konusunda güçlüdür. Bu kombinasyon, ViT'in global bağlamları anlamasıyla, U-Net ve MA-Net'in yerel özellik çıkarımı ve dikkat mekanizmalarını birleştirerek, bölütleme ve diğer görevlerde daha yüksek doğruluk ve hassasiyet sağlar. Böylece, her iki modelin güçlü yönleri, ViT'in geniş bağlamı ile desteklenmiş olur, bu da genel performansı artırır.

ResNet-34: Geleneksel bir evrişimli sinir ağı mimarisi olan ResNet-34, özellikle MA-Net ile etkili sonuçlar vermiş olsa da, genel olarak diğer iki omurga mimarisine kıyasla daha düşük bir performans sergilemiştir. Özellikle PSP-Net ile entegrasyonunda en düşük IoU (0.8515) ve Dice (0.9848) skorlarını elde etmiştir.

4.4. Metriklere Göre Değerlendirme

IoU ve Dice Metrikleri: Vision Transformer omurgası, IoU ve Dice metriklerinde diğer omurga mimarilerine göre üstünlük sağlamıştır. Özellikle U-Net modeliyle entegrasyonu, bölütleme görevinde başarılı bir genel performans sunmuştur. Bu durum, Vision Transformer'ın dikkat mekanizmasının, el bölütleme gibi karmaşık görevlerde nesnelerin sınırlarını daha doğru bir şekilde belirlemede etkili olduğunu göstermektedir.

F1 Skoru ve Doğruluk: EfficientNet-B5 omurgası, F1 skoru ve doğruluk metriği açısından dengeli bir performans sunmuştur. Çoğu modelle kombinasyonunda Vision Transformer'ın hemen arkasından gelerek, el bölütleme görevinde yüksek bir doğruluk ve hassasiyet sağlamıştır. ResNet-34 omurgası ise MA-Net dışında genellikle daha düşük F1 skoru ile diğer iki omurganın gerisinde kalmıştır. Makine öğreniminde çok sınıflı veya dengesiz veri kümeleri için kullanılan metrikleri hesaplarken averaging (ortalama alma) yöntemleri büyük önem taşır. Özellikle F1 skoru, Dice skoru, IoU gibi metriklerde macro ve micro ortalama farkı, sonuçların değişmesine neden olabilir. Bölütleme görevinde, F1 skoru ve Dice skoru aslında aynı formüle sahiptir, ancak ortalama yöntemi farklı olursa hesaplama farklı olabilir. Macro ortalama; Her sınıfın performansını eşit ağırlıkta değerlendirir. Önce her sınıf için ayrı ayrı metrik (örneğin, precision, recall, F1 skoru) hesaplanır, ardından bu değerlerin ortalaması alınır. Bu yöntem, küçük sınıfların performansını daha iyi yansıtır, çünkü her sınıf eşit öneme sahiptir. Micro ortalama; Tüm sınıfları birleştirerek tek bir metrik hesaplar. Örneğin, tüm sınıflar için doğru pozitifler, yanlış pozitifler ve yanlış negatifler toplanır ve bu değerler üzerinden metrik hesaplanır. Bu yöntem, büyük sınıfların performansını daha fazla yansıtır, çünkü büyük sınıflar daha fazla örnek içerir. F1 skoru, precision ve recall değerlerinin harmonik ortalamasıdır, bu da hem doğru pozitiflerin hem de yanlış pozitif ve negatiflerin etkisini dengeler. Diğer taraftan, Dice katsayısı, doğru pozitiflerin, yanlış pozitifler ve yanlış negatiflerle olan oranını ölçer. Bu, Dice katsayısının bölütleme gibi görevlerde daha büyük doğru pozitifleri vurgulamasına neden olur. Bölütleme görevlerinde, Dice katsayısı genellikle daha yüksek çıkar çünkü bu metrik, modelin doğru tahmin ettiği pikselleri (true positives) daha fazla vurgular. Bu, özellikle sınıf dengesizliği olan veri setlerinde, modelin küçük hatalardan ziyade büyük doğru tahminlere odaklandığını gösterir. Dice katsayısı, macro ortalamaya benzer şekilde, her sınıfın doğruluğuna daha fazla dikkat çeker ve bu nedenle küçük sınıfların performansını daha iyi yansıtabilir. Micro ortalama ise, tüm sınıfları birleştirerek tek bir metrik hesaplar. Örneğin, tüm sınıflar için doğru pozitifler, yanlış pozitifler ve yanlış negatifler toplanır ve bu değerler üzerinden metrik hesaplanır. Bu yöntem, büyük sınıfların performansını daha fazla yansıtır çünkü büyük sınıflar daha

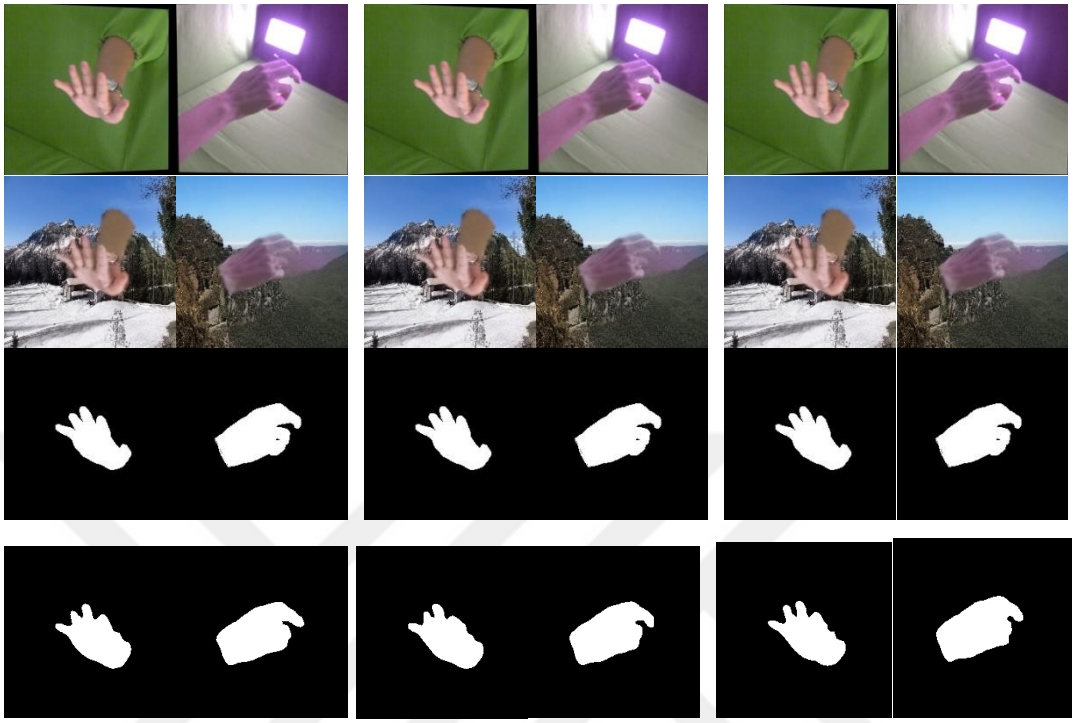
fazla örnek içerir. Bu nedenle, bölütleme ve sınıf dengesizliğine sahip veri setlerinde, Dice katsayısının F1 skorundan daha yüksek olması, modelin küçük hatalardan ziyade daha büyük doğru tahminlere odaklandığını gösterebilir.

