

# Derin Öğrenmeyle Bir Fotoğrafın Yerini Bulma

## Finding Location of A Photograph with Deep Learning

Yunus Emre Özköse, Tarık Ayberk Yılıkoğlu, Levent Karacan ve Aykut Erdem  
Bilgisayar Mühendisliği  
Hacettepe Üniversitesi  
Ankara, Türkiye

{aliozkose, tariyilikoglu} @hacettepe.edu.tr {karacan, aykut} @cs.hacettepe.edu.tr

**Özetçe** —Verilen bir fotoğrafın çekildiği yerin bulunması zorlu bir bilgisayarlı görü problemidir. Bu bildiriye, derin öğrenme teknikleri kullanılarak ilk defa Türkiye’deki şehirler için fotoğrafların hangi şehirde çekildiğini bulan farklı derin öğrenme modelleri üzerinde çalışılmıştır. Bu amaç için Türkiye’deki 15 şehirden fotoğrafların yer aldığı yeni bir veri kümesi oluşturulmuş ve büyük veri kümeleri üzerinde önceden eğitilmiş derin evrimsel sinir ağları(ESA) kullanılarak aktarım öğrenimi(transfer learning) gerçekleştirilmiştir. Kullanılan önceden eğitilmiş modellerin önerilen veri kümesi üzerindeki yer bulma başarımları karşılaştırılmıştır.

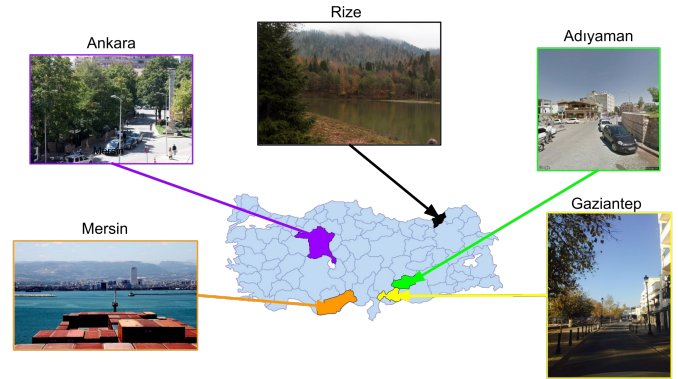
**Anahtar Kelimeler**—derin öğrenme, görüntü coğrafik konumlama, aktarım öğrenimi

**Abstract**—Finding location of a photograph is a challenging computer vision problem. In this paper, different deep learning models which finds the city where a photo was taken were studied for the cities of Turkey first time. For this purpose, a new dataset in which there are photos from 15 cities of Turkey was collected and transfer learning was performed using pre-trained deep convolutional neural network(CNN) models for image recognition on big visual datasets and those pre-trained models are compared according to their prediction performances on proposed dataset.

**Keywords**—deep learning, image geo-localization, transfer learning

### I. GİRİŞ

Biz insanlar için verilen bir fotoğrafın nerede çekildiğini bulmak kolay gibi görünsede aslında çoğu durumda oldukça zorlu bir problemdir. Çünkü dünya üzerinde birçok yer benzer özellikler göstermektedir ve bizim benzer özellikler gösteren fotoğrafları ayırt etmeye yarayacak görsel ipuçlarına ihtiyacımız vardır. Biz görüntüleri ayırt etmemize yarayan bu çeşitli görsel ipuçlarını zaman içerisinde gözlemleyerek öğreniriz. Öğrendiğimiz bu görsel ipuçlarını kullanarak fotoğrafları kolayca tanıyabilir ve nerede çekildiklerini tahmin edebiliriz. Fotoğrafın yere göre özgünlüğü ve ayırt edici özelliğinin artması fotoğrafı daha kolay tanımamızı sağlar. Örneğin, İstanbul’da bulunan tarihi Galata Kulesi oldukça özgün bir yapıdır ve fotoğraflardan tanınması oldukça kolaydır. Ancak pek çok yerde benzerini gördüğümüz okul, hastane ve cami gibi çeşitli yapılara ait fotoğrafların nereye ait olduğunu tahmin etmek zordur. Ayrıca görüntülerdeki ışık, çekim açısı gibi değişkenler



