

Mask R-CNN İle Uydu Görüntülerinde Gemi Tespiti

Nuri Erkin ÖÇER, Eskişehir Teknik Üniversitesi, Yer ve Uzay Bilimleri Enstitüsü, Eskişehir, Türkiye,
neocer@eskisehir.edu.tr.

Uğur, AVDAN, Eskişehir Teknik Üniversitesi, Yer ve Uzay Bilimleri Enstitüsü, Eskişehir, Türkiye,
uavdan@eskisehir.edu.tr

Öz

Derin öğrenme alanında son yıllarda gerçekleşen atılımın katkısıyla uzaktan algılama görüntülerinde gemi tespiti konusunda önemli ilerlemeler kaydedilmiştir. Özellikle, nesne tespiti ve sınıflandırması için geliştirilen yaklaşımlardan olan konvolüsyonel sinir ağları (CNN) gemi tespiti için de başarılı ve yaygın bir şekilde kullanılır hale gelmiştir. Bunun yanında, uydu görüntülerinin niteliklerinin gelişmesi bu verilerden gemilerin ve hatta daha küçük nesnelerin algılanıp ayırt edilebilmesinin önünü açmıştır. Bu çalışmada, optik uydu görüntülerindeki gemileri tespit etmek üzere bölge-tabanlı konvolüsyonel sinir ağı modellerinden biri olan Mask R-CNN yöntemi kullanılmıştır. Çalışmadaki temel amaç, kullanılan modelin uydu görüntülerindeki gemi tespit performansını ve sınırlarını incelemektir. Bunun için, gemilerin yoğun olarak bulunduğu alanların 1 metre mekânsal çözünürlüğe sahip 1838 adet uydu görüntüsü indirilmiş ve içerikteki gemiler bir GIS yazılımı aracılığıyla maskelerle etiketlenerek veri setleri oluşturulmuştur. Elde edilen sonuçlar, önerilen yöntemin zorlu içeriklerde bile gemileri başarıyla tespit etme kabiliyetinde olduğunu fakat mevcut haliyle modelde bazı kısıtlamalar bulunduğunu göstermiştir. Model özellikle birbirine yakın konumlanmış gemilerin tespitinde yanılğıya düşmektedir.

Anahtar Kelimeler: Derin Öğrenme, Gemi Tespiti, Mask R-CNN, Uydu Görüntüleri

Ship Detection in Satellite Images with Mask R-CNN

Abstract

The recent breakthrough in the field of deep learning has led to significant progress in the detection of ships in remote sensing images. In particular, convolutional neural networks (CNN), an approach developed for object detection and classification, has been successfully and widely used for ship detection. In addition, the improved quality of satellite imagery has made it possible to detect and distinguish ships and even smaller objects from this data. In this study, Mask R-CNN, a region-based convolutional neural network model, is used to detect ships in optical satellite images. The main objective of the study is to investigate the performance and limitations of the model in detecting ships in satellite images. For this purpose, 1838 satellite images with a spatial resolution of 1 meter of the areas where ships are densely populated were downloaded and data sets were created by labelling the ships in the content with masks through a GIS software. The results show that the proposed method is capable of successfully detecting ships even in challenging contexts, but there are some limitations in the model in its current form. In particular, the model is inaccurate in detecting ships located close to each other.

Keywords: Deep Learning, Mask R-CNN, Satellite Images, Ship Detection

1. Giriş

Kullanıldığı alanlara bakıldığında, uzaktan algılama görüntülerinde gemi tespiti ülke karasularının gözetimi, güvenliği, yönetimi ve deniz taşımacılığı gibi birçok alanda kendine yer bulmaktadır. Bu bağlamda gemi tespiti, eldeki hava veya uydu fotoğrafında yer alan gemilerin görüntü içerisinde ayırte edilmesini ve konumlarının tespitini içermektedir. Son yıllarda sayısı ve niteliği giderek artan yüksek çözünürlüklü uydu görüntülerinin yeni algılama modelleriyle beraber kullanılması gemilerin içeriklerden otomatik olarak çıkarılmasının önünü açmıştır.

Başarılı, yeni algılama modellerinin büyük çoğunluğu makine öğrenimine dayalı yöntemlerdir. Bu yaklaşımlarda, sahip olunan etiketli bilgiden yararlanılarak nesne özelliklerinin otomatik olarak öğrenilmesi sağlanmaktadır (Hinton & Salakhutdinov, 2006). Makine öğrenimine dayalı obje tespit modelleri kendinden önceki yöntemlere göre daha hızlı ve doğru sonuçlar üretmiştir (Han vd., 2015). Ancak, SVM (Destek Vektör Makineleri) gibi geleneksel makine öğrenme algoritmalarında girdilerin eğitim standardından sapması durumunda söz konusu olan keskin performans düşüşü önemli bir dezavantajdır (Xia vd., 2011).

Makine öğrenimi dallarından biri olan ve insandaki öğrenme davranışını makineye uyarlayan konvolüsyonel sinir ağları (Convolutional Neural Networks-CNN) son yıllarda sınıflandırma alanında en başarılı araçlardan biri haline gelmiştir (Krizhevsky vd., 2012; Szegedy vd., 2014). CNN, nesnenin içeriği büyük oranda kapladığı görüntüleri işleme (sınıflandırma) yeteneğindedir. Ancak, nesnenin görüntüyü kaplamadığı ve bu nesnenin konumuyla beraber görüntü içerisinde çıkarılması, yani nesne tespiti görevi söz konusu olduğunda bir uyarlamaya ihtiyaç vardır.

