

LinkNet Mimarisi ile Görüntü Bölütleme

Şilan Fidan Vural
Bilgisayar Mühendisliği
Yıldız Teknik Üniversitesi
fsilanvural@gmail.com

Abstract—Bilgisayarlı görü (computer vision) yapay zekanın alt alanlarından biri olup kapsamına görüntü sınıflandırma, nesne tespiti ve görüntü bölütleme gibi görevler girmektedir. Bu çalışmada görüntü bölütleme yöntemlerinden bir olan semantik segmentasyon (semantic segmentation) üzerinde durulmuştur. En temel işlemlerden olan görüntü sınıflandırmada bütün resim tek bir etikete atanırken görüntü bölütlemeye ise her bir piksele bir etiket atanmaktadır. Böylece semantik segmentasyon ile görüntü hakkında daha fazla bilgi elde edilmesi sağlanmış olur. Bu çalışmada LinkNet mimarisinin implementasyonu ile Oxford-IIIT Pet veri seti üzerinde görüntü bölütleme işlemi gerçekleştirilmiştir. Bu işlem sonucunda test seti üzerinde %85 oranında bir başarı elde edilmiştir.

Index Terms—görüntü bölütleme, linkNet, semantik segmentasyon

I. GİRİŞ

Bu çalışmada LinkNet [1] mimarisi kullanılarak Oxford-IIIT Pet veri seti üzerinde bir semantik segmentasyon işlemi gerçekleştirilmiştir. Kullanılan Oxford-IIIT Pet veri seti içerisinde 37 tane farklı sınıftan oluşan evcil hayvan görselleri, hayvanların etiketleri ve piksel bazında maskeleri içermektedir. Maskeler, her piksel için olan etiketlerdir ve her piksele 3 kategoriden biri verilmektedir. Bu üç kategori evcil hayvana ait piksel, evcil hayvanı çevreleyen piksel ve bunların dışında kalan kısımdan oluşmaktadır.

II. YÖNTEM

Görüntü bölütleme işlemine geçmeden önce veri seti hazır hale getirilmiştir. Veri seti içerisindeki görseller normalize edilerek her resim belirli bir boyuta getirilmiştir. Veri seti eğitim(train), doğrulama(validation) ve test olmak üzere 3'e ayrılmıştır. Bu işlemlerin tamamlanmasının ardından model kısmına geçilmiştir. Görüntü bölütleme gerçekleştirebilmek için kullanılan LinkNet [1] mimarisi encoder decoder şeklindeki bir derin öğrenme modelidir. Encoder yapısı görüntüyü alarak yüksek boyutlu özellik vektörü oluşturur. Decoder yapısı ise yüksek boyutlu özellik vektörünü alarak bir semantik segmentasyon maskesi oluşturur. Bu mimarinin adımları giriş katmanı, encoder blokları, decoder blokları ve çıkış katmanı olmak üzere 4 kısımdan oluşmaktadır. Mimari Fig. 1'de gösterilmektedir.

Makale incelendiğinde giriş katmanı filtre sayısı 64, kernel büyüklüğü 7x7 ve stride değeri 2 olan bir konvolüsyon katmanından, devamında ise kernel büyüklüğü 3x3 ve stride değeri 2 olan bir max-pooling katmanından oluşmaktadır. Encoder kısmında geldiğimizde art arda 4 adet encoder katmanı bulunmaktadır. Her bir encoder katmanı filtre sayısı değişken

ancak filtre büyüklüğü 3x3 olan 4 adet konvolüsyon katmanının art arda gelmesinden oluşmaktadır. Encoder katmanı residual bir bloktur. Residual bloktan kastedilen bir katman çıktısının daha derindeki bir başka katmana eklenmesidir. Bu encoder yapısında da 2 konvolüsyon işleminden önceki katman çıktısı sonraki katman çıktısına eklenir.

