

Nesne Tanımda Derin Öğrenme Mimarilerinin Performans Karşılaştırması

Eda Dimyanoğlu¹[Y230234292]

¹ İzmir Katip Çelebi Üniversitesi
enedademir@gmail.com

Özet. Bu çalışma, nesne tanıma için yaygın olarak kullanılan farklı derin öğrenme modellerinin performansını, veri augmentasyonu eklemeyi içeren bir yöntemle karşılaştırmaktadır. Kullanılan modeller arasında YOLO (You Only Look Once), Faster R-CNN (Region-based Convolutional Neural Networks), SSD (Single Shot Multibox Detector), ResNet50 ve RetinaNet yer almaktadır. Çalışma, COCO veri seti üzerinde bu modellerin nesne tespiti başarılarını, orijinal ve augmentasyona tabi tutulmuş veri setlerinde test ederek karşılaştırmaktadır. Veri augmentasyonu ile model doğruluğunda gözle görülür iyileşmeler sağlanmıştır. Bu çalışma, nesne tespiti alanında en verimli derin öğrenme mimarisinin tespit edilmesini amaçlamaktadır.

Anahtar Kelimeler: Nesne tanıma, derin öğrenme, veri arttırma, YOLO, Faster R-CNN, SSD, ResNet50, RetinaNet, COCO veri seti.

1 Giriş

Nesne tanıma, bilgisayarla görü alanındaki önemli bir araştırma konusudur. Derin öğrenme yöntemleri, nesne tanıma görevlerinde büyük başarılar elde etmiştir.[1] Bu alandaki en popüler modellerden bazıları, YOLO v3 Faster R-CNN, SSD, ResNet50 ve RetinaNet gibi ağlardır. Ancak bu modellerin performansları, veri setlerinin büyüklüğü ve çeşitliliğiyle doğrudan ilişkilidir. Bu nedenle, bu çalışmada nesne tespiti başarısını değerlendiren bir performans karşılaştırması yapılmaktadır.

COCO veri seti, bu alanda yaygın olarak kullanılan ve üzerinde nesne tanıma görevlerinin uygulandığı geniş çeşitlilikteki görüntüleri içeren bir veri kümesidir. Bu çalışmada, COCO veri setinden rastgele seçilen 500 görüntü ile başlanan denemelerde, veri arttırma tekniklerinin modellerin doğruluğuna olan etkileri analiz edilmiştir. Bu yaklaşım sayesinde veri arttırma tekniklerinin, modellerin doğruluğu üzerinde nasıl bir etkisi olduğu ve her bir modelin veri arttırımı sonrası sağladığı tespit başarı oranları ayrıntılı bir şekilde incelenmiştir. Çalışma, nesne tespiti alanında hangi derin öğrenme mimarisinin daha verimli olduğunu karşılaştırmalı olarak ortaya koymaktadır.

2 Yöntem

2.1 Veri Seti Hazırlığı

COCO veri seti, farklı etiketlenmiş nesnelerin bulunduğu yaklaşık 200.000 görüntü içermektedir. Başlangıçta, veri seti üzerinden rastgele seçilen 500 görüntü kullanılarak nesne tespiti yapılmıştır. Bu görüntülerde tespit edilen çeşitli nesnelerin isimleri ile konumları (bounding box) kaydedilmiştir.

2.2 Veri Arttırma İşlemleri

Başlangıçta orijinal veri setinde COCO içerisinde rastgele seçilmiş 500 görüntü yer almaktadır. Sonrasında, veri arttırma teknikleri devreye girmiştir. Arttırma için her bir görsel üzerine çeşitli işlemler uygulanmıştır:

1. Gürültü ekleme
2. Görüntü döndürme
3. Görüntü kaydırma
4. Yakınlaştırma/uzaklaştırma

Her bir işlem uygulanırken, işlemin etkisi rastgele seçilmiş; örneğin, yakınlaştırma yüzdesi veya döndürme açısı rastgele belirlenerek çeşitlilik daha da artırılmıştır. Veri arttırma sonrası toplamda 2500 görsel ulaşılmıştır. Böylece veri setinin çeşitliliği artırılmış ve her modelin daha farklı senaryolarla karşılaşması sağlanmıştır.

2.3 Nesne Tespiti

Çalışmada karşılaştırılan derin öğrenme modelleri şunlardır:

- YOLO: Gerçek zamanlı nesne tespiti için geliştirilmiş bir modeldir ve hızlı çalışmasıyla dikkat çeker.
- Faster R-CNN: Bölge tabanlı bir yaklaşım kullanarak nesneleri tanır.
- SSD: Tek adımda çoklu nesne tespiti yapan bir modeldir.
- ResNet50: Derin ağlardan biri olan ResNet, özellikle derin katmanlarda bilgi kaybını önleyerek başarısını artırmaktadır.
- RetinaNet: Dengesiz veri problemini çözmeye yönelik olarak geliştirilmiş bir modeldir.

