|  |  |
| --- | --- |
|  | **T.C.**  **KIRŞEHİR AHİ EVRAN ÜNİVERSİTESİ**  **MÜHENDİSLİK MİMARLIK FAKÜLTESİ**  **BİLGİSAYAR MÜHENDİLİĞİ**  **LİSANS** |

**DeepLabV3+ Tabanlı Çoklu Backbone Kullanarak Görüntü Segmentasyonu**

**Öğrenci No 212511039**

**Öğrenci Adı Soyadı Ayşe BAŞ**

**LİSANS / DERS PROJESİ**

**KIRŞEHİR**

**2026**

# İÇİNDEKİLER DİZİNİ

**Sayfa No**

[İÇİNDEKİLER DİZİNİ II](#_Toc219566076)

[ÖZET III](#_Toc219566077)

[TABLOLAR DİZİNİ IV](#_Toc219566078)

[ŞEKİLLER DİZİNİ V](#_Toc219566079)

[SİMGELER VE KISALTMALAR DİZİNİ VI](#_Toc219566080)

[1. GİRİŞ 7](#_Toc219566081)

[1.1. Amaç 8](#_Toc219566082)

[2. ÖNCEKİ ÇALIŞMALAR 9](#_Toc219566083)

[2.1. DeepLabV3+ 9](#_Toc219566085)

[2.2. Deep Residual Learning (ResNet) 9](#_Toc219566086)

[2.3. MobileNet (Howard et al., 2017) 10](#_Toc219566087)

[3. MATERYAL VE METOT 11](#_Toc219566088)

[3.1. Materyal 11](#_Toc219566090)

[3.1.1. Donanım 11](#_Toc219566091)

[3.1.2. Yazılım ve Kütüphaneler 11](#_Toc219566092)

[3.1.3. Kayıtlar ve Çıktılar 12](#_Toc219566093)

[3.1.4. Veri Seti 13](#_Toc219566094)

[3.2. Metot 14](#_Toc219566095)

[3.2.1. Veri Ön İşleme ve Bölme Yöntemi 14](#_Toc219566096)

[3.2.2. Model Mimarisi 14](#_Toc219566097)

[3.2.3. Eğitim Süreci 15](#_Toc219566098)

[3.2.4. Eğitim Geri Çağırmaları (Callbacks) 16](#_Toc219566099)

[4. BULGULAR VE TARTIŞMA 17](#_Toc219566100)

[4.1. Modelin Eğitim ve Doğrulama Sonuçları 17](#_Toc219566102)

[4.2. Test Verisi Üzerindeki Değerlendirme 18](#_Toc219566103)

[5. SONUÇ VE ÖNERİLER 20](#_Toc219566104)

[KAYNAKLAR 22](#_Toc219566105)

[EKLER 23](#_Toc219566106)

[**EK A. 30 Epoch Eğitim Sonuçları** 23](#_Toc219566107)

# ÖZET

Bu çalışmada, derin öğrenme tabanlı semantik segmentasyon yaklaşımları kullanılarak insan segmentasyonu problemi ele alınmıştır. Çalışmada, Kaggle platformundan elde edilen People Segmentation veri seti kullanılmış ve görüntüler eğitim, doğrulama ve test kümeleri olacak şekilde yapılandırılmıştır. Segmentasyon sürecinde, güncel ve yüksek performanslı bir mimari olan DeepLabV3+ tercih edilmiş; farklı encoder (backbone) mimarileri yapılandırmalarının (ResNet-50, ResNet-101 ve MobileNetV2) model başarımı üzerindeki etkileri karşılaştırmalı olarak incelenmiştir.

Model eğitimi PyTorch ve Segmentation Models PyTorch kütüphaneleri kullanılarak gerçekleştirilmiştir. Eğitim aşamasında, sınıf dengesizliği ve sınır hassasiyetini iyileştirmek amacıyla Dice Loss ve Focal Binary Cross Entropy kayıplarının birleşiminden oluşan özel bir kayıp fonksiyonu kullanılmıştır. Model performansı; Intersection over Union (IoU), doğruluk, hassasiyet ve F1-skoru gibi yaygın değerlendirme metrikleri üzerinden analiz edilmiştir. Deneysel sonuçlar, ResNet tabanlı encoderların, insan segmentasyonu probleminde daha istikrarlı ve yüksek doğruluk sağlayarak diğer encoder yapılarına kıyasla üstün performans sergilediğini ortaya koymuştur.

**Anahtar Kelimeler:**  Derin Öğrenme, Semantik Segmentasyon, İnsan Segmentasyonu, DeepLabV3+, ResNet

# TABLOLAR DİZİNİ

**Sayfa No**

[Tablo 1 Backbone Karşılaştırması 17](#_Toc219566176)

.

.

.

# ŞEKİLLER DİZİNİ

**Sayfa No**

[Şekil 1 DeepLabV3+ mimarisinin genel yapısı [3]. 9](#_Toc219566237)

[Şekil 2 MobileNetV2 tabanlı DeepLabV3+ modelinin 50 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi. 12](#_Toc219566238)

[Şekil 3 ResNet50 tabanlı DeepLabV3+ modelinin 50 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi. 13](#_Toc219566239)

[Şekil 4 ResNet101 tabanlı DeepLabV3+ modelinin 50 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi. 13](#_Toc219566240)

[Şekil 5 Orjinal Veri Seti 14](#_Toc219566241)

[Şekil 6 Akış Şeması 16](#_Toc219566242)

[Şekil 7 resnet50 30 epoch 18](#_Toc219566243)

[Şekil 8 resnet50 50 epoch 19](#_Toc219566244)

[Şekil 9 resnet101 30 epoch 19](#_Toc219566245)

[Şekil 10 resnet101 50 epoch 19](#_Toc219566246)

[Şekil 11 MobileNetV2 30 epoch 20](#_Toc219566247)

[Şekil 12 MobileNetV2 50 epoch 20](#_Toc219566248)

[Şekil 13 MobileNetV2 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi. 23](#_Toc219566249)

[Şekil 14 ResNet50 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi. 24](#_Toc219566250)

[Şekil 15 ResNet101 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi. 24](#_Toc219566251)

[Şekil 16 split\_dataset.py 25](#_Toc219566252)