4.5. Model ve Omurga Uyumu

Modeller ve omurga mimarileri arasındaki uyum, el bölütleme performansı üzerinde belirgin bir etkiye sahiptir. Örneğin, Vision Transformer ile U-Net birleşimi en yüksek IoU ve Dice değerlerini sunarken, PSP-Net ve ResNet-34 kombinasyonu bölütleme görevinde beklenenden düşük bir performans göstermiştir.Örneğin PSP-Net mimarisi ResNet-34 omurgası ile çok fazla uyum göstermemiştir. ResNet-34 ün omurga olarak özellik çıkarma yetisi, PSP-Net in yapısındaki havuzlama modülünün özellik çıkarma özelliğine etkisine fazla olmamıştır. Bu sebeble model başarısı diğer modellerin başarısından geride kalmıştır. FPN modeli ile kullandığımız omurgalarda EfficientNet-B5 omurgası dengeli bir performans sunsada, diğer omurgalarda başarı oranı düşük kalmıştır. Model omurga uyumsuzluğu burdada görülmektedir. Diğer yandan U-net modelinin Vision transformer yapısı ile başarısı, U-Net'in ince detayları öğrenme özelliği, basit ve etkili yapısı, Vision transformer'ın dikkat mekanizması ile yüksek oranda model ile uyum sağlayarak yüksek başarı elde etmiştir. Burada genel anlamda Vision transformer yapısının kullandığımız tüm modellerde yüksek bir başarı elde etmesinin anlamı, bu omurganın tüm kullandığımız modellerin yapısı ile uyumlu olarak çalıştığını göstermektedir. MA-Net modelinin ResNet-34 omurgası ile başarısı MA-Net' in çok ölçekli özellik çıkarımı ve sahip olduğu dikkat mekanizmasının, ResNet-34 omurgasının derin sinir ağı ile uyumlu çalıştığını göstermektedir. ResNet-34 omurgası MA-Net modeli dışında genel anlamda en düşük performans elde etmesi onun diğer modeller ile uyumunun pekiyi olmadığını göstermektedir. EfficientNet-B5 omurgası literatürde dengeli bir model olarak anılmaktadır. Bizim çalışmamızda bu omurganın dengeli skorlar elde ettiği görülmüştür. Bu omurga kullanılan tüm modeller ile uyumlu bir yapıda olduğunu dengeli bir performans göstererek kanıtlamıştır. Sonuç olarak el bölütleme gibi özelleşmiş görevlerin, model mimarisi ve omurga seçiminin, önemli olduğunu ve kullanılan model ile omurganın her zaman uyumlu olmadığını bize göstermektedir. Bu durum göz önünde bulundurularak gerekirse modelin yada omurganın değiştirilmesi yada modelin ve omurganın optimize edilmesi gerektiğini ortaya koymaktadır.

4.6. Çalışmanın Görsel Sonuçları

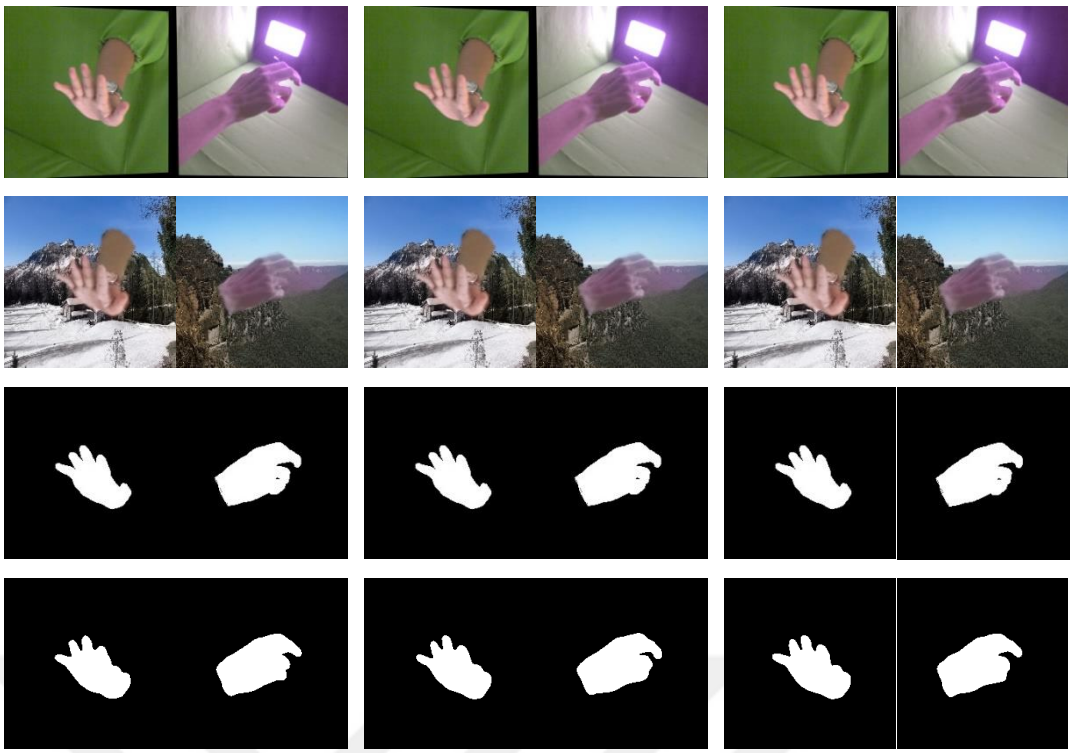
Tez çalışması kapsamında verilerin elde edilmesi ve bu verilen işlenmesi ile ortaya çıkan sonuçların görselleri, Şekil 8-9-10-11'deki gibi gösterilmiştir.



