

UZAKTAN ALGILAMADA NESNE TESPİTİ İÇİN AÇIKLANABİLİR YAPAY ZEKA YÖNTEMLERİNİN KATMAN-BAZLI DEĞERLENDİRİLMESİ

Ebru KOÇDOR¹, Alp ERTÜRK²,

¹ Kocaeli Üniversitesi, Elektronik ve Haberleşme Mühendisliği Bölümü, 41001, İzmit, Kocaeli, 200207001@kocaeli.edu.tr

²Doç. Dr., Kocaeli Üniversitesi, Elektronik ve Haberleşme Mühendisliği Bölümü, 41001, İzmit, Kocaeli, alp.erturk@kocaeli.edu.tr

ÖZET

Derin öğrenme yaklaşımları, uzaktan algılamada nesne tespiti de dahil olmak üzere birçok veri işleme görevinde performans açısından önemli ilerlemeler kaydetmiştir. Ancak, "kara kutu" doğaları ve yorumlanabilirlik eksiklikleri, ortaya çıktıkları günden beri haklı bir eleştiri konusu olmuştur. Derin öğrenme süreçlerinin ve bunların girdi-çıkış ilişkilerinin yorumlanabilirliğini artırmak için son zamanlarda Açıklanabilir Yapay Zeka (AYZ) yaklaşımları ortaya çıkmıştır. AYZ, model kararlarına daha net bir anlayış sağlayarak, makine öğrenimi modellerinde adalet, hesap verebilirlik ve şeffaflığa önemli ölçüde katkıda bulunma potansiyeline sahiptir. Bu çalışmada, uzaktan algılama verileri üzerinde nesne tespiti için You-Only-Look-Once (YOLO) modeli kullanılmış ve modelin yorumlanabilirliğini artırmak amacıyla Grad-CAM, Grad-CAM++, XGrad-CAM, Eigen-CAM ve Layer-CAM AYZ yöntemleri uygulanmıştır. Bu yöntemlerin YOLOv5 ağındaki çeşitli katmanlarda nasıl performans gösterdiği, Detection in Optical Remote Sensing (DIOR) veri kümesi kullanılarak sistematik bir şekilde incelenmiştir. Her bir AYZ yönteminin hangi katmanda daha anlamlı ve yorumlanabilir görseller ürettiği tespit edilmiştir. Katman bazlı analizde, AYZ yöntemlerinin derin sinir ağlarının düşük seviyeli kenar tespitinden yüksek seviyeli nesne temsillerine kadar farklı soyutlama düzeylerini yakaladığı gözlemlenmiştir. Niteliksel olarak, AYZ yöntemleri tarafından üretilen ısı haritaları görselleştirilmiş ve nesne tespiti ile olan uyumları değerlendirilmiştir. Niceliksel değerlendirme kapsamında, sınıflandırma amaçlı bir çalışmadan benimsenen bir AYZ değerlendirme çerçevesi kullanılarak her bir ısı haritası Otsu eşikleme yöntemiyle otomatik olarak eşiklenmiş ve eşiklenmiş görüntülerde hedef tespiti gerçekleştirildiğinde ortalama hassasiyet (mean average precision - mAP) değerlerindeki değişiklikler ölçülmüştür. Daha bilgilendirici ısı haritaları, eşik üstünde kalan alanların maskelendiği yaklaşımda (M1) daha büyük hassasiyet düşüşü ve arka planın maskelendiği yaklaşımda (M2) daha az hassasiyet düşüşü ile sonuçlanmalıdır. Çalışmada, Grad-CAM, Grad-CAM++ ve XGrad-CAM için derin evrimsel katmanlar ile mekansal piramit havuzlama katmanının daha anlamlı açıklamalar sağladığı, Eigen-CAM ve Layer-CAM için ise alt-ve-orta seviye öznelikler elde eden sığ katmanlarda daha etkili olduğu ortaya konmuştur. Bu sonuçlar, AYZ yöntemlerinin katman bazında uygulanmasının model açıklanabilirliğini ve güvenilirliğini artırmada kritik bir öneme sahip olduğunu göstermektedir.

Anahtar Sözcükler: Açıklanabilir yapay zeka (AYZ), derin öğrenme, katman, nesne tespiti, sınıf aktivasyon haritaları.

ABSTRACT

LAYER-WISE EVALUATION OF EXPLAINABLE ARTIFICIAL INTELLIGENCE METHODS FOR OBJECT DETECTION IN REMOTE SENSING

Deep learning approaches have made significant progress in performance for many data processing tasks, including object detection in remote sensing. However, their "black-box" nature and lack of interpretability have been a rightful point of criticism since their inception. Recently, Explainable Artificial Intelligence (XAI) approaches have emerged to enhance the interpretability of deep learning processes and their input-output relationships. XAI has the potential to significantly contribute to fairness, accountability, and transparency in machine learning models by providing a clearer understanding of model decisions. In this study, the You-Only-Look-Once (YOLO) model was used for object detection on remote sensing data, and Grad-CAM, Grad-CAM++, XGrad-CAM, Eigen-CAM and Layer-CAM XAI methods were applied to improve the interpretability of the model. The performance of these methods across various layers of the YOLOv5 network was systematically analyzed using the Detection in Optical Remote Sensing (DIOR) dataset. It was determined which XAI method produced more meaningful and interpretable visualizations at different layers. Layer-wise analysis revealed that XAI methods capture different levels of abstraction, from low-level edge detection to high-level object representations in deep neural networks. Qualitatively, the heatmaps generated by XAI methods were visualized and their alignment with object detection was evaluated. For quantitative assessment, an XAI evaluation framework adapted from a classification study was used, in which each heatmap was automatically thresholded using the Otsu thresholding method, and the changes in mean average precision (mAP) values were measured when object detection was performed on the thresholded images. More informative heatmaps should result in a greater drop in precision in the approach where the regions above the threshold are masked (M1) and a smaller drop in precision in the approach where the background is masked (M2). The study found that deeper convolutional layers and the spatial pyramid pooling layer provided more meaningful explanations for Grad-CAM, Grad-CAM++ and XGrad-CAM, while Eigen-CAM and Layer-CAM were more effective in the shallow layers that captured low- and mid-level features. These findings highlight the critical importance of applying XAI methods in a layer-wise manner to enhance the interpretability and reliability of models.

Keywords: Class activation mapping (CAM), deep learning, explainable artificial intelligence (XAI), layer, object detection.

1. GİRİŞ

Uzaktan algılama, yeryüzü ve atmosferi gözlemleyerek çok çeşitli uygulamalar için veri temelli analizleri mümkün kılan bir teknoloji alanıdır. Son yıllarda gelişen sensör teknolojileri ve sayıca artan alıcılar, büyük miktarda veri üretmiş ve bu verilerin işlenerek anlamlı bilgilere dönüştürülmesi kritik bir öneme sahip olmuştur. Bu süreçte, yapay zeka ve makine öğrenimi tabanlı yöntemler, uzaktan algılama verilerinde nesne tespiti gibi karmaşık görevlerin daha etkin bir şekilde gerçekleştirilmesini sağlamaktadır. Özellikle derin öğrenme yaklaşımları, uzaktan algılama görüntülerinde nesne tespitinde yüksek başarı oranlarıyla öne çıkarak bu alandaki birçok karmaşık görevi daha verimli ve etkili bir şekilde çözümlenmektedir.