[Şekil 17 model.py 26](#_Toc219566253)

[Şekil 18 train.py 27](#_Toc219566254)

[Şekil 19 test.py 28](#_Toc219566255)

# SİMGELER VE KISALTMALAR DİZİNİ

|  |  |  |
| --- | --- | --- |
| **Simgeler** |  | **Açıklama** |
| Dice | **:** | Dice Katsayısı |
| IoU | **:** | Kesişim Bölü Birlik (Intersection over Union) |
| TP | **:** | Doğru Pozitif (True Positive) |
| FP | **:** | Yanlış Pozitif (False Positive) |
| FN | **:** | Yanlış Negatif (False Negative) |
| TN | **:** | Doğru Negatif (True Negative) |

|  |  |  |
| --- | --- | --- |
| **Kısaltmalar** |  | **Açıklama** |
| **AI** | **:** | Yapay Zekâ (Artificial Intelligence) |
| **DL** | **:** | Derin Öğrenme (Deep Learning) |
| **CNN** | **:** | Evrişimli Sinir Ağı (Convolutional Neural Network) |
| **ResNet** | **:** | Residual Network (Artık Bağlantılı Omurga Mimarisi) |
| **GPU** | **:** | Grafik İşlem Birimi (Graphics Processing Unit) |
| **BCE** | **:** | Binary Cross Entropy |

# GİRİŞ

Günümüzde bilgisayarla görü (Computer Vision) alanındaki hızlı gelişmeler, özellikle derin öğrenme (Deep Learning) temelli yaklaşımların yaygınlaşmasıyla birlikte önemli bir ivme kazanmıştır. Görüntülerden anlamlı bilgi çıkarılması, nesnelerin algılanması ve sınıflandırılması gibi problemler; sağlık, güvenlik, otonom sistemler ve insan-bilgisayar etkileşimi başta olmak üzere birçok alanda kritik uygulamalara sahiptir [1]. Bu bağlamda, bir görüntü içerisindeki her bir pikselin ait olduğu sınıfa göre etiketlenmesini amaçlayan anlamsal segmentasyon (semantic segmentation) problemi, bilgisayarla görü alanının temel araştırma konularından biri hâline gelmiştir [2].

Anlamsal segmentasyonun önemli alt problemlerinden biri olan insan segmentasyonu, görüntülerde yer alan insan bölgelerinin arka plandan ayrıştırılmasını hedeflemektedir. İnsan segmentasyonu; video gözetim sistemleri, akıllı güvenlik uygulamaları, artırılmış gerçeklik, otonom araçlar ve sağlık teknolojileri gibi birçok alanda yaygın olarak kullanılmaktadır [2]. Ancak insan vücudunun farklı pozisyonlar alabilmesi, karmaşık arka planlar, aydınlatma koşullarındaki değişkenlikler ve nesneler arası örtüşmeler gibi faktörler, bu problemin çözümünü zorlaştırmaktadır.

Son yıllarda, Evrişimli Sinir Ağları (Convolutional Neural Networks – CNN) tabanlı derin öğrenme modelleri, segmentasyon problemlerinde geleneksel görüntü işleme yöntemlerine kıyasla önemli başarılar elde etmiştir. Özellikle Fully Convolutional Networks (FCN), U-Net ve DeepLab ailesi gibi mimariler, piksel seviyesinde yüksek doğruluk oranları sunarak literatürde öne çıkmaktadır [3]. Bu mimariler arasında yer alan DeepLabV3+, atrous (delikli) evrişim yapısı ve Atrous Spatial Pyramid Pooling (ASPP) mekanizması sayesinde çok ölçekli bağlamsal bilgiyi etkin bir biçimde kullanabilmekte ve ayrıntılı segmentasyon sonuçları üretebilmektedir [3].

DeepLabV3+ mimarisinde kullanılan farklı encoder (backbone) yapıları, modelin öğrenme kapasitesi ve hesaplama maliyeti üzerinde doğrudan etkili olmaktadır. Özellikle ResNet tabanlı omurgalar, derin ağ yapıları sayesinde daha zengin öznitelik temsilleri sunarken [4]; MobileNet gibi hafif mimariler, düşük hesaplama maliyeti ile gerçek zamanlı uygulamalar için avantaj sağlamaktadır [5].

Bu çalışmada, insan segmentasyonu problemi için DeepLabV3+ mimarisi temel alınarak farklı encoder mimarilerinin model performansı üzerindeki etkisi incelenmiştir. Çalışma kapsamında, Kaggle platformundan temin edilen People Segmentation veri seti kullanılarak; veri ön işleme, model eğitimi, doğrulama ve test aşamaları uçtan uca bir sistem yapısı içerisinde gerçekleştirilmiştir [6]. Farklı omurga mimarileri ve eğitim parametreleri ile yapılan deneyler sonucunda elde edilen performans çıktıları karşılaştırmalı olarak analiz edilerek, insan segmentasyonu problemi için en uygun yapıların belirlenmesi amaçlanmıştır.

## Amaç

Bu çalışmanın temel amacı, derin öğrenme tabanlı anlamsal segmentasyon yöntemleri kullanılarak insan segmentasyonu problemini etkili bir şekilde çözebilen bir model geliştirmek ve farklı derin ağ mimarilerinin performanslarını karşılaştırmalı olarak incelemektir. Bu doğrultuda, DeepLabV3+ mimarisi temel alınarak farklı encoder (backbone) mimarileri yapıların insan segmentasyonu üzerindeki etkileri analiz edilmiştir.

Çalışma kapsamında, Kaggle platformundan temin edilen People Segmentation veri seti kullanılarak, veri setinin eğitim, doğrulama ve test alt kümelerine ayrılması, uygun veri ön işleme ve veri artırma tekniklerinin uygulanması amaçlanmıştır. Ayrıca, geliştirilen modelin genelleme yeteneğini artırmak için farklı eğitim parametreleri, epoch sayıları ve kayıp fonksiyonları denenmiştir.

Bu çalışmanın bir diğer amacı, Dice Loss ve Focal Binary Cross-Entropy kayıplarının birleşiminden oluşan hibrit bir kayıp fonksiyonunun, insan segmentasyonu performansı üzerindeki etkisini incelemektir. Modelin başarımı; doğruluk (Accuracy), kesişim-birlik oranı (IoU), kesinlik (Precision), geri çağırma (Recall), F1-Skoru ve özgüllük (Specificity) gibi performans metrikleri kullanılarak değerlendirilmiştir.