Her modelin doğruluğu, precision, recall, Jaccard index ve Intersection over Union (IoU) gibi metriklerle ölçülmüştür.

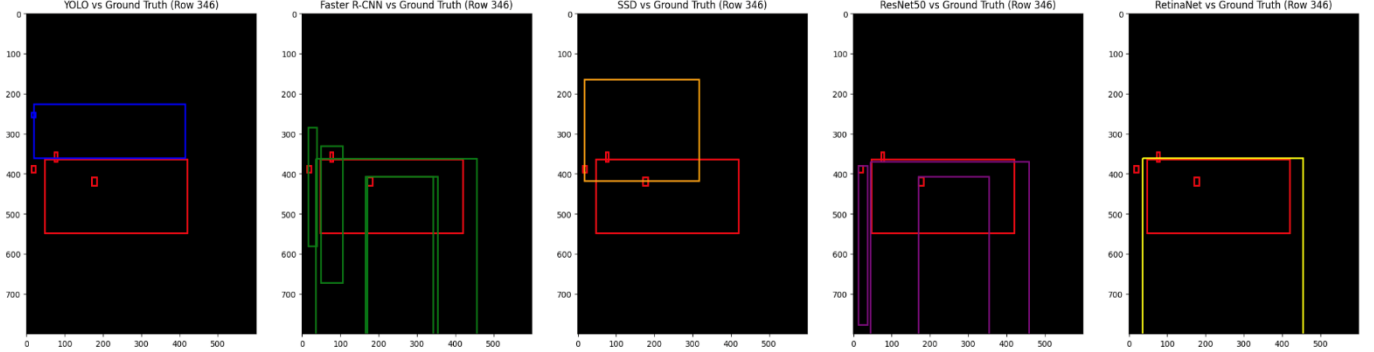
Orijinal Veri Kümesi Üzerindeki Model Performansları

Orijinal veri kümesi üzerinde yapılan karşılaştırmalı ölçümlere aşağıda yer verilmiştir. Yapılan ölçümlerde Faster R-CNN modelinin %64.9 doğruluk oranıyla en iyi performansı verdiği görülmektedir.

Tablo 1. Nesne tespitinde performans ölçümleri.

Model	Jaccard (Accuracy)	Precision	Recall	Intersection over Union
YOLO v3	%53.8	%93.9	%54.5	%5.9
<i>Faster R-CNN</i>	%64.9	%78.1	%79.1	%7.4
Single Shot Detector	%39.1	%89.8	%40.5	%2.7
ResNet50	%58.6	%72.5	%75.1	%7.6
RetinaNet	%51.9	%90.1	%54.9	%8.7

Intersection over Union görselleştirmesi için, modellerin tahmin ettikleri sınırlayıcı kutular, gerçek değerlerle karşılaştırılarak çizilen grafikler elde edilmiştir. Rastgele seçilen beş sonuç için grafikler çizilmiş ve modellerin doğruluğunun görsel olarak karşılaştırılması amaçlanmıştır. Örnek olarak seçilen bir grafiğe aşağıda yer verilmiştir.



Arttırılmış Veri Kümesi Üzerindeki Model Performansları

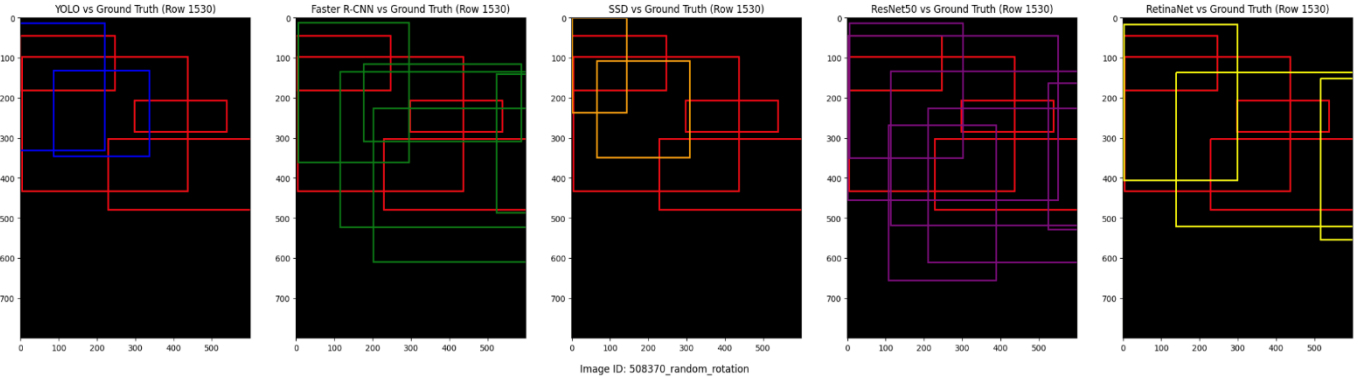
Arttırılmış veri kümesi üzerinde yapılan karşılaştırmalı ölçümlere aşağıda yer verilmiştir. Veri arttırımı sırasında original resimler üzerinde oynamalar yapıldığı için tüm modellerde performans düşüşü beklenmiştir.