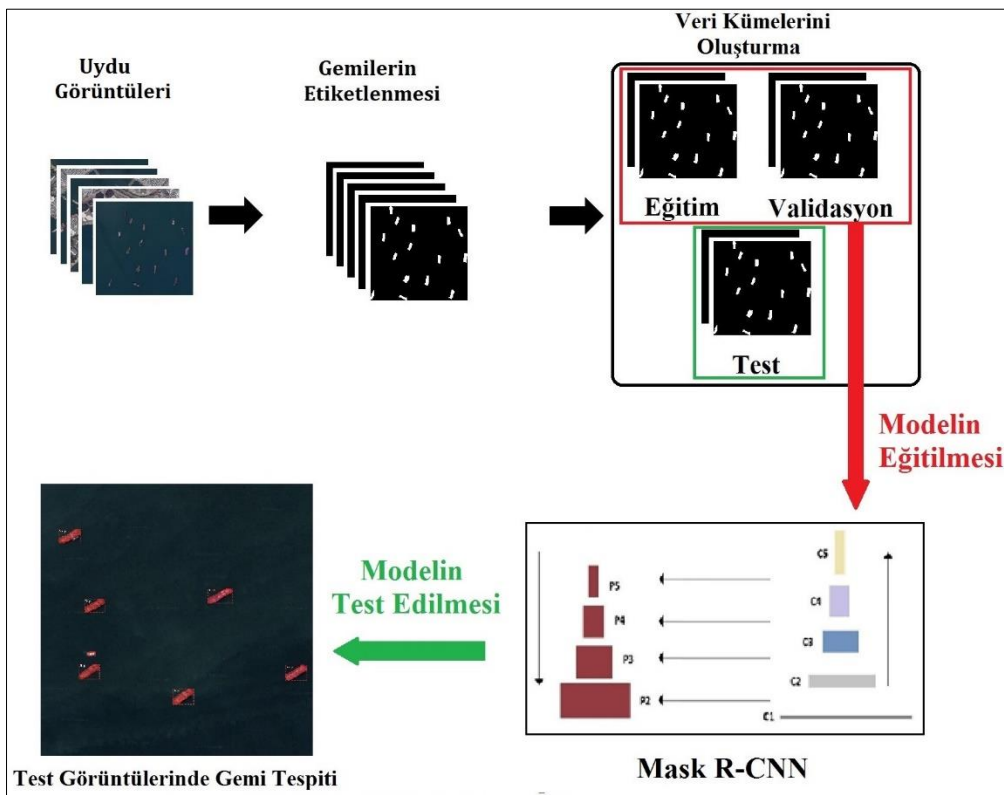
Başlangıçta bu problem, görüntünün ağırlı işleyebileceği boyutta parçalara ayrılıp ağırlı bu parçalarla beslenmesi ve her bir görüntü parçası üzerinde sınıflandırmanın tek tek yapılmasıyla çözülmeye çalışılmıştır. Kayan pencereler (Sliding Windows) adı verilen bu yaklaşımla hem optik hem SAR görüntülerinde birçok nesnenin tespiti gerçekleştirilmiştir (Lee vd., 2017; Huang & Huang, 2016). Lakin nesne boyutları çeşitlilik arz ettiğinde bu yöntemde tek bir tarama pencere boyutu ve en-boy oranı yeterli olmamakta, çok sayıda pencerenin görüntü üzerinde gezdirilmesi ve pencerelerin kaydırma oranlarının çok dikkatli belirlenmesi gerekmektedir. Oldukça fazla işlem zamanına mal olan bu yaklaşım derin öğrenmenin potansiyelinden yeteri kadar faydalanamamakta ve pratiklikten uzak kalmaktadır. Bu sorunun çözümüne dönük olarak CNN-tabanlı iki farklı türde tespit yaklaşımı önerilip kullanılmıştır. Bunlardan biri, iki aşamalı bir süreç içinde önce görüntüler içerisindeki nesne olma potansiyeli yüksek bölgelerin önerilip sonrasında sınıflandırma ve sınırlayıcı kutular kümesinin üretildiği Faster R-CNN (Ren vd., 2016) veya Mask R-CNN (He vd., 2018) gibi bölge-tabanlı CNN yaklaşımlarıdır. Diğer yaklaşım ise bölge öneri ağına ihtiyaç duymayan, nesne tespitini doğrudan görüntü özelliklerinden tahmin eden YOLO (You Only Look Once) (Redmon vd., 2015) ve SSD (Single Shot MultiBox Detector) (Liu vd., 2016) gibi yöntemlerdir.

Bölge-tabanlı konvolüsyonel sinir ağları (R-CNN) yönteminde, görüntüde nesne aday bölgeler belirlenmekte ve özellik çıkarımı konvolüsyonel katmanlar aracılığıyla gerçekleştirilmektedir. Bu türdeki tespit yöntemlerinden olan Mask R-CNN'de nesne konumlarının belirlenmesi de öğrenme yoluyla gerçekleştirilmektedir. Bu modelde nesne olması muhtemel bölgeler bölge öneri ağı (Region Proposal Network - RPN) tarafından üretilmektedir. Mask R-CNN'de görüntüler içerisindeki nesnelere sınırlayıcı kutular ve maskelerle işaretlenmektedir. Maske, objenin görünür dış hatlarına uygun şekilde çizilen bir poligondur.