Son olarak, elde edilen deneysel sonuçlar doğrultusunda, kullanılan mimarilerin güçlü ve zayıf yönlerinin ortaya konulması ve insan segmentasyonu problemleri için uygulanabilir, ölçeklenebilir ve güvenilir bir derin öğrenme yaklaşımının sunulması hedeflenmektedir.

# ÖNCEKİ ÇALIŞMALAR



## DeepLabV3+

DeepLabV3+, Google araştırma ekibi tarafından geliştirilen ve anlamsal segmentasyon problemleri için yüksek performans sunan bir derin öğrenme mimarisidir [3]. Bu yapı, önceki DeepLab sürümlerinin geliştirilmiş bir versiyonu olup, özellikle atrous (delikli) evrişim ve Atrous Spatial Pyramid Pooling (ASPP) mekanizmaları sayesinde çok ölçekli bağlamsal bilgiyi etkili bir şekilde yakalayabilmektedir.

DeepLabV3+ mimarisi, kodlayıcı–çözücü (encoder–decoder) yapısı üzerine kuruludur. Kodlayıcı bölümünde, giriş görüntüsünden yüksek seviyeli öznitelikler çıkarılırken, çözücü kısmında bu öznitelikler daha yüksek uzamsal çözünürlüğe sahip segmentasyon haritalarına dönüştürülmektedir. ASPP modülü, farklı atrous oranları kullanarak çoklu ölçeklerde özellik çıkarımı yapılmasını sağlamakta ve bu sayede nesnelerin farklı boyutlardaki temsilleri daha başarılı bir şekilde öğrenilebilmektedir [3].

Literatürde yapılan birçok çalışmada, DeepLabV3+ mimarisinin insan segmentasyonu, yol algılama, tıbbi görüntü analizi ve uydu görüntü segmentasyonu gibi çeşitli uygulamalarda yüksek doğruluk sağladığı gösterilmiştir [3]. Özellikle karmaşık arka planlara sahip görüntülerde, bu mimarinin bağlamsal bilgiyi etkili kullanması önemli bir avantaj sunmaktadır.

.

metin, diyagram, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

**Şekil 1 DeepLabV3+ mimarisinin genel yapısı [3].**

## Deep Residual Learning (ResNet)

Derin Artık Öğrenme (Deep Residual Learning) yaklaşımı, çok derin sinir ağlarının eğitiminde karşılaşılan gradyan kaybolması (vanishing gradient) problemini çözmek amacıyla geliştirilmiştir [4]. He ve arkadaşları tarafından önerilen ResNet mimarisi, ağ katmanları arasına eklenen artık (residual) bağlantılar sayesinde, daha derin ağların etkin bir şekilde eğitilmesine olanak tanımaktadır.

ResNet mimarisinin temel fikri, bir katmanın öğrenmesi gereken fonksiyonun doğrudan giriş yerine, giriş ile çıkış arasındaki fark (residual) üzerinden modellenmesidir. Bu yapı sayesinde, ağ derinliği arttıkça performansın düşmesi yerine, daha anlamlı özniteliklerin öğrenilmesi mümkün hâle gelmiştir. ResNet50 ve ResNet101 gibi farklı derinliklere sahip varyantlar, segmentasyon ve sınıflandırma problemlerinde yaygın olarak kullanılmaktadır [4].

Anlamsal segmentasyon mimarilerinde ResNet yapıları genellikle encoder (backbone) olarak tercih edilmektedir. DeepLabV3+ gibi modellerde ResNet omurgasının kullanılması, derin ve zengin özniteliklerin çıkarılmasını sağlayarak segmentasyon başarımını artırmaktadır. Literatürdeki çalışmalar, daha derin ResNet varyantlarının genellikle daha yüksek doğruluk sunduğunu, ancak bunun hesaplama maliyeti ve eğitim süresi açısından bir denge gerektirdiğini göstermektedir [4].

## MobileNet (Howard et al., 2017)

MobileNet, hesaplama maliyeti ve bellek gereksinimleri düşük olacak şekilde tasarlanmış, hafif ve verimli bir derin öğrenme mimarisidir [5]. Özellikle mobil cihazlar, gömülü sistemler ve sınırlı donanım kaynaklarına sahip ortamlar için geliştirilmiştir. MobileNet mimarisinin temel özelliği, klasik evrişim işlemleri yerine derinlik ayrılabilir evrişimler (depthwise separable convolutions) kullanmasıdır.

Derinlik ayrılabilir evrişim yapısı, standart evrişim işlemini iki aşamaya ayırmaktadır. İlk aşamada her kanal için ayrı ayrı derinlik evrişimi (depthwise convolution) uygulanırken, ikinci aşamada 1×1 noktasal evrişim (pointwise convolution) ile kanallar birleştirilmektedir. Bu yaklaşım, hesaplama yükünü ve parametre sayısını önemli ölçüde azaltırken, kabul edilebilir bir doğruluk seviyesinin korunmasını sağlamaktadır [5].

MobileNet mimarisinin farklı sürümleri bulunmakta olup, bu çalışmada MobileNetV2 yapısı kullanılmıştır. MobileNetV2, ters artık bloklar (inverted residual blocks) ve doğrusal daraltma (linear bottleneck) yapıları sayesinde, daha az parametre ile daha derin ve etkili özellik temsilleri sunmaktadır. Bu özellikler, MobileNetV2’yi anlamsal segmentasyon gibi piksel düzeyinde hassasiyet gerektiren görevler için uygun bir encoder mimarisi hâline getirmektedir [5].

Literatürde yapılan çalışmalar, MobileNet tabanlı modellerin ResNet gibi daha ağır mimarilere kıyasla daha düşük hesaplama maliyetine sahip olduğunu, ancak buna rağmen rekabetçi segmentasyon performansları sunabildiğini göstermektedir [5]. Bu nedenle MobileNetV2, bu çalışmada DeepLabV3+ mimarisi içerisinde hafif bir kodlayıcı alternatifi olarak değerlendirilmiş ve performansı diğer omurga mimarileri ile karşılaştırmalı olarak incelenmiştir.