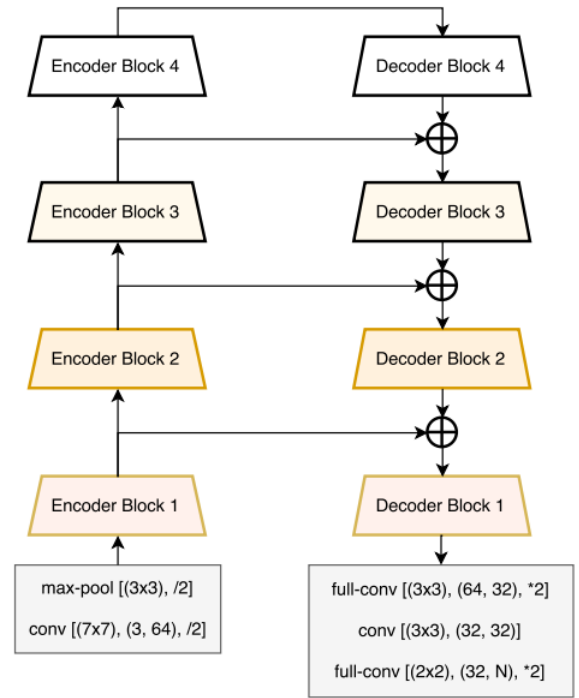


Fig. 1. LinkNet mimarisi

Encoder katmanlarının ardından 4 adet decoder katmanı gelmektedir. Her bir decoder katmanı içerisinde sırasıyla konvolüsyon, transpose konvolüsyon ve tekrar konvolüsyon işlemleri gerçekleştirilmektedir. Konvolüsyon katmanları içerisindeki kernel büyüklüğü 1x1 iken transpose katmanındaki kernel büyüklüğü 3x3'tür. Decoder katmanlarının ardından son katman olan çıkış katmanı gelmektedir. Çıkış katmanı ise sırasıyla transpose konvolüsyon, konvolüsyon ve tekrar transpose konvolüsyon aşamalarından oluşmaktadır. Son transpose konvolüsyon katmanının içerisine girdi olarak segmentasyon sınıf sayısı verilmektedir. LinkNet mimarisinin daha detaylı incelemesi orijinal makaleden [1] incelenebilir.

Genel mimari yapı bu şekilde olup en iyi modelin be-

lirlenmesi için girdi görüntü büyüklüğü ve hiperparametreler üzerinde değişiklikler yapılmıştır. Her model 20 epoch eğitilmiş olup optimizer olarak "adam" ve kayıp fonksiyonu olarak da "categorical cross entropy" kullanılmıştır. İlk aşamada aktivasyon fonksiyonu olarak ReLu kullanılmıştır. Bunun nedeni Relu fonksiyonunun literatürde en çok tercih edilen aktivasyonlardan biri olmasıdır. Bunun sonucunda elde edilen modelde doğrulama(validation) başarıları %82 olarak elde edilmiştir. Literatürde inceleniğinde ELU aktivasyon fonksiyonunun ReLu'ya göre yavaş çalıştığı ancak daha çabuk yakınsadığı ve daha genelleştirilmiş bir performansa sahip olduğu görülmüştür. Bu nedenle ikinci modelde ELU aktivasyon fonksiyonu kullanılmıştır. Bunun sonucunda yine yaklaşık olarak %82'lik bir başarı elde edilmiştir. Bir başka modelde ise aktivasyon fonksiyonu olarak yine ReLu kullanılırken görüntü boyutu 128'den 256'ya çıkarılmıştır. Bu durumda doğrulama başarıları %85 civarına ulaşmıştır. Bu model en iyi model olarak seçilerek test seti üzerinden test edilip %85.56 başarı elde edilmiştir.

Eğitim aşamasında kullanılan platform Google Colab olmuştur. GPU ya da TPU desteği sağlaması nedeniyle tercih edilmiştir.

III. UYGULAMA

Bu kısımda elde edilen test başarısına göre metriklerin yorumlanması gerçekleştirilmiştir. İlk olarak Fig. 2'de train seti için başarı(accuracy) ve kayıp(loss) değerleri eğitim boyunca takip edilmiştir. Test setinin accuracy değeri ise %85 olarak bulunmuştur.

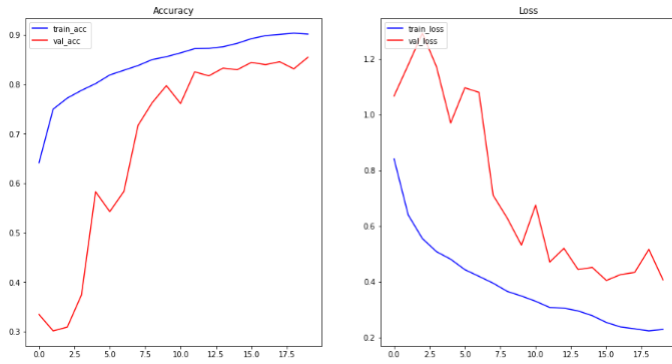


Fig. 2. Accuracy ve Loss değerleri

Görüntü bölütlemeye kullanılan bir başka metrik ise dice katsayısı(dice coefficient) metriğidir. Dice katsayısı tahmin edilen segmentasyon ile ground truth değerleri arasındaki piksel bazında bir karşılaştırmadır. Daha kolay anlaşılabilmesi için formülün görselleştirilmiş hali Fig. 3'te verilmiştir. Dice katsayısının eğitim aşamasında her bir epoch'taki değişimi ise Fig. 4'te gösterilmiştir. Dice katsayısı değeri 1'e ne kadar yaklaşırsa ground truth değerleri ile tahmin edilen segmentasyon değerleri arasında örtüşme o kadar artar denilebilmektedir.

