

Evrişimli Sinir Ağları ve AlexNet Mimarisi Kullanarak Fotoğraf Sınıflandırma (Haziran 2022)

Ahmet Ata Şentürk¹, Şulenur Çörez²

¹Bilgisayar Mühendisliği Osmangazi Üniversitesi, Tepebaşı Eskişehir 26170 TR

²Bilgisayar Mühendisliği Osmangazi Üniversitesi, Tepebaşı Eskişehir 26170 TR

Sorumlu Yazar: Ahmet Ata Şentürk (e-posta: atasentrk4526@gmail.com)

Sorumlu Yazar: Şulenur Çörez (e-posta: sulenurcorez@gmail.com)

ÖZET Günümüzde fotoğraf sınıflandırma, yüz tanıma, öneri sistemi gibi sistem ve teknolojileri çok büyük bir yer kaplıyor. Örneğin, sosyal medya mecralarına bir kişi ile fotoğraf koyulduğunda “X kişisini etiketlemek ister misiniz?” sorusu bunun en büyük kanıtlarından biridir.[1] Ya da telefonlarda albüme girildiğinde kişileri sınıflandırması ve o kişilerin olduğu fotoğrafları ayıklayabilmesinin ardında çok kuvvetli bir yapay zeka mekanizması işler. Daha da net olmak gerekirse, arkasında evrişimli sinir ağı (Convolutional Neural Network, CNN) vardır. CNN, verilen görüntüleri ayırt etmek için benzersiz özelliklerini kullanır. Bir uçak ya da yılan görseli gördüğünde arasındaki farkı anlayıp hangisi olduğunu ayırt edebilir. Aslında bu süreç de insan beyninde bilinçsizce gelişir. İnsan bir resme baktığında onun ne olduğunu kolaylıkla anlayabilir. Cnn’de ise bu süreç köşeler, pikseller gibi ilk önce alt düzey bilgilerden, bunları soyut bilgilere dönüştürür.[2] Çalışma, Kaggle platformunda erişime açık bir şekilde sunulan 5000 kedi ve köpek fotoğrafından oluşan bir veri seti ile sinir ağını eğitme sonucu, hiç görmediği bir fotoğrafın kedi veya köpek olduğunu ayırt etme üzerinedir. Ayrıca çalışmada 8 katmandan oluşan ve CNN mimarilerinden biri olan AlexNet mimarisi kullanılmıştır.

Anahtar kelimeler: Kedi, Köpek, Fotoğraf Sınıflandırması, AlexNet, Sinir Ağları, Evrişimli Sinir Ağları

I. GİRİŞ

Görüntü sınıflandırma, bir bütün olarak resmi kavramaya çalışan ve ona etiket atayarak sınıflama işlemidir. Bir resimde birden fazla nesnenin de bulunabileceğini göz önünde bulundurarak, gerçekçi durumlarda da resimleri sınıflandırmak ve analiz etmek için kullanılır.

Örnek olarak aşağıdaki fotoğrafları ele alalım.



Şekil 1. Güvercin



Şekil 2. Saksığan

Resimlerde görülen iki farklı kuş türünü insan beyni kolay bir şekilde anlar. Yapay zekanın bu resimleri sınıflandırma yapabilmesi için de teknolojiler geliştirilmiştir. Bunlardan en etkili yöntemlerden biri ise derin öğrenme ve Cnn’dir.

Bilindiği üzere veri çağında yaşıyoruz Yapay zekanın gelişmesiyle artık çok büyük miktarlarda veri üretiyoruz. Bu görüntü verileri veya video verileri küresel bir veri oluşturmada büyük bir rol oynuyor. Yapay zekanın ise görüntüleri anlama ihtiyacının çıkış sebebi şudur; İnsan gücünden tasarruf ve gerçekten de bir insanın yapabileceğinden daha fazlasını elde etme isteğidir. Artık görüntüleri verimli bir şekilde işlemede makineler ve gelişmiş algoritmalara güveniyoruz. Görüntü sınıflandırma da dijital görüntü analizinin en önemli parçasıdır. Görüntü sınıflandırma ise bir görüntü içindeki piksel veya vektör gruplarına etiketler atama ve kategorize etme işlemidir.[3] Bu çalışmada da açık bir veri seti kullanılarak bir görüntü sınıflandırma için sinir ağı eğitilmiş ve eğitimde de bir Cnn Alexnet mimarisi kullanılmıştır. Genel olarak çalışmanın amacı kedi ve köpek fotoğraflarını ayırt edebilmektir.