# MATERYAL VE METOT

Araştırma sürecinde kullanılan materyaller ile insan segmentasyonu problemi için izlenen yöntem adım adım açıklanmış, eğitim ve değerlendirme sürecinde kullanılan teknikler ayrıntılı olarak ele alınmıştır.



## Materyal

Bu çalışmada deneyler, Apple M2 işlemciye sahip bir macOS tabanlı kişisel bilgisayar üzerinde gerçekleştirilmiştir. Derin öğrenme modellerinin eğitimi sırasında, Apple Metal Performance Shaders (MPS) desteği kullanılarak GPU hızlandırmasından faydalanılmıştır. Bu sayede eğitim süreci daha verimli hâle getirilmiştir.

Kullanılan donanım altyapısı, orta ölçekli veri setleri üzerinde çoklu deneylerin gerçekleştirilmesine ve farklı model mimarilerinin karşılaştırılmasına olanak sağlamıştır.

### Donanım

Bu çalışmada deneyler, Apple M2 işlemciye sahip bir macOS tabanlı kişisel bilgisayar üzerinde gerçekleştirilmiştir. Derin öğrenme modellerinin eğitimi sırasında, Apple Metal Performance Shaders (MPS) desteği kullanılarak GPU hızlandırmasından faydalanılmıştır. Bu sayede eğitim süreci daha verimli hâle getirilmiştir.

Kullanılan donanım altyapısı, orta ölçekli veri setleri üzerinde çoklu deneylerin gerçekleştirilmesine ve farklı model mimarilerinin karşılaştırılmasına olanak sağlamıştır.

Bu donanım altyapısı, farklı omurga mimarileri ve epoch konfigürasyonları için çoklu deneylerin yürütülmesine olanak sağlamıştır.

### Yazılım ve Kütüphaneler

Çalışmada Python programlama dili kullanılmıştır. Derin öğrenme modellerinin geliştirilmesi ve eğitimi için PyTorch framework’ü tercih edilmiştir. Anlamsal segmentasyon mimarilerinin oluşturulmasında Segmentation Models PyTorch (SMP) kütüphanesinden yararlanılmıştır.

Veri artırma ve ön işleme aşamalarında Albumentations kütüphanesi kullanılmış, görüntü işlemleri için OpenCV (cv2) tercih edilmiştir. Eğitim sürecinde performans takibi, grafik çizimleri ve sonuçların kaydedilmesi amacıyla NumPy, Pandas ve Matplotlib kütüphanelerinden faydalanılmıştır. Kod geliştirme ve deneylerin yürütülmesi PyCharm geliştirme ortamında gerçekleştirilmiştir.

### Kayıtlar ve Çıktılar

Model eğitimi sürecinde elde edilen kayıp (loss), doğruluk (accuracy) ve IoU değerleri her epoch sonunda kaydedilerek CSV dosyaları hâlinde saklanmıştır. Bu kayıtlar, eğitim sürecinin nicel olarak izlenebilmesi ve deneylerin karşılaştırılabilirliğinin sağlanması amacıyla oluşturulmuştur. Ayrıca, eğitim ve doğrulama kümelerine ait kayıp, doğruluk ve IoU değişimleri grafikler hâlinde görselleştirilmiştir.

Şekil 2–4’te, farklı kodlayıcı mimariler kullanılarak 50 epoch boyunca eğitilen modellerin eğitim ve doğrulama IoU eğrileri gösterilmektedir. Bu grafikler, modellerin öğrenme davranışlarını, yakınsama hızlarını ve genelleme performanslarını karşılaştırmalı olarak analiz etmeye imkân tanımaktadır.

Test aşamasında ise, her bir test görüntüsü için model tarafından üretilen segmentasyon maskeleri görselleştirilmiş, ayrıca IoU, Dice/F1, doğruluk ve hassasiyet (precision/recall) gibi performans metrikleri ayrı CSV dosyalarında saklanarak analiz edilmiştir.

Ayrıca, 30 epoch boyunca gerçekleştirilen eğitimlere ait IoU eğrileri **Ekler bölümünde** sunulmuş olup, modellerin erken öğrenme davranışlarının incelenmesi amacıyla kullanılmıştır.

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram, metin içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil MobileNetV2 tabanlı DeepLabV3+ modelinin 50 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi.

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil ResNet50 tabanlı DeepLabV3+ modelinin 50 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi.

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram, metin içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil ResNet101 tabanlı DeepLabV3+ modelinin 50 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi.

### Veri Seti

Bu çalışmada, Kaggle platformundan elde edilen People Segmentation veri seti kullanılmıştır. Veri seti, insan içeren renkli görüntüler ve bu görüntülere karşılık gelen ikili (binary) maske etiketlerinden oluşmaktadır. Orijinal veri seti içerisinde görüntüler, maskeler ve kolajlardan oluşan farklı klasör yapıları bulunmaktadır.

Çalışma kapsamında veri seti yeniden düzenlenmiş ve new\_data adlı bir klasör yapısı oluşturulmuştur. Görüntüler ve maskeler, eğitim (train), doğrulama (validation) ve test olmak üzere üç alt kümeye ayrılmıştır. Bu işlem, veri bütünlüğünü koruyacak şekilde, her görüntünün kendi maskesi ile eşleşmesini sağlayan özel bir Python betiği kullanılarak gerçekleştirilmiştir.

giyim, ayakkabı, kişi, şahıs, sandalye içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.heykel, siluet, siyah beyaz, sanat içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil Orjinal Veri Seti

## Metot

Bu bölümde, insan segmentasyonu problemi için izlenen yöntem, model mimarisi, eğitim süreci ve değerlendirme yaklaşımı açıklanmaktadır.

### Veri Ön İşleme ve Bölme Yöntemi

Model eğitiminden önce, veri seti rastgele karıştırılarak eğitim, doğrulama ve test alt kümelerine ayrılmıştır. Veri bölme işlemi %80 eğitim, %10 doğrulama ve %10 test oranlarında gerçekleştirilmiştir. Tekrarlanabilir sonuçlar elde edebilmek amacıyla sabit bir rastgelelik tohumu (random seed) kullanılmıştır.

