# DERİN ÖĞRENME İLE OBJE TANIMA İŞLEMİ ÜZERİNE BİR İNCELEME

B. Bayram [a], B. Kılıç [a, 1]*, F. Özoğlu [b], F. Erdem [c], S. Sivri [a], A. Delen [d], O. C. Bayrak [a]

[a] Yıldız Teknik Üniversitesi, İnşaat Fakültesi, Harita Mühendisliği Bölümü, 34220 Davutpaşa, İstanbul – (bayram, batuhank)@yildiz.edu.tr, (sinan.sivri1, onurcbayrak)@gmail.com
[b] Coğrafi Bilgi Sistemleri Müdürlüğü, İstanbul Büyükşehir Belediyesi, 34440 Kasımpaşa, İstanbul, furkan.ozoglu@ibb.gov.tr
[c] Eskişehir Teknik Üniversitesi, Yer ve Uzay Bilimleri Enstitüsü, 26555 Tepebaşı, Eskişehir, firaterdem@eskisehir.edu.tr
[d] Gaziosmanpaşa Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Harita Mühendisliği, 60250 Taşlıçiftlik, Tokat, ahmet.delen@gop.edu.tr

**ANAHTAR KELİMELER:** ResNet50, Obje Tanıma, Makine Öğrenmesi, Derin Öğrenme, Tarihi Yapı, Görüntü İşleme

**ÖZET:**

Son yıllarda gelişen teknoloji ile birlikte, özellikle tarihi eserlere ait görüntülerinin hacmi muazzam bir şekilde artmakta ve çevrimiçi sosyal görüntülerin önemli bir bölümünü oluşturmaktadır. İnsanlar, farklı sosyal medya uygulamaları ve internet arama motorları kullanarak bu verileri gönüllü olarak paylaşmaya başlamışlardır. Aynı şekilde, cep telefonu ve kamera sistemlerindeki son gelişmeler, özellikle turistik amaçlı kullanılan görüntülerin sayısının artmasına neden olmaktadır. Turistik amaçlı çekilen görüntüler üzerinden el yapımı veya doğal nesneleri ve bu görüntülerin nereye ait olduğunu tam olarak doğru tespit etmek zordur. Bu bağlamda, bu çalışmada son teknoloji ürünü bir makine öğrenme yöntemi ve büyük veri analizi alanında hızla büyüyen bir teknik olan derin öğrenme ile obje tanıma işleminin gerçekleştirilmesi amaçlanmaktadır. Çalışmada bir evrişimsel (konvolüsyonel) sinir ağı (Convolutional Neural Network - CNN) olan ResNet50 mimarisinin turistik yapıların bulunduğu görüntülerinin tanınması problemindeki başarısı test edilmektedir. Veri setlerinin oluşturulması amacıyla, İstanbul ili sınırları içerisinde bulunan 10 adet tarihi yapı (Kız Kulesi, Sultan Ahmet Camii, Galata Kulesi, Ayasofya Müzesi, Ortaköy Camii, Topkapı Sarayı, Bozdoğan Su Kemeri, Dolmabahçe Sarayı, Dikilitaş ve Dolmabahçe Saat Kulesi) belirlenmiş ve her bir yapı için 500 olmak üzere toplamda 5000 görüntü farklı web sitelerinden toplanmıştır. Bu görüntüler ResNet50 derin öğrenme ağı ile eğitilerek yapılar üzerindeki başarı sonuçları tespit edilmiştir. Görüntülerin %70' i eğitim amaçlı kullanılmıştır. ResNet50 mimarisi ile eğitim ve test işlemleri sonucunda % 96.3 oranında doğruluk başarısı elde edilmiş ve turistik amaçlar için obje tanıma problemlerinde kullanılabilirliği sonucuna varılmıştır.