Bu çalışmada, uydu görüntülerinde yer alan gemilerin tespiti için Mask R-CNN yöntemi kullanılarak nesne işaretlemesinde maske kullanımının tespit sonuçları üzerindeki etkileri incelenecektir. Modelin eğitiminde kullanılmak üzere gemilerin maskelerle etiketlendiği veri setleri oluşturulacak ve yapılacak deneylerle önerilen model için en uygun çalışma parametreleri belirlenecektir. Elde edilen tespit sonuçları çeşitli metrikler kullanılarak değerlendirilecek ve modelin başarılı ve başarısız olduğu durumlar tartışılacaktır.

2. Materyal ve Yöntem

Bu çalışmada Şekil 1'de gösterilen bir akış süreci izlenmiştir. Ham görüntülerden başlayan süreçte öncelikle görüntüler işlenip etiketlenmekte ve veri setleri oluşturulmaktadır. Ardından bu veriler kullanılarak model eğitilmekte ve test görüntülerdeki gemilerin tespitiyle süreç tamamlanmaktadır. Bu başlıkta sırasıyla çalışmada kullanılan verilere, gemi tespiti için önerilen modele, modelin eğitimine dair detaylara ve tespit değerlendirme metriklerine yer verilmektedir.



Şekil 1.Çalışmadaki Gemi Tespit Prosedürü

2.1. Veri Setleri

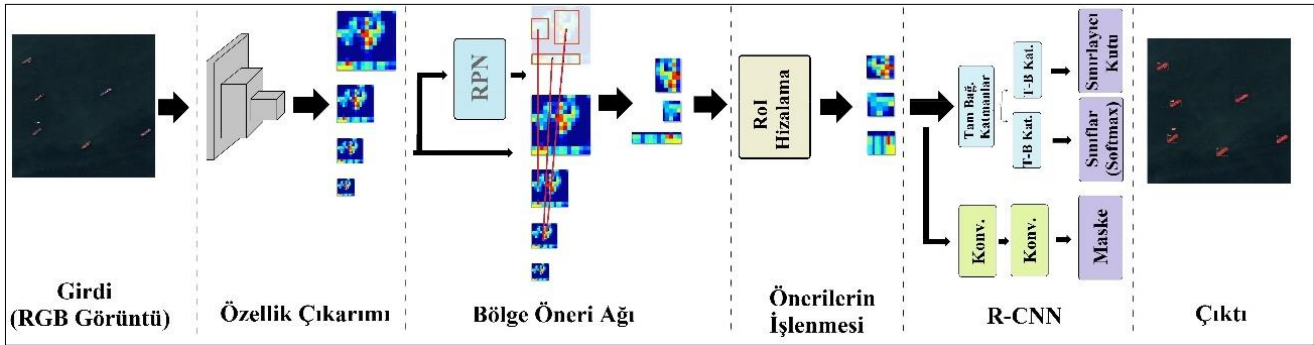
Çalışmada kullanılan veriler Google Earth tarafından sağlanan 1 metre mekânsal çözünürlüklü, geotif formatındaki RGB uydu görüntüleridir. Açık denizlerden, kıyı yakınlarından ve iç sulardan örnekler bulunmaktadır. Süreçte öncelikle her biri 768x768 piksel büyüklüğünde toplam 1838 görüntü elde edilmiştir. Bunlar kullanılarak eğitim, validasyon ve test veri setleri oluşturulmuştur (Tablo 1). Görüntülerdeki toplam 3279 gemi bir GIS yazılımı kullanılarak sayısallaştırılmıştır. Oluşturulan eğitim ve validasyon setleri kullanılarak model eğitilmiş ve doğrulaması yapılmıştır. Sonrasında, eğitilen model test görüntüleri üzerinde çalıştırılarak tespit performansı hesaplanmıştır.

Tablo 1. Çalışma Kapsamında Kullanılan Veri Kümelerine Dair Bilgiler

	Eğitim Kümesi	Validasyon Kümesi	Test Kümesi	Toplam
Görüntü Sayısı	1224	320	294	1838
Gemi Sayısı	2096	579	604	3279

2.2. Yöntem

Bu çalışmada uydu görüntülerinde yer alan gemileri tespit etmek üzere Mask R-CNN modeli kullanılmıştır. Girdi görüntüsü bu modelde Şekil 2’de gösterilen aşamalardan geçerek işlenmektedir. Mask R-CNN çıktı olarak, tespit ettiği gemilerin maskelerini üretmekte, sınırlayıcı kutularını oluşturmakta ve her tespitin olasılık değerini hesaplamaktadır.