a **b** **c**

a) FPN – Resnet-34 b) FPN – Vision Transformer c) FPN – Efficientnet-b5 (en üstten başlayarak; yeşil arka plan görüntü, arka plan değiştirilmiş görüntü, gerçek maskeler, tahmin maskeleri)

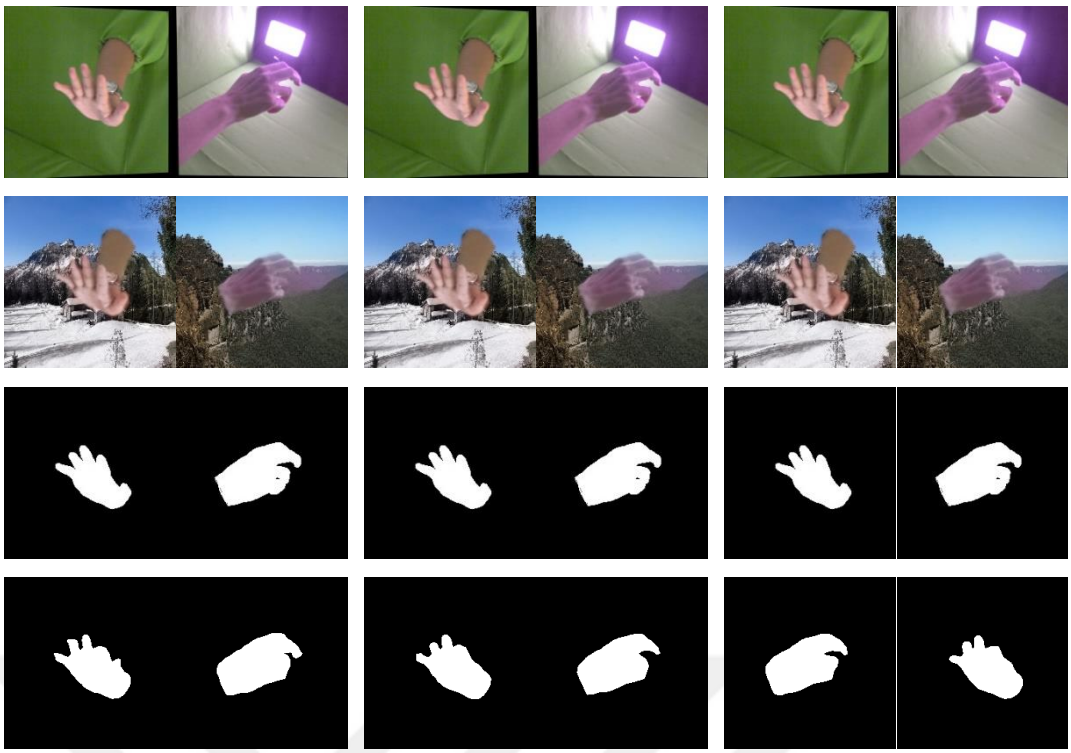
Şekil 15. FPN ve farklı omurgalardan elde edilen sonuçlar



a **b** **c**

a) MA-Net – EfficientNet-B5 b) MA-Net – Resnet-34 c) MA-Net – Vision Transformer (en üstten başlayarak; yeşil arka plan görüntü, arka plan değiştirilmiş görüntü, gerçek maskeler, tahmin maskeleri)

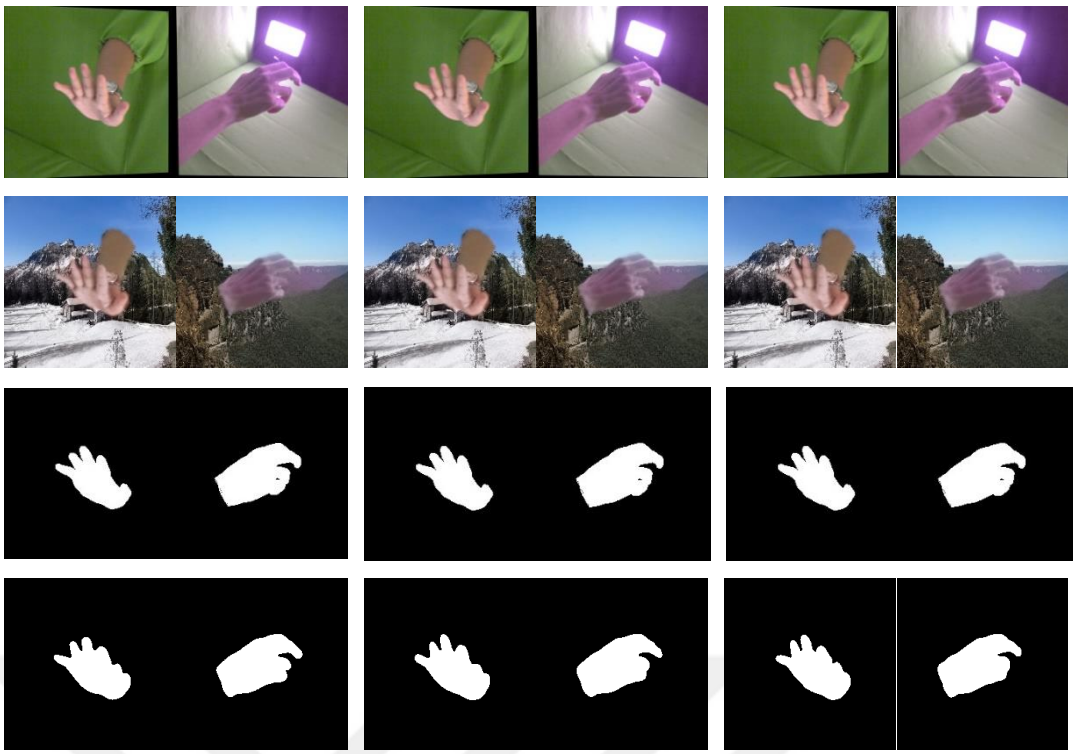
Şekil 16. MA-Net ve farklı omurgalardan elde edilen sonuçlar



a b c

a) PSP-Net – Efficientnet-B5 b) PSP-Net – Resnet-34 c) PSP-Net – Vision Transformer (en üstten başlayarak; yeşil arka plan görüntü, arka plan değiştirilmiş görüntü, gerçek maskeler, tahmin maskeleri)