# A STUDY ON OBJECT RECOGNITION WITH DEEP LEARNING

**KEY WORDS:** ResNet50, Object Recognition, Machine Learning, Deep Learning, Historical Building, Image Processing

**ABSTRACT:**

Recently, the volume of the images of historical heritages grew due to developments of the camera and smart phone technology and these images became an important part of the online social networks. Especially, the number of shared images for touristic purposes increased. People began to share this data voluntarily using different social media applications and internet search engines. However, to obtain semantic and location information accurately from these touristic images may be very difficult. Deep learning is a state-of-art technology and its application areas are growing rapidly in the field of big data analysis. Therefore, in this study, it is aimed to realize object recognition process using deep learning technique. In this study, the performance of ResNet50 convolutional neural network (CNN) has been tested for recognizing of touristic structures from images. For this purpose, 10 historical buildings within the city of Istanbul (Maiden's Tower, Blue Mosque, Galata Tower, Hagia Sophia, Ortakoy Mosque, Topkapi Palace, Valens Aqueduct, Dolmabahçe Palace, Obelisk Theodosius, and Dolmabahçe Clock Tower) were selected. The established system has been trained for recognition of these landmarks by using deep learning. Five hundred images of each monument, totally five thousand images were collected from different websites. %70 of these images have been used for training of ResNet50 architecture. According to accuracy assessment results, %96.3 accuracy has been obtained with ResNet50 architecture.

## 1. GİRİŞ

Derin öğrenme (DÖ), bilgisayarların insanlara benzer bir şekilde basit kavramların hiyerarşisinden yararlanarak öğrenmelerine dayanan bir makine öğrenmesi yöntemi olarak adlandırılmakta ve günümüzde büyük veri analizinde yeni ve hızlı şekilde büyüyen bir çalışma alanı olarak ifade edilmektedir (Goodfellow vd., 2016). DÖ' nin başlangıcının 1940'lı yıllara dayandığı görülmektedir (Pitts & McCulloch, 1947). Bu nedenle, derin öğrenme geçmişten günümüze gelişerek ulaşmış ve halen gelişmeye devam etmektedir (Zhu vd., 2017). Bilgisayar ve grafik işleme birimlerindeki (GPU'ların) gelişmeler sonucunda, DÖ geniş bir kullanım alanına kavuşmuştur. 2013 yılında en gelişmiş araştırma alanlarından biri olarak kabul edilmiştir (MIT Technology Review, 2019). Etkili görsel tanıma işlevini sağlayan süreçlerden biri, insanın çeşitli yer ve örnek kümelerini öğrenme ve hatırlama kapasitesidir. İnsan beyni dünyayı saniyede birkaç kez örnekleyerek kısa bir süre için bile yeni girdiler kaydetmekte ve yaşam boyunca bu işlem milyonlarca kez tekrarlanmaktadır (Zhou vd., 2014).

Son yıllarda gelişen teknoloji ile birlikte, çevrimiçi sosyal medya önemli bir veri altyapısı konumuna ulaşmıştır. Bu büyük verilerin toplanması, kullanıcılar tarafından depolanması ve iletilmesi, çeşitli uygulamalar (örneğin Flickr, Facebook ve Instagram) ve internet arama motorları tarafından sağlanmaktadır. Yer tespiti (Hays & Efros, 2008), yüz ve manzara tanıma (Parkhi vd., 2015; Zhou vd., 2018) gibi çeşitli çalışmalar ve uygulamalar bu tür büyük veri tabanları üzerinden gerçekleştirilmektedir. Bunların dışında, söz konusu arama motorlarındaki ve uygulamalardaki fotoğrafların büyük bir kısmı genellikle tarihi eserlere ilişkin turistik fotoğraflardır. Bir fotoğraftaki objeleri veya yapıları tanımak bilgisayarlı tanı işleminin en önemli problemlerinden biridir. DÖ, özellikle açık veri tabanları ve internet kaynakları kullanılarak, görüntü ve bilgi analizi için çok kullanışlı bir araç haline gelmiştir (Tzelepi & Tefas, 2018). Jiang vd. (2017), DÖ ile gerçek zamanlı bir internet üzerinden görüntü tanıma sistemi geliştirmişlerdir. Huang vd. (2018), derin öğrenmeyi, MIR Flickr 2011 ve NUS-WID veri setini kullanarak çoklu kavram tabanlı görüntü tanıma problemlerini çözmede bir araç olarak sunmuştur. Xu vd. (2019) aynı veri setini kullanmış ve derin öğrenme tekniklerini kullanarak semantik görüntü tanıma sistemi geliştirmiştir. Huang vd. (2018) DÖ ile model tabanlı görüntü tanıma sistemi geliştirmişlerdir.

