

Yüksek çözünürlüklü uydu görüntülerinden daha hızlı bölge tabanlı derin öğrenme modeli ile bina tespiti

Building detection from high-resolution satellite images with faster regional based deep learning model

Ekrem SARALIOĞLU*^{1,a}, Oğuz GÜNGÖR^{2,b}

¹ Artvin Çoruh Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, 08000, Artvin

² Ankara Üniversitesi, Uygulamalı Bilimler Fakültesi, Gayrimenkul Geliştirme ve Yönetimi Bölümü, 6000, Ankara

• Geliş tarihi / Received: 20.10.2021

• Düzeltilecek geliş tarihi / Received in revised form: 11.02.2022

• Kabul tarihi / Accepted: 19.02.2022

Öz

İnsan beynindeki öğrenme ve analiz mekanizmasını taklit ederek geniş bir veri kümesinden özellikleri otomatik olarak öğrenmeye çalışan derin öğrenme algoritmaları, yüksek hesaplama gerektiren problemleri çözmeye bazen insanlardan daha başarılı olabilmektedir. Derin öğrenme tabanlı yöntemlerin çeşitli alanlarda başarı ile kullanımı bu yöntemlerin uzaktan algılama alanında da kullanımını arttırmaktadır. Bu çalışmada, yüksek mekânsal çözünürlüğe sahip uydu görüntülerinden derin öğrenme ile otomatik bina tespiti yapılması amaçlanmıştır. Bina tespiti için, ilk olarak yüksek mekânsal çözünürlüklü Worldview-2 uydu görüntüsüne görüntü kaynaştırma işlemi yapılarak mekânsal olarak detayların daha belirgin olduğu kaynaştırılmış bir görüntü elde edilmiştir. Daha sonra çalışma bölgesine ait kaynaştırılmış görüntü bina detaylarının yoğun olduğu bölgeleri içerecek şekilde parçalara ayrılmıştır. Bu görüntü parçalarındaki bina nesnelere etiketlenerek test ve eğitim veri seti oluşturulmuştur. Oluşturulan veri seti ile Faster R-CNN modeli üzerinde ince ayar yapılarak model eğitimi %94 F1 skor ve %88 doğruluk değerlerinde gerçekleştirilmiştir. Sonuç olarak Faster R-CNN modeli ile (genel kullanımının dışında) küçük nesne boyutlu uzaktan algılanmış görüntülerden bina tespiti ortalama %88.6 doğrulukta gerçekleştirilmiştir.

Anahtar kelimeler: Bina tespiti, Derin öğrenme, Faster R-CNN, Yapay zeka

Abstract

Deep learning algorithms, which try to automatically learn features from a large data set to mimic the learning and analysis mechanism in the human brain, have sometimes started to be more successful than humans in solving problems that require high computation. The successful use of deep learning-based methods in various fields also increases its use in remote sensing. This study, it is aimed to make automatic building detection by deep learning from satellite images with high spatial resolution. First, a fused image with more spatial details was obtained by fusing the image to the high spatial resolution Worldview-2 satellite image. Then, the fused image of the study area was divided into parts, including the areas where building details are concentrated. The test and training data set was created by labeling the building objects in these image fragments. Finally, the Faster R-CNN model was trained with the prepared data set, enabling building detection from high spatial resolution satellite images. Building detection was performed with an average accuracy of 88.6% from high-resolution satellite image fragments containing fragments from different regions.

Keywords: Building detection, Deep learning, Faster R-CNN, Artificial intelligence

*a Ekrem SARALIOĞLU; ekremsaralioglu@artvin.edu.tr, Tel: (0539)521 27 94, orcid.org/0000-0002-0609-3338

^b orcid.org/0000-0002-3280-5466

1. Giriş

1. Introduction

Uydu görüntülerinden nesne tespiti; savunma, askeri uygulamalar, coğrafi verilerin oluşturulması, güncellenmesi, şehir çalışmaları, planlama ve izleme gibi çok çeşitli alanlarda büyük önem taşımaktadır (Stankov vd., 2014; Zhang vd., 2015; Cheng & Hang, 2016). Nesne tespiti, belirli bir hava veya uydu görüntüsünde ilgilenilen nesnelerin sınırlayıcı bir şekil kullanılarak (poligon, dikdörtgen, çember vb.) her bir nesnenin konumunun tespiti ve etiketlenmesi ile sağlanmaktadır. Uzaktan algılama alanında uydu görüntülerinden nesne tespiti 1980'li yıllardaki orta mekânsal çözünürlüklü Landsat görüntülerinden günümüze kadar çok yoğun bir uğraş alanı olmuştur (Zhu vd., 2012). Bununla birlikte daha önceki uydu görüntülerinin düşük uzamsal çözünürlüğü bina, araba, ağaç gibi nesnelerin tespitine izin vermemektedir. Bu nedenle araştırmacılar daha çok bu görüntülerden bölge özelliklerini çıkarmaya odaklanmışlardır (Chen & Hang, 2016). Gelişen sensor ve platform teknolojileri sayesinde metre ve metre altı mekânsal çözünürlüğe sahip IKONOS, WorldView-2 (WV-2), QuickBird gibi uydu görüntüleri ve hava fotoğrafları üzerinden daha detaylı doku ve mekânsal detayların elde edilmesi mümkün olmaktadır. Bu görüntülerinin yaygınlaşması ile çok daha fazla yöntem ve algoritma nesne tespiti için geliştirilmiş ve kullanılmıştır (Guo et al., 2018; Stuparu vd., 2020; Arévalo vd., 2006; Han vd., 2014; Barsi ve Heipke, 2003; Senaras vd., 2013; Shufelt, 1996; Kembhavi vd., 2010; Cao vd., 2017; Yu vd., 2016; Cheng vd., 2013).

Nesnelerin tespitinde renk, şekil, doku gibi sadece düşük ve orta seviye özellikleri kullanan Viola-Jones Detektörü (Viola vd., 2001), Deforme Edilebilir Parça Tabanlı Model (Felzenszwalb vd., 2009), Yönlendirilmiş Gradyanların Histogramı (Dalal vd., 2005) gibi geleneksel yöntemlerin genelleştirme yetenekleri sınırlı olmaktadır (Wu vd., 2020; Xiao vd., 2020). Ayrıca bu yöntemlerin farklı uydu görüntülerine uygulanması aşamasında her görüntü için baştan eğitim ve parametre ayarlamaları gerekmektedir. Geleneksel yöntemler göreceli homojen nesnelere içeren küçük görüntülerde umut verici performans sunmalarına