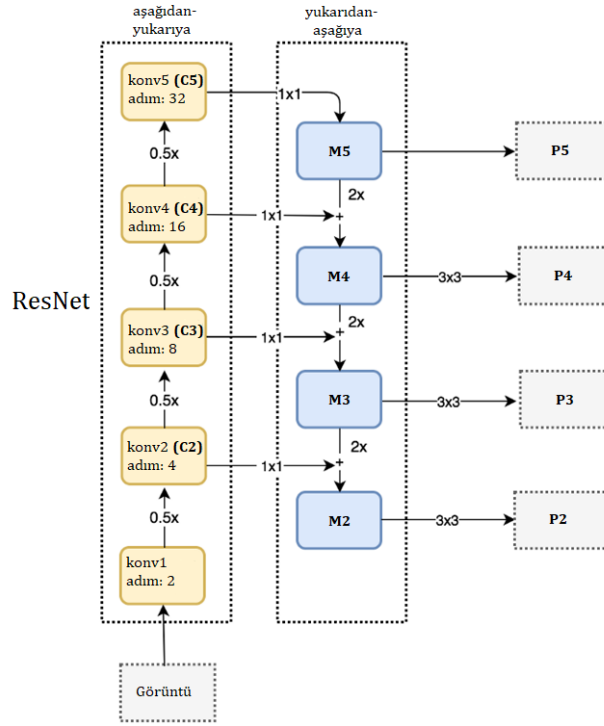


Şekil 2. Bir Girdi Görüntüsünün Mask R-CNN Modelinde İşlenerek İçeriğindeki Gemilerin Tespit Edilmesi Süreci

Bölge önerisine dayalı bir nesne tespit algoritması olan Mask R-CNN dört modülden oluşmaktadır: özellik çıkarım ağı, bölge öneri ağı, öneri işleme modülü ve bölge-tabanlı konvolüsyonel sinir ağı (R-CNN).

Özellik çıkarımı, ön-egitimli bir konvolüsyonel ağdan geçirilen Yükseklik × Genişlik × Derinlik tensörü formatındaki girdi görüntüsünden konvolüsyonel özellik haritalarının elde edilmesidir. Özellik çıkarımı elde edilen veriden içeriğe dair bilginin derinleştirilip (daha kullanışlı veri gruplarına evriltilip) mekânsal boyutlarının indirgenildiği bir süreçtir. Bu ağlar çok katmandan meydana gelmekte ve her konvolüsyonel katman kendinden önceki bilgilere dayanarak soyutlamalar oluşturmakta, özellik haritaları üretmektedir. Özellik haritasının genişliği ve yüksekliği, konvolüsyonel katmanlar arasında uygulanan havuzlama nedeniyle azalmakta ve öğrendiği filtre sayısına bağlı olarak derinliği artmaktadır.

Bu çalışmada kullanılan Mask R-CNN sürümünde özellik çıkarımı Özellik Piramit Ağı (Feature Pyramid Network – FPN) tarafından yapılmaktadır. FPN, özellik çıkarım ağındaki yüksek çözünürlüklü ancak sığ semantik değere sahip katman özelliklerini semantik değeri yüksek derin katman özellikleriyle kaynaştırmak üzere geliştirilmiş yanal bağlantılara sahip, bölge-temelli CNN sistemlerine uyumlu bir mimaridir (Şekil 3).



Şekil 3.Özellik piramit ağında yukarıdan-aşağıya tanımlanan yeni yolda mekânsal çözünürlüğü düşük, semantik değeri yüksek özellik haritalarının, alt katmanla kurulan yanıl bağlantılar ve üst-örneklemeyle kaynaştırılarak semantik değeri korunmakta ve mekânsal boyutu artırılmaktadır.

FPN'nin ilk aşaması olan aşağıdan yukarıya yolda, 5 ayrı blokta özellik özellik haritası (C_i) üretilmektedir. Haritaların mekânsal boyutları bir önceki blok çıktısına göre yarıya inerken, derinliği (semantik değeri) artmaktadır. İkinci aşamada, C_i'ler kullanılarak yukarıdan-aşağıya yol inşa edilmektedir. Yukarıdan-aşağıya yolda ilk oluşturulan özellik haritası M5'tir. M5'i oluşturmak üzere, 1x1'lik bir konvolüsyon filtresi uygulanarak C5'in kanal derinliği 256 kanala düşürülmektedir. Bu, nesne tahmininde kullanıma hazır ilk özellik haritası katmanıdır. Yukarıdan aşağıya inilirken, en yakın komşuluk üst-örnekleme kullanılarak üst katmanın mekânsal çözünürlüğü 2 katına çıkarılmaktadır. Alt katmanda karşılık gelen ilgili özellik haritasına 1x1'lik konvolüsyon uygulanmasından sonra mekânsal boyutlarıyla beraber derinlikleri de denk hale getirilen iki katman birbirleriyle kaynaştırılmaktadır. Bozucu (aliasing) etkileri hafifletmek üzere kaynaştırılmış tüm katmanlara ayrıca 3 × 3'lük bir konvolüsyon da uygulanmaktadır. Aynı işlemler P3 ve P2 için de tekrar edilmektedir. Süreç, P2'de durdurulmaktadır. Her özellik haritası çıktısı için son (tam-bağlı katman yapısında olan) R-CNN modülünde aynı sınıflandırıcı ve kutu regresörü kullanıldığından tüm piramit özellik haritalarının {P5, P4, P3, P2} derinlikleri aynı (d=256) tutulmaktadır.