Tarihi ve turistik yapılar, fiziksel özellikleri nedeniyle fotoğraflarını çevrimiçi sosyal uygulamalar yoluyla aile ve/veya arkadaşlarıyla paylaşmaya meraklı birçok yabancı/yerel ziyaretçi için ilgi odağı olmaktadır (Cheng & Shen, 2016). Milyarlarca kullanıcı seyahat resimlerini sosyal medya platformlarında paylaşsa da çoğu bir etiket olmadan paylaşılır. Yabancı ve hatta yerel turistler ziyaret edilen bölgeye ait konumsal ve semantik bilgiye sosyal medya platformlarında paylaşılan görüntülerden çoğunlukla sahip olamamaktadırlar. Dolayısı ile ziyaret edilen bölge ve eser ile ilgili isim, önemi, kısa tarihi gibi çeşitli genel bilgiye ulaşmakta da zorlanabilmektedirler. Bu nedenle, bu çalışmada, tarihi yapıların otomatik olarak tanınarak, bu yapılar hakkında konum ve semantik bilgilerin elde edilebileceği DÖ tabanlı bir sistem önerilmiştir. Bu amaçla bir evrişimsel sinir ağı (ESA) (Convolutional Neural Network - CNN) olan ResNet50 mimarisinin İstanbul ili sınırları içerisinde bulunan çeşitli karakteristik özelliklere sahip on yapıyı tanımadaki başarısı irdelenmiştir. Çalışma Python programlama dilinde, Keras kütüphanesi kullanılarak gerçekleştirilmiştir.

## 2. ÇALIŞMA ALANI VE VERİ SETİ

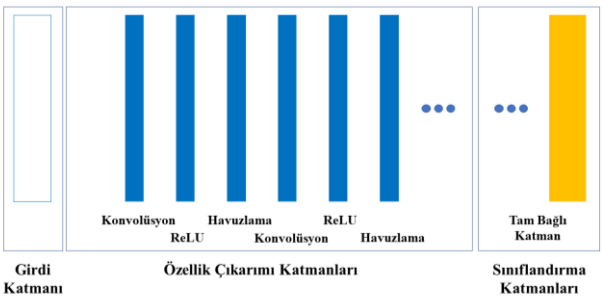
Sunulan çalışmada veri seti oluşturmak amacıyla, İstanbul kentinde bulunan 10 adet tarihi yapı belirlenmiştir. Belirlenen tarihi yapılar; Kız Kulesi, Sultan Ahmet Cami, Galata Kulesi, Ayasofya, Ortaköy Cami, Topkapı Sarayı, Valens Su Kemeri, Dolmabahçe Sarayı, Dikilitaş ve Dolmabahçe Saat Kulesi'dir. Özellikle bu noktaların seçilmesinin nedeni, Türkiye ve İstanbul'un önemli kültürel miraslarından olan ve yabancı ziyaretçilerin oldukça yoğun ilgi gösterdiği tarihi yapılar olmasıdır.

Belirlenen tarihi yapılar için Bing, Foursquare ve Yandex web sitelerinden indirilen her bir tarihi yapı için 500 adet olmak üzere toplamda 5000 görüntüden oluşan bir veri seti hazırlanmıştır. Veri setlerinin % 70'i eğitim, % 10'u doğrulama, % 20'si ise test için ayrılmıştır. Tüm görüntüler, 224 x 224 boyutlarında yeniden ölçeklendirilmiştir. Olası deformasyonlardan kaçınmak için kare görüntüler tercih edilmiştir. Şekil 1'de veri setine ait örnekler gösterilmektedir.