rağmen, gerçek dünya veri kümelerinde daha az başarılı olmaktadır (Fischer vd., 2014). Derin öğrenme tabanlı yöntemlerde ise oluşturulan model doğrudan görüntülerden, metinden veya sestan öğrenme işlemini gerçekleştirmektedir (Arel vd., 2010; Li vd., 2018). Bu sayede insan müdahalelerine aşırı bağımlılık azalmakta ve daha genelleştirilebilir modeller oluşturulmaktadır. Literatürde Tekrarlayan Sinir Ağları (Recurrent Neural Networks), Derin İnanç Ağları (Deep Belief Networks), Çekişmeli Üretici Ağlar (Generative Adversarial Networks), Yığın Otokodlayıcılar (Stack Autoencoders), Kapsül Ağları (Capsule Networks) ve Evrimsel Sinir Ağları (Convolutional Neural Networks (CNN)) tabanlı pek çok derin öğrenme yapısı mevcuttur. Görüntüler üzerinden bilgi çıkarma işlemleri için en çok tercih edilen derin öğrenme yöntemlerini CNN tabanlı yöntemler oluşturmaktadır. CNN ile çok sayıda filtre kullanılarak eğitim veri kümesinden oldukça spesifik özellikler otomatik olarak öğrenilmektedir. Öğrenme gerçekleştikçe filtrelerde bulunan ağırlıklarda sürekli olarak değişmektedir. Temel bir CNN yapısı içerisinde genellikle evrişim katmanları, havuzlama katmanları ve tamamen bağlı katmanlar bulunmaktadır. CNN tabanlı yöntemlerden bölge tabanlı (R-CNN, Fast R-CNN, Faster R-CNN gibi) ve tespit bölge önerisi gerektirmeyen yöntemler (Sadece Bir Kere Bak (You Look Only Once (YOLO)), Tek Atış Çoklu Kutu Dedektörü (Single Shot Multi-Box Detector (SSD)) nesne tespiti aşamasında görüntüler üzerinde çok başarılı bir şekilde kullanılabilir (Guo vd., 2018). Bu yöntemlerin en önemli avantajlarından birisi gerçek zamanlı nesne tespitini mümkün kılmalarıdır. Bu hızlı ve bölgeye dayalı evrimsel sinir ağları PASCAL VOC, Microsoft COCO gibi veri setlerinde büyük boyutlu nesnelerin tespiti alanında önemli bir ilerleme sağlamıştır. PASCAL VOC-2007 (The PASCAL Visual, 2007) veri seti 20 sınıftan oluşmaktadır ve Şekil 1(a) da görüldüğü gibi sınıflar büyük boyutludur. PASCAL VOC veri setindeki sınıfları da içeren Microsoft COCO (Common Objects, 2021) veri seti ise 80 sınıftan oluşmaktadır. Bu veri setlerinden farklı olarak uzaktan algılanmış veriler Şekil 1.(b) de görüldüğü gibi küçük nesne boyutuna sahip, spektral anlamda daha fazla bant içerebilen ve dik iz düşümlü görüntülerdir.



Şekil 1. (a) PASCAL VOC-2007 veri seti, (b) yüksek uzamsal çözünürlüklü uydu görüntüsü
Figure 1. (a) PASCAL VOC-2007 dataset, (b) high spatial resolution satellite imagery

Uzaktan algılama alanında da [Cao vd. \(2016\)](#) uçak tespiti için R-CNN yapısını kullanmıştır. [Ammour vd. \(2017\)](#) R-CNN yaklaşımına benzer olarak CNN'leri ve destek vektör makinelerini (DVM) birleştirerek araç tespitini gerçekleştirmiştir. [Zhang, vd. \(2019\)](#) tarafından ise R-CNN tabanlı gemi tespiti yapılmıştır. R-CNN, Fast R-CNN ve Faster R-CNN şeklinde ilerleyen gelişim süreci sonrası Faster R-CNN birçok alanda nesne tespiti için kullanılmıştır ([Ren vd., 2016](#); [Zhang vd., 2018](#)). [Yang vd., \(2017\)](#) tarafından yapılan çalışmada Faster R-CNN modeli kullanılarak %86 F1 skorunda mikroskop ile görüntülenen hücrelerin tespiti yapılmıştır. [Sardoğan vd., \(2020\)](#) tarafından yapılan çalışmada Faster R-CNN modeli hastalıklı elma yapraklarının tespitinde kullanılmış ve %86 F1 skoru ile bu işlem gerçekleştirilmiştir. [Wan & Goudos \(2020\)](#) tarafından yapılan çalışmada robotik görme sistemleri kullanılarak Faster R-CNN ile çok sınıflı meyvelerin tespiti yapılmıştır. Faster R-CNN modeli uzaktan algılanmış görüntüler üzerinde de kullanılmaya başlamıştır. Uzaktan algılanmış görüntülerindeki nesnelere tespitini klasik bilgisayarla görme (computer vision) alanında kullanılan kırmızı, yeşil ve mavi (RGB) bantlardan oluşan renkli görüntülere oranla çok daha fazla zorluk ortaya çıkarmaktadır ([Zhu vd., 2017](#)). Uzaktan algılanmış görüntüler yüksek irtifalardan elde edildiği için atmosferde bulunan

küçük parçacıkların etkisi, bakış açısı farklılığı, arka plan dağınıklığı, aydınlatma farklılıkları ve küçük nesne boyutu gibi ek zorluklara sahiptir. Bunların dışında uydu görüntüleri klasik RGB görüntülerden daha fazla bant sayısına sahiptir. Genelleştirme yeteneğine sahip yüksek boyutlu derin öğrenme modelleri yüksek doğruluk için oldukça fazla etiketli veriye ihtiyaç duymaktadır. Fakat bilgisayarla görme alanında kullanılan klasik RGB görüntüler ile karşılaştırıldığında uzaktan algılanmış verilerde etiketli veri miktarı oldukça azdır ([Saralioglu & Gungor, 2020](#)). Yeni veri kümesi derin bir ağından baştan eğitilebilmesi için yeterli olmadığı durumda büyük veri setleri ile eğitilmiş ve başarı kazanmış derin öğrenme yapılarında transfer öğrenme ile başarılı sonuçlar elde edilmektedir.

Bu çalışmada yüksek çözünürlüklü uydu görüntülerinden Faster R-CNN modeli ile binaların otomatik tespiti amaçlanmıştır. İnsan faaliyetlerinin çoğunun gerçekleştiği binalar, uzaktan algılanmış görüntüler üzerinde tespit edilebilen en önemli nesnelere biridir. Binalar, bir şehrin oluşumunda temel bir unsur olmaktadır. ([Lafarge vd., 2008](#)) Sürdürülebilir kalkınma ile ilgili araştırmaların yürütülmesi için bina bilgilerinin çıkarılması büyük önem taşımaktadır. Çıkarılan bina bilgileri, insanların yaşam koşullarının değerlendirilmesi, bina koşullarının

izlenmesi, afet risklerinin tahmin edilmesi, 3 boyutlu şehir modelleri, kaçak yapıların tespiti gibi pek çok araştırma için temel bir veri kaynağı olmaktadır (Nahhas vd., 2018). Coğrafi Bilgi Sistemleri ile yapılan pek çok analizde de bina bilgisine ihtiyaç duyulmaktadır. Örneğin iletişim şirketlerinin, daha geniş bir kapsama alanı elde etmek için verici istasyonlarını konumlandırma konusunda uygun yer seçimi yapmaları gerekmektedir. En uygun konumun bulunması için sayısal arazi modeli gibi katmanların yanında bina katmanı da gerekli olmaktadır.

