

# YÜKSEK ÇÖZÜNÜRLÜKLÜ UYDU GÖRÜNTÜLERİNDEN OTOMATİK BİNA ÇIKARIMI İÇİN DERİN ÖĞRENME MİMARİLERİNİN PERFORMANS ANALİZİ

Elif Özlem YILMAZ<sup>1</sup>, Taşkın KAVZOĞLU<sup>2</sup>,

<sup>1</sup>Arş. Gör., Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, eoymaz@gtu.edu.tr

<sup>2</sup>Prof. Dr., Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, kavzoglu@gtu.edu.tr

## ÖZET

Bina çıkarımı, kentsel planlama, afet müdahalesi ve akıllı şehir teknolojileri gibi birçok araştırma için kritik öneme sahiptir. Uzaktan algılama teknolojileri kullanılarak binaların hassas bir şekilde tanımlanması, şehircilikle ilgili birçok araştırma için önemli bilgiler sağlamaktadır. Binaları çıkarımı için kullanılan geleneksel yöntemlerin yetersiz gelmesi sebebiyle, görüntü işleme için derin öğrenme algoritmaları bina çıkarımı alanında gelişmeler sağlamaktadır. Bu çalışmada, bina çıkarımı için U-Net, U-Net++, DeepLabV3 ve DeepLabV3+ modelleri kullanılmıştır. U-Net ve U-Net++ modelleri, hassas bina algılama ve segmentasyon yetenekleri sunarak mükemmel doğruluk sağlarken, DeepLabV3 ve DeepLabV3+ ise bilgi kaybını azaltmak için özel evrimsel katmanlar kullanarak bina sınırlarının doğru bir şekilde tespit edilmesini sağlamaktadır. Söz konusu derin öğrenme modellerinin performanslarını değerlendirmek için, coğrafi konumları, fiziksel şekilleri ve yapısal bileşimleri açısından çok çeşitli binaları kapsayan Massachusetts bina veri seti kullanılmıştır. Çalışma kapsamında, derin öğrenme modellerinin eğitiminden önce veri setini eğitime uygun hale getirilmiştir. Eğitim amacıyla 810 adet görüntü-maske çifti oluşturulurken, sırasıyla validasyon ve test veri seti için de 270 ve 279 adet görüntü-maske çifti oluşturulmuştur. Ayrıca, eğitim işlemlerinden önce veri artırma yöntemleri kullanılmıştır. Uygulama aşamasında, her derin öğrenme modeli için kodlayıcı olacak ResNet34 seçilmiş, ImageNet ile ön eğitime tabi tutulan ağırlıklar kullanılmıştır. Tüm modellerin performansının değerlendirilmesi için doğruluk, birleşim üzerinden kesişme (IoU), F-skoru, hassasiyet ve geri çağırma ölçütleri hesaplanmıştır. Elde edilen bulgulara göre, U-Net modeli, yüksek F-Skor (0,9356) ve IoU metriği (0,8819) ile güçlü bir performans göstermektedir. U-Net++ modeli, 0,9349 F-Skoru ve 0,8810 IoU değeri U-Net modeli ile karşılaştırılabilir bir performans sergilemiştir. DeepLabV3+ modeli, 0,9292 F-Skoru ve 0,8712 IoU değeri ile DeepLabV3'ten daha iyi performansa sahip olduğu tespit edilmiştir. DeepLabV3 modeli, F1 skoru 0,9227, IoU metriği 0,8602 ile dört model arasında en düşük performansı göstermektedir. Bu çalışmalarda, derin öğrenme modellerinin performansının artırılması için yeni veri artırma teknikleri, optimize edilmiş model mimarileri ve çeşitli veri setleri kullanılabilir.

**Anahtar Sözcükler:** Bina Çıkarımı, Derin Öğrenme, Uydu Görüntüsü, Transfer Öğrenme.

## ABSTRACT

### PERFORMANCE ANALYSIS OF DEEP LEARNING ARCHITECTURES FOR AUTOMATED BUILDING EXTRACTION FROM HIGH RESOLUTION SATELLITE IMAGES

Building extraction is critical information for many research areas such as urban planning, disaster response and smart city technologies. Accurate identification of buildings using remote sensing technologies provides useful data for many fields related to urban planning. Due to the inadequacy of traditional building extraction methods, deep learning image processing algorithms have made progress in the field of building extraction. In this study, U-Net, U-Net++, DeepLabV3 and DeepLabV3+ models are used for building extraction. U-Net and U-Net++ models provide excellent accuracy by offering precise building detection and segmentation capabilities, while DeepLabV3 and DeepLabV3+ provide accurate building boundary detection by using special convolutional layers to reduce information loss. To evaluate the performance of these deep learning models, the Massachusetts Building Dataset was used, which covers a wide range of buildings in terms of geographic location, physical shape and structural composition. As part of the study, the dataset was prepared for training before training the deep learning models. While 810 image-mask pairs were created for training, 270 and 279 image-mask pairs were created for the validation and test datasets, respectively. In addition, data augmentation methods were applied prior to training. In the implementation stage, ResNet34 was chosen as the encoder for each deep learning model, and weights pre-trained with ImageNet were used. Accuracy, intersection over union (IoU), F-score, precision and recall metrics were calculated to evaluate the performance of all models. According to the results, the U-Net model shows a strong performance with high F-Score (0.9356) and IoU metric (0.8819). The U-Net++ model performed similarly to the U-Net model with an F-score of 0.9349 and an IoU value of 0.8810. The DeepLabV3+ model performed better than DeepLabV3 with an F-score of 0.9292 and an IoU value of 0.8712. The DeepLabV3 model showed the lowest performance of the four models with an F1 score of 0.9227 and an IoU metric of 0.8602. In these studies, new data augmentation techniques, optimized model architectures and different data sets can be used to improve the performance of deep learning models.