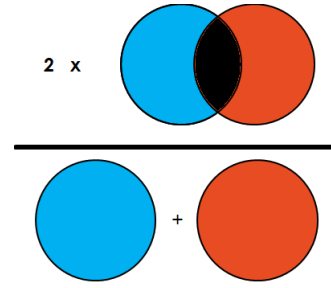


Fig. 3. Dice katsayısı formülü

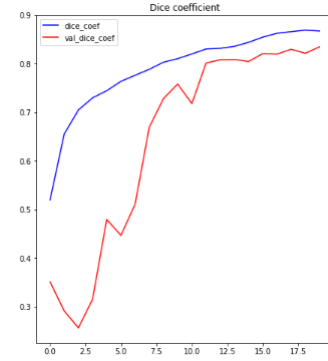


Fig. 4. Dice katsayısı değişimi

Test veri seti içerisinde 3 görüntü seçilerek bu görüntülerin kendileri, ground truth yani gerçek değerleri ve model tarafından tahmin edilen segmentasyon değerleri Fig. 5'te görselleştirilmiştir. Şekilden de görülebileceği üzere başarılı sonuçlar elde edilebilmiştir.

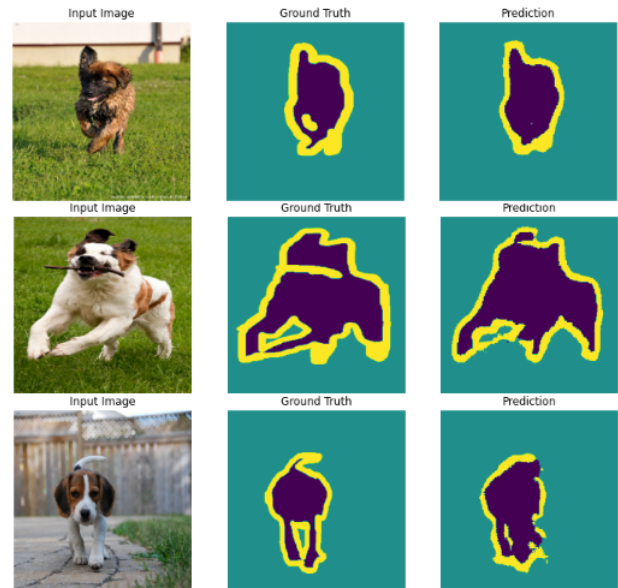


Fig. 5. Test setinden örnekler

IV. SONUÇ

Görüntü bölütleme işlemi derin öğrenme yöntemlerinden önce klasik yöntemlerle gerçekleştirilmekteydi. Bu klasik yöntemlere örnek olarak threshold ve kümeleme yöntemlerini verebiliriz. Threshold yönteminde nesneleri ayırmak için belirli eşik değerleri belirlenir ve buna göre ayırma işlemi gerçekleştirilir. Bu yöntemin en temelde sıkıntısı threshold değerinin ne olacağının belirlenmesidir. Kümeleme yönteminde ise K-means algoritması ile bir küme değeri belirlenerek görüntüdeki pikseller o küme değeri kadar sınıflara ayrılır. Kümeleme ile sınıflandırma temelde basit ve kullanışlı dursa da eğitim süreci ve kümelerin belirlenmesi uzun sürmektedir. Bu nedenle artık tercih edilen bir yöntem değildir. Derin öğrenme yöntemlerine geldiğimizde hem doğruluk olarak hem de performans açısından çok daha avantajlıdırlar. Özellikle bu makalede de incelenen konvolüsyonel encoder-decoder mimarileri ile ciddi başarılar elde edilmektedir. Bu aşamada en çok kullanılan bir diğer konvolüsyonel encoder-decoder mimarisi de U-Net'tir [2]. U-Net mimarisi ile LinkNet mimarisi bir makalede karşılaştırılmıştır ve bunun sonucunda da LinkNet mimarisinin daha başarılı olduğu görülmüştür [3].

V. KOD

Proje kapsamında yapılan çalışmanın kodlarına aşağıdaki GitHub linkinden erişebilirsiniz.

<https://github.com/FidanVural/Digital-Image-Processing/tree/main/Image-Segmentation>

REFERENCES

- [1] Abhishek Chaurasia and Eugenio Culurciello, "LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation," Computer Science & Computer Vision and Pattern Recognition, Wed, 14 Jun 2017.
- [2] Olaf Ronneberger, Philipp Fischer and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Computer Science & Computer Vision and Pattern Recognition, Mon, 18 May 2015.
- [3] V. Anantha Natarajan, M. Sunil Kumar, Rizwan Patan, Suresh Kallam and Mohamed Yasin Noor Mohamed, "Segmentation of Nuclei in Histopathology images using Fully Convolutional Deep Neural Architecture," 2020 International Conference on Computing and Information Technology (ICIT-1441), September 2020.