Araştırmacılar tarafından pek çok bina tespiti ve çıkarımı yöntemi önerilmiştir. Bunlardan bazılarını eşik değer kullanımı, morfolojik görüntü işleme, bağlantılı bileşen etiketleme (Maas & Vosselman 1999; Vu vd. 2009; Yu vd. 2010) geometrik kriterlerin kullanımı (Gamba & Houshmand 2000) oluşturmaktadır. Bunların dışında Liow ve Pavlidis (1990) ve Nevatia vd. (1999) tarafından yapılan çalışmalarda gölge analizi tabanlı algoritmalar önerilmiştir. Normal Vektörler, Graf Tabanlı algoritmalar ve Çoklu Çözünürlüklü (Multi-Resolution) analizler (Satari vd., 2012) bina tespiti ve çıkarımında kullanılmıştır. Benzer şekilde ISODATA (Haala & Brenner 1999), Dempster Shafer (Rottensteiner vd., 2007), Rastgele Orman (Guo et al. 2011) ve diğer sınıflandırma tabanlı algoritmalarda bina tespiti ve çıkarımı uygulamalarında kullanılmıştır. Çoğu bina tespit ve çıkarma yöntemleri belirli veri kümeleriyle (Wang & Tenhunen, 2004) iyi performans gösterse de genelleştirilebilme yetenekleri zayıf olmaktadır. Bu yöntemler tipik olarak binaların dokusunu, şeklini ve diğer özelliklerini tam olarak kullanamadığından binalar ile arka planları arasındaki mekânsal ilişkiyi modelleme yetenekleri yetersiz olmaktadır. Aynı zamanda, bina kategorisi altındaki farklı binaların farklı özellikleri ve hatta bölgeler arasındaki belirgin farklılıklar göz önüne alındığında bu yöntemlerin kararlılığı ve genelleştirilebilmeleri zayıf kalmaktadır (Han vd., 2021).

Son zamanlarda derin öğrenme tabanlı yöntemler ile nesne tespiti alanında önemli ilerleme kaydedilmiştir. Derin ağlar ön veri işleme ve özelliklerin tek tek çıkarımına yönelik ek süreçleri içermemektedir. Derin öğrenme ile yüksek seviye özellikler veriler üzerinden otomatik olarak öğrenilmektedir. Bu sayede daha iyi doğruluk ve genelleştirilebilme yeteneklerine sahip olmaktadır. CNN tabanlı yaklaşıma odaklanan bina tespit görevlerinde etkileyici başarı gösteren araştırmalar bulunmaktadır. Örneğin Chen vd. (2017) tarafından yapılan çalışmada bina tespiti

CNN modeli ile yapılmıştır. Sun vd. (2017) tarafından yapılan çalışmada iki seviyeli oluşturulan yeni bir CNN modeli bina tespiti için önerilmiştir. Chen vd., (2019) tarafından yapılan çalışmada bina tespiti için çoklu filtre boyutu ile çoklu görüntü ölççekleri birleşimi ile oluşturulan yeni bir CNN modeli önerilmiştir.

Son araştırmalarda obje tespitinde bölge tabanlı R-CNN (Ren vd., 2015), Fast R-CNN (Girshick 2015), Faster R-CNN ve Mask R-CNN gibi ezberleme sorununa daha dirençli olan ve hızlı tespit yöntemleri kullanılmaktadır. Buna rağmen bu yöntemler ile yüksek çözünürlüklü uydu görüntüleri üzerinde bina tespitinin yapıldığı çok az çalışma mevcuttur. Faster R-CNN modelinin uzaktan algılanmış veriler üzerinde kullanımının genelde hava fotoğrafları (Ammar vd., 2020; Xu vd., 2017) ve Google Earth' den indirilen görüntüler (Yin vd., 2020; Yao vd., 2017; Tao vd., 2021; Zeren vd., 2020; Peng vd., 2020; Han vd., 2020; Liu vd., 2020) üzerinde olduğu görülmektedir. Faster R-CNN modeli şimdiye kadar, çok bantlı uydu görüntüleri üzerinde bina tespiti için doğrudan kullanılmamıştır.

Bu çalışmada Faster R-CNN tabanlı derin öğrenme modelinin yüksek çözünürlüklü uydu görüntülerinden bina tespitinde uygulanması gerçekleştirilerek literatürde önemli bir boşluk doldurulmaya çalışılmış ve uydu görüntüleri üzerinde bölgesel tabanlı hızlı derin öğrenme yöntemleri ile obje tespiti dikkat çekilerek yeni çalışmalar için bir kapı açmıştır. Faster R-CNN modelinin yaygın kullanımı büyük boyutlu objeler içeren görüntülerde obje tespitini içeren çalışmalar olmaktadır. Bu çalışmada Faster R-CNN modeli, küçük boyutlu objelerin tespit edilmesine imkan verecek şekilde düzenlenmiş ve model oluşturulan yeni bir veri seti ile eğitilerek uydu görüntülerinden bina tespiti gerçekleştirilmiştir. Faster R-CNN gibi hızlı nesne tespitine imkân sağlayan bu tarz modellerin uydu görüntüleri üzerinde kullanımı anlık nesne tespiti gibi önemli bir avantaj sağlar. Bunun için çalışma bölgesine ait WV-2 uydu görüntüsü ilk olarak siyah-beyaz (pankromatik) bandı ile kaynaştırılarak yüksek mekânsal çözünürlüklü pan keskinleştirilmiş görüntü elde edilmiştir. Daha sonra kaynaştırılmış görüntü 500x500 piksel boyutlarında parçalara bölünmüştür. Oluşturulan parçalar üzerinde binaların etiketleme işlemleri yapılmış modelin eğitiminde kullanılacak veri seti eğitim ve test olarak ayrılmıştır. Son olarak oluşturulan veri seti ile eğitilen Faster R-CNN modeli ile WV-2 görüntüleri üzerinde otomatik bina tespiti gerçekleştirilmiştir.

2. Yöntem

2. Method

2.1. Çalışma alanı ve veri

2.1. Study area and data

Bu çalışmada Trabzon iline bağlı bir ilçe olan (Kuzeydoğu Türkiye'de Karadeniz kıyısında yer alan) Sürmene'ye ait WV-2 görüntüsü kullanılmıştır. WV-2 uydusu, 0,46 m mekânsal çözünürlükle 450 ila 800 nm spektrumu kapsayan siyah-beyaz bir bant sağlamaktadır. Ayrıca, 1,84 m mekânsal çözünürlükte, 400 ila 1050 nm arasındaki spektral aralığı kapsayan sekiz adet MS bandı (yani kırmızı, yeşil, mavi ve yakın kızılötesi-1 olmak

üzere dört standart bant ve kıyı, sarı, kırmızı kenar ve yakın kızılötesi 2 olmak üzere dört ek bant) içermektedir (Padwick vd., 2010). Bölge çok engebeli ve dağlık bir araziye sahip olduğundan evler genellikle birbirinden uzak olarak inşa edilmiştir. Trabzon merkeze göre daha fazla yağışlı olan Sürmene'de her mevsim yağışlıdır ve yağış miktarı 1500 mm'yi bulmaktadır. Beton malzeme yıl boyunca yağışlara daha dayanıklı olduğu için yollar genellikle betondan yapılmıştır. Bu durum piksel tabanlı sınıflandırma algoritmaları kullanıldığında çatısız evlerin yol ve çıplak toprak sınıfları ile karıştırılmasına neden olmaktadır (Saralioğlu & Gungor., 2022). Şekil 2'de çalışma alanı gösterilmektedir.



Şekil 2. Çalışma alanı (WV-2 görüntüsü)

Figure 2. Study area (WV-2 image)

2.2. Görüntü kaynaştırma

2.2. Image fusion

Çok bantlı uydu görüntülerinin çoğu renkli bantlarından daha yüksek mekânsal çözünürlükte bir siyah-beyaz bant da sunmaktadır. Bu çalışmada siyah-beyaz bantın mekânsal çözünürlüğünde renkli görüntü elde edebilmek için Gram-Schmidt görüntü kaynaştırma yöntemi uygulanmıştır. Gram-Schmidt yöntemi etkili görüntü kaynaştırma yöntemlerinden biridir (Karathanassi vd., 2007; Klonus & Ehlers, 2009; Yuhendra & Kuze, 2011).