Şekil 1: Bu çalışmada Türkiye’de çekilmiş bir fotoğrafın hangi şehirden çekildiğini bulan derin evrimsel sinir ağı modeli eğitiyoruz.

problemi daha da zorlu hale getirmektedir. Örneğin, Şekil 1’de verilen görüntülerin çekildiği şehirleri tahmin edebilmek için çeşitli görsel ipuçlarına ihtiyaç vardır. Daha açık bir ifadeyle birden fazla şehirde deniz görüntüsü olabilir veya cadde görüntüleri her şehirden elde edilebilecek benzer görüntülerdir. Önemli olan nokta bu görüntülerin farklı konumlar için ayırt edilebilmelerini sağlayan görsel ipuçlarıdır. Bunlar bir trafik levhası, ağaç türü, bina şekilleri gibi birçok görsel bilgi olabilir. Bu zorlukların aşılması için bilgisayarların hesaplama güçlerinden faydalanılarak fotoğrafların konumunu belirleyen ve çoğu durumda insanlar için bulunması zor olan görsel ipuçlarının öğrenilmesini sağlayan teknikler geliştirilmektedir. Özellikle son yıllarda akıllı telefonların yaygınlaşmasıyla artan görsel veriler problemin çözülmesi için öğrenmeye dayalı yöntemleri ön plana çıkarmaktadır.

Geçmişte bu problem çoğunlukla coğrafi konum bilgilerinin bulunduğu çok büyük veri kümeleri içerisinde görüntü arama problemi olarak ele alınmıştır [1], [3], [4], [8], [9], [12]. Son yıllarda derin evrimsel sinir ağlarının birçok problemdeki başarısından dolayı derin öğrenmeye dayalı yöntemler bir görüntünün çekildiği yeri belirlemede kullanılmaya başlanmıştır [7], [10], [13]–[15].

Bu çalışmada ilk kez, Türkiye’deki şehirler için bir fotoğrafın yerini bulmayı bir sınıflandırma problemi olarak ele aldık ve bunun için farklı derin evrimsel sinir ağları (ESA)

modelini eğiterek inceledik. Bu amaçla, Google Street View ve Flickr'dan Türkiye'deki farklı şehirlerden çekilmiş fotoğrafları toplayıp Türkiye15 isimli bir veri kümesi oluşturduk. Veri kümemizin belirtilen problem için uygunluğunu belirlemek için çeşitli özneteliklerle testler gerçekleştirdik. Bu işlemle ilgili ayrıntılı bilgiler Bölüm 3'te verilmektedir.

Oluşturduğumuz veri kümesi üzerinde görüntü sınıflandırma problemi için büyük görsel veriler üzerinde önceden eğitilmiş AlexNet [6] ve ResNet [5] derin sinir ağı modellerinden aktarım öğrenimi gerçekleştiriyoruz. Ayrıca bu ağ yapılarının Places365 [16] ve ImageNet [2] büyük görsel veri kümelerinden öğrenilen modellerinden yapılan aktarım öğrenimi başarımlarını da karşılaştırıyoruz. Gerçekleştirilen aktarım öğrenimi ile ilgili ayrıntılar Bölüm 4 ve Bölüm 5'te verilmektedir.

## II. İLGİLİ ÇALIŞMALAR

Görüntülerin çekildiği yerleri bulmak için Schindler vd. [12] şehir ölçeğinde büyük cadde görüntüleri verisi içerisinde tanımlanan el yapımı öznetelikler ile bir girdi görüntüsünün yerini tanıma problemi üzerinde çalışmışlardır. Bu çalışmada etkili bir şekilde öznetelik uzayında arama gerçekleştirebilmek için sözlük ağaçları yaklaşımı benimsenmiştir.