**Keywords:** Building Extraction, Deep Learning, Satellite Imagery, Transfer Learning.

## 1. GİRİŞ

Uzaktan algılamadaki temel uygulama alanlarından biri, uzaktan algılamış görüntülerden otomatik bina çıkarımıdır. Bina çıkarımı ile elde edilen bilgiler, şehir planlama, çevresel izleme, nüfus tahmini ve afet riski değerlendirme gibi birçok alanda kritik bir rol oynamaktadır (Pasquinelli vd., 2019; Mora vd., 2017; Agung, Van Loenen, ve Van Oosterom, 2019). Uzaktan algılama teknolojileri, yer yüzeyine ilişkin geniş veri yelpazesi sunmakla birlikte, mekânsal çözünürlüğün artmasıyla daha ayrıntılı bilgilere erişim sağlamaktadır. Artan mekânsal çözünürlük, kentsel alanlardaki farklı nesnelere ayırt etme yeteneğine katkı sağlamakta ve bireysel bina bilgisinin çıkarılmasına olanak tanımaktadır. Bu bağlamda, çok yüksek çözünürlüklü görüntülerden bina tespiti, uzaktan algılama alanında günümüzde popüler bir araştırma konusu haline gelmiştir. Ancak, bu görüntülerden doğru bina tespiti yapmak, binaların ve çevrelerinin karmaşıklığı nedeniyle halen zorluk teşkil etmektedir (Luo vd., 2020). Binalar, boyutları, şekilleri ve işlevleri açısından önemli farklılıklar göstermektedir. Ayrıca, yüksek çözünürlüklü görüntülerde aydınlatma koşulları, görüş açısı ve gölgelerden kaynaklanan büyük değişkenlikler mevcuttur. Ek olarak, yol, çıplak zemin ve otopark gibi spektral olarak benzer nesnelere dolu karmaşık kentsel oluşumlar, doğru bina tespitini daha da zorlaştırmaktadır (Gavankar ve Ghosh, 2018). Söz konusu problemler için çok sayıda araştırma gerçekleştirilerek farklı perspektiflere sahip birçok yöntem önerilmiştir (Notarangelo vd., 2021).

Önceki çalışmalarda, bina çıkarımı için özellik öğrenme yöntemlerine odaklanarak semantik segmentasyonun performansını artırmak için bazı makine öğrenme yöntemleri kullanılmıştır (Schlosser vd., 2020; Zho vd., 2018). Ancak bu yöntemler, büyük miktarda eğitim verisi gerektirmekte ve karmaşık bina özelliklerinin yanlış sınıflandırılması gibi sorunlarla karşılaşmaktadır. Nesne tabanlı sınıflandırma yöntemleri, otomatik ve yarı otomatik çıkarım süreçlerinde kullanılmış, ancak manuel düzenleme gereksinimleri nedeniyle sınırlamalar yaşamıştır (Benz vd., 2004; Gavankar ve Ghosh, 2018). Son yıllarda, derin öğrenme teknikleri, özellikle çok yüksek çözünürlüklü görüntülerde bina çıkarımı için önemli bir araç haline gelmiştir. Derin öğrenme, görüntü sınıflandırma ve semantik segmentasyon gibi uzaktan algılama uygulamalarında yüksek doğruluk oranlarına ulaşmakta, hızlı ve doğru bina çıkarımı sağlamaktadır.

Grafik işlem birimlerinin (GPU) donanımsal gelişmeleri, derin öğrenme tabanlı semantik segmentasyon için uçtan uca ağ mimarilerinin geliştirilmesini sağlamış ve bu gelişmeler neticesinde nesne sınıflandırmada önemli ilerlemeler kaydedilmiştir. Derin öğrenme yöntemleriyle bina çıkarımı, geleneksel yöntemlere göre daha hızlı ve doğru bir şekilde gerçekleştirilebilmektedir (Bakirman vd., 2022). Bu kapsamda, bina çıkarımı uygulamaları, bina sınırları ile arka plana ait objelerin algılandığı ikili sınıflandırma problemi olarak ele alınmaktadır. Derin öğrenme tabanlı anlamsal segmentasyon modelleri, mümkün olduğunca gerçeğe yakın ikili sınıflandırma maskelerinin üretilmesini sağlamaktadır. Bu maskeler, binaların yer gerçekliği görüntülerini ifade eder ve ikili sınıflandırma maskeleri, bina ve arka plan olarak temsil edilmektedir (Amirgan ve Erener, 2024).