Şekil 17. PSP-Net ve farklı omurgalardan elde edilen sonuçlar



a b c

a) U-Net – Efficientnet-B5 b) U-Net – Resnet-34 c) U-Net – Vision Transformer (en üstten başlayarak; yeşil arka plan görüntü, arka plan değiştirilmiş görüntü, gerçek maskeler, tahmin maskeleri)

Şekil 18. U-Net ve farklı omurgalardan elde edilen sonuçlar

5. SONUÇ

Bu çalışma, oto kodlayıcı tabanlı bölütleme modellerinin ve farklı omurga mimarilerinin el bölütleme görevindeki performansını kapsamlı bir şekilde değerlendirmeyi amaçlamıştır. Yapılan analizler, hem model seçimlerinin hem de kullanılan omurga mimarilerinin bölütleme başarımı üzerinde önemli etkileri olduğunu açıkça göstermektedir. U-Net modeli, Vision Transformer ile entegre edildiğinde, diğer tüm kombinasyonlara kıyasla en yüksek bölütleme başarımını elde etmiştir. Bu sonuç, hem doğruluk hem de bölütleme odaklı metrikler (IoU, F1 skoru, Dice katsayısı) açısından gözlemlenmiştir. Örneğin, U-Net + Vision Transformer kombinasyonu, 0.9892 doğruluk ve 0.8917 IoU skoruyla dikkat çekmiştir. Bu da, bu entegrasyonun yalnızca biyometrik tanıma sistemlerinde değil, aynı zamanda medikal görüntü işleme gibi diğer uygulamalarda da güçlü bir seçenek olabileceğini göstermektedir.

MA-Net, özellikle dikkat mekanizmalarının etkisiyle, Vision Transformer ve ResNet-34 kombinasyonlarında güçlü bir alternatif sunmuştur. Özellikle MA-Net + ResNet-34, 0.9921 doğruluk ve 0.9584 F1 skoru ile en iyi sonuçları vermiştir. Bu durum, dikkat mekanizmalarının belirli omurga mimarileriyle optimize edildiğinde önemli performans artışlarına yol açabileceğini ortaya koymaktadır. Bu kombinasyonun başarısı, düşük çözünürlüklü ya da gürültülü verilerde dahi dikkat mekanizmalarının etkinliğini kanıtlamaktadır. Diğer yandan, FPN ve PSP-Net modelleri, özellikle daha karmaşık bölütleme görevlerinde beklenenden düşük performans göstermiştir. Bu durum, bu modellerin hem omurga mimarisi kombinasyonlarına hem de veri setinin özelliklerine daha az uyum sağladığını göstermektedir. Örneğin, PSP-Net + EfficientNet-B5 kombinasyonu, yalnızca 0.9853 doğruluk ve 0.8579 IoU skoruna ulaşmıştır. Bu bulgu, FPN ve PSP-Net'in daha fazla optimizasyon gerektirdiğine işaret etmektedir.

Omurga düzeyinde yapılan karşılaştırmalarda, Vision Transformer, genel olarak en yüksek performansı sunan yapı olarak öne çıkmıştır. Hem MA-Net hem de U-Net ile olan kombinasyonlarında elde edilen yüksek doğruluk ve segmentasyon skorları, bu mimarinin el bölütleme gibi karmaşık görevlerde etkili bir çözüm olduğunu göstermektedir. Bunun yanı sıra, EfficientNet-B5, dengeli bir alternatif olarak öne çıkmış ve yüksek doğruluk ile düşük işlem maliyeti sunmuştur. Bu özellikleri, EfficientNet-B5'i pratik uygulamalar için cazip bir seçenek haline getirmektedir. Bu çalışmanın bulguları, el bölütleme gibi spesifik görevler için model ve omurga seçiminin dikkatli bir şekilde optimize edilmesi gerektiğini vurgulamaktadır. Örneğin, biyometrik tanıma sistemleri

bağlamında, U-Net + Vision Transformer ve MA-Net + ResNet-34 kombinasyonları, hem doğruluk hem de güvenilirlik açısından güçlü adaylar olarak değerlendirilebilir. Bu kombinasyonlar, avuç içi izi biyometrik sistemlerinde el bölütleme sürecini daha doğru ve verimli hale getirme potansiyeline sahiptir. Son olarak, gelecekteki çalışmalar, bu modellerin daha geniş veri setleri ve farklı uygulama senaryolarında test edilmesine odaklanabilir. Bu tür analizler, model performansının gerçek dünya koşullarında daha derinlemesine incelenmesine olanak sağlayacaktır. Aynı zamanda, yeni geliştirilecek omurga mimarileri ve model entegrasyonlarının mevcut yöntemlerle kıyaslanması, bölütleme teknolojisinin daha da ilerlemesine katkı sunabilir.

6. ÖNERİLER VE TARTIŞMA

Çalışmamızda MA-Net, FPN, U-Net ve PSP-Net gibi farklı oto kodlayıcı tabanlı modellerin, EfficientNet-B5, Vision Transformer ve ResNet-34 gibi omurga mimarileri ile birlikte el bölütleme performansına etkisi incelenmiştir. Elde edilen sonuçlar, Vision Transformer omurga mimarisinin genel olarak daha yüksek doğruluk ve IoU skorları sağladığını görülmüştür. Performansı daha da iyileştirmek adına, dikkat mekanizmalarının optimizasyonu, model birleştirme teknikleri ve hibrit model geliştirme gibi farklı stratejiler öneriyoruz. Bu stratejiler, el bölütleme alanında daha yüksek doğruluk ve verimlilik elde etmek için gelecekteki çalışmalara yol gösterebilir.

MA-Net ve PSP-Net gibi modellerin dikkat mekanizmalarını daha etkin kullanması için kanal ve mekânsal dikkat mekanizmaları iyileştirilebilir. Kanal dikkat mekanizmaları, SE (Squeeze-and-Excitation) blokları gibi yöntemlerle her kanalın önemini belirleyerek modelin ilgili özelliklere odaklanmasını sağlar. Mekânsal dikkat mekanizmaları ise, CBAM (Convolutional Block Attention Module) gibi yöntemlerle önemli konumlara ağırlık vererek modelin ilgili bölgelere odaklanmasını sağlar. Bu mekanizmaların iyileştirilmesi için daha gelişmiş dikkat modülleri kullanılabilir, dikkat mekanizmaları modelin farklı katmanlarına yerleştirilebilir ve parametreleri optimize edilebilir. Bu sayede modellerin özellik çıkarımı ve temsil yetenekleri geliştirilerek daha doğru ve güvenilir sonuçlar elde edilebilir.