Şekil 1. Veri Setinden Örnekler

## 3. METODOLOJİ

Görüntü işleme problemlerinde en çok kullanılan DÖ modellerinden biri ESA'lardır. Bir ESA, genellikle konvolüsyon, aktivasyon fonksiyonu, pooling, fully-connected layer ve softmax katmanlarından oluşmaktadır. Konvolüsyon katmanı, çekirdek matris kullanılarak girdi verilerinden özellik haritalarının oluşturulması için; aktivasyon fonksiyonu, konvolüsyon sonucunda üretilen özellik haritasına doğrusal olmayan bir yapı kazandırmak için kullanılmaktadır. Örneğin ReLU aktivasyon fonksiyonunda, çıkarılan özelliklerinden pozitif olanlar bir sonraki katmana direkt olarak, negatif olanlar ise sıfır değeri ile aktarılır. Maksimum havuzlama katmanı ise gereksiz bilgilerin kullanımını engelleyerek hesaplama maliyeti azaltmak için kullanılmaktadır (Patterson & Gibson, 2017). Konvolüsyon ve havuzlama katmanlarında yüksek seviyeli özelliklerin çıkarımıyla birlikte, ağın sonunda genellikle Tam Bağlı Katman (Fully Connected Layer) bulunmaktadır. Bu katmandaki tüm nöronlar ile önceki katman arasında bağ bulunmaktadır. Tam Bağlı Katman kullanılarak özellik haritaları ile Softmax katmanı arasında bağlantı kurulur. Son olarak Softmax fonksiyonu ile her bir girdi verisinin belirli bir sınıfa ait olma olasılığı hesaplanmaktadır (Goodfellow vd., 2016). Bir konvolüsyonel sinir ağının temel bileşenleri Şekil 2'de gösterilmektedir.

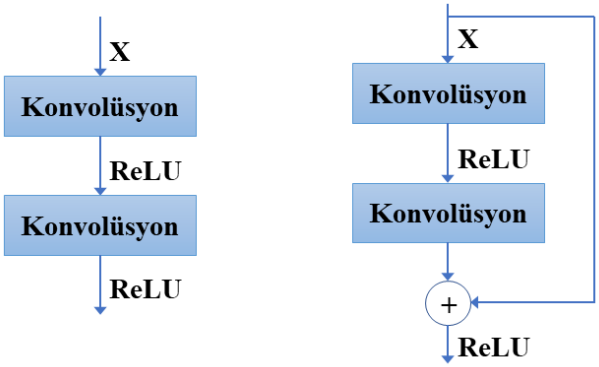



Şekil 2. Bir konvolüsyonel sinir ağının temel bileşenleri.

Çalışmada, görüntü sınıflandırma amacıyla geliştirilen ResNet50 DÖ mimarisi kullanılmıştır. ResNet mimarileri, Microsoft araştırma ekibi tarafından derinliği fazla olan sinir ağlarının eğitimindeki zorluğu azaltmak amacıyla geliştirilen mimarilerdir. ResNet'in 18, 34, 50, 101 ve 152 ağırlık katmanından oluşan çeşitli türleri bulunmaktadır. Sunulan

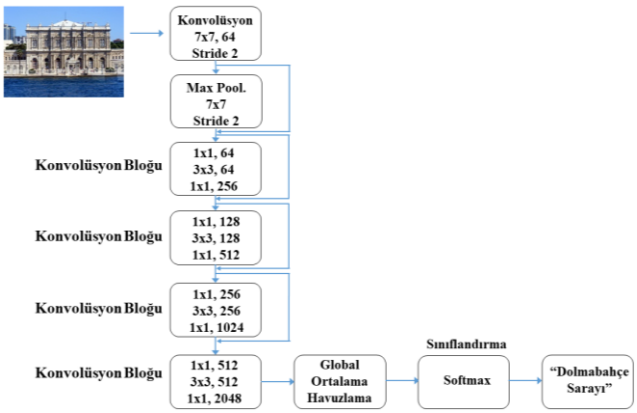
çalışmada 50 ağırlık katmanına sahip ResNet50 mimarisi kullanılmıştır.

ResNet mimarilerinde standart ESA'lardan farklı olarak kısa yol bağlantıları kullanılmaktadır. Kısa yol bağlantıları, ekstra parametre içermemekte ve hesaplama karmaşıklığına neden olmamaktadır (He vd., 2016). Kısa yol bağlantılarının kullanılmasıyla önceki katmandaki önemli bilgiler sonraki katmanlara aktarılabilir. Şekil 3.'de standart bir ESA ile ResNet mimarilerinde kullanılan kısa yol bağlantılarının açıklaması verilmiştir.