Bu çalışmada, bina çıkarımı uygulamaları için U-Net, U-Net++, DeepLabV3 ve DeepLabV3+ derin öğrenme modellerinin performansları, Massachusetts Bina Veri Seti üzerinde karşılaştırılmıştır. Performans değerlendirmesi, doğruluk, birleşim üzerinden kesişme (IoU), F-Skoru, hassasiyet ve geri çağırma ölçütleri üzerinden yapılmıştır. Massachusetts bina veri seti, yüksek çözünürlüklü ve zengin içerikli bir veri seti olarak, bina çıkarımında derin öğrenme modellerinin performansını değerlendirmek için kullanılmıştır. Bu bağlamda, yapılan çalışma ile derin öğrenme modellerinin bina çıkarımı üzerindeki etkisini kapsamlı bir şekilde ele almayı amaçlanmıştır. Ayrıca, bu çalışma, uzaktan algılama alanında derin öğrenme tabanlı bina çıkarım yöntemlerinin etkili bir şekilde kullanılmasına yönelik önemli bir katkı sunmayı hedeflemektedir.

## 2. VERİ SETİ & ÇALIŞMA ALANI

Çalışma kapsamında kullanılan Massachusetts veri seti, ABD'nin Boston şehrine ait 151 adet yüksek çözünürlüklü üç bantlı (kırmızı-yeşil-mavi) görüntülerden oluşmaktadır (Mnih, 2013). Her bir görüntü 1500x1500 pikselden oluşmakta olup 1 metre mekânsal ve 8 bit radyometrik çözünürlüğe sahiptir. Bina etiketleri OpenStreetMap projesinden elde edilmiştir. Bu veri kümesi, görüntü ve etiket arasında var olmayan yapılar ve eksik bina etiketleri de dahil olmak üzere hatalı veriler de içermektedir. Çalışma alanına ait görüntü parçası örneği ve buna ait bina etiketi **Şekil 1**'de gösterilmiştir.

Derin öğrenme modellerinin eğitilmesi için tüm veriler birleştirilerek 151 görüntüden ve her görüntünün yer doğrulama verisiyle diğer bir deyişle bina etiketi içeren verilerle birlikte 128 piksel örtüşmesiyle 512x512 boyutlarında görüntü parçacıkları elde edilmiştir. Eğitim için 810 adet görüntü-maske çifti oluşturulurken, sırasıyla validasyon ve test veri seti için de 270 ve 279 adet görüntü-maske çifti oluşturulmuştur. Ayrıca, eğitim işlemlerinden önce veri artırma amacıyla rastgele kırpma, yatay çevirme, dikey çevirme ve rastgele döndürme yöntemleri kullanılmıştır.



Şekil 1. Çalışma alanına ait veri setinden seçilmiş görüntü-maske örneği.

### 3. METODOLOJİ

#### 3.1. U-Net Modeli

U-Net modeli 2015 yılında özellikle görüntü segmentasyonu için geliştirilmiş bir derin öğrenme modelidir. Bu model, girdi görüntüsünün özelliklerini çıkaran bir kodlayıcı ve bu kodlanmış bilgiyi girdi görüntüsünün orijinal boyutuna döndüren bir kod çözücü ile simetrik bir yapıya sahiptir. Kodlayıcı bölümde genellikle konvolüsyonel katmanlar kullanılırken, kod çözücü kısmında ise üst örnekleme işlemleri uygulanarak girdi görüntüsünün boyutu geri kazanılır. Ayrıca, model içerisinde kodlayıcı ve kod çözücüler arasındaki geçişlerde bağlantılar kullanılmaktadır. Bu bağlantılar, modelin daha düşük seviyeli özellikleri daha yüksek seviyeli özelliklerle birleştirerek ayrıntılı segmentasyon sonuçları üretmesini sağlamaktadır (Ronneberger vd., 2015). Yapılan çalışmanın amacına göre U-Net modelinde genellikle segmentasyon problemleri için Sigmoid veya Softmax gibi uygun aktivasyon fonksiyonları ile sonlandırılır.

#### 3.2. U-Net++ Modeli

U-Net++ modeli, U-Net mimarisinden türetilmiş bir semantik segmentasyon mimarisidir. Klasik U-Net modelinin tasarımı, iki önemli özellik ile geliştirilmiştir: bunlar yığılmış yoğun atlama bağlantıları ve derin denetim özellikleridir. Yığılmış yoğun atlama bağlantılarının kullanılmasının, kodlayıcı ve kod çözücü özellik haritaları arasındaki anlamsal eşitsizliği etkili bir şekilde azalttığını ve bunun da gelişmiş gradyan akışı ile sonuçlandığını tespit edilmiştir. Ayrıca, derin denetim kullanımıyla eğitim sırasında ağa bir düzenleme mekanizması dahil ederek modelin performansını artırdığı ortaya çıkmıştır (Zhou vd., 2018).