FPN'nin çoklu ölçekli özellik çıkarma yeteneğini geliştirmek için, daha yüksek çözünürlüklü katman birleştirme teknikleri eklenebilir. Bu, daha fazla sayıda katmanı birleştirerek, deconvolution veya bilinear interpolation gibi gelişmiş yukarı örnekleme tekniklerini kullanarak, dikkat mekanizmalarıyla önemli özelliklere odaklanarak ve atlamalı bağlantılarla farklı ölçeklerdeki bilgileri birleştirerek elde edilebilir. Bu yöntemler, daha detaylı ve zengin özellik haritaları oluşturarak nesne tespiti ve görüntü bölütleme gibi görevlerin performansını artırır (örnek PANet (Path Aggregation Network)).

Kullanılan metriklerin genel değerlendirmesinde modellerin dice puanı yüksek ona kıyasla IoU puanın fazla düşük olduğu görülmektedir. Bunun olmasının nedenlerinden biri küçük nesne etkisi dediğimiz, el bölgesinin görüntünün büyük bir bölümünü kapsayan arka plana göre küçük kalması ve bu sebeble dice skoru kesişimleri 2 kat daha ağırlıklandırdığı için dice puanı daha yüksek çıkmıştır. IoU ise birleşim kümesi daha fazla odaklandığı için daha düşük çıkmıştır. Bir diğer sebeb yanlış pozitif ve yanlış negatif

piksel sayısı. Eğer tahmin maskesi gerçek el bölgesinden daha fazla pikseli kapsıyorsa veya maske gerçek el bölgesinden eksik pikseller içeriyorsa bu durumda dice değeri daha az etkilenecek, IoU değeri ise daha fazla etkilenecektir. Bir diğer sebeb sınır problemi durumu. Model elin sınırlarını doğru tahmin edemiyorsa yani sınırda eksik yada taşmalar oluyorsa bu durumda IoU değeri ciddi bir şekilde düşecektir. Dice değeri bu durumda daha az etkilecektir çünki kesişimleri iki kat ağırlandırdığı için sınır probleminden daha az etkilenir. Bu durumu iyileştirmek için IoU değerine odaklanan IoU Loss fonksiyonu yada model eğitiminde kullandığımız Tversky Loss fonksiyonunun α ve β parametreleri ile hatalı pozitif ve hatalı negatif ağırlıklar kontrol edilebilir. Çözüm için bir diğer yöntem olarak Focall Loss yada Weighted Cross-Entropy gibi yöntemler ile el piksellerine daha fazla ağırlık verilebilir. Model sonuçlarında doğruluk puanının çok yüksek olduğu fakat IoU değerinin daha düşük olduğu görülmektedir. Burada veride genel dengesizlik olduğunun belirtileridir. Zira sınıflar arasında dengesizlikten dolayı model çoğunluk sınıfını ki bizim çalışmamız da bu arkaplan olmaktadır, doğru tahmin ederek yüksek doğruluk elde etmiştir. Ancak IoU düşük kalmıştır çünkü model azınlık sınıfını veya daha küçük objeleri düzgün bir şekilde tanıyamamıştır. Sorunun çözümü için veri artırma yada sınıf ağırlıklarında değişikliğe gidilebilir. Bir diğer sebeb model nesnelerin kenarlarını hatalı bölütleme yapıyorsa yani nesne sınırlarını düzgün bir şekilde tanımıyorsa IoU değeri düşük, doğruluk yüksek olabilir. Sorunun çözümü için farklı mimariler yada veri ön işleme teknikleri kullanılabilir.

Modellerin güçlü yönlerini birleştirmek, derin öğrenme tabanlı uygulamalarda performansı artırmak için etkili bir stratejidir. Örneğin, U-Net ve MA-Net modellerinin birleşimi, her iki mimarinin avantajlarını kullanarak daha yüksek doğruluk sağlar. U-Net'in ayrıntılı encoder-decoder yapısı, küçük nesnelerin ve ince ayrıntıların yakalanmasında etkiliyken, MA-Net'in kanal ve mekânsal dikkat mekanizmaları, modelin kritik bölgelere odaklanma yeteneğini güçlendirir. Ayrıca, bu birleşimler hibrit dikkat mekanizmaları ve adaptif öğrenme yöntemleriyle geliştirilerek doğruluk ve verimliliği daha da artırabilir. (örnek attention U-net) Birleştirme modelleri, farklı omurga mimarileriyle eğitilmiş modellerin çıktılarını birleştirerek daha geniş bir genelleme kapasitesine ulaşmayı amaçlar. Bu yaklaşım, her omurga modelinin kendine özgü avantajlarından faydalanarak daha zengin özellik temsillerini ve daha doğru sonuçlar elde edilmesini sağlar. Örneğin, ResNet-34 gibi derin mimarilerden gelen yüksek seviyeli semantik bilgiler, EfficientNet-B5 gibi daha hafif ve optimize edilmiş mimarilerin düşük seviyeli ayrıntılarıyla birleşebilir. Böyle bir yapı, modelin hem ayrıntılı hem de küresel bağlam bilgisini daha iyi öğrenmesini sağlar. Böylece model,

farklı omurga ile eğitilmiş modellerin çıktılarını birleştirerek genelleme kapasitesini artırabilir.

Vision Transformer, bölütleme görevlerinde üstünlük sağlayarak dikkat mekanizmalarının el bölütleme gibi karmaşık görevlerde etkili olduğunu kanıtlamıştır. Özellikle U-Net ve MA-Net ile entegrasyonunda elde edilen 0.8917 IoU ve 0.9892 Dice skorları, nesne sınırlarının doğru bir şekilde belirlenmesini desteklemiştir. Transformer'ı daha etkili kullanmak için dikkat mekanizmalarına ince ayar yapılabilir. Özellikle önceden eğitilmiş Vision Transformer modellerinin bölütleme için yeniden eğitilmesi performansı artırabilir. Evrişimsel sinir ağı + Transformer Kombinasyonu**:** CvT (Convolutional Vision Transformer), hem evrişimsel sinir ağlarının verimliliğini hem de Vision Transformer 'ların yüksek performansını bir araya getiren hibrit bir görüntü tanıma modelidir. Geleneksel Vision Transformer 'lardan farklı olarak, CvT, görüntüleri örtüşmeyen parçalara ayırmak yerine konvolüsyonel katmanlar kullanarak token üretir. Bu sayede daha incelikli bir özellik çıkarımı ve daha iyi bir uzamsal bilgi korunumu sağlanır. Ayrıca, Transformer bloklarındaki tamamen bağlı katmanlar yerine konvolüsyonel katmanlar kullanarak hesaplama maliyetini düşürür. CvT, hiyerarşik yapısı sayesinde farklı ölçeklerde özellikler öğrenir ve daha iyi genelleme yeteneği gösterir. Bu iki yöntemin birleştirilmesi performansı artırabilir