Nesne tespiti görevlerinde kullanılan derin öğrenme modelleri çok katmanlı yapay sinir ağlarından oluşur. Bu ağlar, görüntülerdeki özellikleri hiyerarşik bir şekilde öğrenir; ilk katmanlar düşük seviyeli özellikleri (örneğin kenarlar, renkler), orta katmanlar daha soyut yapıları (şekiller, yüzeyler), son katmanlar ise sınıf veya nesne gibi spesifik bilgileri öğrenir. Ancak, bu katmanlı yapının iç işleyişi bir "kara kutu" olarak değerlendirilir. Derin öğrenme modellerinin karar mekanizmalarını anlamak ve açıklayabilmek için Açıklanabilir Yapay Zeka (AYZ) yaklaşımları geliştirilmiştir. Gevaert vd. (2022), yapay zeka ve makine öğrenimi yöntemlerinin yeryüzü gözlemi ve uzaktan algılama alanındaki kullanımını inceleyerek, derin öğrenme modellerinin karmaşıklığı ve "kara kutu" doğası nedeniyle anlaşılabilirliğinin zor olduğunu ve bu durumun model kararlarının şeffaflığını ve güvenilirliğini azalttığını ortaya koymuştur. Taşkın vd. (2023), derin öğrenme yöntemlerinin uzaktan algılama ve yeryüzü gözlemi alanındaki performansını ve bu yöntemlerin açıklanabilirlik eksikliğini ele alarak, mevcut açıklanabilir yapay zeka tekniklerinin durumunu, açık zorlukları ve fırsatları kapsamlı bir şekilde incelemişlerdir. AYZ yöntemlerinin uzaktan algılama uygulamalarında daha geniş çapta benimsenmesi gerektiği ve bu sayede model kararlarının daha anlaşılır ve güvenilir hale geleceği sonucuna varılmıştır.

AYZ yöntemleri, modelin karar alma süreçlerini daha şeffaf hale getirerek, yorumlanabilirliğin ve açıklanabilirliğin artırılmasını hedeflemektedir. Höhl vd. (2024), derin öğrenme modellerinin karar alma süreçlerinin daha şeffaf hale getirilmesi gerektiğini vurgulamışlardır ve AYZ yöntemlerinin kullanıcıların model kararlarına olan güvenini artırdığı sonucuna ulaşmışlardır. AYZ yöntemleri derin öğrenme modellerinin yorumlanabilirliği için son yıllarda farklı alanlarda yoğun bir şekilde kullanılmasına rağmen, uzaktan algılamada nesne tespiti özelinde literatürde bu alanda sınırlı sayıda çalışma yer almaktadır. Hogan ve Aouf (2021), insansız hava araçları ile alınan optik verilerde YOLOv5 ile nesne tespiti için sınıf aktivasyon haritaları (Class Activation Maps - CAM) temelli Grad-CAM AYZ yöntemi ile ısı haritaları elde etmiştir. Hogan vd. (2022), yine YOLOv5 ile aynı veri tabanında nesne tespitinde yorumlanabilirlik için bu sefer KernelSHAP AYZ yöntemini kullanmışlardır. Kawauchi ve Fuse (2022), YOLOv3 ve Mask R-CNN ile nesne tespitinde açıklanabilirlik için GradientSHAP yöntemini kullanmışlardır.

Bu çalışma kapsamında, optik uzaktan algılama verilerinde derin öğrenme temelli nesne tespitinde yorumlanabilirlik için CAM tabanlı AYZ yöntemlerinin performansı nitel ve nicel açıdan değerlendirilmiştir. CAM tabanlı yöntemler, kullanılan derin öğrenme modeli üzerinde değişiklik yapılmasına gerek olmadan, ısı haritaları üreterek, kullanıcıya modelin hangi görüntü bölgelerine dayanarak karar verdiğini gösterir. Literatürde son yıllarda çok sayıda CAM tabanlı AYZ yöntemi geliştirilmiştir. Ancak bu yöntemlerin başarımlarını, uzaktan algılama verilerinde nesne tespitinde yorumlanabilirlik için, nitel ve nicel olarak karşılaştıran bir çalışma yakın zamana kadar yapılmamış durumdadır. Bu çalışmada, Detection in Optical Remote Sensing (DIOR) veri tabanında (Li vd., 2020) You Look Only Once (YOLO) (Redmon vd., 2016) yönteminin bir versiyonu olan YOLOv5 derin öğrenme modeli ile nesne tespiti gerçekleştirilmiştir. Daha sonrasında Grad-CAM (Selvaraju vd., 2017), Grad-CAM++ (Chattopadhyay vd., 2017), XGradCAM (Fu vd., 2020), Eigen-CAM (Muhammed ve Yaesin, 2020) ve Layer-CAM (Jiang vd., 2021) AYZ yöntemleri ile derin öğrenme modelinin kapalı kutu doğasına yorumlanabilirlik sağlanmıştır. AYZ yöntemleri tarafından üretilen ısı haritaları görselleştirilmiş ve AYZ yöntemlerinin başarımları nitel olarak değerlendirilmiştir. Ancak DIOR ve benzeri büyük veri tabanlarında görsel başarımların analizleri ile nesnel bir değerlendirme yapmak mümkün olmamaktadır. Bu nedenle, AYZ yöntemlerinin başarımlarının nicel olarak değerlendirilmesi için Gizzini vd. (2023)'ten uyarlanan bir değerlendirme mekanizması kullanılmıştır. Bu değerlendirme mekanizmasında, ısı haritaları otomatik eşiklemeden geçirilerek, eşğin üstünde kalanlar atılıp altında kalan kısımlar (arka plan) bırakıldığında (M1 stratejisi) nesne tespiti başarımlarındaki düşüş ve eşğin altında kalan kısımlar (arka plan) atılıp üstünde kalan kısımlar bırakıldığında (M2 stratejisi) nesne tespiti başarımlarındaki düşüş hesaplanmıştır. Nesne konumları ile uyumlu ısı haritalarının M1 stratejisi için başarımlarında yüksek, M2 stratejisi için ise düşük başarımların düşüşü gözlenmesi beklenmelidir. Çalışma kapsamında AYZ yöntemleri YOLOv5 modelinin farklı katmanlarına uygulanarak analizlerin katman-bazlı gerçekleştirilmesi sağlanmıştır. Deneysel sonuçlar AYZ yöntemlerinin hangi katmanda uygulandığının elde edilen ısı haritaları ve dolayısıyla AYZ yöntemlerinin ışık tuttuğu açıklamalar açısından kritik bir rol oynadığı göstermektedir. Çalışma kapsamında elde edilen bulgular nesne tespitinde model güvenilirliğinde ve yapay zeka yaklaşımlarının adil, hesap verebilir ve şeffaf olması yolunda büyük önem taşıyan AYZ yaklaşımlarının performansına ve kullanılan derin ağ modelinin farklı katmanları ile ilişkilerine dair önemli ipuçları sunmaktadır.

2. MATERYAL VE METOT

2.1 YOLOv5 Modeli

Bu çalışmada nesne tespiti için YOLOv5 modeli kullanılmıştır. YOLO algoritması, yüksek hızda ve doğrulukla nesne tespiti yapabilen bir derin öğrenme modelidir. YOLOv5, büyük ve karmaşık veri kümelerinde başarımını doğruluk ve hız açısından kanıtlamış bir modeldir. YOLOv5 mimarisi, derin öğrenme tabanlı nesne tespiti için optimize edilmiş üç ana bileşenden oluşur: backbone, neck ve head. Bu yapı, görüntülerdeki nesnelere farklı ölçeklerde tespit edebilmek için çoklu çözünürlüklerden gelen özellikleri etkili bir şekilde işler. YOLOv5'in esnek ve modüler yapısı, hem küçük hem de büyük nesnelere tespiti sırasında modelin öğrenme kapasitesini artırırken, işlem verimliliğini de korur.