#### 3.3. DeepLabV3 Modeli

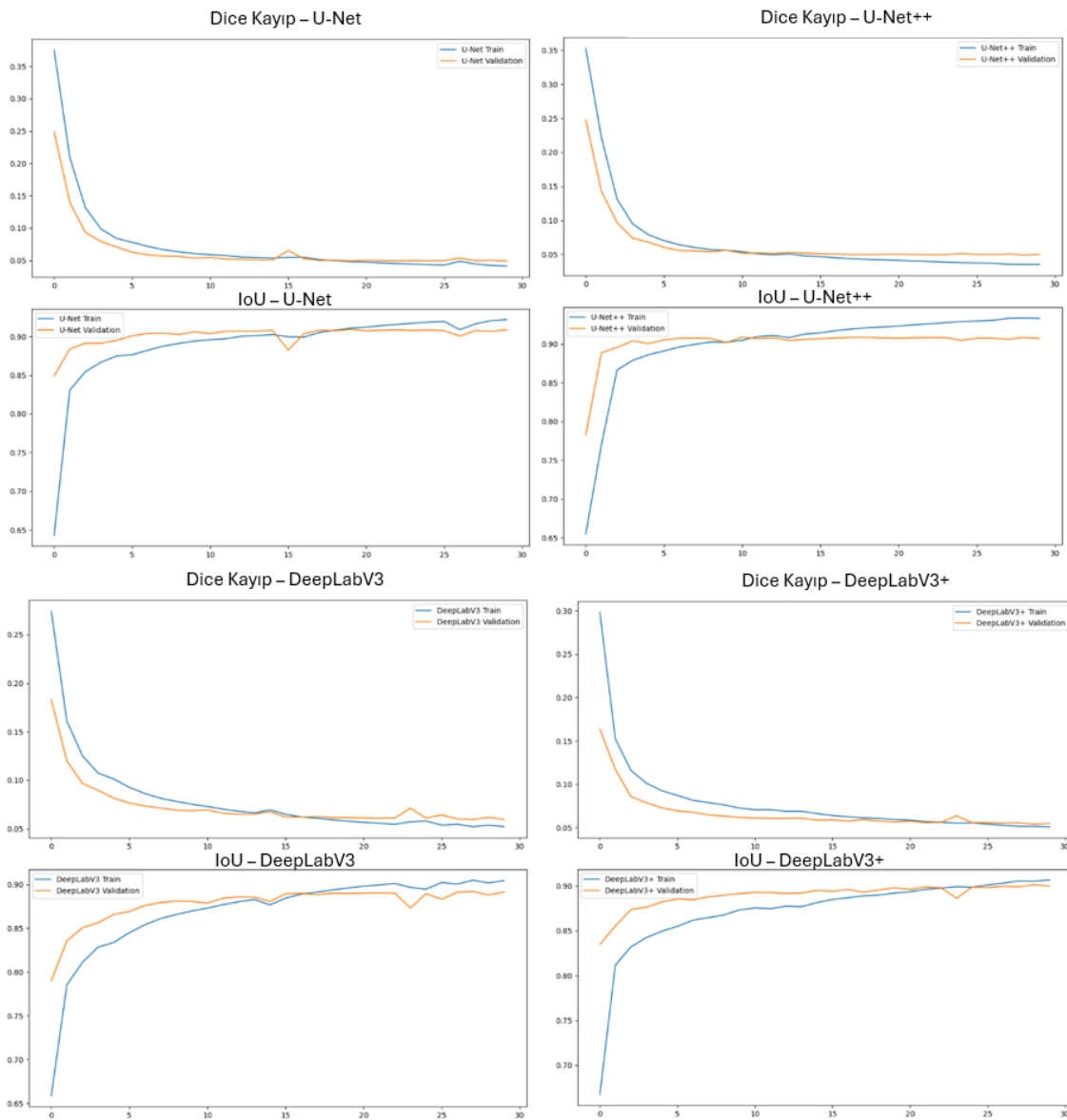
DeepLabV3 modeli, derin öğrenme tekniklerini kullanan bir görüntü segmentasyon modelidir. Özellikle, Atrous Spatial Pyramid Pooling (ASPP) bloğunu kullanarak farklı boyutlardaki özellik haritalarını verimli bir şekilde birleştirir. Bu blok, alıcı alanı genişletmek için atrous (genişletilmiş) konvolüsyonları kullanır. Ayrıca, çeşitli boyutlardaki görüntülerdeki nesnelerin hassas bir şekilde tanımlanmasını ve segmentasyonunu sağlamaktadır. DeepLabV3, büyük ve karmaşık yapıları doğru bir şekilde segmentlere ayırmada başarılı olduğu için yüksek çözünürlüklü görüntüler de bile iyi sonuçlar elde etmektedir (Chen, 2017). Bu özellikler, özellikle uzaktan algılama, otonom navigasyon ve haritalama gibi alanlarda yaygın olarak kullanılmasına katkıda bulunur.