Eğitim sürecinde modelin genelleme yeteneğini artırmak amacıyla veri artırma teknikleri uygulanmıştır. Bu kapsamda, yatay çevirme, parlaklık ve kontrast değişimi gibi işlemler kullanılmıştır. Tüm görüntüler, model girişine uygun olacak şekilde yeniden boyutlandırılmış ve normalize edilmiştir.

### Model Mimarisi

Bu çalışmada, insan segmentasyonu problemi için DeepLabV3**+** mimarisi temel alınmıştır. DeepLabV3+, kodlayıcı–çözücü (encoder–decoder) yapısı sayesinde hem yüksek seviyeli anlamsal bilgiyi hem de uzamsal detayları etkili bir şekilde öğrenebilen bir anlamsal segmentasyon modelidir. Modelin kodlayıcı (encoder) kısmında, giriş görüntüsünden çok ölçekli öznitelik çıkarımı yapılırken, çözücü (decoder) kısmında bu öznitelikler daha yüksek çözünürlüklü segmentasyon haritalarına dönüştürülmektedir.

Kodlayıcı bölümünde farklı omurga (backbone) mimarilerinin performansa etkisini incelemek amacıyla ResNet50, ResNet101 ve MobileNetV2 olmak üzere üç farklı yapı kullanılmıştır. ResNet tabanlı mimariler, artık (residual) bağlantılar sayesinde derin ağların daha etkili şekilde eğitilmesine olanak tanırken, MobileNetV2 mimarisi derinlik ayrılabilir evrişimler ile daha düşük hesaplama maliyeti sunmaktadır.

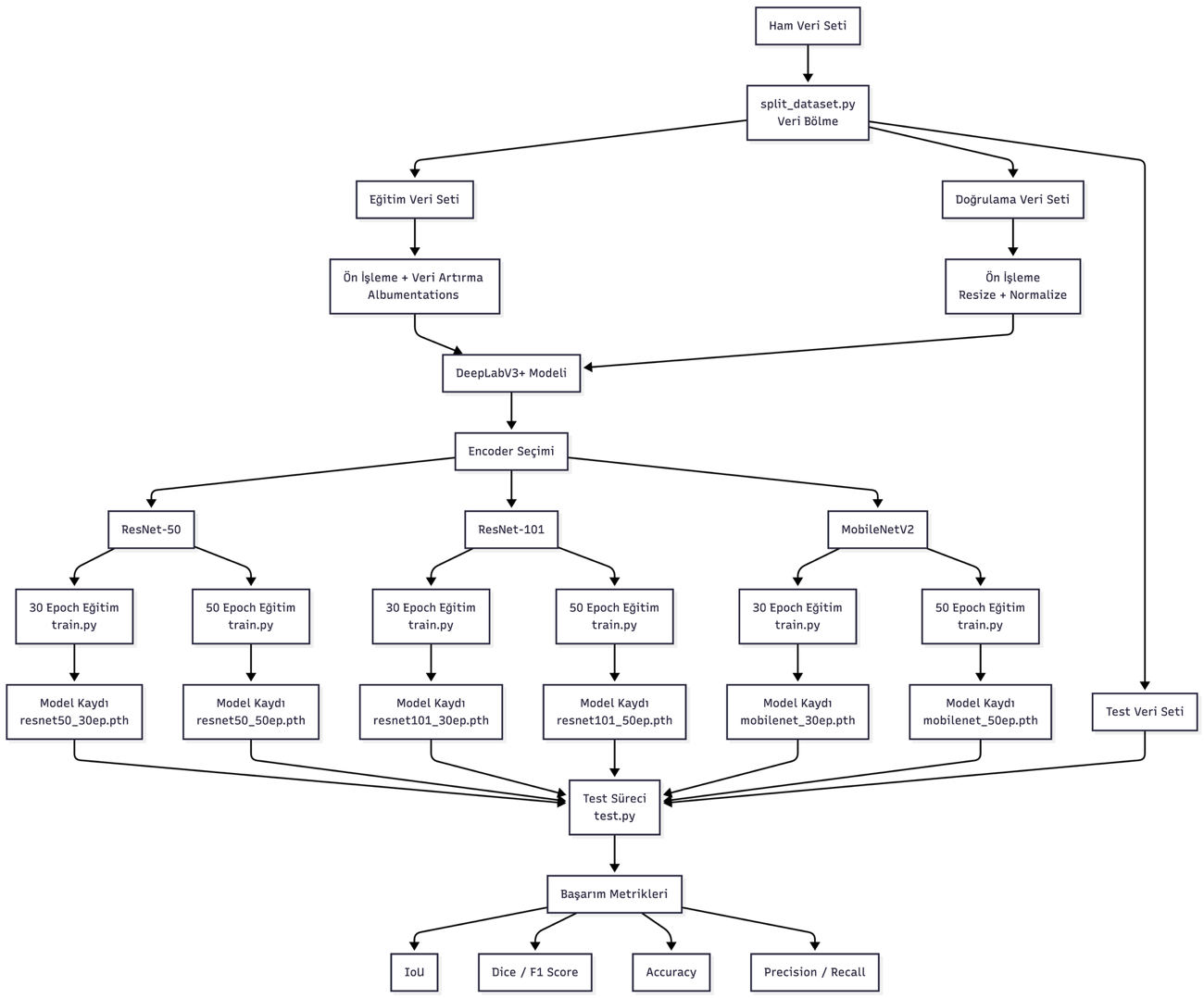
DeepLabV3+ mimarisinde yer alan Atrous Spatial Pyramid Pooling (ASPP) modülü, farklı atrous oranları ile çok ölçekli bağlamsal bilgilerin yakalanmasını sağlamaktadır. Bu sayede, görüntü içerisindeki insan nesneleri farklı boyut ve şekillerde olsa dahi daha başarılı bir segmentasyon elde edilmektedir. Modelin çıkış katmanında tek kanallı (binary) bir maske üretilmiş ve piksel seviyesinde insan–arka plan ayrımı gerçekleştirilmiştir.

### Eğitim Süreci

Model eğitimi sürecinde, giriş görüntüleri ve bunlara karşılık gelen maske etiketleri kullanılarak denetimli öğrenme yaklaşımı benimsenmiştir. Eğitim öncesinde tüm görüntüler, model girişine uygun olacak şekilde 256×256 boyutuna yeniden ölçeklendirilmiş ve ImageNet istatistiklerine göre normalize edilmiştir. Ayrıca, modelin genelleme yeteneğini artırmak amacıyla eğitim verileri üzerinde çeşitli veri artırma (data augmentation) teknikleri uygulanmıştır.