Verilen bir girdi görüntüsünün coğrafi konumunu bulan IM2GPS [3], [4] bu alandaki öncü çalışmalardan birisidir. Veriye dayalı bir yaklaşımı kullanarak coğrafi konumları etiketlenmiş büyük görüntü veri kümesi üzerinde girdi görüntüsüne benzeyen görüntüler çeşitli öznetelikler kullanılarak bulunmaya çalışılır. Girdi görüntüsüne en yakın komşu görüntüler bulunarak tüm dünya yüzeyi üzerinde girdi görüntüsünün nerede çekilmiş olabileceğini gösteren olasılık haritası çıkarılmaktadır.

Li vd. [8] öznetelikleri SIFT ile göstermeyi ve sorgu resmindeki özneteliklerle veri kümesindeki resimlerin özneteliklerini karşılıklı olarak eşleştirmeyi önermiştir, fakat eşleştirme sadece öncelikli özneteliklerle yapılmıştır. Sadece resimlerdeki bilgi verebilecek noktaları saklamışlardır. Bu yolla hesaplama maliyeti azaltılmıştır.

Veri kümesindeki görüntüler yeryüzünde çoğunlukla belli noktalardan çekilmektedir. Büyük veri kümeleri içinde görüntü bulma yaklaşımlarının başarısı ilgili örneklerle ilişkili görüntülerin veri kümesinde bulunmasına bağlıdır. Bu problemin üstesinden gelmek için farklı görüntü alanları arasında çevrim işlemi yapılarak veri kümesinde örnek görüntüler olmadığı durumlar için de konum belirleme yapılmaya çalışılmıştır [9].

Qian vd. [11] sadece sözcük torbası (bag of words) yerine herbir resim için görsel kelime grupları oluşturmayı önermiştir. Yer tahminleri için adaylar seçilirken evrensel öznetelikler, öznetelik çıkarımı içinse yerel öznetelikler kullanılmıştır. Kendi oluşturdukları görsel kelime grup oluşturucuları ile resimleri saflaştırdıktan sonra kelime grupları oluşturulmuştur.

Derin öğrenmenin yaygınlaşmasına neden olan görüntü tanıma çalışması [6] büyük veri kümesi ImageNet [2] üzerinde çok başarılı sonuçlar almıştır. Bu çalışmayla birlikte evrişimsel sinir ağları yapısı birçok bilgisayarlı görü alanında kullanılmaya başlanmıştır. Ardışıl evrişim işlemlerinden oluşan bu yapıyı önceki çalışmalardan daha etkili kılan el yapımı



Şekil 2: Veri kümesinde farklı özellikte fotoğraflar yer almaktadır.

öznetelikler yerine ilgili probleme uygun sıradüzenli özneteliklerin ham verinin kendisinden öğrenilmesidir. Bu çalışmanın ardından evrişimsel sinir ağları yapısını daha etkili yapabilmek için farklı ağ yapıları da önerilmiştir. Diğer büyük iyileştirmeyi yapan ResNet [5], ardışıl evrişim katmanları arasında bilgi akışını kaybetmemek için katmanlar arasındaki artık bilgileri öğrenmeyi amaçlamıştır.

Derin öğrenme, birçok bilgisayarlı görü problemindeki başarısıyla birlikte görüntülerin konumunu belirleme problemi için de kullanılmaya başlanmıştır. Lin vd. [10] yerden ve havadan çekilen görüntü çiftlerinden, yerden çekilen bir görüntünün havadan konumunu belirleyen bir evrişimsel sinir ağı yapısı önermişlerdir. Evrişimsel sinir ağlarını aynı yerin farklı alanlardaki (yerden ve havadan) görüntüleri arasında ortak temsili öğrenmeye çalışan benzer bir çalışma Workman vd. [15] tarafından önerilmiştir.

Lee vd. [7] görüntülerin yerleri yerine çekildikleri yerlere ait yükselti, nüfus yoğunluğu, gelir düzeyi gibi farklı nitelikleri tahmin edebilen evrişimsel sinir ağı eğitimi amaçlamışlardır.

PlaNet [14] çalışması ile verilen bir görüntünün yerini bulma problemi yeryüzü çok ölçekli coğrafik hücelere bölünerek bir sınıflandırma problemine dönüştürülmüştür. Daha sonra milyon büyüklüğündeki veri üzerinden derin evrişimsel sinir ağı eğitilerek verilen bir görüntünün hangi hücreye ait olduğu bulunmaya çalışılmıştır.