Bu yöntemin önemli avantajlarından biri, bu teknik kullanılarak tek seferde işlenebilecek bant sayısında bir sınırlama olmamasıdır. İkinci olarak, orijinal MS verilerinin spektral özellikleri birleştirilmiş görüntüde korunmaktadır (Liv vd., 2004). WV-2 görüntüsüne ait renkli (1.84 m) ve siyah-beyaz (0.46 m) bantlar kaynaştırılarak 0.46 m mekânsal çözünürlükte kaynaştırılmış renkli görüntü elde edilmiştir. Şekil 3'te renkli, siyah-beyaz ve kaynaştırılmış görüntü örneği gösterilmektedir.



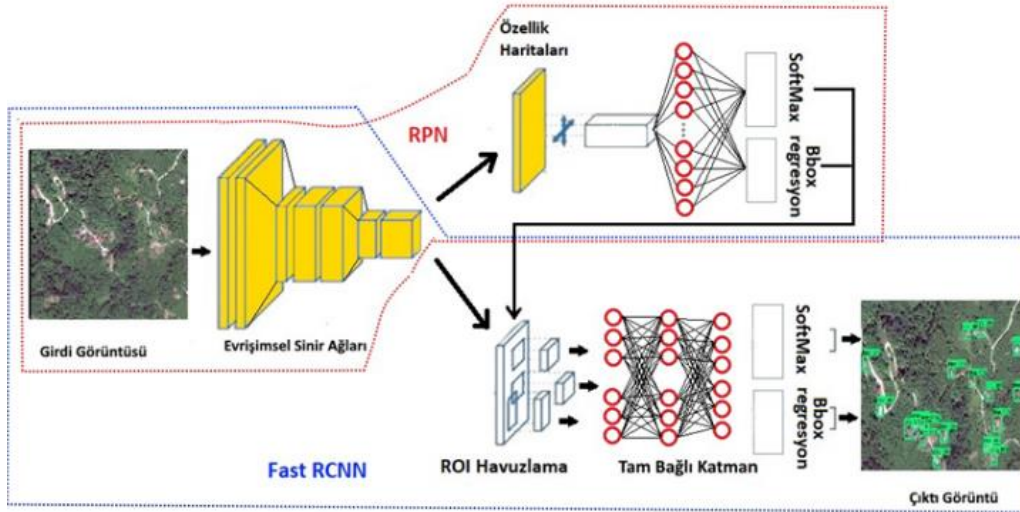
Şekil 3. (a) siyah-beyaz bant, (b) doğru renkli görüntü, (c) kaynaştırılmış görüntü
Figure 3. (a) panchromatic band, (b) true color image, (c) pansharpened image

2.3. Faster R-CNN modeli

2.3.1. Faster R-CNN model

Faster R-CNN, Şekil 4'te görüldüğü gibi bir bölge öneri ağı (Region Proposal Network (RPN)) (Ren vd., 2016) ve Fast R-CNN (Girshick, 2015) olmak üzere iki modülden oluşmaktadır. RPN'in getirdiği yenilik, doğrudan örnekleme katmanına bağlanılabilesidir. Böylece Faster R-CNN, güçlü yapısı ile görüntülerde uçtan uca nesne tespitini sağlayabilecek bir yapıda olmaktadır (Ren vd.,

2016). RPN bölgeleri ve hedefleri tahmin skorları ile birlikte üreten bir evrimsel sinir ağıdır. Fast R-CNN ise boyutu tanımlayan ve hedefleri sınıflandıran bir ağıdır. Önerilen bölgelerin sınıflandırma katmanı ve skor değerleri tahmin katmanı aşamasında Fast R-CNN ağına birleştirilir. Fast R-CNN ağının iki çıkış katmanı bulunmaktadır. Bunlardan ilki Softmax sınıflayıcı katmanı diğeri ise tespit edilen bölgenin tespit edilme doğruluğunu veren regresyon katmanıdır (Yang vd., 2017).



Şekil 4. Faster R-CNN ile bina tespit süreci, kırmızı çizgi ile çevrili bölge RPN, mavi çizgi ile çevrili bölge Fast R-CNN

Figure 4. Building detection process with Faster R-CNN, area surrounded by red line RPN, area surrounded by blue line Fast R-CNN

Faster R-CNN işleyişi temelde 4 aşamadan oluşmaktadır. İlk aşamada evrimsel sinir ağları ile girdi görüntüsünden özellikler çıkartarak özellik haritaları oluşturulmaktadır. İkinci aşamada üretilen özellik haritaları RPN ile bölge önerilerine dönüştürür. Bölge önerileri, içerisinde nesne olma olasılığı yüksek olacak şekilde oluşturulan

bölgeleri içermektedir. Üçüncü aşamada bölge önerilerinin boyutlarının aynı olmasını sağlamak amacıyla ilgi bölgesi (region of interest (ROI)) havuzlama katmanı uygulanmaktadır. Son aşamada önerilen nesnelere sınıflandırmak için bir Softmax katmanı ve nesnelere sınırlarını belirlemek için bir regresyon katmanını içeren

tam bağlı ağ kullanılarak sonuç ürün elde edilmektedir. Bu çalışmada kullanılan Faster R-CNN modelinin bir Inception ResNet (Szegey vd., 2016) omurgası vardır. Bu özellik çıkarımı için yaygın kullanılan CNN yapılarından bir tanesidir.

2.4. Eğitim sınıflarının oluşturulması ve model ayarları

2.4. Producing of training classes and model settings

Faster R-CNN modelinin yüksek mekânsal çözünürlüklü görüntülerde bina tespitinde kullanılabilmesi için WV-2 kaynaştırılmış görüntüsü eğitim için hazırlanmıştır. Kaynaştırılmış görüntünün RGB bantları modelin eğitimi için kullanılmıştır. İlk olarak binaları içeren görüntü parçaları Şekil 5'te görüldüğü gibi parçalara ayrılarak oluşturulmuştur. Görüntü boyutları tespit edilmek istenen nesnenin

boyutundan küçük olmamalıdır. Bunun dışında çok büyük görüntü parçaları üzerinde çalışmak işlem yükünü önemli derecede arttırarak uzun zaman almaktadır. Bu çalışma için birkaç tane binayı rahat bir şekilde içerisine alabilecek şekilde 500x500 piksel boyutlarında görüntü parçaları oluşturulmuştur. Görüntü parçalarında oluşan boş alanlar ve bina içermeyen parçalar silinmiştir. Bu görüntüler %80'i eğitim ve %20'si test olacak şekilde iki gruba ayrılmıştır. Eğitim verisi 2350 bina nesnesi içeren toplamda 96 görüntü parçasından oluşurken test verisi de 365 bina nesnesi içeren 24 görüntü parçasından oluşmaktadır. Ön hazırlık işlemlerinden sonra binaların etiketleme işlemleri LabelImg (Tzutalin, 2015) ile yapılmıştır. LabelImg Github'daki açık kaynaklı grafik görüntü etiketleme yazılımıdır. Görüntülerdeki nesnelere sınırlayıcı kutucuklar çizilerek açıklama eklenerek bina sınıfları oluşturulmuştur.