#### 3.4. DeepLabV3+ Modeli

DeepLabV3 modelinde de olduğu gibi DeepLabV3+ modelinde, çok ölçekli bağlamsal bilgileri kodlamak için Atrous Spatial Pyramid Pooling (ASPP) modülünü kullanılmaktadır. Bu özellik, daha keskin nesne sınırlarını yakalamak için bir kodlayıcı-kod çözücü mimarisinin avantajlarını kullanmaktadır. Kodlayıcı kısmı farklı ölçeklerde özellik haritaları oluştururken, kod çözücü kısmı daha iyi segmentasyon sonuçları elde etmeye yardımcı olmak için özellik haritalarını iyileştirir. Bu özellikler sayesinde, DeepLabV3+'ın en önemli avantajlarından biri güçlü nesne segmentasyon performansısıdır. DeepLabV3+ modelinin DeepLabV3'ten farkı, bir kodlayıcı-kod çözücü yapısının eklenmesi olduğu söylenebilir. Bu yapı segmentasyon sonuçlarını iyileştirerek daha keskin nesne sınırları elde edilmesini sağlamaktadır (Chen vd., 2018).

### 4. UYGULAMA

Derin öğrenme modellerinin eğitimine başlamadan önce, veri seti üzerinde kapsamlı ön hazırlık işlemleri gerçekleştirilmiştir. İlk aşamada, Massachusetts veri setine ait üç bantlı (kırmızı-yeşil-mavi) görüntüler ve bunlara karşılık gelen maske görüntüleri, 128 piksel örtüşme sağlanacak şekilde 512x512 piksel boyutlarında parçalara bölünmüştür. Bu işlem sırasında, her bir görüntü parçasının kenarlarında 128 piksellik bir kesişim bölgesi bırakılmıştır. Üretilen veri seti, rastgele bir şekilde %60'ı eğitim, %20'si test ve %20'si validasyon olmak üzere üç gruba ayrılmıştır. Validasyon seti haricindeki tüm veri üzerinde, veri çeşitliliğini artırmak amacıyla yatay ve dikey çevirme, rastgele döndürme gibi veri artırma işlemleri uygulanmıştır. Bu işlemler için Albumentations kütüphanesi kullanılmıştır. Tüm modellerde, kodlayıcı olarak ResNet34 mimarisi kullanılmış ve kodlayıcı ağırlıkları ImageNet modeli üzerinden yüklenmiştir. Modellerin çıkış katmanlarında ise Sigmoid aktivasyon fonksiyonu kullanılmıştır. Tüm modellerde optimizasyon olarak Adam algoritması tercih edilmiş, öğrenme hızı 0.0001 ve epok sayısı 30 olarak belirlenmiştir. Eğitim sürecinde PyTorch kütüphanesi kullanılmıştır. Modellerin performans değerlendirmesi amacıyla IoU, F-skor, hassasiyet, geri çağırma ve doğruluk gibi metrikler hesaplanmıştır. Ayrıca, aşırı öğrenme durumunu kontrol etmek amacıyla Dice kayıp fonksiyonu da değerlendirmeye dahil edilmiştir. Derin öğrenme modellerinin eğitim işlemleri tamamlandıktan sonra modellere ait öğrenme eğrileri çizdirilmiştir. U-Net, U-Net++, DeepLabV3 ve DeepLabV3+ modellerinin eğitim ve doğrulama setlerindeki performanslarını Dice kaybı ve IoU metrikleri açısından karşılaştırılması Şekil 2'de gösterilmiştir.



Şekil 2. Çalışmada kullanılan modellere ait hem doğruluk hem de kayıp için oluşturulmuş öğrenme eğrileri.

U-Net, U-Net++, DeepLabV3 ve DeepLabV3+ modellerinin Dice kaybı eğrileri incelendiğinde, tüm modellerin eğitim süreci boyunca Dice kaybı değerlerinde belirgin bir düşüş gözlemlenmiştir. U-Net ve U-Net++ modelleri, başlangıçta daha yüksek kayıplara sahipken, hızla bu kayıplarını azaltmış ve eğitim sonuna doğru stabil bir seviyeye