Şekil 3. Standart ESA (sol); ResNet mimarilerinde kullanılan kısa yol bağlantıları (sağ).

ResNet mimarilerinde, ağın sonunda Global Ortalama Havuzlama (Global Average Pooling) katmanı ve bir tam bağlı katman bulunmaktadır. global ortalama havuzlama işleminde, her bir özellik haritasındaki ortalama değer bir sonraki katmana aktarılmaktadır (Lin vd., 2013). Şekil 4.'de sunulan çalışmada kullanılan ResNet50 mimarisinin katmanları verilmiştir.




Şekil 4. ResNet50 mimarisinin katmanları.

ResNet50 derin öğrenme mimarisi Python Keras kütüphanesinde (Chollet, 2019) oluşturulmuştur. Adaptif öğrenme oranlarını kullanan optimizasyon algoritmalarının diğer optimizasyon algoritmalarına göre oldukça sağlam (robust) performans gösterdiği fakat standart bir "en iyi optimizasyon algoritmasının" mevcut olmadığı belirtilmiştir (Schaul vd., 2013). Bu sebeple sunulan çalışmada kullanılan DÖ modelinin eğitim işlemi adaptif öğrenme oranını kullanan Adadelta optimizasyon algoritması ile gerçekleştirilmiştir. Bu yöntem, öğrenme hızının manuel ayarlanmasına gereksinim duymamaktadır. Hesaplama maliyeti düşük bir optimizasyon algoritmasıdır (Zeiler, 2012).

Sunulan çalışmada derin sinir ağlarının eğitim işlemi kullanılan hiper parametreler Tablo 1.'de verilmiştir. Kullanılan hiper parametreler deneysel olarak seçilmiştir.

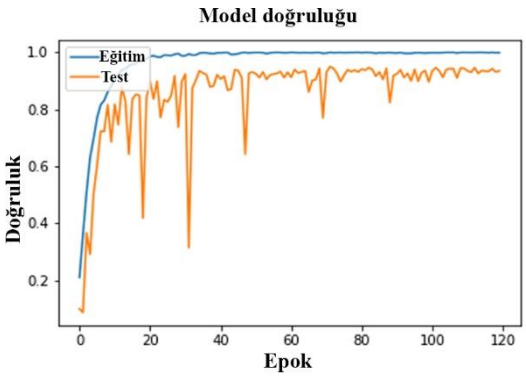
| Parametre | Parametre Seçimi |
|---|---|
| Başlangıç Öğrenme Oranı | 1 |
| Rho (ρ) | 0.95 |
| Epok Sayısı | 120 |
| Küme Boyutu | 16 |

Tablo 1. Kullanılan hiper parametreler

Öğrenme oranı (learning rate), ağ parametrelerinin ne kadar hızlı güncellendiğini göstermektedir. Adaptif optimizasyon algoritmalarının kullanılması durumunda bu değer öğrenme sırasında otomatik olarak öğrenilmektedir ve sürekli değişim halindedir. Sabit ρ değeri, ağın geri yayılımı aşamasında parametrelerin güncellenmesinde kullanılan bir ölçüttür (Patterson ve Gibson, 2017). Epok sayısı, iterasyon sayısıdır (Patterson ve Gibson, 2017). Küme boyutu ise; eğitim aşamasında bir epoktaki tek bir adımda işleme tabi tutulan eğitim verisi sayısını göstermektedir (Soon vd., 2018).

## 4. SONUÇLAR

Eğitim sürecindeki başarının değerlendirilmesinde, her bir çalışma epoğunda elde edilen doğruluk Şekil 5' te verilmiştir. Eğitim süreci için elde edilen doğruluk grafiği incelendiğinde özellikle ilk epoklarda kararsız bir eğitim süreci gözlemlenmiştir. Daha sonraki epoklarda daha stabil değerler elde edilmiş, ve 120. epoğun sonunda % 93'lere ulaşan bir test doğruluğu elde edilmiştir.




Şekil 5. Eğitim sürecine ait elde edilen doğruluk grafiği.