EfficientNet-B5'in performansını artırmak için yapısal değişiklikler ve optimizasyonlar uygulanabilir. Modelin mevcut SE (Squeeze-and-Excitation) bloklarına CBAM gibi daha gelişmiş dikkat mekanizmaları eklenerek kanal ve mekânsal dikkati artırabilir; ayrıca Transformer tabanlı dikkat modülleriyle küresel bağlam bilgisi güçlendirilebilir. Daha derin veya daha geniş katmanlar ekleyerek modelin özellik çıkarma kapasitesi artırılabilir. FPN gibi hiyerarşik özellik birleştirme yöntemleriyle farklı seviyelerden daha zengin bilgi çıkarılabilir.

ResNet-34'ün daha derin varyantları (örneğin ResNet-50 veya ResNet-101) tercih edilebilir. ResNet-50, ResNet-34'te kullanılan temel yapı taşının (building block) daha karmaşık bir versiyonunu kullanır. ResNet-34, her artık blokta iki evrişimli katman kullanırken, ResNet-50, darboğaz adı verilen üç katmanlı bir yapı kullanır. Bu yapı, 1x1, 3x3 ve 1x1 evrişimli katmanlardan oluşur ve hesaplama maliyetini azaltırken performansı artırmaya yardımcı olur. Ayrıca ResNet-50 nin sahip olduğu daha fazla katman sayesinde daha fazla doğruluk elde edilebilir.

KAYNAKÇA

Abedalla, A., Abdullah, M., Al-Ayyoub, M. ve Benkhelifa, E. (2021). Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures. *PeerJ Computer Science*, 7, e607.

Abraham, N. ve Khan, N. M. (2019, April). A novel focal tversky loss function with improved attention u-net for lesion segmentation. *In 2019 IEEE 16th international symposium on biomedical imaging* (ISBI 2019) (pp. 683-687). IEEE.

Alausa, D. W., Adetiba, E., Badejo, J. A., Davidson, I. E., Obiyemi, O., Buraimoh, E., ... Oshin, O. (2022). Contactless palmprint recognition system: a survey. *IEEE Access*, 10, 132483-132505.

Alshakree, F., Akbas, A. ve Rahebi, J. (2024). Human identification using palm print images based on deep learning methods and gray wolf optimization algorithm. *Signal, Image and Video Processing*, 18(1), 961-973.

Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lucic, M., & Schmid, C. (2021). ViViT: A Video Vision Transformer. *In 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE.

Bambach, S., Lee, S., Crandall, D. J. ve Yu, C. (2015). Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. *In Proceedings of the IEEE international conference on computer vision* (pp. 1949-1957).

Bank, D., Koenigstein, N. ve Giryes, R. (2023). *Autoencoders*. Machine learning for data science handbook: data mining and knowledge discovery handbook, 353-374.

Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 17-36.

Bingöl, Ö. ve Ekinci, M. (2017). Stereo-based palmprint recognition in various 3D postures. *Expert Systems with Applications*, 78, 74-88.

Ebenezer, A. S., Kanmani, S. D., Sivakumar, M. ve Priya, S. J. (2022). Effect of image transformation on EfficientNet model for COVID-19 CT image classification. *Materials Today: Proceedings*, 51, 2512-2519.

Elharrouss, O., Akbari, Y., Almaadeed, N. ve Al-Maadeed, S. (2022). Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv*: 2206.08016.

Zhang, R., Du, L., Xiao, Q. ve Liu, J. (2020, May). Comparison of backbones for semantic segmentation network. *In Journal of Physics: Conference Series* (Vol. 1544, No. 1, p. 012196). IOP Publishing.

Nasalwai, N., Punn, N. S., Sonbhadra, S. K. ve Agarwal, S. (2021, May). Addressing the class imbalance problem in medical image segmentation via accelerated tversky loss function. *In Pacific-Asia conference on knowledge discovery and data mining* (pp. 390-402). Cham: Springer International Publishing.

Bojja, A. K., Mueller, F., Malireddi, S. R., Oberweger, M., Lepetit, V., Theobalt, C., ... Tagliasacchi, A. (2019, May). Handseg: An automatically labeled dataset for hand segmentation from depth images. *In 2019 16th conference on computer and robot vision* (CRV) (pp. 151-158). IEEE.

Cai, M., Lu, F. ve Sato, Y. (2020). Generalizing hand segmentation in egocentric videos with uncertainty-guided model adaptation. *In Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 14392-14401).

Cicek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T. ve Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 424-432. doi.org/10.1007/978-3-319-46723-8_49

Chai, T., Prasad, S., Yan, J. ve Zhang, Z. (2023). Contactless palmprint biometrics using DeepNet with dedicated assistant layers. *The Visual Computer*, 39(9), 4029-4047.

Chen, Z., Xie, L., Niu, J., Liu, X., Wei, L. ve Tian, Q. (2021). Visformer: The vision-friendly transformer. *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 589-598).

Csurka, G., Larlus, D., Perronnin, F. ve Meylan, F. (2013, September). What is a good evaluation measure for semantic segmentation?. *In Bmvc* (Vol. 27, No. 2013, pp. 10-5244).

Dargan, S. ve Kumar, M. (2020). A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. *Expert Systems with Applications*, 143, 113114.

Davis, J. ve Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233-240.

Demir, Y. ve Bingöl, Ö. (2021, June). Pneumonia detection from pediatric lung x-ray images with convolutional neural network method. *In 2021 29th Signal Processing and Communications Applications Conference (SIU)* (s. 1-4). IEEE.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. ve Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.

Devlin, J., Chang, M. W., Lee, K. ve Toutanova, K. (2018): BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv*:1810.04805.

Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv*:1606.05908.

Dogan, R. O., Dogan, H., Bayrak, C. ve Kayikcioglu, T. (2021). A two-phase approach using mask R-CNN and 3D U-Net for high-accuracy automatic segmentation of pancreas in CT imaging. *Computer Methods and Programs in Biomedicine*, 207, 106141.

Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv*:2010.11929.

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S. ve Pal, C. (2016). The importance of skip connections in biomedical image segmentation. *In Deep Learning and Data Labeling for Medical Applications* (pp. 179-187). Springer, Cham.

Ebenezer, A. S., Kanmani, S. D., Sivakumar, M. ve Priya, S. J. (2022). Effect of image transformation on EfficientNet model for COVID-19 CT image classification. *Materials Today: Proceedings*, 51, 2512-2519.

F. van Beers, A. Lindström, E. Okafor, and M. A. Wiering, "Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation", *In Proceedings of the 8th international conference on pattern recognition applications and methods* (pp. 438-445). SciTePress

Fan, T., Wang, G., Li, Y. ve Wang, H. (2020). Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8, 179656-179665.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.

Gao, Q., Liu, J. ve Ju, Z. (2020). Robust real-time hand detection and localization for space human–robot interaction based on deep learning. *Neurocomputing*, 390, 198-206.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V. ve Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv*:1704.06857.

Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep Learning*. MIT Press

Guo, Y., Liu, Y., Georgiou, T. ve Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval*, *7*, 87-93.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., ... Tao, D. (2022). A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(1), 87-110.

Han, Y., Sun, Z., Wang, F. ve Tan, T. (2007). Palmprint recognition under unconstrained scenes. In *Computer Vision–ACCV 2007: 8th Asian Conference on Computer Vision*, Tokyo, Japan, November 18-22, 2007, Proceedings, *Part II 8* (pp. 1-11). Springer Berlin Heidelberg.

He, H. ve Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.

Ho, Y. ve Wookey, S. (2019). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE access*, *8*, 4806-4813.

Hossin, M. ve Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining Knowledge Management Process*, 5(2), 1.

Huh, M., Agrawal, P. ve Efros, A. A. (2016). What makes imagenet good for transfer learning? *arXiv preprint arXiv*:1608.08614.

I. M. Alsaadi, "Study on most popular behavioral biometrics, advan-tages, disadvantages and recent applications: A review," *Int. J. Sci. Technol. Res*, 10(1), 2021.

Ikromjanov, K., Bhattacharjee, S., Sumon, R. I., Hwang, Y. B., Rahman, H., Lee, M. J., ... Choi, H. K. (2023). Region segmentation of whole-slide images for analyzing histological differentiation of prostate adenocarcinoma using ensemble efficientnetb2 u-net with transfer learning mechanism. *Cancers*, *15*(3), 762.

Jadon, S. (2020, October). A survey of loss functions for semantic segmentation. *In 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)* (pp. 1-7). IEEE.

Japkowicz, N. ve Shah, M. (2011). *Evaluating learning algorithms: a classification perspective*. Cambridge University Press.

Jia, W., Liu, M. ve Rehg, J. M. (2022, October). Generative adversarial network for future hand segmentation from egocentric video. *In European Conference on Computer Vision* (pp. 639-656). Cham: Springer Nature Switzerland.

Jia, W., Ren, Q., Zhao, Y., Li, S., Min, H. ve Chen, Y. (2022). EEPNet: An efficient and effective convolutional neural network for palmprint recognition. *Pattern Recognition Letters*, 159, 140-149.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S. ve Shah, M. (2022). Transformers in vision: A survey. *ACM computing surveys* (CSUR), 54(10s), 1-41.

Khan, F. S., Mohd, M. N. H., Soomro, D. M., Bagchi, S. ve Khan, M. D. (2021). 3D hand gestures segmentation and optimized classification using deep learning. *IEEE Access*, 9, 131614-131624.

Kingma, D. P. ve Welling, M. (2019). An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4), 307-392.

Kitrungrotsakul, T. ve Jirapatnakul, A. (2023). Evaluation Metrics for Medical Image Segmentation: A Comprehensive Review. *Diagnostics*, 13(16), 2689

Krizhevsky, A., Sutskever, I. ve Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B. ve Belongie, S. (2017). Feature pyramid networks for object detection. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

Lin, F., Price, B. ve Martinez, T. (2020). Ego2hands: A dataset for egocentric two-hand segmentation and detection. *arXiv preprint arXiv*:2011.07252.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).

Lu, H., Fu, Y., Feng, J., Knight, T., Xiao, Y. ve Yang, L. (2017). Deep learning for semantic segmentation of high-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), 4168-4183.

Ma, S., Hu, Q., Zhao, S., Chen, S. ve Jiang, L. (2024). SYEnet: Simple yet effective network for palmprint recognition. *Information Sciences*, 669, 120518.

Mahasin, M. ve Dewi, I. A. (2022). Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 backbones on YOLO v4 as object detector. *International journal of engineering, science and information technology*, 2(3), 64-72.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I. ve Frey, B. (2015). Adversarial autoencoders. *arXiv preprint arXiv*:1511.05644.

Minaee, S., Abdolrashidi, A., Su, H., Bennamoun, M. ve Zhang, D. (2023). Biometrics recognition using deep learning: A survey. *Artificial Intelligence Review*, 56(8), 8647-8695.

Montalbo, F. J. P., Hernandez, L. R. T., Palad, L. P., Castillo, R. C., Alon, A. S. ve De Ocampo, A. L. P. (2023, October). Performance Analysis of Lightweight Vision Transformers and Deep Convolutional Neural Networks in Detecting Brain Tumors in MRI Scans: An Empirical Approach. *In Proceedings of the 2023 8th International Conference on Biomedical Imaging, Signal Processing* (pp. 17-25).

Tomar, N. K., Jha, D., Riegler, M. A., Johansen, H. D., Johansen, D., Rittscher, J., Ali, S. (2022). Fanet: A feedback attention network for improved biomedical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(11), 9375-9388.

Neven, D., Brabandere, B. D., Proesmans, M. ve Van Gool, L. (2018). Towards end-to-end lane detection: an instance segmentation approach. *In 2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 286-291). IEEE.

Neverova, N., Wolf, C., Taylor, G. W. ve Nebout, F. (2015). Hand segmentation with structured convolutional learning. In Computer Vision--ACCV 2014: *12th Asian Conference on Computer Vision*, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part III 12 (pp. 687-702). Springer International Publishing.

Ovi, T. B., Bashree, N., Mukherjee, P., Mosharrof, S. ve Parthima, M. A. (2023, July). Performance Analysis of Various EfficientNet-Based U-Net++ Architecture for Automatic Building Extraction from High Resolution Satellite Images. *In International Conference on Human-Centric Smart Computing* (pp. 385-399). Singapore: Springer Nature Singapore.