Eğitim süreci, farklı epoch sayıları (30 ve 50 epoch) ile farklı kodlayıcı (backbone) mimarileri kullanılarak ayrı ayrı gerçekleştirilmiştir. Her bir deneyde, eğitim ve doğrulama kümeleri üzerindeki kayıp değerleri ve performans metrikleri düzenli olarak izlenmiştir. Kayıp fonksiyonu olarak, Dice Loss ve Focal Binary Cross-Entropy kayıplarının birleşiminden oluşan Combined Loss fonksiyonu kullanılmıştır. Bu tercih, sınıf dengesizliği problemini azaltmayı ve segmentasyon sınırlarının daha hassas öğrenilmesini amaçlamaktadır.

Optimizasyon aşamasında AdamW optimizasyon algoritması kullanılmış ve öğrenme oranı eğitim sürecine bağlı olarak dinamik şekilde güncellenmiştir. Eğitim sırasında, doğrulama veri seti üzerinde en yüksek IoU değerini elde eden model ağırlıkları en iyi model (best model) olarak kaydedilmiştir. Eğitim süreci boyunca kayıp, doğruluk ve IoU değerleri kayıt altına alınmış ve analiz amacıyla grafikler hâlinde görselleştirilmiştir.



Şekil Akış Şeması

### Eğitim Geri Çağırmaları (Callbacks)

Model eğitim sürecinin daha verimli ve kararlı bir şekilde ilerlemesi amacıyla çeşitli geri çağırma (callback) mekanizmaları kullanılmıştır. Bu kapsamda, doğrulama kaybına bağlı olarak öğrenme oranını azaltan ReduceLROnPlateau yöntemi uygulanmıştır. Doğrulama kaybında belirli bir süre iyileşme gözlenmediğinde, öğrenme oranı otomatik olarak düşürülerek modelin daha hassas öğrenmesi sağlanmıştır.

Ayrıca, modelin aşırı öğrenmesini (overfitting) önlemek amacıyla erken durdurma (early stopping) stratejisi kullanılmıştır. Doğrulama IoU değerinde belirli bir sayıda epoch boyunca iyileşme görülmediğinde eğitim süreci sonlandırılmıştır. Bu sayede, gereksiz epoch’ların önüne geçilmiş ve eğitim süresi optimize edilmiştir.

Eğitim süreci boyunca en iyi performansı gösteren model ağırlıkları otomatik olarak kaydedilmiş ve test aşamasında bu modeller kullanılmıştır. Bu geri çağırma mekanizmaları, modelin hem daha kararlı bir şekilde eğitilmesine hem de genelleme performansının artırılmasına katkı sağlamıştır.

# BULGULAR VE TARTIŞMA

Bu bölümde, DeepLabV3+ mimarisi kullanılarak gerçekleştirilen deneyler sonucunda elde edilen performans çıktıları sunulmuş ve analiz edilmiştir. Eğitim, doğrulama ve test aşamalarında elde edilen sonuçlar tablo ve grafikler yardımıyla değerlendirilmiş, farklı encoder (backbone) mimarilerinin model başarımı üzerindeki etkileri karşılaştırmalı olarak incelenmiştir.



## Modelin Eğitim ve Doğrulama Sonuçları

Model eğitimi sürecinde, farklı epoch sayıları (30 ve 50) ve farklı kodlayıcı mimariler (ResNet50, ResNet101 ve MobileNetV2) kullanılarak çoklu deneyler gerçekleştirilmiştir. Her bir deney için eğitim ve doğrulama kayıpları, doğruluk (Accuracy) ve kesişim-birlik oranı (IoU) değerleri epoch bazında kaydedilmiş ve grafikler hâlinde sunulmuştur.

Elde edilen eğitim eğrileri incelendiğinde, tüm mimarilerde epoch sayısının artmasıyla birlikte eğitim kaybının azaldığı ve doğrulama performansının genel olarak arttığı gözlemlenmiştir. Özellikle ResNet101 omurgasının kullanıldığı modellerde doğrulama IoU değerlerinin daha istikrarlı bir artış gösterdiği görülmüştür. MobileNetV2 tabanlı modeller ise daha düşük parametre sayısına sahip olmasına rağmen rekabetçi sonuçlar üretmiş ve daha kısa sürede yakınsama sağlamıştır.

Eğitim ve doğrulama sonuçları, kullanılan geri çağırma mekanizmalarının (öğrenme oranı azaltma ve erken durdurma) modelin aşırı öğrenmesini büyük ölçüde engellediğini göstermektedir.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Encoder (Backbone)** | **Derinlik** | **Parametre Sayısı** | **Hesaplama Maliyeti** | **Çalışma Hızı** | **Doğruluk Potansiyeli** | **Kullanım Amacı** |
| **ResNet-101** | Çok derin | Çok yüksek | Çok yüksek | Düşük | Çok yüksek | Yüksek doğruluk gerektiren, güçlü donanımlı sistemler |
| **ResNet-50** | Derin | Yüksek | Yüksek | Orta | Yüksek | Doğruluk ve performans dengesi gereken uygulamalar |
| **MobileNetV2** | Sığ | Düşük | Düşük | Yüksek | Orta | Gerçek zamanlı, düşük maliyetli ve mobil uygulamalar |

Tablo Backbone Karşılaştırması

## Test Verisi Üzerindeki Değerlendirme

Eğitim süreci sonunda, her bir deney için doğrulama kümesinde en yüksek IoU değerini elde eden modeller seçilmiş ve test veri kümesi üzerinde değerlendirilmiştir. Test aşamasında; doğruluk (Accuracy), kesinlik (Precision), geri çağırma (Recall), F1-Skoru, özgüllük (Specificity) ve IoU metrikleri kullanılmıştır.