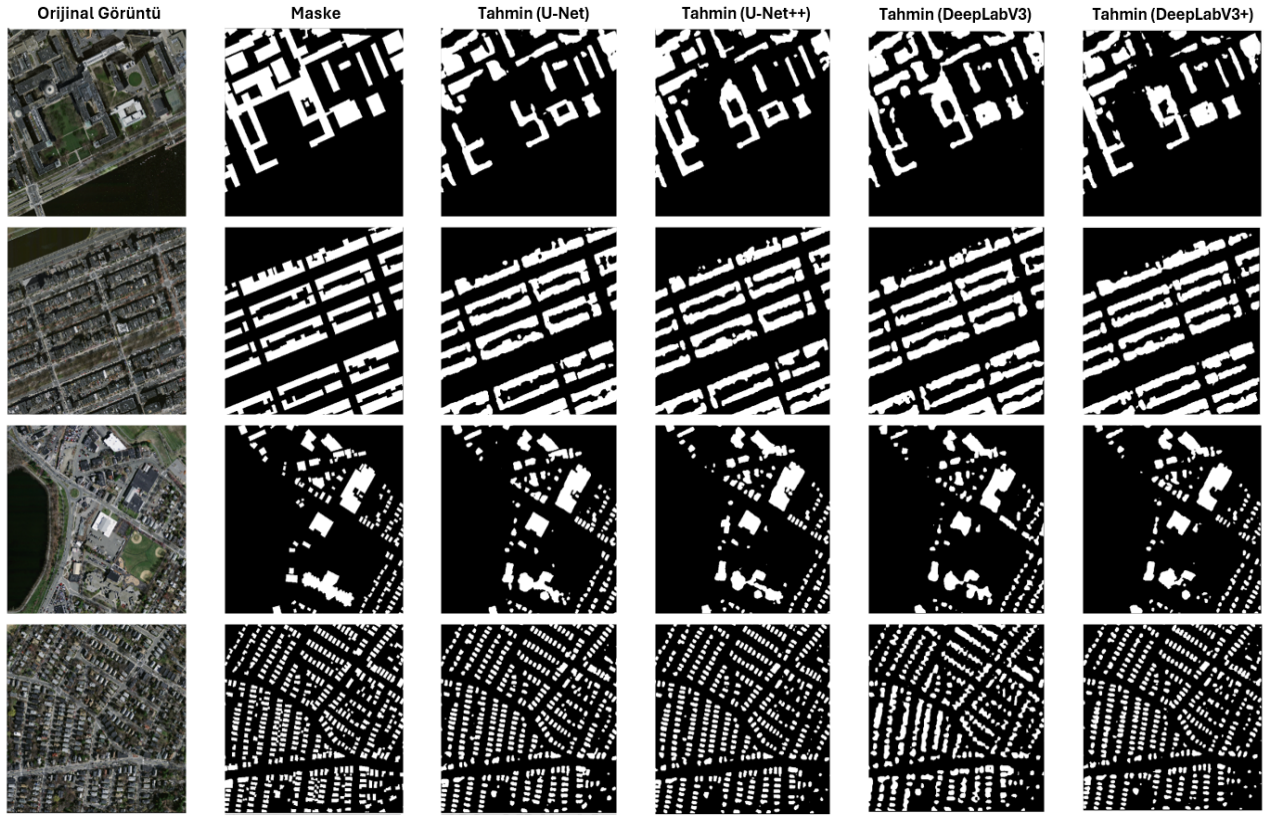
ulaşmıştır. DeepLabV3 ve DeepLabV3+ modelleri de benzer bir iyileşme göstermiş, ancak genel olarak U-Net tabanlı modeller kadar düşük kayıp değerlerine ulaşmamıştır. Bu durum, U-Net tabanlı modellerin daha etkin bir şekilde öğrenme sağladığını ve daha iyi genelleştirme performansı gösterdiğini ortaya koymaktadır. Ayrıca, IoU metrikleri açısından değerlendirildiğinde, U-Net ve U-Net++ modelleri, eğitim sürecinin erken aşamalarında yüksek bir performans göstermiş ve bu performansı eğitim süresi boyunca korumuştur. IoU eğrilerinin stabil bir şekilde yüksek seviyelerde seyretmesi, bu modellerin bina tespiti görevinde yüksek doğrulukta segmentasyon sağladığını göstermektedir. DeepLabV3 ve DeepLabV3+ modelleri ise daha düşük başlangıç IoU değerlerine sahip olmuş, ancak eğitim süreci boyunca performanslarını artırmışlardır. Yine de U-Net tabanlı modellerin performansını yakalayamamışlardır. Bu sonuçlar, U-Net ve U-Net++ modellerinin bina tespiti için semantik segmentasyon görevlerinde daha üstün performans gösterdiğini, özellikle Dice kaybı ve IoU metriklerinde diğer modellere göre daha iyi sonuçlar elde ettiğini göstermektedir.