Süreçte bir sonraki aşama, oluşturulan özellik haritalarının Bölge Öneri Ağında (RPN) işlenmesidir. Bu modülde, görüntüde nesne olması muhtemel bölgeler önerilmektedir. Özellik haritalarında çapaların oluşturulmasıyla başlanıp paralel yürütülen iki konvolüsyon işlemiyle görüntüdeki olası nesnelere için sınırlayıcı kutular oluşturulmakta ve bu kutuları görüntüdeki gerçek nesnelere oturtmak üzere konum düzeltme işlemleri (regresyon) uygulanmaktadır.

Mask R-CNN'de FPN'de üretilmiş farklı ölçeklerdeki özellik haritaları kümesi {P2, P3, P4 ve P5} Bölge Öneri Ağı (RPN) modülüne girilmekte ve aday nesnelere dair sınırlayıcı kutular veya İlgi Alanları (Region of Interest - RoI) önerilmektedir. Önerilen RoI'nin boyutuna (genişlik: w ve yükseklik: h) bağlı olarak, özellik yamalarını çıkarmak için en uygun ölçekteki özellik haritası katmanı seçilmektedir. Bu seçim şu denkleme göre yapılmaktadır:

$$k = 4 + \log_2 \left(\frac{\sqrt{wh}}{224} \right) \quad (1)$$

Burada k, özellik yamasını oluşturmada kullanılacak FPN'deki Pk katmanıdır. Bu denklemin çözümü sonucunda elde edilen k değerine bağlı olarak hangi katmanın kullanılacağı belirlenmektedir. Yani k=4 ise sürecin geri kalanında kullanılacak özellik haritası olarak P4 seçilmektedir. Diğer bir deyişle, nesne önerileri tek bir özellik haritasından değil, önerilen nesnenin büyüklüğüne bağlı olarak uygun ölçekteki özellik haritasından çıkarılmaktadır. Böylelikle görüntü içerisindeki nesne (ve buna dair öneri); göreceli olarak küçükse yukarıdan-aşağıya yolda üretilen mekânsal olarak daha büyük boyuttaki özellik haritası (M2 gibi), göreceli olarak büyükse daha küçük boyuttaki özellik haritasından faydalanılmaktadır.

Modelin RPN modülünde ayrıca, üretilen önerilerin nesne olup olmadığına dair bir güven eşiği tanımlıdır. Öneri için üretilen olasılık bu eşiğin altında olduğu takdirde öneri "arka plan" olarak kabul edilmekte, üstünde olduğu takdirde "gemi" olarak atanmaktadır. Bu eşik için en uygun değer yapılacak deneylerle belirlenmektedir.

Mask R-CNN'nin son modülünde tam-bağlantılı katmanlar yer almaktadır. Tam-bağlantılı katmanlar ancak büyüklükleri belirli tensörleri işleyebilmektedir. Bu sebeple bu katman girdilerinin aynı boyutta olması gerekmektedir. Dolayısıyla RPN aracılığıyla elde edilen önerilerin işlenmesine gerek vardır.

Bu işlem, RPN'de üretilen önerilerin görüntüden kesilip hizalanarak bir havuzda toplanmasıyla gerçekleştirilmektedir (RoI Hizalama). Bu katmanda RPN'den gelen her öneri bölgesi, konvolüsyonel özellik haritasından hizalanarak kırılıp çıkarılmakta ve sonrasında tüm parçalar bilineer enterpolasyon kullanılarak $14 \times 14 \times \text{konv_derinliđi}$ boyutlarına getirilmektedir. Sonrasında, her öneri bölgesinden $7 \times 7 \times \text{konv_derinliđi}$ boyutunda nihai bir özellik haritası elde etmek üzere 2×2 'lik bir çekirdekle en büyük değer havuzlaması (max-pooling) uygulanmaktadır. Bu şekilde, tekliflerin tümü belirli bir kalıba uydurulmakta böylece çıktılar tam-bağlantılı katmanların kullanıldığı bir sonraki modülde kullanıma hazır hale gelmektedirler. Orijinal çalışmada (He vd., 2018), doğal görüntülerdeki nesne tespitinde RoI Hizalamasının doğruluđu RoI Havuzlamasına göre yüzde 20'ye varan oranda iyileştirdiđi belirtilmiştir.

Mask R-CNN modelindeki son işlem adımı bölge-temelli konvolüsyonel sinir ađıdır (R-CNN). Bu modül, RPN önerilerinden uygun olanların gemi sınıfına atanmasını, sınırlayıcı kutuların yer doğruluđuna göre daha düzgün ayarlanmasını ve maskelenmesini sağlamaktadır.

Böylece, dört aşama sonrasında bir girdi görüntüsünden; (1) özellik haritaları oluşturulmuş, (2) RPN ile nesne önerileri çıkarılmış ve (3) bunlar RoI hizalaması vasıtasıyla belirli bir formata getirilerek tam-bağlantılı katmanları da içeren R-CNN modülünde işlenmeye hazır hale getirilmişler ve (4) son modülde görüntüdeki nesnelere konumları ve maskeleriyle beraber tespit edilebilir hale gelmişlerdir.