Modelin backbone kısmı, görüntüden yüksek seviye özelliklerin çıkarıldığı aşamadır ve bu özellikler, nesnelere farklı boyutlarda tanınabilmesi için çeşitli uzamsal çözünürlük seviyelerinde (P1, P2, P3 vb.) öğrenilir. Backbone'un ilk katmanları kenarlar ve dokular gibi temel görsel özellikleri öğrenirken, derin katmanlarda daha soyut ve karmaşık özellikler öğrenilir. YOLOv5'te kullanılan C3 katmanları, ResNet'ten türetilmiş bottleneck yapısını içerir. C3 [128] katmanı, iki evrimsel yol arasında geçiş oluşturup parametre verimliliğini artırırken, C3 [256] ve C3 [512] katmanları daha büyük nesnelere tespitine yönelik derin soyutlamalar çıkarır. Backbone'un son aşamasında yer alan uzamsal piramit havuzlama (Spatial Pyramid Pooling-Fast - SPPF) katmanı ise, çoklu çözünürlüklerdeki bilgiyi sabit boyutlarda birleştirerek, farklı ölçeklerdeki nesnelere tespiti için bilgi yoğunlaşmasını sağlar. SPPF'nin havuzlama işlemi, geniş kapsamlı uzamsal bilginin korunarak yüksek soyutlamaya ulaşılmasına katkıda bulunur.

Neck kısmı, çoklu ölçekli nesne tespit yeteneklerini geliştiren ve farklı çözünürlük seviyelerindeki özellikleri birleştiren bir yapıdır. YOLOv5'in neck bölümü, Feature Pyramid Network (FPN) ve Path Aggregation Network (PAN) yapılarından esinlenmiştir. Bu bölüm, alt ve üst seviyelerdeki özelliklerin birbirine aktarılmasını sağlayarak, farklı boyutlardaki nesnelere etkili bir şekilde tanınmasını mümkün kılar. Özellikle Upsample katmanı, düşük çözünürlükteki özellik haritalarını üst katmanlara ileterek büyük nesnelere daha doğru tespit edilmesini sağlar. Concat katmanları ise farklı çözünürlüklerdeki (P3, P4) özellik haritalarını birleştirerek zengin bilgi akışı sunar ve bu da çok ölçekli nesne tespiti performansını artırır.

Head kısmı, nesnelere tespit ve sınıflandırılmasının yapıldığı bölümdür. YOLOv5'te tespit modülü, farklı çözünürlük seviyelerinde (P3, P4, P5) çalışarak küçük, orta ve büyük boyuttaki nesnelere tespitini sağlar. Tespit katmanları (17, 20, 23), farklı boyutlardaki nesnelere için üç ayrı ölçekte çıkışlar verir ve her bir ölçek için anchor box adı verilen sınırlı kutuları kullanır. Bu kutular görüntüdeki nesnelere boyutlarına uygun kutular oluşturularak, modelin farklı boyutlardaki nesnelere tespit etmesini ve bu nesnelere için sınırlayıcı kutular oluşturabilmesini sağlar.

YOLOv5'in bu üç bileşeni katman bazında değerlendirildiğinde, modelin uzaktan algılama gibi yüksek çözünürlüklü görüntülerde verimliliğini artıran mekanizmalar sunduğu görülmektedir. Backbone, erken aşamalarda temel görsel özellikleri öğrenirken, neck ve head kısımları çoklu çözünürlük seviyelerindeki bilgileri birleştirerek nesnelere farklı boyutlarda doğru şekilde tanınmasını sağlar. Böylece, modelin uzaktan algılama verilerinde yüksek doğrulukla küçük, orta ve büyük boyuttaki nesnelere tespit edebilmesi ve modelin zorlu veri kümelerinde başarılı sonuçlar elde edebilmesi mümkün olur.

Bu çalışma kapsamında, DIOR veri tabanı üzerinden YOLOv5 ile nesne tespiti gerçekleştirildikten sonra, modelin içsel karar mekanizmalarına ışık tutacak çeşitli CAM yöntemleri, modelin 8., 9., 13., 17., 19., 20., 22. (-3. katman) ve 23. (-2. katman) katmanlarına uygulanmıştır. Bu katmanlar, derin öğrenme modelinin özellik çıkarma ve sınıflandırma süreçlerindeki önemli aşamaları temsil eder.

- 8. ve 9. katmanlar, modelin daha erken aşamalarında, genellikle düşük seviyeli özelliklerin (örneğin, kenarlar, doku detayları) çıkarıldığı katmanlardır. Bu katmanlar, görüntüdeki temel yapıları öğrenir ve sonraki katmanlara iletir.
- 13. katman, daha orta seviyeli özelliklerin öğrenildiği ve nesnelere daha soyut yapılarının (örneğin, şekiller, yüzeyler) tespit edildiği bir aşamadır. Bu katman, modelin nesnelere birbirinden ayırma yeteneğini güçlendirir.
- 17. ve 19. katmanlar, YOLOv5 mimarisindeki "neck" kısmında yer alır ve çoklu ölçekli nesne tespiti için gerekli olan özellikleri birleştiren katmanlardır. Bu katmanlar, farklı çözünürlüklerdeki nesne bilgilerini bir araya getirir.
- 20. katman, modelin daha derin kısımlarında yer alan ve giderek daha yüksek seviyeli, sınıfa özgü özelliklerin öğrenildiği bir katmandır.

- 22. ve 23. katmanlar, "head" kısmında yer alır ve nesne tespiti ve sınıflandırmanın doğrudan yapıldığı katmanlardır. Bu katmanlar, modelin son aşamalarında çıktılarını oluşturan sınıflandırma ve sınır kutusu regresyon katmanlarıdır.

2.3 CAM Yöntemleri

Bu çalışma kapsamında CAM tabanlı Grad-CAM, Grad-CAM++, XGrad-CAM, Eigen-CAM ve Layer-CAM AYZ yöntemleri kullanılmıştır. Grad-CAM (Selvaraju vd., 2017), derin öğrenme modellerinin özellikle evrişimli sinir ağı tabanlı olanlarının karar alma süreçlerini görselleştirmek amacıyla kullanılan bir tekniktir. Grad-CAM, modelin tahmin ederken hangi görüntü bölgelerine dikkat ettiğini ortaya koyarak görsel açıklamalar sağlar. Bu yöntemin standart kullanımında, modelin son evrişim katmanındaki öznelik haritaları ve bu haritaların sınıf skorlarına olan gradyanları hesaplanır. Gradyanların ağırlıklı ortalamaları alınarak bir ısı haritası oluşturulur. Bu ısı haritası, modelin belirli bir sınıf tahminine en çok katkıda bulunan alanları vurgular. Grad-CAM, derin öğrenme modelinde mimari değişiklikler gerektirmediği için geniş bir kullanım alanına sahiptir ve tıbbi görüntüleme, otonom araçlar ve güvenlik gibi çeşitli uygulamalarda kullanılır. Ayrıca, yorumlanabilirlik açısından kullanıcıların modellerin kararlarını anlamasını kolaylaştırır.

Grad-CAM++ (Chattopadhyay vd., 2017), Grad-CAM'in geliştirilmiş bir versiyonudur. Grad-CAM++, gradyanların pozitif kısımlarını ağırlıklandırarak her bir öznelik haritasının sınıf skoruna katkısını daha hassas bir şekilde hesaplar. Chattopadhyay vd. (2017), birden fazla nesne bulunan görüntülerde Grad-CAM++ ile Grad-CAM'e göre daha doğru ve detaylı ısı haritaları elde edildiğini göstermişlerdir.