Padilla, R., Netto, S. L. ve Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. *In 2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237-242). IEEE.

Pan, S. J. ve Yang, Q. (2010): A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359

Parvaiz, A., Khalid, M. A., Zafar, R., Ameer, H., Ali, M. ve Fraz, M. M. (2023). Vision Transformers in medical computer vision—A contemplative retrospection. *Engineering Applications of Artificial Intelligence*, 122, 106126.

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

Pinaya, W. H. L., Vieira, S., Garcia-Dias, R. ve Mechelli, A. (2020*). Autoencoders. In Machine learning* (pp. 193-208). Academic Press.

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv*:2010.16061.

Priesnitz, J., Rathgeb, C., Buchmann, N. ve Busch, C. (2021). Deep learning-based semantic segmentation for touchless fingerprint recognition. *In Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January* 10-15, 2021, Proceedings, Part VIII (pp. 154-168). Springer International Publishing.

Provost, F. ve Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51-59.

Raghu, M., Zhang, C., Kleinberg, J. ve Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, 32.

Rahman, M. A. ve Wang, Y. (2018). Optimizing intersection-over-union in deep neural networks for image segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 284-294).

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I. ve Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 658-666).

Ronneberger, O., Fischer, P. ve Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: *18th international conference*, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

Ruby, U. ve Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng*, *9*(10).

Salehi, S. S. M., Erdogmus, D. ve Gholipour, A. (2017, September). Tversky loss function for image segmentation using 3D fully convolutional deep networks. *In International workshop on machine learning in medical imaging* (pp. 379-387). Cham: Springer International Publishing.

Sevim, Z., Dogan, H., Demir, Z., Sezen, F. S. ve Dogan, R. O. (2023, June). An Extensive Study: Creation of A New Inverted Microscope Image Data Set and Improving Auto-Encoder Models for Higher Accuracy Segmentation of HaCaT Cell Culture Line. *In 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1-6). IEEE.

Sokolova, M. ve Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing Management*, 45(4), 427-437.

Taha, A. A. ve Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1), 1-28.

Tan, P. N., Steinbach, M. ve Kumar, V. (2005). *Introduction to data mining*. Boston, MA: Pearson Education.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. ve Liu, C. (2018): A survey on deep transfer learning. *International Conference on Artificial Neural Networks*.

Tan, M. ve Le, Q. V. (2019): EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the International Conference on Machine Learning*

Tan, M. ve Le, Q. V. Efficientnetv2: Smaller models and faster training. *arXiv 2021*. *arXiv preprint arXiv*:2104.00298.

Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A. ve Chavez-Urbiola, E. A. (2023). Loss functions and metrics in deep learning. A review. *arXiv preprint arXiv*:2307.02694.

Trabelsi, S., Samai, D., Dornaika, F., Benlamoudi, A., Bensid, K. ve Taleb-Ahmed, A. (2022). Efficient palmprint biometric identification systems using deep learning and feature selection methods. *Neural Computing and Applications*, 34(14), 12119-12141.

Urooj, A. ve Borji, A. (2018). Analysis of hand segmentation in the wild. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4710-4719).

Varior, R. R., Shuai, B., Tighe, J. ve Modolo, D. (2019). Multi-scale attention network for crowd counting. *arXiv preprint arXiv*:1901.06026.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998-6008.

Vijayan, M. (2023). A regression-based approach to diabetic retinopathy diagnosis using efficientnet. *Diagnostics*, 13(4), 774.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P. A. ve Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11, (pp. 3371-3408).

Vincent, P., Larochelle, H., Bengio, Y. ve Manzagol, P. A. (2008, July). Extracting and composing robust features with denoising autoencoders. *In Proceedings of the 25th international conference on Machine learning* (pp. 1096-1103).

Vodopivec, T., Lepetit, V. ve Peer, P. (2016). Fine hand segmentation using convolutional neural networks. *arXiv preprint arXiv*:1608.07454.

Wang, X., Girshick, R., Gupta, A. ve He, K. (2018). Non-local neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7794-7803.

Wang, Z., Chen, J. ve Hoi, S. C. (2024). Multi-scale attention network for single image super-resolution. *IEEE Transactions on Image Processing*, 33, 1321-1334.

Wu, J., Zhou, S., Zuo, S., Chen, Y., Sun, W., Luo, J., ... Wang, D. (2021). U-Net combined with multi-scale attention mechanism for liver segmentation in CT images. *BMC Medical Informatics and Decision Making*, 21, 1-12.

Zhu, X., Cheng, Z., Wang, S., Chen, X., & Lu, G. (2021). Coronary angiography image segmentation based on PSPNet. *Computer Methods and Programs in Biomedicine*, *200*, 105897.

Yang, W., Wang, S., Hu, J., Zheng, G. ve Valli, C. (2019). Security and accuracy of fingerprint-based biometrics: A review. *Symmetry*, 11(2), 141.

Yosinski, J., Clune, J., Bengio, Y. ve Lipson, H. (2014): How transferable are features in deep neural networks?. *Advances in Neural Information Processing Systems*, 27.

Yuan, Y., Huang, L., Guo, J., Zhang, C., Chen, X. ve Wang, J. (2018). Ocnet: Object context network for scene parsing. *arXiv preprint arXiv*:1809.00916.

Zhao, H., Shi, J., Qi, X., Wang, X. ve Jia, J. (2017). Pyramid scene parsing network. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881-2890).

Zhu, M., Tang, Y. ve Han, K. (2021). Vision transformer pruning. *arXiv preprint arXiv*:2104.08500.

Zimmermann, C., Ceylan, D., Yang, J., Russell, B., Argus, M. ve Brox, T. (2019). Freihand: A dataset for markerless capture of hand pose and shape from single rgb images. *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 813-822).

Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D. ve Chen, H. (2018). Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection. *In International conference on learning representations*.

ÖZGEÇMİŞ

İlkokulu Balıkçıl ilköğretim okulunda okudu. Ortaokul ve liseyi Elbistan Gazi Mustafa Kemal Lisesinde okudu. 2021 yılında Gümüşhane Üniversitesi Elektrik Elektronik Mühendisliği bölümünden mezun oldu. Halen Gümüşhane Üniversitesi yapı işleri ve teknik daire başkanlığında elektrik elektronik mühendisi olarak görev yapmaktadır.