Bu çalışmada, bina tespiti için U-Net, U-Net++, DeepLabV3 ve DeepLabV3+ gibi önde gelen semantik segmentasyon ağlarının performansları kapsamlı bir şekilde değerlendirilmiştir (**Çizelge 1**). Sonuçlar, U-Net'in genel olarak en yüksek performansı sunduğunu göstermektedir. Özellikle, U-Net'in F-skoru 0,9356, IoU değeri 0,8819, hassasiyet oranı 0,9337, doğruluk değeri 0,9354 ve geri çağırma oranı 0,9374 olarak ölçülmüştür. Dice kaybı ise 0,0654 ile en düşük seviyede gerçekleşmiştir. Bu bulgular, U-Net'in bina tespiti görevlerinde son derece başarılı olduğunu ve segmentasyon doğruluğu açısından diğer mimarilere kıyasla üstünlük sağladığını ortaya koymaktadır. U-Net++ ise, U-Net'e oldukça yakın sonuçlar sunmuş, ancak tüm metriklerde U-Net'in biraz gerisinde kalmıştır. U-Net++'ın F-skoru 0,9349, IoU değeri 0,8810, hassasiyeti 0,9347, doğruluğu 0,9347, geri çağırma oranı 0,9377 ve Dice kaybı 0,0680 olarak kaydedilmiştir. Bu sonuçlar, U-Net++'ın da güçlü bir alternatif olduğunu, ancak U-Net'in üstün performansını yakalayamadığını göstermektedir. DeepLabV3 ve DeepLabV3+ ağları ise U-Net ve U-Net++'a göre nispeten daha düşük performans sergilemiştir. DeepLabV3'ün F-skoru 0,9227, IoU değeri 0,8602, hassasiyeti 0,9197, doğruluğu 0,9224, geri çağırma oranı 0,9257 ve Dice kaybı 0,0786 olarak belirlenmiştir. Benzer şekilde, DeepLabV3+'ın F-skoru 0,9292, IoU değeri 0,8712, hassasiyeti 0,9262, doğruluğu 0,9289, geri çağırma oranı 0,9322 ve Dice kaybı 0,0721 olarak elde edilmiştir. Bu sonuçlar, DeepLabV3 ve DeepLabV3+'ın bina tespiti görevlerinde başarılı olduğunu, ancak U-Net tabanlı mimarilerin gerisinde kaldığını ortaya koymaktadır. Sonuç olarak, U-Net'in bina tespiti için semantik segmentasyon görevlerinde en uygun ağ mimarisi olduğu, U-Net++'ın da güçlü bir alternatif sunduğu, ancak DeepLabV3 ve DeepLabV3+'ın U-Net tabanlı yaklaşımların performansını yakalayamadığı sonucuna varılmıştır. Bu bulgular, bina tespiti ve diğer benzer segmentasyon görevlerinde U-Net'in tercih edilmesini önermektedir.

**Çizelge 1.** Derin öğrenme modellerine ait doğruluk değerlendirmesi.

	<b>U-Net</b>	<b>U-Net++</b>	<b>DeepLabV3</b>	<b>DeepLabV3+</b>
<b>F-skor</b>	0,9356	0,9349	0,9227	0,9292
<b>IoU</b>	0,8819	0,8810	0,8602	0,8712
<b>Hassasiyet</b>	0,9337	0,9347	0,9197	0,9262
<b>Doğruluk</b>	0,9354	0,9347	0,9224	0,9289
<b>Geri Çağırma</b>	0,9374	0,9377	0,9257	0,9322
<b>Dice (Kayıp)</b>	0,0654	0,0680	0,0786	0,0721

Derin öğrenme modellerinin binaların tespiti amacıyla kullanılan çeşitli mimariler üzerindeki performanslarını görsel olarak karşılaştırılmıştır. Orijinal görüntüler ve bunlara karşılık gelen referans maskeler, sırasıyla U-Net, U-Net++, DeepLabV3 ve DeepLabV3+ modelleri tarafından üretilen bazı tahminler ile sunulmuştur (**Şekil 3**). U-Net ve U-Net++ modelleri, genellikle daha net ve keskin tahminler üretirken, özellikle küçük ve ince yapılar üzerinde başarılıdır. U-Net modelinin, U-Net++'a kıyasla bazı detaylarda daha başarılı olduğu, bu durumun özellikle küçük yapıların tespitinde kendini gösterdiği gözlemlenebilir. DeepLabV3 ve DeepLabV3+ modelleri ise, daha geniş ölçekli ve genelleştirilmiş tahminler sunmakta, ancak ince detaylarda, özellikle karmaşık kentsel alanlarda, daha az keskinlik sağlayabilmektedir. Ancak, DeepLabV3+ modelinin DeepLabV3'e göre genel olarak daha başarılı sonuçlar verdiği, bu durumun ince yapıların daha net tespitinde etkili olduğu görülmüştür. Bu sonuçlar, modellerin mimarilerindeki farklılıkların, binaların segmentasyonunda belirgin etkiler yarattığını ve her bir modelin belirli durumlarda diğerine üstünlük sağlayabileceğini göstermiştir. U-Net ve U-Net++ modelleri, daha ince detayları ve küçük yapıları başarıyla yakalarken, DeepLabV3 ve DeepLabV3+ modelleri daha genel ve bütüncül tahminlerde öne çıkmaktadır.



Şekil 3. Massachusetts veri seti üzerinde bina tespiti amacıyla kullanılan çeşitli derin öğrenme modellerinin (U-Net, U-Net++, DeepLabV3, DeepLabV3+) veri setinden rastgele seçilen tahmin sonuçlarının karşılaştırması.