XGrad-CAM (Fu vd., 2020) yöntemi, evrişimli sinir ağlarının belirli bir katmanındaki özellik haritalarının gradyanlarını hesaplayarak başlar. Bu gradyanlar, her özellik haritasının sınıfa olan katkısını belirlemek için ağırlıklandırma faktörü olarak kullanılır. XGrad-CAM, duyarlılık ve korunum aksiyomlarına dayanır. Duyarlılık aksiyomu, bir özellik haritasının çıkarıldığında sınıf puanında büyük bir değişiklik meydana geliyorsa bu haritanın önemli olduğunu belirtir. Korunum aksiyomu ise tüm özellik haritalarının katkılarının sınıf puanının toplamına eşit olması gerektiğini ifade eder. Yöntem, bu aksiyomları yerine getirmek için bir optimizasyon problemi çözerek her bir özellik haritasını yeniden ağırlıklandırır. Gradyanlarla ağırlıklandırılan bu haritalar, sınıf aktivasyon haritasını oluşturur. Oluşturulan bu harita, pozitif değerlerle sınırlandırılır ve ReLU aktivasyonu ile negatiflerden arındırılır. Son aşamada, harita giriş görüntüsünün boyutuna getirilerek, hangi bölgelerin sınıf tahmininde etkili olduğu görselleştirilir. Fu vd. (2020), bu süreç ile XGrad-CAM'in özellikle zayıf denetimli görevlerde nesne lokalizasyonu gibi görevlerde daha iyi sonuçlar verdiği ifade etmiştir.

Eigen-CAM (Muhammed ve Yaesin, 2020) ise önceki yöntemlerden farklı olarak, gradyan tabanlı değildir. Eigen-CAM yönteminde ısı haritalarının elde edilmesi için temel bileşen analizi kullanılır. Muhammed ve Yaesin (2020), Eigen-CAM'in Grad-CAM'e kıyasla daha küresel bir bakış sunduğunu, çünkü yalnızca belirli sınıfların değil, modelin genel aktivasyonlarının hangi alanlarda yoğunlaştığını gösterdiğini ifade etmişler ve savlarını deneysel çalışmalar ile desteklemişlerdir.

Layer-CAM (Jiang vd., 2021), evrişimli sinir ağının farklı katmanlarındaki aktivasyon haritalarını kullanarak daha hassas ve ayrıntılı nesne lokalizasyonu sağlamayı amaçlar. Yöntem, sınıfa özgü gradyanları kullanarak her bir pikselin hedef sınıfa katkısını hesaplar. Gradyanların pozitif değerleri, her konumun sınıfa olan etkisini belirlerken, negatif gradyanlar ReLU aktivasyonu ile sıfırlanır. Bu sayede yalnızca hedef sınıfa olumlu etkisi olan bölgeler dikkate alınır. Layer-CAM'in en önemli yeniliği, evrişimli sinir ağının hem derin hem de sığ katmanlarından sınıf aktivasyon haritaları üretmesidir. Derin katmanlar genellikle nesnenin kaba konumunu belirlerken, sığ katmanlar daha yüksek çözünürlükte ayrıntılar sağlar. Bu iki bilgi kaynağını birleştirerek, hem nesnenin genel konumunu hem de ince detaylarını içeren daha doğru bir aktivasyon haritası elde edilir. Aktivasyon haritaları birleştirilirken, sığ katmanlardan gelen haritaların değerleri düşük olduğu için bir ölçekleme fonksiyonu kullanılarak dengeleme sağlanır.

2.4 CAM Yöntemleri Değerlendirme Mekanizması

Bu çalışma kapsamında farklı katmanlara uygulanmış çeşitli CAM tabanlı AYZ yöntemlerinin başarımının sayısal analizi için Gizzini vd. (2023)'den uyarlanan bir çerçeve yaklaşım kullanılmıştır. Bu yaklaşımda, CAM yöntemleri ile ısı haritaları elde edildikten sonra, ilk aşamada bu ısı haritalarının eşiklenerek ikili görsellere dönüştürülmesi gerekmektedir. Bu çalışma kapsamında, bu amaçla, ısı haritalarına histogram tabanlı Otsu eşikleme yöntemi (Otsu, 1979) uygulanarak, eşik değerleri otomatik olarak belirlenmiştir.

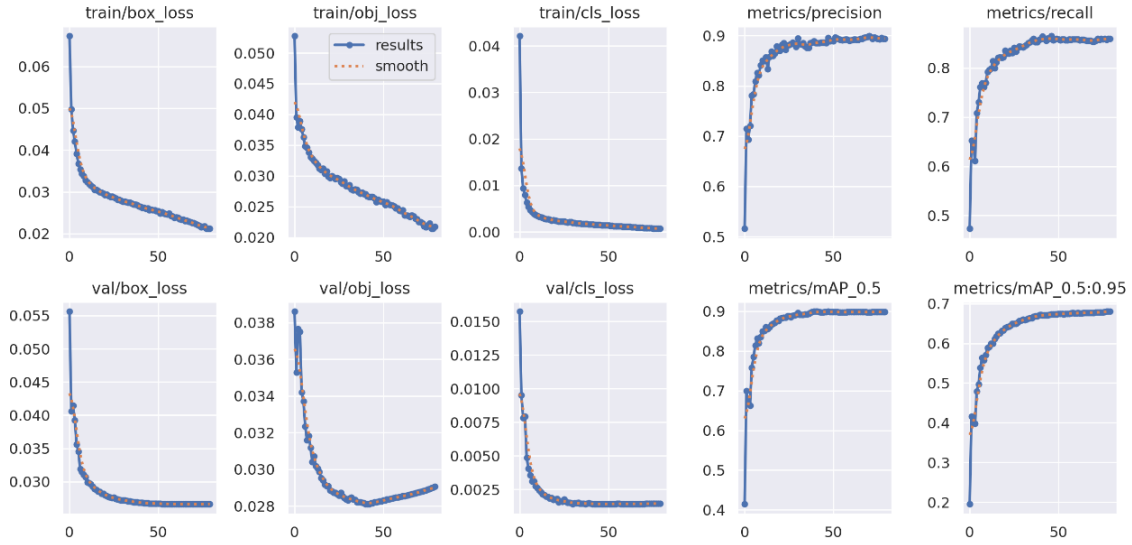
Çerçeve yaklaşım altında iki farklı nicel değerlendirme mekanizması kullanılmıştır. İlk yaklaşımda (M1), AYZ yöntemlerinin ısı haritası çıktılarında eşik üstünde yer alan değerlere sahip bölgeler orijinal görüntüden maskelenir ve bu maskelenmiş görüntülerden oluşan bir veri kümesi oluşturulur. Orijinal görüntüler üzerinden eğitilen model maskelenmiş görüntülerden oluşan bu veri kümesi üzerinde tekrar nesne tespiti için kullanılır. Sağladığı ısı haritaları

hedef konumları ile uyumlu AYZ yöntemleri için, maskelenmiş görüntüler üzerinden nesne tespiti başarımı, orijinal nesne tespiti başarımına göre yüksek bir düşüş gösterecektir. Bu nedenle, M1 ölçütünde, maskelenmiş ve maskelenmemiş görüntülerde nesne tespitinde en yüksek mAP değeri farkına neden olan AYZ yöntemi daha başarılı kabul edilir. İkinci yaklaşımda (M2) ise AYZ yöntemleri ile elde edilen ısı haritalarına arka plan (eşiğin altında kalan bölgeler) maskelenir. Maskeleye sonrası, orijinal görüntülerde eğitilen derin öğrenme modeli, maskelenmiş görüntülerde nesne tespiti için kullanılır. Sağladığı ısı haritaları hedef konumları ile uyumlu AYZ yöntemleri için, M2 yaklaşımına göre maskelenmiş görüntülerde nesne tespiti başarımı, orijinal nesne tespiti başarımına göre yüksek bir düşüş göstermeyecektir. Bu nedenle, M2 ölçütünde, maskelenmiş ve maskelenmemiş görüntülerde nesne tespitinde en düşük mAP değeri farkına neden olan AYZ yöntemi daha başarılı kabul edilir.