Yapılan ölçümlerde Faster R-CNN modelinin %60.4 doğruluk oranıyla en iyi performansı verdiği görülmektedir. Bunun yanı sıra tüm metriklerde tahmin edildiği üzere düşüşler yaşanmıştır. Her bir metriğin düşüş miktarı ölçümün yanında ayrıca belirtilmiştir.

Tablo 2. Arttırılmış veri kümesinde yapılan performans ölçümleri.

Model	Jaccard (Accuracy)	Precision	Recall	Intersection over Union
YOLO v3	%48.1 (-5.7%)	%92.1 (-1.8%)	%49.1 (-5.4%)	%4.9 (-1.0%)
<i>Faster R-CNN</i>	%60.4 (-4.5%)	%75.7 (-2.4%)	%74.9 (-4.2%)	%6.3 (-1.1%)
Single Shot Detector	%37.3 (-1.8%)	%88.6 (-1.2%)	%38.5 (-2.0%)	%2.2 (-0.5%)
ResNet50	%54.2 (-4.4%)	%70.0 (-2.5%)	%70.7 (-4.4%)	%6.5 (-1.1%)
RetinaNet	%48.1 (-3.8%)	%88.5 (-1.6%)	%51.3 (-3.6%)	%7.2 (-1.5%)

Intersection over Union görselleştirmesi için, modellerin tahmin ettikleri sınırlayıcı kutular, gerçek değerlerle karşılaştırılarak çizilen grafikler elde edilmiştir. Rastgele seçilen beş sonuç için grafikler çizilmiş ve modellerin doğruluğunun görsel olarak karşılaştırılması amaçlanmıştır. Örnek olarak seçilen bir grafiğe aşağıda yer verilmiştir.



3 Sonuç

Bu çalışmada, nesne tanıma modellerinin performansları, orijinal ve arttırılmış veri kümesi üzerinde karşılaştırılmıştır. İlk olarak, orijinal veri kümesi kullanılarak yapılan testlerde, Faster R-CNN modelinin %64.9 doğruluk oranı ile en yüksek başarıyı elde ettiği gözlemlenmiştir. YOLO v3, ResNet50, RetinaNet ve Single Shot Detector modellerinin de sırasıyla aynı veri kümesi üzerindeki performansları test edilmiştir.

Veri arttırımı uygulanan arttırılmış veri kümesi üzerinde yapılan testler, nesne tanıma sırasında model başarısının belirgin bir şekilde düştüğünü göstermiştir. Özellikle, tüm modellerde doğruluk oranında düşüşler görülmüş, ancak Faster R-CNN modeli hâlâ en iyi performansı %60.4 doğruluk oranı ile sunmuştur. Arttırılmış veri setiyle test edilen YOLO v3, Single Shot Detector, ResNet50 ve RetinaNet modelleri, orijinal küme üzerinde elde edilen sonuçlardan daha düşük performans göstermiştir. Faster R-CNN modelinde gözlenen küçük performans kaybı, onun veri arttırımı gibi farklı veri işleme tekniklerine karşı daha dayanıklı olduğuna işaret etmektedir.

Bu sonuçlar ışığında, Faster R-CNN modelinin daha yüksek doğruluk oranıyla (yani, %64.9 ve arttırılmış veri kümesinde de %60.4) en iyi performansı gösterdiği görülmektedir. Bu nedenle, veri arttırma sonrası performans düşüşlerine rağmen, Faster R-CNN yüksek doğruluk oranı ve diğer metriklerde daha az kayıp yaşamış olması dolayısıyla en uygun model olarak önerilebilir. Özellikle Intersection over Union ve Recall metriklerinde güçlü bir performans sergileyen Faster R-CNN, nesne tespiti konusunda etkili sonuçlar sunmaktadır.

Referanslar

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.