2.3. Model Eğitimi

Bu çalışmada Mask R-CNN modeli için Waleed'in (2017) Tensorflow ve Keras derin öğrenme çerçevelerinde çalışan GitHub uyarlaması kullanılmıştır. Modelde omurga ađ olarak Common Objects in Context (COCO) veri setiyle ön-eđitim almış ResNet-101 kullanılmıştır. Modelde ikisi RPN, üçü R-CNN için olmak üzere toplam 5 adet kayıp fonksiyonu tanımlıdır. Model, başlangıç değeri 0,001 olan öğrenme oranı kullanılarak toplam 80 bin iterasyonda eğitilmiştir. Eğitimde 0,0001'lik bir ağırlık bozunumu (weight decay) ve 0,9'luk momentum değerleriyle momentumlu Stokastik Gradyan İniş yöntemi ve geri-yayılım algoritmaları kullanılmıştır. Deneyler, Ubuntu 16.04 işletim sisteminde Intel i7-6850k işlemcili, 64 GB RAM'e sahip, Nvidia GeForce GTX 1080 Ti ekran kartlı bir bilgisayar kullanılarak gerçekleştirilmiştir.

2.4. Tespit Değerlendirme Metrikleri

Modelin tespit performansını değerlendirmek için yaygın olarak kullanılan kesinlik, geri getirme ve F1-skoru metrikleri kullanılmıştır. Bunlar vasıtasıyla tespit modellerinin performansı sadece doğru tespitlerle (True positives - TP) değil, üretilen yanlış alarmlar (False Positives - FP) ve gözden kaçan nesnelere (False Negatives - FN) de göz önüne alınarak hesaplanmaktadır.

Kesinlik, doğru tahminlerin toplam tahminler içerisindeki oranını (TP'lerin toplam pozitif tahminlere oranını) göstermektedir. Bu metrik, üretilen yanlış alarmların nicel olarak değerlendirilmesine yardımcı olmaktadır. Şu şekilde tanımlanmıştır:

$$Kesinlik = \frac{TP}{TP + FP} \quad (2)$$

Geri getirme ise yapılan tespitlerdeki doğru pozitiflerin toplam yer doğruluklarına oranıdır ve tespit doğruluğu olarak da geçmektedir. Tanımından da anlaşılacağı üzere bu metrik, tespiti hedeflenen nesnelere ne oranda belirlendiğini nicel olarak değerlendirmeye yardımcı olmaktadır. Tanımı şöyledir:

$$Geri\ getirme = \frac{TP}{TP + FN} \quad (3)$$

F1 skoru, kesinlik ve geri getirme metriklerinin harmonik ortalamasıdır. Bu sebeple tespit performansına dair daha etraflı bir değerlendirme imkânı vermektedir ve şöyle tanımlanmıştır:

$$F1 = 2 \cdot \frac{kesinlik \cdot geri\ getirme}{kesinlik + geri\ getirme} = \frac{2TP}{2TP + FN + FP} \quad (4)$$

Bu çalışma kapsamında, modelce üretilen herhangi bir tespit tahmini, eğer yer-doğruluğuyla olan IoU değeri 0,5'ten büyükse gerçek doğru (TP), aksi halde yanlış tespit (FP) olarak değerlendirilmiştir.

3. Bulgular ve Tartışma

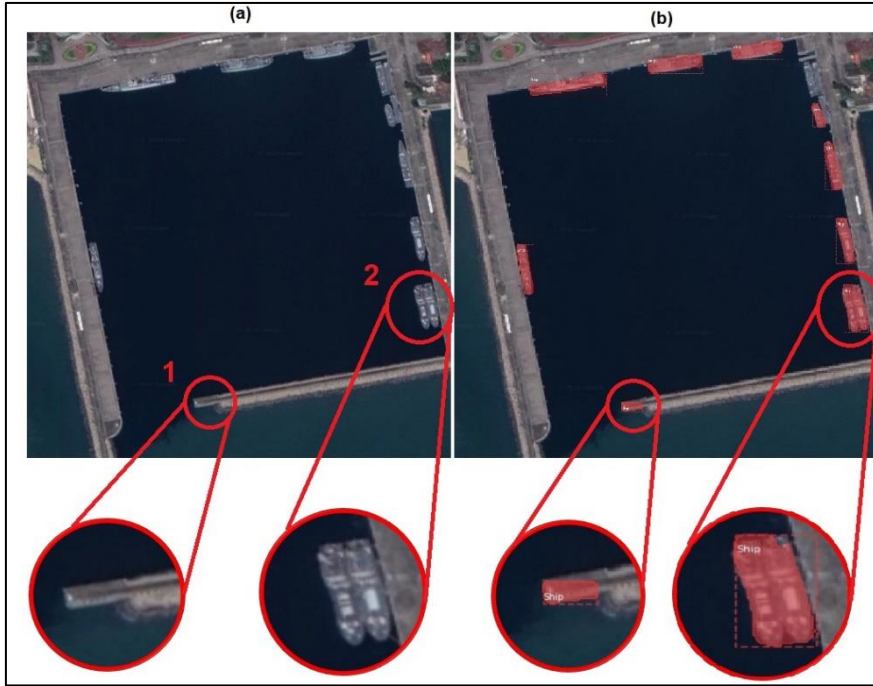
Çalışmada öncelikle model için en uygun güven eşiğinin belirlenmesine yönelik deneyler gerçekleştirilmiştir. Bu kapsamda, modelin güven eşik değeri 0'dan başlatılarak 1'e kadar 0,1'lik adımlarla artırılmış ve her değer için tespit metrikleri hesaplanmıştır. Sonuçlara göre optimal değer 0,5 olduğu bulunmuştur. Eldeki veri seti için doğru tespitler ve yanlış alarmlar arasında en dengeli tespit sonuçları bu değerle sağlanmıştır. Bu değer için model, test kümesindeki 604 yer doğruluğunun 558'ini tespit etmiş, 46'sını gözden kaçırmış 58 adet de yanlış alarm üretmiştir. Buna göre, test verisi için modelin geri getirmesi 0,9238; kesinliği 0,9058; F1-skoru ise 0,9148 olmuştur.