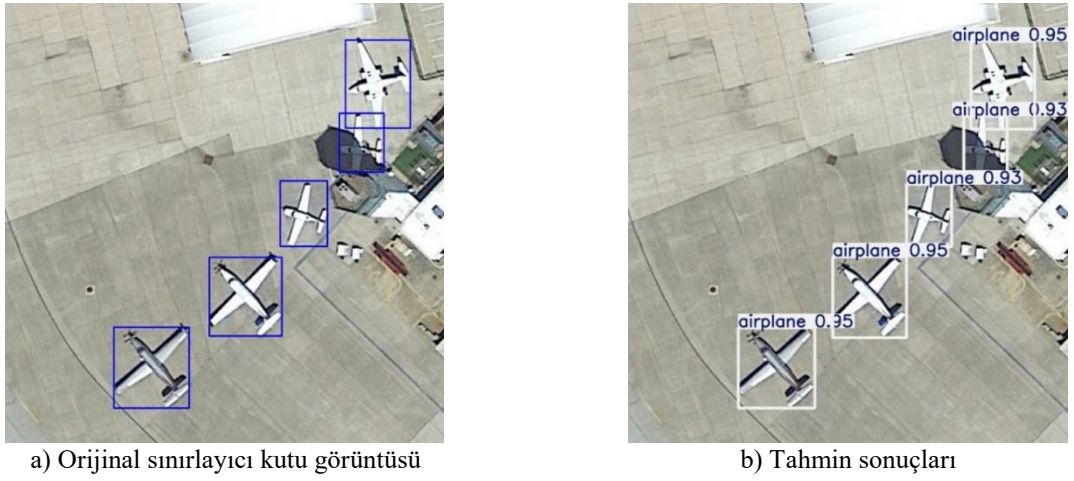
3. ARAŞTIRMA BULGULARI

3.1 Veri Kümesi ile Elde Edilen Sonuçlar

Bu çalışmada uzaktan algılamada nesne tespiti için DIOR veri kümesi kullanılmıştır. DIOR veri kümesi, çeşitli hava koşulları ve görüntüleme açılarında alınmış 20 farklı sınıftan oluşan geniş bir veri kümesidir. Bu veri kümesi, yüksek çözünürlüklü 23463 görüntü içerir. Veri kümesindeki uçaklar, arabalar, gemiler, binalar ve köprüler gibi farklı boyut ve şekillere sahip nesnelere bulunmaktadır. Bu çalışmada nesne tespiti için YOLOv5x mimarisi kullanılmıştır. DIOR veri kümesi ile gerçekleştirilen çalışmalar, Google Colab Pro ortamında NVIDIA Tesla A100 GPU kullanılarak yapılmıştır. Veri kümesi, %80 eğitim, %10 doğrulama ve %10 test için kullanılacak şekilde bölünmüştür. YOLOv5x modeli toplam 80 epoch boyunca eğitilmiştir. Eğitim süresi boyunca modelin doğruluk oranı artış göstermiş ve yaklaşık %89 seviyelerinde kalmıştır. Tüm sınıflar için test verisi üzerinden ortalama kesinlik (precision) ve duyarlılık (recall) değerleri üzerinden hesaplanan $mAP@0.5$ değeri %89.9 olarak elde edilmiştir. Bu sonuç, modelin genel olarak yüksek bir performans sergilediğini göstermektedir. Şekil 1’de modelin kayıp ve başarımları grafikleri sunulmuştur. Şekil 2’de veri kümesinde yer alan örnek bir görüntü için nesne konumları ve tahmin sonuçları görülmektedir.



Şekil 1. Nesne tespiti kayıp ve başarımları grafikleri.



Şekil 2. YOLOv5 modeli ile tahmin edilen hedeflerin orijinal sınırlayıcı kutular ile karşılaştırılması

3.2 AYZ Analizi Sonuçları

Bu çalışmada karşılaştırılan CAM tabanlı AYZ yöntemlerinin M1 ve M2 ölçütlerine göre başarımlarının elde edilmesi için veri kümesinde yer alan 1000 adet görüntü için ısı haritaları kullanılmıştır. Maskeleye için kullanılacak eşik değerini belirlemek için Otsu yöntemi uygulanmıştır. Öncelikle derin öğrenme modeli üzerinden AYZ yöntemi çıktılarının histogramları hesaplanmıştır ve bu histogramlar birleştirilerek Otsu yöntemiyle optimal eşik değeri elde edilmiştir. Eşik değerine göre maskeleye işlemi, M1 için AYZ tarafından vurgulanan bölgelerin maskelenmesi, M2 için ise AYZ tarafından vurgulanan bölgeler dışında kalan bölgelerin yani arka planın maskelenmesi anlamına gelmektedir. Bu yaklaşımda, M1 için maskelenen görüntülerde hedef tespiti sonucunda elde edilen mAP değerinin düşük, M2 için ise hedef tespiti sonucunda elde edilen mAP değerinin yüksek olması beklenmektedir. Şekil 3'de veri kümesindeki örnek bir görüntü için bir AYZ yönteminin sağladığı ısı haritası ve bu haritasının M1 ve M2 yaklaşımlarına göre maskelenmiş halleri sunulmuştur.