Test sonuçları incelendiğinde, ResNet101 tabanlı DeepLabV3+ modelinin genel olarak en yüksek IoU ve F1-Skoru değerlerini elde ettiği görülmüştür. ResNet50 mimarisi, hesaplama maliyeti ile performans arasında dengeli bir yapı sunarken, MobileNetV2 mimarisi daha hafif bir model olmasına rağmen tatmin edici sonuçlar üretmiştir. Bu durum, MobileNetV2’nin sınırlı donanım kaynaklarına sahip sistemlerde uygulanabilir bir alternatif olduğunu göstermektedir.

Ayrıca, test aşamasında elde edilen görsel çıktılar incelendiğinde, modellerin insan siluetlerini büyük oranda doğru bir şekilde ayırt edebildiği, ancak karmaşık arka planlar ve ince detaylar içeren görüntülerde kısmi hataların oluşabildiği gözlemlenmiştir. Buna rağmen, genel olarak DeepLabV3+ mimarisinin insan segmentasyonu problemi için güçlü ve güvenilir sonuçlar sunduğu söylenebilir.

1. **Test Sonuçlarına Ait Görsel Değerlendirme**

ResNet50 omurgası kullanılan DeepLabV3+ modelinin **test veri kümesinden seçilen örnek görüntüler** için, **30 ve 50 epoch ile eğitilmiş modellerin segmentasyon sonuçları** gösterilmektedir. (Şekil 8 ) 30 epoch eğitim sonunda elde edilen modelin test çıktıları, (Şekil 9) 50 epoch eğitim sonunda elde edilen modelin test çıktıları sunulmuştur. Görseller incelendiğinde, epoch sayısının artmasıyla birlikte segmentasyon sınırlarının daha belirgin hâle geldiği ve maske bütünlüğünde iyileşmeler olduğu gözlemlenmektedir.

ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil resnet50 ****30 epoch****

ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil resnet50 50 ****epoch****

ResNet101 tabanlı DeepLabV3+ modelinin **test veri kümesinden seçilen örnek görüntüler** için segmentasyon sonuçları gösterilmektedir. (Şekil 10) **30 epoch ile eğitilmiş modelin test çıktıları**, (Şekil 11) **50 epoch ile eğitilmiş modelin test çıktıları** sunulmuştur. ResNet101 mimarisinin daha derin yapısı sayesinde özellikle karmaşık arka planlara sahip görüntülerde daha tutarlı ve kararlı segmentasyon sonuçları ürettiği gözlemlenmektedir.

ekran görüntüsü, dans içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil resnet101 30 ****epoch****

ekran görüntüsü, dans içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil resnet101 50 ****epoch****

MobileNetV2 omurgası kullanılan DeepLabV3+ modelinin **test veri kümesi üzerindeki segmentasyon sonuçları** sunulmaktadır. (Şekil 12) **30 epoch ile eğitilmiş modelin test çıktıları**, (Şekil 13) **50 epoch ile eğitilmiş modelin test çıktıları** gösterilmektedir. MobileNetV2 mimarisinin daha düşük parametre sayısına sahip olmasına rağmen, insan segmentasyonu görevinde kabul edilebilir düzeyde performans sergilediği görülmektedir.

ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil MobileNetV2 30 ****epoch****

ekran görüntüsü, dans içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil MobileNetV2 50 ****epoch****

# SONUÇ VE ÖNERİLER

Bu çalışmada, insan segmentasyonu problemi için DeepLabV3+ mimarisi kullanılarak farklı encoder (backbone) mimarileri ağ yapılarının model performansı üzerindeki etkileri incelenmiştir. ResNet50, ResNet101 ve MobileNetV2 omurgaları ile oluşturulan modeller; eğitim, doğrulama ve test veri kümeleri üzerinde değerlendirilmiştir. Performans ölçütü olarak doğruluk (Accuracy), IoU, F1-Skoru, kesinlik (Precision) ve geri çağırma (Recall) metrikleri kullanılmıştır.

Elde edilen sonuçlar, DeepLabV3+ mimarisinin insan segmentasyonu görevinde başarılı ve kararlı sonuçlar ürettiğini göstermektedir. Özellikle ResNet101 tabanlı model, yüksek özellik çıkarım kapasitesi sayesinde en yüksek IoU ve F1-Skoru değerlerine ulaşarak en başarılı performansı sergilemiştir. ResNet50 mimarisi, hesaplama maliyeti ve başarı oranı arasında dengeli bir yapı sunmuş ve pratik uygulamalar için uygun bir alternatif olduğunu göstermiştir. MobileNetV2 tabanlı model ise daha az parametreye sahip olmasına rağmen kabul edilebilir başarı düzeyi sunarak, sınırlı donanım kaynaklarına sahip sistemler için uygulanabilir bir çözüm olduğunu ortaya koymuştur.

Çalışma kapsamında uygulanan veri ön işleme adımları ve eğitim sürecinde kullanılan geri çağırma mekanizmaları (erken durdurma ve öğrenme oranı azaltma), modelin aşırı öğrenmesini önlemiş ve genelleme başarımını artırmıştır. Eğitim ve doğrulama eğrileri, modelin istikrarlı bir şekilde öğrenme gerçekleştirdiğini ve epoch sayısının artırılmasıyla birlikte performansın belirli bir noktadan sonra doygunluğa ulaştığını göstermektedir.

Gelecek çalışmalar kapsamında, daha büyük ve çeşitli veri kümeleri kullanılarak modelin genelleme kabiliyetinin artırılması hedeflenebilir. Ayrıca, farklı veri artırma tekniklerinin uygulanması, attention tabanlı mekanizmaların entegrasyonu ve transformer tabanlı segmentasyon modelleriyle karşılaştırmalı analizler yapılması önerilmektedir. Gerçek zamanlı uygulamalar için MobileNet tabanlı modellerin donanım hızlandırıcılar (GPU, TPU) üzerinde optimize edilmesi de önemli bir geliştirme alanı olarak değerlendirilebilir.

Sonuç olarak, bu çalışma DeepLabV3+ mimarisinin insan segmentasyonu probleminde etkili bir yöntem olduğunu ortaya koymuş; farklı encoder (backbone) mimari yapıların performans, hesaplama maliyeti ve uygulama alanları açısından sunduğu avantajları karşılaştırmalı olarak göstermiştir.

# KAYNAKLAR

[1] I. Goodfellow, Y. Bengio ve A. Courville, *Deep Learning*. Cambridge, MA, ABD: MIT Press, 2016.