## 5. SONUÇLAR

Bu çalışma kapsamında, Massachusetts veri seti üzerinde bina tespiti için dört farklı derin öğrenme mimarisi (U-Net, U-Net++, DeepLabV3, DeepLabV3+) değerlendirilmiştir. Modellerin performansı, IoU, F-skor, hassasiyet, geri çağırma, doğruluk ve Dice kayıp fonksiyonu gibi çeşitli metrikler kullanılarak kapsamlı bir şekilde analiz edilmiştir. Sonuçlar, U-Net ve U-Net++ modellerinin daha yüksek F-skor ve IoU değerleri ile öne çıktığını, özellikle küçük ve karmaşık yapıları tespit etmede üstün performans sergilediğini göstermektedir. DeepLabV3 ve DeepLabV3+ modelleri ise daha geniş ölçekli yapılar üzerinde genelleştirilmiş sonuçlar sunmakla birlikte, detaylı segmentasyon gerektiren alanlarda U-Net tabanlı modellere kıyasla daha düşük performans sergilemiştir. Ayrıca, eğitim sürecinde elde edilen IoU ve Dice kayıp eğrileri, modellerin hem eğitim hem de validasyon verileri üzerinde kararlı bir öğrenme süreci geçirdiğini göstermektedir. Görsel karşılaştırmalar, her bir modelin segmentasyon başarısını açıkça ortaya koymuş ve U-Net tabanlı modellerin daha detaylı tahminler sunduğunu doğrulamıştır. Bu bulgular, bina tespiti gibi yüksek hassasiyet gerektiren görevler için U-Net ve U-Net++ mimarilerinin daha uygun olduğunu, ancak daha büyük ölçekli genelleme gerektiren görevler için DeepLabV3 ve DeepLabV3+ modellerinin tercih edilebileceğini göstermektedir. Diğer bir deyişle, elde edilen sonuçlar model seçiminde uygulamanın gereksinimlerine göre dikkatli bir değerlendirme yapılması gerektiğini önermektedir.

## KAYNAKLAR

**Amirgan, B. ve Erener, A., 2024.** Semantic segmentation of satellite images with different building types using deep learning methods. *Remote Sensing Applications: Society and Environment*, 34:101176.

**Bakirman, T., Komurcu, I. ve Sertel, E., 2022.** Comparative analysis of deep learning-based building extraction methods with the new VHR Istanbul dataset. *Expert Systems with Applications*, 202: 117346.

**Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. ve Heynen, M., 2004.** Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4):239-258.

**Chen, L. C., 2017.** Rethinking atrous convolution for semantic image segmentation. *arXiv preprint*

*arXiv:1706.05587.*

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

**Gavankar, N. L. ve Ghosh, S. K.,** 2018. Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing*, 51(1):182-193.

**Indrajit, A., Van Loenen, B. ve Van Oosterom, P.,** 2019. Assessing spatial information themes in the spatial information infrastructure for participatory urban planning monitoring: Indonesian cities. *ISPRS International Journal of Geo-Information*, 8(7):305.

**Luo, S., Li, H. ve Shen, H.,** 2020. Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167:443-457.

**Mnih, V.,** 2013. Machine learning for aerial image labeling. PhD Thesis, University of Toronto.

**Mora, H., Gilart-Iglesias, V., Pérez-del Hoyo, R. ve Andújar-Montoya, M. D.,** 2017. A comprehensive system for monitoring urban accessibility in smart cities. *Sensors*, 17(8):1834.

**Notarangelo, N. M., Mazzariello, A., Albano, R. ve Sole, A.,** 2021. Comparing three machine learning techniques for building extraction from a digital surface model. *Applied Sciences*, 11(13):6072.

**Pasquinelli, A., Agugiaro, G., Tagliabue, L. C., Scaioni, M. ve Guzzetti, F.,** 2019. Exploiting the potential of integrated public building data: Energy performance assessment of the building stock in a case study in Northern Italy. *ISPRS International Journal of Geo-Information*, 8(1):27.

**Ronneberger, O., Fischer, P. ve Brox, T.,** 2015. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of 18th International Conference of Medical Image Computing and Computer-assisted Intervention–MICCAI 2015, Munich, Germany, October 5-9, 2015, part III 18*, syf: 234-241.

**Schlosser, A. D., Szabó, G., Bertalan, L., Varga, Z., Enyedi, P. ve Szabó, S.,** 2020. Building extraction using orthophotos and dense point cloud derived from visual band aerial imagery based on machine learning and segmentation. *Remote Sensing*, 12(15): 2397.

**Zhou, Y., Zhang, R., Wang, S. ve Wang, F.,** 2018. Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy. *Sensors*, 18(7): 2013.

**Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. ve Liang, J.,** 2018. Unet++: A nested u-net architecture for medical image segmentation. In *Proceedings of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20*, syf: 3-11.