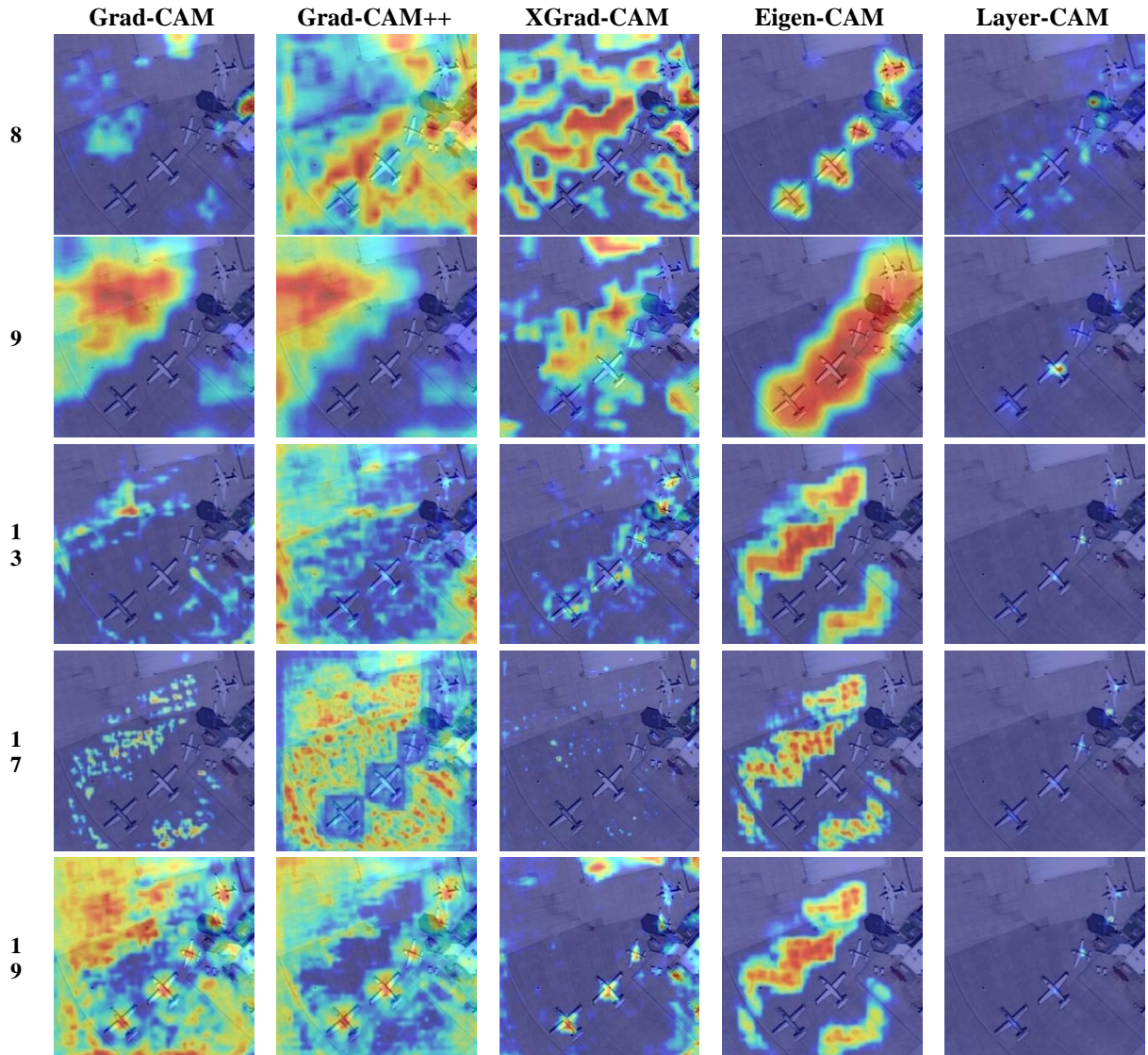
M1 ve M2 yaklaşımlarına göre maskeleye işlemi, Grad-CAM, Grad-CAM++, XGrad-CAM, Eigen-CAM ve Layer-CAM yöntemlerinden elde edilen görüntüleri ayrı ayrı uygulanarak toplam 10 tane test kümesi hazırlanmıştır. Bu görüntüler, DIOR veri kümesi ile eğitilmiş YOLOv5 modeline beslenerek tekrar nesne tespiti yapılmıştır. M1 ve M2 metriklerini hesaplamak için, maskelenen görüntülerde hedef tespiti yapılarak elde edilen mAP değeri, CAM yöntemlerini uygulamak için oluşturulan orijinal test veri kümesinde hedef tespiti yapılarak elde edilen mAP değerinden çıkarılmıştır. Karşılaştırma yapılırken en yüksek M1 değerine ve en düşük M2 değerine sahip yöntemi en başarılı kabul edilmelidir. Bu analiz sonucunda elde edilen sayısal değerler Çizelge 1'de sunulmuştur.