Vo vd. [13] derin öğrenmenin gücünü IM2GPS [3] çalışmasıyla birleştirerek, PlaNet [14] çalışmasına benzer bir evrişimsel sinir ağı modelini sınıflandırma amaçlı kullanarak, test aşamasında eğitilen ağdan elde edilen öznetelikleri görüntü aramak için kullanmışlardır.

## III. VERİ KÜMESİ

Çalışmamızı Türkiye ölçeğinde sınırlı tuttuğumuz için Flickr üzerinden Türkiye'deki 15 şehirden her şehir için yaklaşık olarak 150 görüntü toplamda 2250 görüntü topladık ve veri kümemize Türkiye15 adını verdik. Verimizin 1500'ünü eğitim kümesi, 375'ini doğrulama kümesi ve 375'ini test kümesi olarak belirledik. Topladığımız veri kümesi Şekil 2'de görüldüğü gibi farklı açılardan çekilmiş ve farklı özelliklere sahip görüntüler içermektedir.

TABLO I: EL YAPIMI ÖZNETELİKLER KULLANARAK BASIT ÖĞRENME YÖNTEMLERİYLE ELDE EDİLMİŞ SINIFLANDIRMA DOĞRULUĞU SONUÇLARI.

Öznitelik	Resim Ölçüsü	Algoritma	Doğruluk
Küçük Resimler	64x64	11-NN	15.47%
Küçük Resimler	64x64	Lojistik Regresyon (500 iter)	11.00%
Küçük Resimler	128x128	1-NN	16.27%
Küçük Resimler	128x128	Lojistik Regresyon (500 iter)	13.00%
HOG Öznitelikleri	32x32	31-NN	20.27%
HOG Öznitelikleri	32x32	Lojistik Regresyon (500 iter)	13.00%
HOG Öznitelikleri	64x64	1-NN	18.93%
HOG Öznitelikleri	128x128	Lojistik Regresyon (500 iter)	18.00%
GIST Öznitelikleri	64x64	71-NN	28.53%
GIST Öznitelikleri	128x128	11-NN	28.53%
GIST Öznitelikleri	256x256	1-NN	28.15%
GIST Öznitelikleri	64x64	Lojistik Regresyon (500 iter)	41.26%

Öncelikle veri kümemizi çeşitli el yapımı öznitelikler ile test ederek verileri oluşturulan veri kümesinin kullanışlı olup olmadığını araştırdık. Bunun için 3 farklı el yapımı öznitelik *Küçük Görüntüler*, *HOG* ve *GIST* özniteliklerini çıkararak Lojistik Regresyon ve KNN öğrenme yöntemleriyle Tablo I’de verilen yer tahmin etme başarılarını gözlemledik. Buna göre her iki yöntemle ve kullanılan özniteliklerle rastgele seçimin üzerine çıkarak öğrenme gerçekleştirebiliyoruz. Sonuçlarda görüldüğü gibi GIST öznitelikleri verilen bir fotoğrafın yerini bulmada daha iyi sonuçlar vermiştir. GIST öznitelikleri sahne tanımlamak için kullanıldığından bu beklenen bir durumdur. Sonuç olarak veri kümemiz üzerinde verilen bir fotoğrafın yerini bulma işlemi basit makine öğrenimi yöntemleriyle bile gerçekleştirebiliyoruz.

#### IV. YÖNTEM

Bu çalışmada Türkiye ölçeğinde verilen bir görüntünün hangi şehirden çekilmiş olduğunu bulan farklı evrimsel sinir ağı modellerini inceliyoruz. PlaNet [14] çalışmasında olduğu gibi problemi bir sınıflandırma problemi olarak ele alıyor ve PlaNet’teki gibi yeryüzünü küçük alanlara bölmek yerine Şekil 1’de verilen Türkiye’nin şehirleri haritasını temel alıyoruz. Böylelikle problemimiz verilen bir girdi görüntüsü  $I$  için bir  $s = f(I; \theta)$  modelindeki  $\theta$  parametrelerinin öğrenilerek verilen görüntünün 15 şehir  $s$  den hangisine ait olduğunu bulmaya karşılık gelmektedir.