Şekil 5. Görüntü parçaları
Figure 5. Image patches

Bu çalışma i7 9750h işlemci (4.50GHz), 16GB DDR4 2666MHz RAM ve NDVI Geforce 1650 ekran kartına sahip bilgisayar ile gerçekleştirilmiştir. Model eğitimi Tensorflow kütüphanesi kullanılarak Python programa dili ile gerçekleştirilmiştir. TensorFlow Nesne Algılama API'si, eğitim ve değerlendirme süreci için protobuf dosyalarını kullanır. Eğitim kümesindeki veriler ön-eğitilmiş bir ağ olan Microsoft COCO veriseti ile eğitilmiş faster_rcnn_inception_v2_coco modeli (GitHub, Tensorflow, 2019) kullanılmıştır. Microsoft COCO veri seti Pascal-VOC veri seti gibi büyük boyutlu klasik RGB görüntüleri içeren büyük kapsamlı bir veri setidir. Uzaktan algılanmış verilerle farklı yapıda olan bu veri seti ile eğitilmiş bu ağa hazırlanan WV-2 eğitim veri setini entegre etmek için aşağıdaki şekilde hiperparametreler düzenlenmiştir.

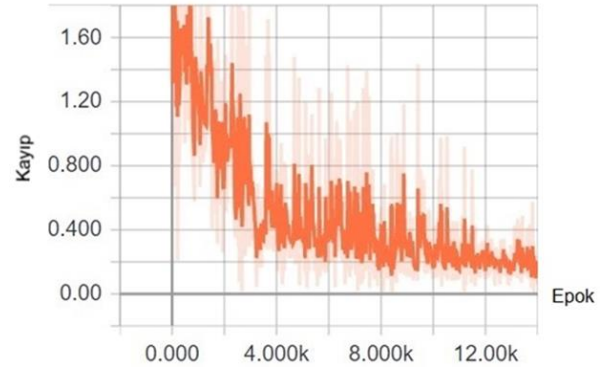
Model eğitimi için görüntü ve binalara ait koordinat bilgilerini içerecek şekilde oluşturulan csv dosyaları tfrecord dosyasına dönüştürülmüştür. Bu dosya eğitim aşamasında girdi olarak modele tanıtılmaktadır. Eğitim sırasında model ile ilgili öğrenilen detaylar "cpkt" uzantılı dosyada tutulmaktadır. Bu sayede eğitim aşamasına daha sonra kalındığı yerden devam edilebilmektedir. Labelmap dosyası modelde hangi sınıfın hangi ID bilgisine denk geldiğini tanıtmak için oluşturulmuştur. Bu çalışmada tek bir sınıf olduğu için sınıf adı: Bina ve ID:1 olacak şekilde dosya hazırlanmıştır. Model_config dosyasında bu çalışmada sadece binalar tespit edileceği için çıkış sınıfı 1 olarak değiştirilmiştir. Sonuç ürünler için yol tanımlanarak son olarak test için kullanılacak veri sayısı girilmiştir. Öğrenme oranı için kademeli olarak azalan öğrenme modeli kullanılmıştır. Bu yöntemde, eğitimin belirli kademelerinde öğrenme hızı azaltılarak modelin başarımı artırılabilir. Bunun için ilk 90000 adım için öğrenme oranı 0.0002 ve sonra 160000'e kadar olan kısım için öğrenme oranı 0.0001 olarak belirlenmiştir. Başlangıç ağırlık değeri 0 olarak seçilmiştir. Sınıflandırma katmanı olarak başarılı sonuçlar vermesi nedeniyle softmax fonksiyonu kullanılmıştır. Kayıp değerleri incelendiğinde çok fazla model hassasiyetinde artış olmaması nedeniyle 15000 adımda eğitim yapılmıştır. Görüntülerin çoğaltılarak daha iyi öğrenme sağlanabilmesi amacıyla veri çoğaltma yöntemi olarak "random_horizontal_flip" kullanılmıştır. Pencere ölçüğü küçük boyutlu bina nesnelere

tespit edilebilmesi amacıyla 0.25, 0.5, 1 ve 2 olarak ayarlanmıştır. Faster R-CNN mimarisi içerisinde Resnet 50 (He vd., 2016) mimarisi kullanılmıştır. Özellik çıkarıcı olarak VGG16 (URL-5), Resnet 101 (He vd., 2016) yerine Resnet50 daha az grafik belleği istemesi ve küçük nesnelere tespitindeki başarısı nedeniyle tercih edilmiştir.

3. Bulgular ve tartışma

3. Results and discussion

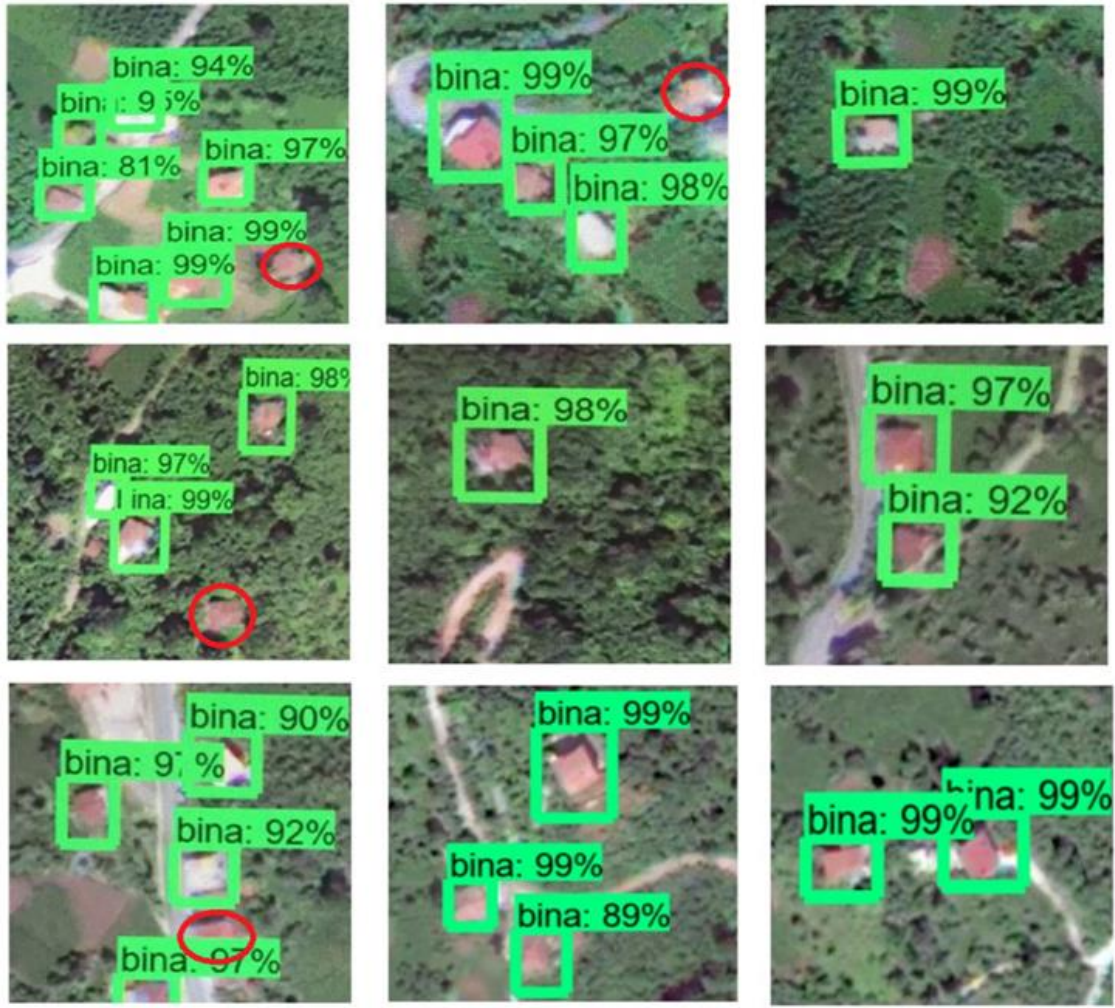
Eğitim aşamasındaki kayıp değerleri anlık olarak Tensorflow görselleştirme aracı olan TensorBoard üzerinden izlenebilmektedir. Modelin eğitimi yaklaşık 32 saat sürmüştür. Eğitim aşamasında oluşan kayıp fonksiyonu değerinin zamanla değişim grafiği Şekil 6' da görülmektedir. Yatay eksen eğitim aşamasındaki adım sayısını gösterirken dikey eksen kayıp değerlerini göstermektedir. Hata değerinin sürekli olarak düşmesi ve sifıra yaklaşması istenen durumdur. Şekil 6'da görüldüğü gibi 1.80 ile başlayan kayıp değerleri 0.1'e kadar düşmüştür. Bu modelin geliştiğini göstermektedir.