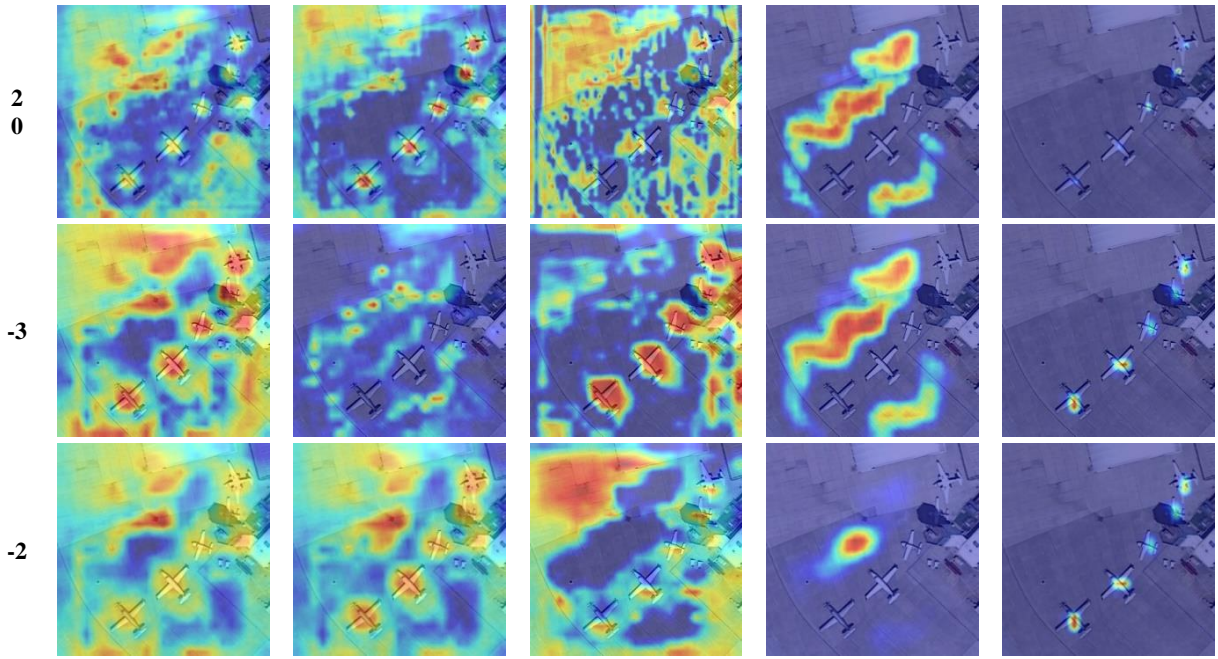


Şekil 3. M1 ve M2 yaklaşımlarına göre maskeleye örneği

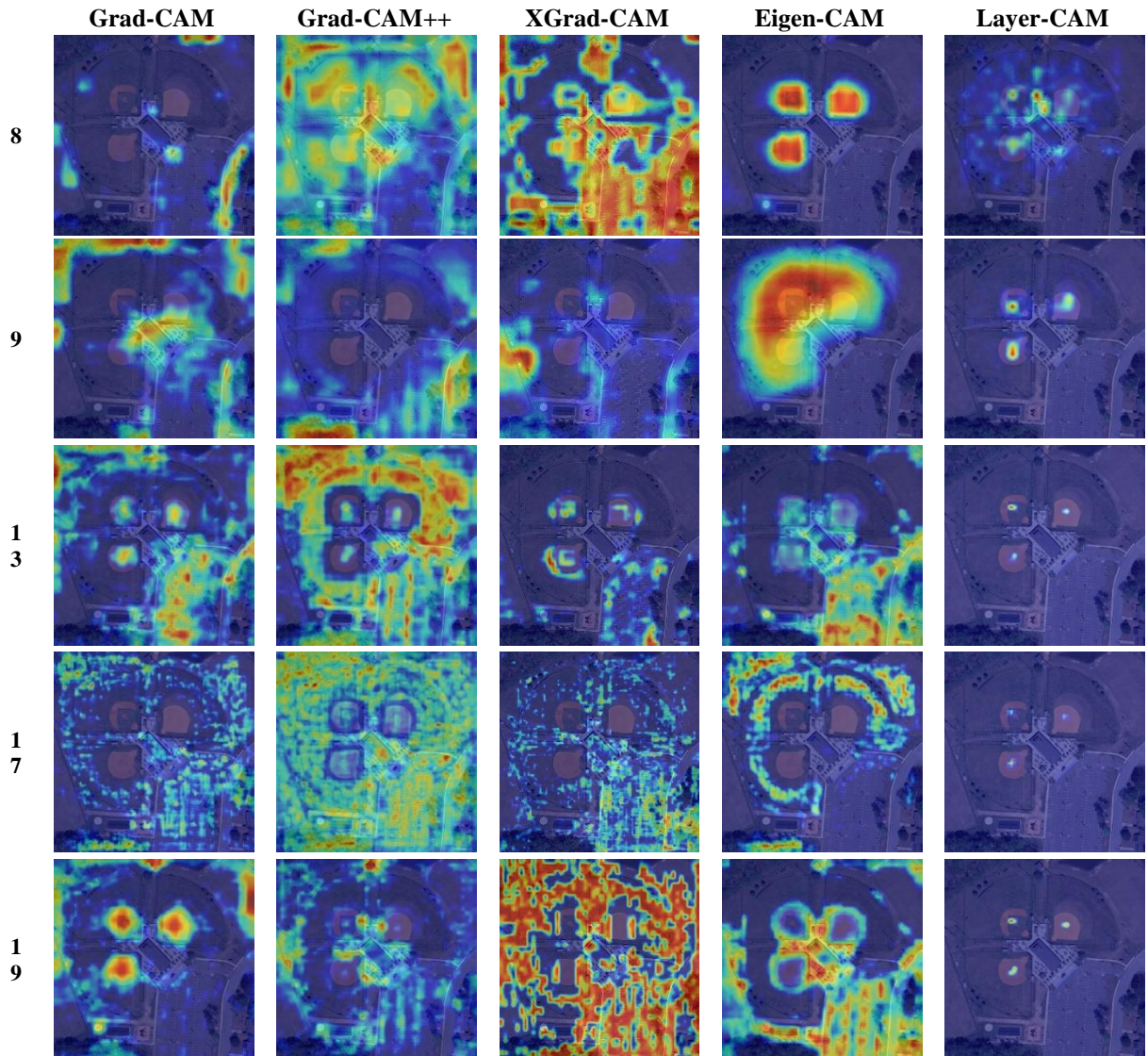
Çizelge 1. CAM yöntemlerinin katman-bazlı performansının sayısal analizi (1000 görüntü üzerinden).

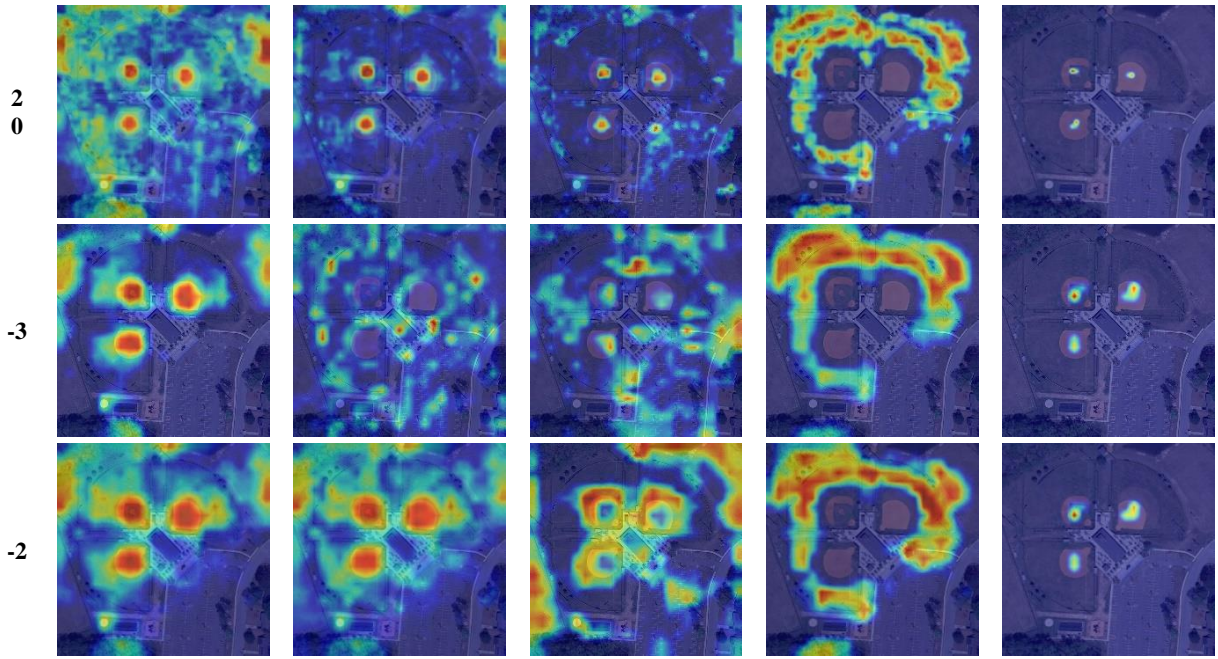
	M1					M2				
	Grad-CAM	Grad-CAM++	XGrad-CAM	Eigen-CAM	Layer-CAM	Grad-CAM	Grad-Cam++	XGrad-CAM	Eigen-CAM	Layer-CAM
Katman 8	0.13	0.42	0.18	0.28	0.11	0.91	0.76	0.92	0.74	0.97
Katman 9	0.32	0.38	0.19	0.69	0.08	0.81	0.7	0.84	0.33	0.97
Katman 13	0.27	0.31	0.17	0.36	0.06	0.83	0.9	0.93	0.67	0.97
Katman 17	0.29	0.48	0.19	0.29	0.04	0.84	0.8	0.93	0.75	0.97
Katman 19	0.46	0.42	0.18	0.29	0.06	0.68	0.82	0.92	0.7	0.9
Katman 20	0.47	0.44	0.26	0.32	0.05	0.66	0.8	0.9	0.73	0.97
Katman 22 (Katman -3)	0.65	0.46	0.2	0.35	0.06	0.38	0.7	0.87	0.64	0.96
Katman 23 (Katman -2)	0.67	0.61	0.51	0.22	0.06	0.36	0.43	0.65	0.82	0.96





Şekil 4. 'Uçak' içeren bir görüntü için AYZ yöntemlerinin katman-bazlı performansının görsel analizi.





Şekil 5. 'Golf sahası' içeren görüntü için AYZ yöntemlerinin katman-bazlı performansının görsel analizi.

Çizelge 1 incelendiğinde, Grad-CAM, Grad-CAM++ ve XGrad-CAM yaklaşımlarının, beklenilebileceği üzere, M1 yaklaşımında en yüksek başarı farkı ve M2 yaklaşımında en düşük başarı farkı ile, en başarılı sonuçlarını eksi ikinci katman olan 23. katmanda gösterdiği görülmektedir. Eigen-CAM AYZ yöntemi ile en iyi başarıyı ise 9. katmanda elde edilmiştir. Grad-CAM tabanlı yöntemlerin (Grad-CAM, Grad-CAM++, XGrad-CAM) son katmanlarda, özellikle nesne tespiti ve sınıflandırma işlemlerinin nihai kararlarının verildiği aşamalarda, daha başarılı olması, bu yöntemlerin sınıf-temelli aktivasyonları vurgulamalarından kaynaklı olabilir. Eigen-CAM ise daha erken aşamalarda, özellikle modelin görüntünün genel özelliklerini çıkardığı katmanlarda, daha etkili olabilir. Bu da Eigen-CAM'in düşük seviyeli özellikler üzerinden analiz yapabilme yeteneğini gösterir. Sonuç olarak, bu bulgu, farklı CAM yöntemlerinin modelin farklı katmanlarında değişik seviyelerde bilgiye duyarlı olduğunu ve katman bazlı analizlerin, modelin karar mekanizmalarını anlamak açısından önemli olduğunu göstermektedir.