İlgili model için AlexNet [6] ve ResNet [5] evrimsel sinir ağı yapıları kullanılmıştır. Son katmanlar sınıf sayımıza göre yeniden düzenlenmiştir. Kullandığımız veri sayısının baştan sona derin öğrenme için küçük olmasından dolayı aktarım öğrenme yaklaşımları benimsenmiştir. Ayrıca modelin aşırı öğrenmesini engellemek için veri arttırımı yapılmıştır.

Öncelikle temel model olarak baştan sona öğrenme gerçekleştirilmiştir. Sonrasında aktarım öğrenimi için öznitelik çıkarma ve ince ayarlama teknikleri kullanılarak önceden öğrenilmiş parametreler üzerinde eğitim yapılmıştır. Eğitimler yapılırken optimizasyon için 0.9 momentum ve 0.001 öğrenme oranı ile dereceli azalmaya göre hata oranı daha büyük çıkma olasılığı olsa bile daha hızlı olduğu için olasılıksal bayır inişi (*stochastic gradient descent*) algoritması kullanılmıştır. Yığın (*batch*) boyutu 4 olarak kullanılmıştır. Kayıp fonksiyonu olarak çapraz düzensizlik (*cross entropy*) kayıp fonksiyonu kullanılmıştır. Her 7 adımda öğrenme oranı 0.1 oranında azaltılmıştır. Ayrıca veri sayısını artırmak için, her öğrenme çevriminde (*epoch*) görüntülerimiz içerisinde rastgele

katman adı	çıkı boyutu	18 - katman
conv1	112x112	7x7, 64, kaydırma 2
Eğitim A	conv2_x	3x3 maksimum havuzlama, kaydırma 2
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
Eğitim B	conv3_x	28x28
Eğitim C	conv4_x	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
		$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
	conv5_x	7x7
		$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
	1x1	ortalama havuzlama, 15-b fc, softmax

Şekil 3: ResNet18 modelinde eğitim A’da sadece A’daki katmanlar, Eğitim B’ de A ve B’deki katmanlar ve Eğitim C’de A, B ve C’deki katmanlar eğitime kapatılarak 3 farklı model öğrenilmiştir.

224 x 224 büyüklüğünde alanlar kırılmış ve rastgele seçilen görüntülerin yarısı yatay olarak ters çevrilmiştir.

Veri sayımızın baştan öğrenme için yeterli olmamasından dolayı aşağıda ayrıntıları verilen 2 farklı aktarım öğrenimi yaklaşımı kullanılmıştır.

**Öznitelik çıkarma:** Places365 ve ImageNet veri kümesi ile eğitilmiş AlexNet ve ResNet18 modellerini kullanarak Türkiye15 veri kümesindeki görüntüler için en son sınıflandırma katmanından önce gelen katman öznitelik çıkarıcı olarak kullanılmıştır. Daha sonra çıkarılan öznitelikler üzerinde basit bir sınıflandırıcı olan Lojistik Regresyon öğrenme algoritması çalıştırılmıştır.

**İnce Ayarlama:** Aktarım öğrenimi için kullanılacak diğer yaklaşım önceden öğrenilmiş modellerin belli katmanlarındaki parametreleri öğrenmeye kapatıp son katmanlardaki belli katmanlar için öğrenmenin gerçekleştirildiği ince ayarlama yöntemidir. Önceden eğitilmiş AlexNet [6] ve ResNet18 [5] modelleri üzerinde öncelikle son katmandaki parametreler öğrenmeye kapatılarak eğitim gerçekleştirilmiştir. Daha sonra ResNet18 modeli üzerinde farklı katmanlardaki parametrelerin öğrenmeye kapatılmasıyla elde edilen sonuçlar incelenmiştir. Şekil 3’de ResNet modeli için eğitime kapatılan katmanlar gösterilmektedir.