Belirlenen 0,5'lik nesne güven eşiği ve ResNet-101 omurga ağı ile çalıştırılan Mask R-CNN modeli test görüntüleri üzerinde çalıştırılarak tespitler alınmıştır. Modelin ürettiği bazı tespit sonuçları Şekil 4-5-6'da gösterilmektedir. Singapur Limanı açıklarından alınmış bir görüntü olan Şekil 4.a'da herhangi bir kara parçası yoktur. Toplam sekiz geminin bulunduğu bu görüntü için model yedi gemiyi doğru tespit etmiş, kırmızı yuvarlakla yakınlaştırılmış alanda yer alan büyük gemiye yanaşmış olan küçük gemiyi gözden kaçırmıştır (Şekil 4b). Şekil 5.a'da ise Hong Kong'daki askeri bir limana yanaşmış gemileri içeren bir uydu görüntüsü görülmektedir. Toplam on geminin bulunduğu bu görüntü için model sekiz gemiyi doğru tespit etmiş, ikisini ise gözden kaçırmıştır (Şekil 5b). Model; Şekil 5.a'da 1 numaralı daire ile gösterilen kara parçasını hatalı şekilde gemi olarak sınıflandırmış, 2 numaralı daire içerisinde gösterilen, yan yana

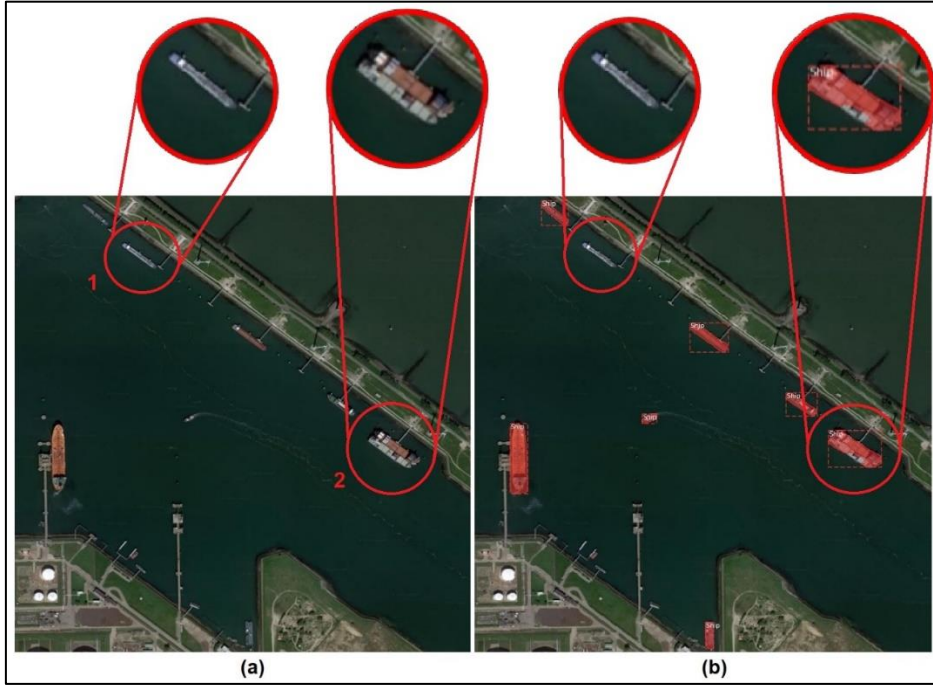
bulunan iki gemiyi ise tek gemi olarak işaretlemiştir. Ayrıca, limanda karaya yanaşmış bir diğer gemi de model tarafından gözden kaçırılmıştır. Rotterdam'daki bir iç limanda bulunan gemileri içeren bir uydu görüntüsünün yer aldığı Şekil 6.a'da ise dokuz gemi bulunmaktadır. Model bu gemilerden yedisini doğru tespit etmiş ikisini ise gözden kaçırmıştır (Şekil 6b). Hatalardan biri Şekil 6.a'da 1 numaralı daire ile gösterilen limana yanaşmış olan gemi, diğeri ise 2 numaralı daire içerisinde gösterilen, yan yana bulunan iki geminin tek gemi olarak işaretlenmesidir.