Eğitilen sinir ağları test verileri kullanılarak incelenmiştir. Sonuçların analizi için test doğruluğu (test accuracy), ortalama kesinlik (average precision), ortalama tanıma (average recall) ve ortalama F1 skor (average F1 score) kalite değerlendirme ölçütleri kullanılmıştır. Elde edilen sonuçlar Tablo 2.'de verilmiştir. ResNet50 derin öğrenme modeli ile 1000 adet test görüntüsü için % 93.20 genel doğruluğa ulaşılmıştır. Kesinlik, tanıma ve F1 skor değerleri sırasıyla %93.43, %93.20 ve %93.22 olarak elde edilmiştir.

| Doğruluk (%) | Ortalama kesinlik (%) | Ortalama Tanıma(%) | Ortalama F1 Skoru(%) |
|---|---|---|---|
| 93.20 | 93.43 | 93.20 | 93.22 |

Tablo 2. Doğruluk analizi sonuçları

Modelin yapmış olduğu yanlış tahminlerin daha iyi analiz edilebilmesi amacıyla hata matrisi oluşturulmuştur (Tablo 3). Hata matrisinde yer alan satırlar gerçek sınıfı, sütunlar ise

tahmin edilen sınıfı göstermektedir. Tablo incelendiğinde en çok hata yapılan sınıfın 4. sınıf olan Ayasofya olduğu görülmektedir. Bu sınıf için 85 test görüntüsünde doğru tahmin yapılırken, 15 test görüntüsünde hatalı tahmin elde edilmiştir. ResNet50 modelinin, Ayasofya'ya ait 8 test görüntüsünde 2. sınıf olan Sultan Ahmet Camii olarak tahmin yaptığı gözlemlenmiştir. Benzer bir durum 5. sınıf olan Ortaköy Camii için de geçerlidir. Bu tarihi yapıların cami olması, dolayısıyla kubbe ve minare gibi benzer karakteristik dokulara sahip olmaları nedeniyle hatanın diğer sonuçlara oranla daha fazla olduğu görülmektedir. Bu sınıflar için eğitim veri setinde daha fazla ayırt edici görüntünün bulunmasıyla hatalı tahminlerin önüne geçilebileceği öngörülmektedir.

| Tarihi Yapı | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 96 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 2 | 0 | 93 | 0 | 1 | 2 | 0 | 1 | 0 | 1 | 2 |
| 3 | 3 | 0 | 95 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 4 | 0 | 8 | 2 | 85 | 1 | 0 | 3 | 1 | 0 | 0 |
| 5 | 0 | 2 | 3 | 1 | 89 | 0 | 0 | 4 | 0 | 1 |
| 6 | 0 | 0 | 0 | 0 | 0 | 97 | 0 | 1 | 0 | 2 |
| 7 | 0 | 0 | 1 | 0 | 0 | 0 | 96 | 1 | 2 | 0 |
| 8 | 3 | 0 | 2 | 0 | 0 | 0 | 0 | 95 | 0 | 0 |
| 9 | 0 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 93 | 1 |
| 10 | 0 | 0 | 3 | 1 | 1 | 0 | 1 | 1 | 0 | 93 |

Tablo 3. Hata matrisi (1. Kız Kulesi, 2. Sultan Ahmet Cami, 3. Galata Kulesi, 4. Ayasofya, 5. Ortaköy Cami, 6. Topkapı Sarayı, 7. Valens Su Kemeri, 8. Dolmabahçe Sarayı, 9. Dikilitaş, 10. Dolmabahçe Saat Kulesi)

## 5. TARTIŞMA

Sunulan çalışmada kültürel objeleri otomatik olarak tanımlayan bir derin öğrenme modeli kullanılmıştır. Yapılan literatür araştırması sonucunda, önerilen sistemin en azından Türkiye'de turizm alanında öncü çalışmalardan biri olduğu görülmüştür.

Seçilen on tarihi yapı, Türkiye ve İstanbul'un önemli kültürel miraslarından olan ve ziyaretçiler tarafından oldukça ilgi gören tarihi yapılardır. Her birinin farklı doku ve şekil özelliklerinin olması sebebiyle veri setinin hazırlanması aşamasında farklı karakteristikteki görüntüleri toplamak mümkün olmuştur.