Farklı AYZ yöntemleri için elde edilen sayısal sonuçlar karşılaştırıldığında ise M1 yaklaşımında en yüksek, M2 yaklaşımında ise en düşük başarı farkını veren, bir başka deyişle bu ölçütlere göre en başarılı olan CAM tabanlı yöntemin Eigen-CAM olduğu görülmektedir. Son olarak Layer-CAM yaklaşımında ise M1'de düşük, M2'de yüksek mAP farkı olduğunu, bir başka deyişle Layer-CAM'ın bu başarı ölçütlerine göre başarısız olduğu görülebilir. Bu durumun nedeni görsel analiz sonrası tekrar irdelenecektir.

Örnek iki veri için AYZ yöntemleri ile elde edilen ısı haritaları Şekil 4 ve Şekil 5'te sunulmuştur. Bu görseller de Eigen-CAM'in erken katmanlarda, Grad-CAM tabanlı yöntemlerin ise ileri katmanlarda daha başarılı sonuçlar verdiğini doğrular durumdadır. Ayrıca dikkatli bir görsel analiz ile, Eigen-CAM yönteminin diğer AYZ yöntemlerine göre daha başarılı sonuçlar sunduğu değerlendirilebilir. Layer-CAM ile elde edilen ısı haritaları incelendiğinde ise, bu yöntemin ısı haritalarının aslında nesne konumlarını vurguladığı görülmektedir. Ancak vurgulanan alanın nesne boyutuna göre çok küçük olması, bu alan maskelendiğinde düşük başarı düşüşüne, arka plan maskelendiğinde ise yüksek başarı düşüşüne neden olmaktadır. Bu durum Layer-CAM yönteminin bir zayıflığı olarak değerlendirilmek yerine, bu çalışmada kullanılan nicel başarı ölçütleri ile Layer-CAM yönteminin çalışma yaklaşımları arasındaki uyumsuzluk olarak değerlendirilebilir.

4. SONUÇLAR

Bu çalışmada, DIOR veri kümesi üzerinde YOLOv5 ile gerçekleştirilen nesne tespit işlemi sonrası CAM tabanlı farklı AYZ yöntemlerinin modelin farklı katmanlarına uygulanması incelenmiştir. Elde edilen sonuçlar, Grad-CAM tabanlı yöntemlerin (Grad-CAM, Grad-CAM++, XGrad-CAM) modelin son katmanlarında, özellikle nesne tespiti ve sınıflandırma aşamalarında en başarılı performansı sergilediğini ortaya koymuştur. Bu durum, söz konusu yöntemlerin sınıf-temelli aktivasyonları öne çıkarmasıyla açıklanabilir. Diğer yandan, Eigen-CAM yönteminin daha erken katmanlarda, özellikle modelin görüntünün genel özelliklerini çıkardığı aşamalarda daha etkili olduğu gözlemlenmiştir. Bu, Eigen-CAM'in düşük seviyeli özellikler üzerinden daha güçlü bir analiz yeteneğine sahip olduğunu göstermektedir. Çalışma kapsamında yapılan analizler, farklı CAM yöntemlerinin modelin farklı katmanlarına farklı duyarlılıklar sergileyerek, modelin karar mekanizmalarını daha iyi anlamaya olanak sağladığını

göstermiştir. Özellikle katman bazlı analizlerin, modelin öğrenme süreçlerini ve hangi bilgilere odaklandığını anlamak açısından kritik olduğu vurgulanmıştır. Bu bulgular, nesne tespiti modellerinin açıklanabilirliğini artırmaya yönelik yeni araştırma alanlarına ışık tutabilir. Gelecek çalışmalarda farklı derin öğrenme mimarilerinde, daha çok sayıda AYZ yöntemi için, daha çeşitli başarımlar ölçütleri kullanılarak kapsamlı deneyler yapılabilir. Bu sayede, her bir yöntemin performansı daha derinlemesine analiz edilebilir ve farklı koşullar altında hangi AYZ yöntemlerinin daha uygun olduğu daha net bir şekilde belirlenebilir.

KAYNAKLAR

Chattopadhyay A., Sarkar, A., Howlader, P. ve Balasubramanian, V., 2017. Grad-CAM++: Generalized Gradient-based Visual Explanations for Deep Convolutional Networks. *IEEE Winter Conference on Applications of Computer Vision (WACV)*.

Fu, R., Hu, Q., Dong, X., Guo, Y., Gao, Y., Li, B., 2020. Axiom-Based Grad-CAM: Towards Accurate Visualization and Explanation of CNNs. arXiv preprint arXiv:2008.02312.

Gevaert, C. M., 2022. Explainable AI for earth observation: A Review Including Societal and Regulatory Perspectives. *International Journal of Applied Earth Observation and Geoinformation*, 112:102869.

Gizzini, A. K., Shukor, M. ve Ghandour, A. J., 2023. Extending CAM-based XAI methods for Remote Sensing Imagery Segmentation. *Environ. Sci. Proc.*

Hogan, M., Aouf, N., 2021. Towards Real Time Interpretable Object Detection for UAV Platform by Saliency Maps. In 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), 1178-1183.

Hogan, M., Aouf, N., Spencer, P., Almond, J., 2022. Explainable Object Detection for Uncrewed Aerial Vehicles using Kernelshap. In 2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), 136-141.

Höhl, A., Obadic, I., Fernandez-Torres, M.-A., Najjar, H., Oliveira, D., Akata, Z., Dengel, A., Zhu, X. X., 2024. Opening the Black-Box: A Systematic Review on Explainable AI in Remote Sensing, *arXiv preprint arXiv:2402.13791*.

Jiang, P.-T., Zhang, C.-B., Hou, Q., Cheng, M.-M., Wei, Y., 2021. LayerCAM: Exploring Hierarchical Class Activation Maps. *IEEE Transactions on Image Processing*, 30: 5875 – 5888.

Kawauchi, H., Fuse, T., 2022. Shap-Based Interpretable Object Detection Method for Satellite Imagery. *Remote Sensing*, 14(9), 1970.

Li, K., Wan, G., Meng, L. Han, J., 2020. Object Detection in Optical Remote Sensing Images: A Survey and A New Benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 296-307.

Li, Y., Zhang, Y., Huang, X., Yuille, A. L., 2018. “Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images,” *ISPRS journal of photogrammetry and remote sensing*, 146, 182–196.

Muhammed, M. B., Yeasin, M., 2020. Eigen-CAM: Class Activation Map using Principal Components. *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 1-7.

Otsu N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1): 62-66.

Redmon, J., Divvala, S., Girshick, R., Farhadi A., 2016. You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 779-788.

Selvaraju, R. R., Cogswell, M., Das A., Vedantam, R., Parikh, D., Batra, D., 2017 Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 618-626.

Taskin, G., Aptoula, E., Ertürk, A., 2024. Explainable AI for Earth observation: Current methods, open challenges, and opportunities. *Advances in Machine Learning and Image Analysis for GeoAI*, 115-152.