#### V. DENEYLER

Bu bölümde gerçekleştirdiğimiz deneylerin detayları ve elde edilen sonuçlar sunulmaktadır. Tüm deneyler önerdiğimiz Türkiye15 veri kümesi üzerinde gerçekleştirilmiştir ve değerlendirme ölçüsü olarak sınıflandırma doğruluğu kullanılmıştır.

Veri kümemizin büyüklüğünün baştan öğrenme için yeterli olmamasından dolayı beklendiği gibi en düşük sınıflandırma doğruluğunu baştan öğrenme için elde ettik. Tablo II’de baştan öğrenme ve aktarım öğrenme yaklaşımlarının farklı deney kurulumlarında elde edilen sonuçlar yer almaktadır. Baştan öğrenmede en iyi sonuç %22.40 ile ResNet18 modelinden elde edilmiştir.

**TABLO II: FARKLI KURULUMLARDA EĞİTİLEN DERİN ÖĞRENME MODELLERİNDEN ELDE EDİLEN SINIFLANDIRMA DOĞRULUĞU SONUÇLARI.**

Yöntem	Model	Ön eğitim	Öğrenme Oranı	Doğruluk
Sıfırdan öğrenme	AlexNet	-	0.001	10.40%
<b>Sıfırdan öğrenme</b>	<b>ResNet-18</b>	-	<b>0.001</b>	<b>22.40%</b>
Sıfırdan öğrenme	ResNet-18	-	0.01	15.73%
Sıfırdan öğrenme	ResNet-18	-	0.005	21.07%
Öznitelik Çıkarma	Resnet-18	Places365	0.001	46.13%
<b>Öznitelik Çıkarma</b>	<b>Resnet-18</b>	<b>ImageNet</b>	<b>0.001</b>	<b>48.00%</b>
<b>İnce Ayar</b>	<b>Resnet-18</b>	<b>Places365</b>	<b>0.001</b>	<b>55.73%</b>
İnce Ayar	Resnet-18	ImageNet	0.001	52.26%
İnce Ayar	AlexNet	Places365	0.001	45.06%
İnce Ayar	AlexNet	ImageNet	0.001	46.13%
Katman dondurma A	ResNet-18	Places365	0.001	56.80%
Katman dondurma B	ResNet-18	Places365	0.001	56.80%
<b>Katman dondurma C</b>	<b>ResNet-18</b>	<b>Places365</b>	<b>0.001</b>	<b>57.60%</b>

Sonuçlarda görüleceği üzere aktarım öğrenimi yaklaşımları baştan öğrenmeye göre daha iyi sonuçlar vermiştir. Aktarım öğrenimi yaklaşımları için önceden belirtildiği gibi ImageNet ve Places365 veri kümeleri üzerinde önceden eğitilmiş AlexNet ve ResNet18 modelleri kullanılmıştır. İnce ayarlama yaklaşımı öznitelik çıkarma yaklaşımından daha iyi sonuçlar vermiştir. Tablo II’de görüldüğü gibi sadece sınıflandırma katmanındaki parametrelerin güncellendiği ince ayarlama en iyi sonucu %55.73’lük doğruluk ile Places365 veri kümesi üzerinde eğitilen ResNet18 modeli vermiştir. ImageNet veri kümesi nesne odaklı bir veri kümesi olduğu için, ImageNet üzerinde eğitilen modeller genel sahne görüntülerinden oluşan Places365 veri kümesi üzerinde eğitilen modellere göre beklediğimiz gibi sınıflandırma doğruluğu bakımından genel olarak daha düşük sonuçlar vermişlerdir.

ResNet18 modeli için farklı katmanlardaki parametrelerin öğrenmeye kapatılması ile eğitilen modellerin sonuçlarına baktığımızda Eğitim C ile elde edilen model %57.60 doğruluk ile en iyi sonuçları üretmiştir. Bu aynı zamanda tüm modeller içinde elde edilen iyi sonuçtur.