Şekil 4. Açık Denizde Bulunan Gemileri İçeren Bir Uydu Görüntüsü (a) ve Modelin Bu Görüntü İçin Ürettiği Tespit Sonucu (b)



Şekil 5. Hong Kong'daki Askeri Bir Limana Yanaşmış Gemileri İçeren Bir Uydu Görüntüsü (a) ve Modelin Bu Görüntü İçin Ürettiği Tespit Sonucu (b)



Şekil 6.Rotterdam'daki Bir İç Limanda Bulunan Gemileri İçeren Bir Uydu Görüntüsü (a) ve Modelin Bu Görüntü İçin Ürettiği Tespit Sonucu (b)

Yukarıda ifade edilen, birbirine yakın konumlanmış gemiler için üretilen tespitlerdeki yanlış sayısının fazlalığı Faster R-CNN gibi bölge-tabanlı nesne tespit algoritmalarının kullanıldığı diğer araştırmalarda da söz konusudur (Yang vd., 2018; Gao vd., 2019). Bu problemin RPN'de öneriler için kullanılan dik sınırlayıcı kutulardan kaynaklandığı ifade edilmektedir. Bu sorunun çözümüne yönelik olarak RPN'de dönük sınırlayıcı kutuların kullanımı önerilmiştir (Gao vd., 2019). Faster R-CNN modeli üzerine inşa edilmiş olan Mask R-CNN için de aynı tartışma geçerlidir.

4. Sonuç

Çalışmada bölge-tabanlı bir nesne tespit modeli olan Mask R-CNN yöntemiyle uydu görüntülerindeki gemiler tespit edilmiş ve çeşitli metriklerle performansı değerlendirilmiştir. Mask R-CNN'nin tespit işleminde kullanılmasıyla önceki yaklaşımlardan farklı olarak gemiler görüntülerde sınırlayıcı kutulara ilave olarak maskelerle de işaretlenmiştir. Sonuçlara göre, 0,5'lik bir güven eşiği ile çalıştırılan Mask R-CNN modeli test görüntülerinde yer alan 604 geminin 558'ini doğru şekilde tespit etmiş ve 58 yanlış alarm üretmiştir. Modelin çeşitli görüntüler için ürettiği tespitler incelendiğinde, özellikle birbirine yakın konumlanmış gemiler söz konusu olduğunda ya hepsinin tek bir gemi olarak işaretlendiği ya da gemilerden birinin veya birkaçının gözden kaçırıldığı görülmektedir. Diğer taraftan, daha az sayıda olmakla beraber bazı karasal bölgelerin kimi zaman model tarafından gemi olarak algılandığı, kimi zaman da karaya yavaşmış bazı gemilerin gözden kaçırıldığı gözlenmiştir. Gelecek çalışmalarda bu hataların altında yatan sebeplerin ortaya çıkarılmasına yönelik deneyler ve Mask R-CNN'nin aynı test verisi üzerinde farklı modellerle karşılaştırılması planlanmaktadır.

Teşekkür

Bu çalışma Nuri Erkin Öçer tarafından Eskişehir Teknik Üniversitesi'nde Profesör Dr. Uğur Avdan danışmanlığında yürütülen "Uzaktan Algılama Görüntülerinde Derin Öğrenme Temelli Yaklaşımlar Kullanarak Nesne Tespiti" başlıklı doktora tezinin bir kısmından faydalanılarak yazılmış ve Eskişehir

Teknik Üniversitesi BAP Komisyonu tarafından kabul edilen 1707F461 numaralı proje kapsamında desteklenmiştir. Yazarlar herhangi bir çıkar çatışması beyan etmemektedirler.

Kaynaklar

- Gao, L., He, Y., Sun, X., Jia, X., & Zhang, B. (2019). Incorporating Negative Sample Training for Ship Detection Based on Deep Learning. *Sensors*.
- Han, J., Zhang, D., Cheng, G., Guo, L., & Ren, J. (2015). Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning And High-Level Feature Learning. *Geosci. Remote Sens.*, 3325-3337.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *In Proceedings of the IEEE international Conference on Computer Vision* (pp. 2961-2969).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning For Image Recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Huang, S., & Huang, W. Z. (2016). A New SAR Image Segmentation Algorithm for the Detection of Target and Shadow Regions. *Scientific Reports*.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet Classification With Deep Convolutional Neural Networks. *Advances in neural information processing systems*, 25.
- Lee, J., Bang, J., & Yang, S. (2017). Object Detection With Sliding Window in Images Including Multiple Similar Objects. *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. Jeju, South Korea: IEEE.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single Shot Multibox Detector. *In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.
- Liu, Y. C., Kuang, Z., & Li, G. (2017). Ship Detection and Classification on Optical Remote Sensing Images Using Deep Learning. *ITM Web of Conferences 12*. ITM.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *In Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition* (pp. 779-788).
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in neural information processing systems*, 28.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going Deeper With Convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- Waleed, A. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. *GitHub repository*.
- Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., & Guo, Z. (2018). Automatic Ship Detection In Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks. *Remote sensing*, 10(1), 132.

Öçer, N.E. & Avdan, U.(2024). Mask R-CNN İle Uydu Görüntülerinde Gemi Tespiti. *GSI Journals Serie C: Advancements in Information Sciences and Technologies (AIST)*,7 (1): 40-50

Xia, Y., Wan, S., & Yue, L. (2011, August). A Novel Algorithm for Ship Detection Based on Dynamic Fusion Model of Multi-Feature And Support Vector Machine. *In 2011 Sixth International Conference on Image and Graphics* (pp. 521-526). IEEE.