Şekil 6. Kayıp değerleri

Figure 6. Loss

Eğitilen Faster R-CNN modeli WV-2 uydu görüntüsü üzerinde Şekil 7'de görüldüğü gibi binaları otomatik olarak tespit edebilmektedir. Şekil 7'de Kırmızı ile çevrili alanlar modelin tespit edemediği binaları göstermektedir. Model tespit işlemini yaklaşık 1sn gibi kısa bir sürede gerçekleştirebilmektedir. Şekil 7'de görüldüğü gibi bina çatılarındaki farklı renk tonlarından etkilenmemekte ve yüksek doğrulukta binaları otomatik tespit edebilmektedir.



Şekil 7. Faster R-CNN modeli ile bina tespiti
Figure 7. Building detection with Faster R-CNN model

Bu çalışmada bina tespit doğruluğunu nicel olarak değerlendirmek için yaygın olarak kullanılan doğruluk ve F1 puan metrikleri kullanılmıştır.

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (1)$$

$$\text{Keskinlik} = \frac{DP}{DP+YP} \quad (2)$$

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (3)$$

$$\text{F1Puanı} = \frac{2 \times \text{Keskinlik} \times \text{Duyarlılık}}{\text{Keskinlik} + \text{Duyarlılık}} \quad (4)$$

Burada DP doğru pozitif, DN doğru negatif, YP yanlış pozitif ve YN yanlış negatiftir. DP ilgilenilen nesnenin görüntüde tespit edilmesi, yani sınırlayıcı kutu gösterilmesidir. YN, görüntüde istenen nesnenin olmasına rağmen modelin bunu tespit edememesidir. DN görüntüde istenen nesne olmadığında nesne tespiti yapılmayan durumu ifade etmektedir. YP, bir nesnenin istenen nesne olmamasına rağmen istenen nesne olarak tespit edilmesidir.

Doğruluk değerlendirme aşamasında tüm görüntü parçalarının %80'i eğitim ve %20'si test olacak şekilde Python programlama dili ile rastgele seçilerek iki gruba ayrılmıştır. Eğitim verisi 96 görüntü içerisinde bulunan 2350 bina içermektedir. Test verisi 24 görüntü ve 365 bina nesnesi içermektedir. Faster R-CNN ile yapılan bina tespiti sonuçları Tablo 1.'de gösterilmektedir.

Tablo 1. Modele ait doğruluk değerlendirme sonuçları

Table 1. Accuracy evaluation results of the model

Doğruluk değerlendirme metrikleri	Sonuçlar
F1-Skor	0.94
mAP	0.91
Doğruluk	0.88
Keskinlik	1.00
Duyarlılık	0.88

Model daha fazla veri ile eğitilmesi durumunda daha da yüksek doğrulukta bina tespiti yapılabilir. Bununla birlikte Faster R-CNN modeli uydu

görüntülerinden nesne tespiti amacıyla üretilmediğinden 3 bant için çalıştırılabilmektedir. Ayrıca model büyük boyutlu nesnelere için eğitildiğinden uzaktan algılanmış görüntülerdeki küçük boyutlu nesnelere zorluk oluşturmaktadır. Bunların yanı sıra modelin eğitim süresi uzun olmaktadır. Bu durum sınıf sayısı arttıkça çok daha uzayacaktır. Fakat tek bir eğitimden sonra saniyeler içinde yüksek doğrulukta nesne tespitine imkan sağlaması çok önemli bir avantajdır.

4. Sonuçlar

4. Conclusion

Bu çalışmada Faster R-CNN modeli küçük boyutlu objelerin tespit edilmesine imkan verecek şekilde düzenlenmiş ve oluşturulan yeni bir veri seti ile eğitilerek uydu görüntülerinden bina tespiti gerçekleştirilmiştir. Çalışmada kullanılan veri setinde binalar etiketlenerek modelin eğitimi gerçekleştirilmiştir. Modelin en önemli avantajlardan bir tanesi yeni bir uydu görüntü parçası verildiğinde tekrardan eğitime ihtiyaç duyulmadan otomatik olarak binaları tespit edebilmesidir. Bunun yanında gerçek zamanlı nesne tespitinde klasik RGB görüntüler üzerinde kullanılan Faster R-CNN modelini uydu görüntüleri üzerinde bina tespiti amacıyla kullanıldığı durumda da yine saniyeler içerisinde bina tespiti yapılabildiği görülmüştür. Bina tespit doğruluğu %88.6 olmaktadır. Ayrıca tespit süresi yaklaşık 1sn gibi çok kısa bir sürede gerçekleştirilmektedir. Bu sonuçlar, yüksek mekânsal çözünürlüklü farklı uydu görüntüleri ile eğitilmesi durumunda Faster-RCNN modelinin bina tespitini yüksek başarı ile yapılabileceğini göstermektedir.

Teşekkür

Acknowledgement

Bu çalışmada kullanılan WorldView-2 uydu görüntüsü 111Y296 numaralı TÜBİTAK 1001 projesinden alınmıştır. Bu nedenle TÜBİTAK'a sağladığı destekten dolayı teşekkür ederiz.

Yazar katkısı

Author Contribution

Bu çalışmada Saralioğlu modelin oluşturulması, eğitim verilerinin oluşturulması, literatür taraması ve makalenin yazımı aşamalarını gerçekleştirmiştir. Güngör makalenin eleştirel incelemesi, geliştirilmesi ve editlenmesini gerçekleştirmiştir.

Etik beyanı

Declaration of ethical code

Bu makalenin yazarları, mevcut çalışmada kullanılan materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel izin gerektirmediğini beyan etmektedir.

Çıkar çatışması beyanı

Conflicts of interest

Mevcut çalışma kapsamında herhangi bir çıkar çatışması olmadığı yazarlar tarafından beyan edilmektedir.