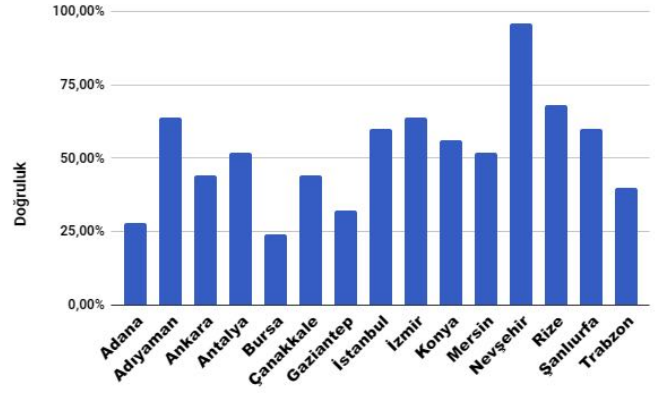
Son olarak en iyi modelimizin Türkiye15 veri kümesindeki 15 farklı şehre ait fotoğrafları sınıflandırma doğruluğu Şekil 4’de gösterilmiştir. Buna göre en iyi sonucu veren model Nevşehir’e ait fotoğrafları daha doğru sınıflandırmıştır. En kötü sonuçlar ise Bursa’ya ait fotoğraflardan alınmıştır. Bu sonucun çıkmasına birçok etken neden olabilir. Nevşehir’deki turistik alanların çokluğu ve özgünlüğü sınıflandırma başarısının artmasına neden olan etkenler olarak sayılabilir.

## VI. SONUÇ

Bu çalışmada ilk defa Türkiye’deki şehirler için bir fotoğraftan çekildiği şehri bulan farklı derin öğrenme modellerini eğiterek inceledik. Bunun için küçük bir veri kümesi topladık. Veri sayısının küçüklüğüne rağmen oldukça başarılı sonuçlar elde ettik. Gelecek çalışmalarda Türkiye’deki tüm şehirlere ait fotoğrafları içeren daha büyük bir veri kümesi üzerinde çalışılabilir ve daha başarılı derin öğrenme modelleri tasarlanabilir. Ayrıca şehirler yerine daha hassas coğrafi konumlar üzerine de çalışılabilir.

## KAYNAKÇA

- [1] S. Cao and N. Snavely. Graph-based discriminative learning for location recognition. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 700–707. IEEE, 2013.



**Şekil 4: En iyi modelimizin her şehre ait fotoğraflar için yer bulma başarımları.**

- [2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [3] J. Hays and A. A. Efros. Im2gps: estimating geographic information from a single image. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [4] J. Hays and A. A. Efros. Large-scale image geolocalization. In *Multimodal Location Estimation of Videos and Images*, pages 41–62. Springer, 2015.
- [5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [7] S. Lee, H. Zhang, and D. J. Crandall. Predicting geo-informative attributes in large-scale image collections using convolutional neural networks. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pages 550–557. IEEE, 2015.
- [8] Y. Li, N. Snavely, and D. P. Huttenlocher. Location recognition using prioritized feature matching. In *European conference on computer vision*, pages 791–804. Springer, 2010.
- [9] T.-Y. Lin, S. Belongie, and J. Hays. Cross-view image geolocalization. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 891–898. IEEE, 2013.
- [10] T.-Y. Lin, Y. Cui, S. Belongie, and J. Hays. Learning deep representations for ground-to-aerial geolocalization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5007–5015, 2015.
- [11] X. Qian, Y. Zhao, and J. Han. Image location estimation by salient region matching. *IEEE Transactions on Image Processing*, 24(11):4348–4358, 2015.
- [12] G. Schindler, M. Brown, and R. Szeliski. City-scale location recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*, pages 1–7. IEEE, 2007.
- [13] N. Vo, N. Jacobs, and J. Hays. Revisiting im2gps in the deep learning era. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2640–2649. IEEE, 2017.
- [14] T. Weyand, I. Kostrikov, and J. Philbin. Planet-photo geolocation with convolutional neural networks. In *European Conference on Computer Vision*, pages 37–55. Springer, 2016.
- [15] S. Workman, R. Souvenir, and N. Jacobs. Wide-area image geolocalization with aerial reference imagery. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3961–3969, 2015.
- [16] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.