Çalışmada karşılaşılan önemli bir kısıt olarak, her bir tarihi yapıda görüntü sayısının yeterli miktarda olmaması söylenebilir. Mevcut sosyal medya sitelerinden her bir tarihi yapı için 500 görüntü temin edilmiştir. Daha fazla sayıda görüntüden oluşan bir veri setinin hazırlanması, sunulan yöntemlerin başarısını artıracaktır. Turist ziyaretçilerin sosyal medya sitelerinde fotoğraf paylaşırken kullandığı etiketler (hashtag) sayesinde görüntü sayısı artırılarak daha geniş bir veri setinin hazırlanması olanaklı olabilecektir.

Çalışmada ResNet50 derin öğrenme modeli kullanılarak %93.22 ortalama F1 skor değeri elde edilmiştir. Hata matrisi incelendiğinde benzer karakteristikteki tarihi yapılarda hata yapıldığı gözlemlenmiştir. Eğitim veri setinde benzer tarihi yapıların ayırt edici özelliklerini içeren görüntülerin bulunması yöntemin başarısını arttıracaktır.

Çalışmanın bir sonraki aşamasında tarihi yapı sayısı ve her tarihi yapı için kullanılacak olan görüntü sayısı arttırılacak ve farklı derin öğrenme modelleri ile eğitim ve test işlemleri gerçekleştirilecektir.

## KAYNAKLAR


Cheng, Z., Shen, J., 2016. On very large scale test collection for landmark image search benchmarking. *Signal Processing*, 124, pp. 13-26.

Chollet, F., https://github.com/fchollet/keras/ (Last Updated: March 27, 2019).

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. *Deep Learning*. MIT Press, Cambridge.

Hays, J., Efros, A. A., 2008. IM2GPS: estimating geographic information from a single image. In: *Computer Vision and Pattern Recognition, IEEE Conference*, pp. 1-8.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778.

Huang, C., Xu, H., Xie, L., Zhu, J., Xu, C., Tang, Y., 2018. Large-scale semantic web image retrieval using bimodal deep learning techniques. *Information Sciences*, 430, pp. 331-348.

Huang, F., Zhang, X., Zhao, Z., Li, Z., He, Y., 2018. Deep multi-view representation learning for social images. *Applied Soft Computing*, 73, pp. 106-118.

Jiang, B., Yang, J., Lv, Z., Tian, K., Meng, Q., Yan, Y., 2017. Internet cross-media retrieval based on deep learning. *Journal of Visual Communication and Image Representation*, 48, pp. 356-366.

Lin, M., Chen, Q., Yan, S., 2013. Network in network. *arXiv preprint*, arXiv:1312.4400.

MIT Technology Review, https://www.technologyreview.com/lists/technologies/2013/ (Last Updated: March 27, 2019).

Parkhi, O. M., Vedaldi, A., Zisserman, A., 2015. Deep face recognition. In: *Proceedings of the British Machine Vision Conference (BMVC)*, 1(3), pp. 6.

Patterson, J., Gibson, A., 2017. *Deep Learning: A Practitioner's Approach*, California, O'Reilly Media.

Pitts, W., McCulloch, W. S., 1947. How we know universals the perception of auditory and visual forms. *The Bulletin of mathematical biophysics*, 9(3), pp. 127-147.

Schaul, T., Antonoglou, I., Silver, D., 2013. Unit tests for stochastic optimization. *arXiv preprint*, arXiv:1312.6055.

Soon, F. C., Khaw, H. Y., Chuah, J. H., Kanesan, J., 2018. Hyper-parameters optimisation of deep CNN architecture for vehicle logo recognition. *IET Intelligent Transport Systems*, 12(8), pp. 939-946.

Tzelepi, M., Tefas, A., 2018. Deep convolutional image retrieval: A general framework. *Signal Processing: Image Communication*, 63, pp. 30-43.

Xu, H., Huang, C., Wang, D., 2019. Enhancing semantic image retrieval with limited labeled examples via deep learning. *Knowledge-Based Systems*, 163, pp. 252-266.

Zeiler, M. D., 2012. ADADELTA: an adaptive learning rate method. *arXiv preprint*, arXiv:1212.5701.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., 2014. Learning deep features for scene recognition using places database. In: *Advances in neural information processing systems*, pp. 487-495.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2018. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6), pp. 1452-1464.

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), pp. 8-36.