Kaynaklar

References

- Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE Computational Intelligence Magazine*, 5(4), 13-18. <https://doi.org/10.1109/MCI.2010.938364>
- Arévalo, V., González, J., Valdes, J., & Ambrosio, G. (2006). Detecting shadows in QuickBird satellite images. *ISPRS Commission VII Midterm Symposium Remote Sensing: From Pixels to Processes*, (pp. 8-11), Enschede, Netherlands.
- Ammar, A., Koubaa, A., Ahmed, M., & Saad, A. (2019, October 16). *Aerial images processing for car detection using convolutional neural networks: Comparison between faster r-cnn and yolov3*. <https://arxiv.org/abs/1910.07234>
- Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., & Zuair, M. (2017). Deep learning approach for car detection in UAV imagery. *Remote Sensing*, 9(4), 1-15. <https://doi.org/10.3390/rs9040312>
- Barsi, A., & Heipke, C. (2003). Artificial neural networks for the detection of road junctions in aerial images. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/W8), 113-118.
- Cao, Y., Niu, X., & Dou, Y. (2016). Region-based convolutional neural networks for object detection in very high resolution remote sensing images. *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (ss. 548-554). Changsha: IEEE. <https://doi.org/10.1109/FSKD.2016.7603232>
- Cao, L., Luo, F., Chen, L., Sheng, Y., Wang, H., Wang, C., & Ji, R. (2017). Weakly supervised vehicle detection in satellite images via multi-instance discriminative learning. *Pattern Recognition*, 64,

- 417-424.
<https://doi.org/10.1016/j.patcog.2016.10.033>
- Chen, C., Gong, W., Hu, Y., Chen, Y., & Ding, Y. (2017). Learning oriented region-based convolutional neural networks for building detection in satellite remote sensing images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 461. <https://doi.org/10.5194/isprs-archives-XLII-1-W1-461-2017>
- Chen, C., Gong, W., Chen, Y., & Li, W. (2019). Learning a two-stage CNN model for multi-sized building detection in remote sensing images. *Remote Sensing Letters*, 10(2), 103-110. <https://doi.org/10.1080/2150704X.2018.1528398>
- Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117, 11-28. <https://doi.org/10.1016/j.isprsjprs.2016.03.014>
- Cheng, G., Han, J., Guo, L., Qian, X., Zhou, P., Yao, X., & Hu, X. (2013). Object detection in remote sensing imagery using a discriminatively trained mixture model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 85, 32-43. <https://doi.org/10.1016/j.isprsjprs.2013.08.001>
- Common Objects in Context (COCO). (2021, April 12). <https://cocodataset.org/#home>.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (pp. 886-893). IEEE. <https://doi.org/10.1109/CVPR.2005.177>
- Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 1627-1645. <https://doi.org/10.1109/TPAMI.2009.167>
- Fischer, P., Dosovitskiy, A., & Brox, T. (2014, May 22). *Descriptor matching with convolutional neural networks: a comparison to sift*. <https://arxiv.org/abs/1405.5769>
- Gamba, P., & Houshmand, B. (2000). Digital surface models and building extraction: A comparison of IFSAR and LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4), 1959-1968. <https://doi.org/10.1109/36.851777>
- GitHub. Tensorflow detection model zoo. (2019, October 10). https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_model_zoo.md
- Guo, W., Yang, W., Zhang, H., & Hua, G. (2018). Geospatial object detection in high resolution satellite images based on multi-scale convolutional neural network. *Remote Sensing*, 10(1), 131. <https://doi.org/10.3390/rs10010131>
- Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), 56-66. <https://doi.org/10.1016/j.isprsjprs.2010.08.007>
- 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* (pp. 91-99), Massachusetts.
- Haala, N., & Brenner, C. (1999). Extraction of buildings and trees in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), 130-137. [https://doi.org/10.1016/S0924-2716\(99\)00010-6](https://doi.org/10.1016/S0924-2716(99)00010-6)
- Han, Y., Ma, S., Zhang, F., & Li, C. (2020). Object detection of remote sensing airport image based on improved Faster R-CNN. *Journal of Physics: Conference Series*. IOP Publishing. <https://doi.org/10.1088/1742-6596/1601/3/032010>
- Han, Q., Yin, Q., Zheng, X., & Chen, Z. (2021). Remote sensing image building detection method based on Mask R-CNN. *Complex & Intelligent Systems*, <https://link.springer.com/article/10.1007/s40747-021-00322-z>
- Han, J., Zhang, D., Cheng, G., Guo, L., & Ren, J. (2014). Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6), 3325-3337. <https://doi.org/10.1109/TGRS.2014.2374218>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- Karathanassi, V., Kolokousis, P., & Ioannidou, S. (2007). A comparison study on fusion methods using evaluation indicators. *International*

- Journal of Remote Sensing*, 28(10), 2309–2341. <https://doi.org/10.1080/01431160600606890>
- Kembhavi, A., Harwood, D., & Davis, L. S. (2010). Vehicle detection using partial least squares. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 33(6), 1250-1265. <https://doi.org/10.1109/TPAMI.2010.182>
- Klonus, S., & Ehlers, M. (2009). Performance of evaluation methods in image fusion. *2009 12th International Conference on Information Fusion* (pp.1409–1416). Seattle, WA: IEEE.
- Lafarge, F., Descombes, X., Zerubia, J., & Pierrot-Deseilligny, M. (2008). Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(3), 365-381. <https://doi.org/10.1016/j.isprsjprs.2007.09.003>
- Li, C., Liu, L., Wang, J., Zhao, C., & Wang, R. (2004, September). Comparison of two methods of the fusion of remote sensing images with fidelity of spectral information. *IGARSS 2004, 2004 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 4, pp. 2561-2564). IEEE. <https://doi.org/10.1109/IGARSS.2004.1369819>
- Li, X., Shang, M., Qin, H., & Chen, L. (2015). Fast accurate fish detection and recognition of underwater images with fast r-cnn. In *OCEANS 2015-MTS/IEEE* Washington (pp. 1-5). IEEE. <https://doi.org/10.23919/OCEANS.2015.7404464>
- Li, Y., Zhang, H., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6), 1-17. <https://doi.org/10.1002/widm.1264>
- Liow, Y. T., & Pavlidis, T. (1990). Use of shadows for extracting buildings in aerial images. *Computer Vision, Graphics, and Image Processing*, 49(2), 242-277. [https://doi.org/10.1016/0734-189X\(90\)90139-M](https://doi.org/10.1016/0734-189X(90)90139-M)
- Liu, X., Ghazali, K. H., Han, F., Mohamed, I. I., Zhao, Y., & Ji, Y. (2020). Oil palm tree detection and counting in aerial images based on Faster R-CNN. In *ECCE2019* (pp. 475-482). Springer, Singapore.
- Maas, H. G., & Vosselman, G. (1999). Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), 153-163. [https://doi.org/10.1016/S0924-2716\(99\)00004-0](https://doi.org/10.1016/S0924-2716(99)00004-0)
- Nahhas, F. H., Shafri, H. Z., Sameen, M. I., Pradhan, B., & Mansor, S. (2018). Deep learning approach for building detection using lidar-orthophoto fusion. *Journal of Sensors*, 2018 (7212307), 1-12. <https://doi.org/10.1155/2018/7212307>
- Nevatia, R., Huertas, A., & Kim, Z. (1999). The MURI Project for rapid feature extraction in urban areas. *International Archives of Photogrammetry and Remote Sensing*, 32(3-2W5), 3–14. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.2460&rep=rep1&type=pdf>
- Padwick, C., Deskevich, M., Pacifici, F., & Smallwood, S. (2010). WorldView-2 pan-sharpening. *Proceedings of the ASPRS 2010 Annual Conference*, San Diego, CA, USA. Vol. 2630, pp. 1-14. <http://www.asprs.org/wp-content/uploads/2013/08/Padwick.pdf>
- Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., & Ye, H. (2020). Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 364-376. <https://doi.org/10.1016/j.isprsjprs.2020.08.026>
- Rottensteiner, F., Trinder, J., Clode, S., & Kubik, K. (2007). Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(2), 135-149. <https://doi.org/10.1016/j.isprsjprs.2007.03.001>
- Saralioğlu, E., & Gungor, O. (2020). Crowdsourcing in remote sensing: a review of applications and future directions. *IEEE Geoscience and Remote Sensing Magazine*, 8(4), 89-110. <https://doi.org/10.1080/01431160600606890> 10.1109/MGRS.2020.2975132
- Saralioğlu, E., & Gungor, O. (2022) Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network. *Geocarto International*, 37(2), 657-677, <https://doi.org/10.1080/10106049.2020.1734871>
- Sardoğan, M., Yunus, Ö., & Tuncer, A. (2020). Detection of Apple Leaf Diseases using Faster R-CNN. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 8(1), 1110-1117. <https://doi.org/10.29130/dubited.648387>
- Satari, M., Samadzadegan, F., Azizi, A., & Maas, H. G. (2012). A multi-resolution hybrid approach for building model reconstruction from lidar data. *The Photogrammetric Record*, 27(139), 330-359. <https://doi.org/10.1111/j.1477-9730.2012.00688.x>