[2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn ve A. Zisserman, “The Pascal Visual Object Classes Challenge: A retrospective,” *International Journal of Computer Vision*, c. 111, s. 1, ss. 98–136, 2015.

[3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff ve H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” *Proceedings of the European Conference on Computer Vision (ECCV)*, ss. 801–818, 2018.

[4] K. He, X. Zhang, S. Ren ve J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ss. 770–778, 2016.

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto ve H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint*, arXiv:1704.04861, 2017.

[6] Kaggle, “People segmentation dataset,” 2023. [Çevrimiçi]. Erişim tarihi: 2025. https://www.kaggle.com/datasets/tapakah68/segmentation-full-body-mads-dataset

[7] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan *vd*., “PyTorch: An imperative style, high-performance deep learning library,” *Advances in Neural Information Processing Systems (NeurIPS)*, c. 32, ss. 8024–8035, 2019.

[8] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean *vd*., “TensorFlow: A system for large-scale machine learning,” *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, ss. 265–283, 2016.

# EKLER

**EK A. 30 Epoch Eğitim Sonuçları**

Bu ekte, **MobileNetV2, ResNet50 ve ResNet101** kodlayıcı mimarileri kullanılarak **30 epoch boyunca eğitilen modellere ait eğitim ve doğrulama IoU eğrileri** sunulmaktadır. Bu grafikler, modellerin erken epoch’lardaki öğrenme davranışlarının ve yakınsama eğilimlerinin incelenmesi amacıyla verilmiştir..

**Şekil A.1.** MobileNetV2 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerleri.  
**Şekil A.2.** ResNet50 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerleri.  
**Şekil A.3.** ResNet101 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerleri.

öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi, diyagram içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil MobileNetV2 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi.

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil ResNet50 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi.

çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram, metin içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil ResNet101 tabanlı DeepLabV3+ modelinin 30 epoch boyunca elde edilen eğitim ve doğrulama IoU değerlerinin değişimi.

**split\_dataset.py**

metin, ekran görüntüsü, yazılım içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

**Şekil 16 split\_dataset.py**

**model.py**

metin, ekran görüntüsü, yazı tipi, yazılım içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

**Şekil 17 model.py**

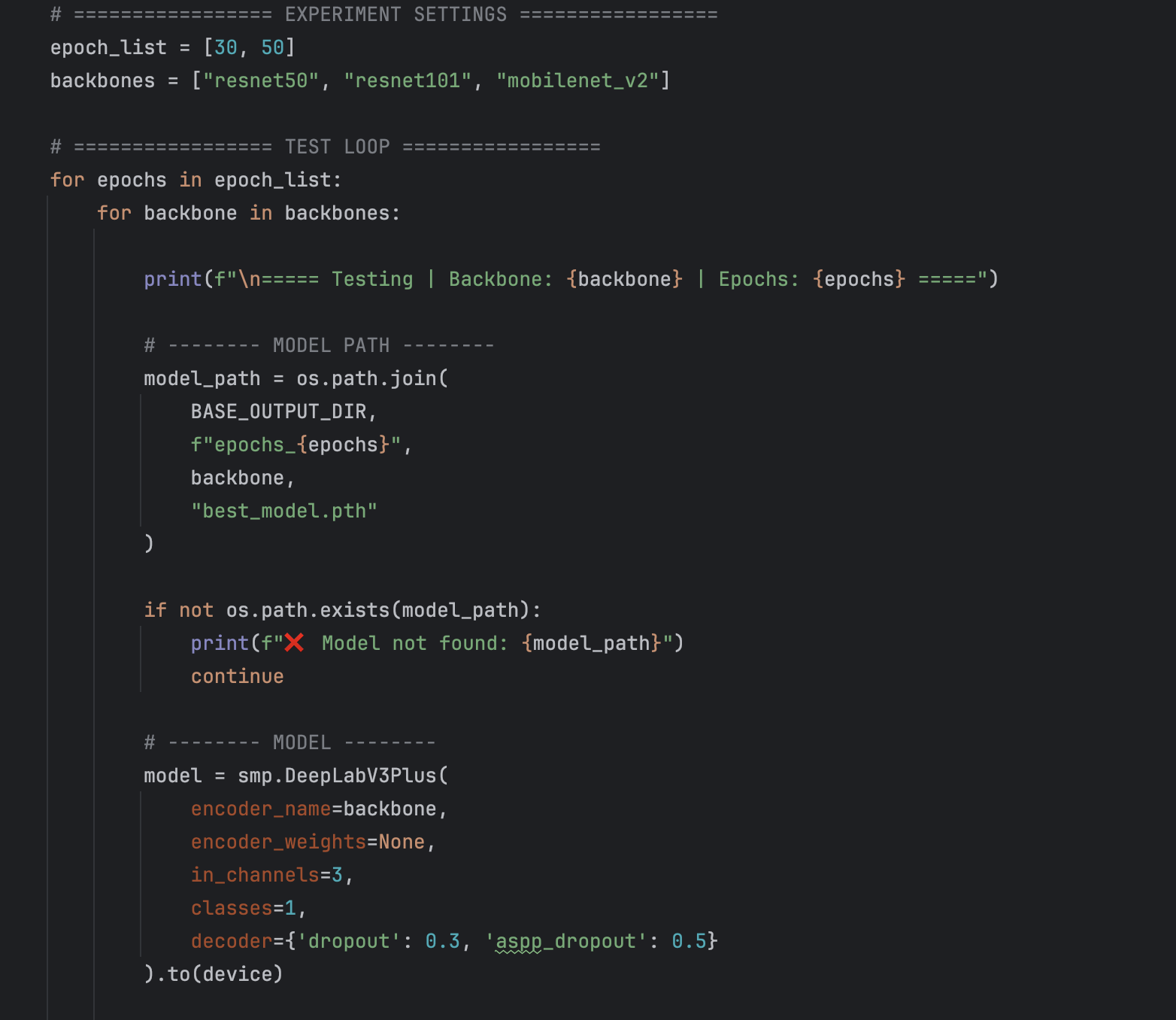
**train.py**

metin, ekran görüntüsü, yazılım içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

**Şekil 18 train.py**

**test.py**



**Şekil 19 test.py**