- Senaras, C., Ozay, M., & Vural, F. T. Y. (2013). Building detection with decision fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3), 1295-1304. <https://doi.org/10.1109/JSTARS.2013.2249498>
- Shufelt, J. A. (1996). Exploiting photogrammetric methods for building extraction in aerial images. *International Archives of Photogrammetry and Remote Sensing*, 31(B6), 74-79.
- Stuparu, D. G., Ciobanu, R. I., & Dobre, C. (2020). Vehicle detection in overhead satellite images using a one-stage object detection model. *Sensors*, 20(22), 6485. <https://doi.org/10.3390/s20226485>
- Sun, L., Tang, Y., & Zhang, L. (2017). Rural building detection in high-resolution imagery based on a two-stage CNN model. *IEEE Geoscience and Remote Sensing Letters*, 14(11), 1998-2002. <https://doi.org/10.1109/LGRS.2017.2745900>
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016, February 23). *Inception-v4, inception-resnet and the impact of residual connections on learning*. <https://arxiv.org/abs/1602.07261>
- Stankov, K., & He, D.C. (2014). Detection of buildings in multispectral very high spatial resolution images using the percentage occupancy hit-or-miss transform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(10), 4069-4080. <https://doi.org/10.1109/JSTARS.2014.2308301>
- Tao, H., Runsheng, L., Qing, H., & Qingqing, K. (2021). Aircraft target recognition in optical remote sensing image with Faster R-CNN. *6th International Symposium of Space Optical Instruments and Applications*. Delft, the Netherlands, September 24-25, 2019 (pp. 55-63). Springer International Publishing. https://doi.org/10.1007/978-3-030-56488-9_6
- The PASCAL Visual Object Classes Challenge 2007 (2021, April 12). <http://host.robots.ox.ac.uk/pascal/VOC/voc2007>
- Tzutalin. LabelImg. Git code (2020 August 12) <https://github.com/tzutalin/labelImg>
- Viola, P., & Jones, M. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition*. CVPR 2001 (Vol. 1, pp. I-I). IEEE. <https://doi.org/10.1109/CVPR.2001.990517>
- Vu, T. T., Yamazaki, F., & Matsuoka, M. (2009). Multi-scale solution for building extraction from LiDAR and image data. *International Journal of Applied Earth Observation and Geoinformation*, 11(4), 281-289. <https://doi.org/10.1016/j.jag.2009.03.005>
- Yu, B., Liu, H., Wu, J., Hu, Y., & Zhang, L. (2010). Automated derivation of urban building density information using airborne LiDAR data and object-based method. *Landscape and Urban Planning*, 98(3-4), 210-219. <https://doi.org/10.1016/j.landurbplan.2010.08.004>
- Yu, Y., Guan, H., Zai, D., & Ji, Z. (2016). Rotation-and-scale-invariant airplane detection in high-resolution satellite images based on deep-Hough-forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 112, 50-64. <https://doi.org/10.1016/j.isprsjprs.2015.04.014>
- Wan, S., & Goudos, S. (2020). Faster R-CNN for multi-class fruit detection using a robotic vision system. *Computer Networks*, 168(107036). <https://doi.org/10.1016/j.comnet.2019.107036>
- Wang, Q., & Tenhunen, J. D. (2004). Vegetation mapping with multitemporal NDVI in North Eastern China transect (NECT). *International Journal of Applied Earth Observation and Geoinformation*, 6(1), 17-31. <https://doi.org/10.1016/j.jag.2004.07.002>
- Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, 396, 39-64. doi: 10.1016/j.neucom.2020.01.085
- Xiao, Y., Tian, Z., Yu, J., Zhang, Y., Liu, S., Du, S., & Lan, X. (2020). A review of object detection based on deep learning. *Multimedia Tools and Applications*, 79(33), 23729-23791. <https://doi.org/10.1007/s11042-020-08976-6>
- Xu, Y., Yu, G., Wang, Y., Wu, X., & Ma, Y. (2017). Car detection from low-altitude UAV imagery with the faster R-CNN. *Journal of Advanced Transportation*, 2017(2823617), 1-11. <https://doi.org/10.1155/2017/2823617>
- Very Deep Convolutional Networks for Large-Scale Visual Recognition (2019 October, 15). http://www.robots.ox.ac.uk/~vgg/research/very_deep.
- Yang, S., Fang, B., Tang, W., Wu, X., Qian, J., & Yang, W. (2017). Faster R-CNN based microscopic cell detection. *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)* (pp. 345-350). IEEE. <https://doi.org/10.1109/SPAC.2017.8304302>
- Yao, Y., Jiang, Z., Zhang, H., Cai, B., Meng, G., & Zuo, D. (2017). Chimney and condensing tower detection based on faster R-CNN in high resolution remote sensing images. *2017 IEEE International Geoscience and Remote Sensing*

- Symposium (IGARSS)* (pp. 3329-3332). IEEE. <https://doi.org/10.1109/IGARSS.2017.8127710>
- Yin, S., Li, H., & Teng, L. (2020). Airport detection based on improved faster RCNN in large scale remote sensing images. *Sensing and Imaging*, 21(1), 1-13. <https://doi.org/10.1007/s11220-020-00314-2>
- Yuhendra, J., & Kuze, H. (2011). Performance analyzing of high resolution pan-sharpening techniques: Increasing image quality for classification using supervised kernel support vector machine. *Research Journal of Information Technology*, 8(1), 12–28.
- Zeren, M. T., Aytulun, S. K., & Kirelli, Y. (2020). Comparison of SSD and Faster R-CNN algorithms to detect the airports with data set which obtained from unmanned aerial vehicles and satellite images. *Avrupa Bilim ve Teknoloji Dergisi*, (19), 643-658. <https://doi.org/10.31590/ejosat.742789>
- Zhang, L., Shi, Z., & Wu, J. (2015). A hierarchical oil tank detector with deep surrounding features for high-resolution optical satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10), 4895-4909. <https://doi.org/10.1109/JSTARS.2015.2467377>
- Zhang, S., Wu, R., Xu, K., Wang, J., & Sun, W. (2019). R-CNN-based ship detection from high resolution remote sensing imagery. *Remote Sensing*, 11(6), 631. <https://doi.org/10.3390/rs11060631>
- Zhang, Q. X., Lin, G. H., Zhang, Y. M., Xu, G., & Wang, J. J. (2018). Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images. *Procedia Engineering*, 211, 441-446. <https://doi.org/10.1016/j.proeng.2017.12.034>
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36. <https://doi.org/10.1109/MGRS.2017.2762307>
- Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83-94. <https://doi.org/10.1016/j.rse.2011.10.028>