II. AMAÇ VE HEDEF

Görüntü sınıflandırma, dijital görüntü analizini için belki de en önemli aşamadır. Bu teknoloji insan gücünün yerini çoktan almaya başlamış ve insan beyninin de başaramayacağı birçok şeyi başarmaya başlamıştır. Bu

çalışmanın amacı ise, kedi veya köpek görselini tanıyan bir yapay sinir ağı geliştirmektir. Görsel sınıflandırma için yapıtaşı sayılabilecek bir çalışma niteliği taşımaktadır.

III. SÜREÇ

İlk olarak veri seti Kaggle üzerinden bulunmuş ve daha sonrasında bu veri seti kullanıma hazır hale getirmek üzere düzenlenmiştir. Bu düzenlemede öğrenmeyi olumsuz etkileyecek ve test aşamasında sıkıntı yaratacak çok fazla parazit içeren fotoğrafları veri setinden çıkarma işlemi uygulanmıştır. Bu sayede daha temiz denilebilecek bir veri seti oluşturulmuş ve daha sonrasında veri setinde kullanıma en uygun olabilecek CNN mimarilerinden biri olan Alexnet kullanımına karar verilmiştir. Python dili ve colab ortamı kullanılarak sinir ağı eğitilmiş ve sonuçlar eklenmiştir.

IV. KULLANILAN VERİ SETİ

Evrimsel sinir ağları(convolutional neural networks) derin öğrenmenin bir alt dalıdır ve genellikle görsel sınıflandırmada kullanılır. Bunun yanında öneri sistemleri, doğal dil işleme gibi alanlarda da kullanılabilir.[6] Adından gelen evrimsel, özel bir lineer işlem türüdür. Katmanlı bir mimaridir ve katmanlarından birinde bile matris çarpımı yerine olması evrimsel basit sinir ağına dönüşmesini sağlar.

Çalışmada materyal olarak Kaggle platformunda açık bir şekilde kullanıma sunulan “Dogs&Cats Images” görsel veri seti kullanılmıştır. Veri setini “CHETANIMRAVAN” adlı kullanıcı 2018 yılında kullanıma sunmuştur ve lisansını ise kullanıma açık bir şekilde belirlemiştir. Veri seti eğitim ve test olarak iki bölüme ayrı şekilde bulunmaktadır. Ayrıca eğitim ve test bölümleri de köpek ve kedi şeklinde iki sınıf içerecek şekilde bölümlenmiştir. Veri seti, eğitim veri setinde toplamda 8000 görsel, test veri setinde ise 2000 görsel olacak şekilde toplamda 10000 görsel barındırmaktadır. Bu çalışma için her şey yarıya indirilecek şekilde, eğitim 4000, test için ise 1000 görsel kullanılmıştır.



Şekil 3. Veri setinden örnek köpek fotoğrafı



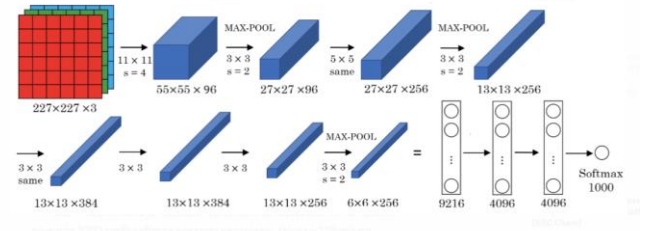
Şekil 3. Veri setinden örnek kedi fotoğrafı

V. YÖNTEM

Yapılan çalışmada CNN mimarilerinden biri olan Alexnet kullanılmıştır. Alexnet mimarisi, 30 Eylül 2012’de ImageNet Large Scale Visual Recognition Challenge’da yarıştırmıştır. Alex Krizhevsky, Ilya Sutskever ve Geoffrey Hinton tarafından geliştirilmiş bir mimaridir. Modelin derinliği ve hesaplamanın pahalı olduğu, yüksek GPU hızları gerektiren bir mimaridir. GPU destekli çalışması gerekir. Google Academics’e göre 2021 yılı itibarıyla 80000’den fazla atıf yapılmıştır.[4]

Alexnet 8 katman içerir ve ilk beşi evrimsel(convolutional) katmandır. Diğer 3 katman ise tam bağlantılı(fully connected) katmandan oluşur. Aktivasyon fonksiyonu olarak ReLU, pooling katmanı olarak ise max-pooling kullanılmaktadır. 60 milyona yakın parametre hesaplama süreci vardır.[5]. Bahsi geçen yarışmada %74.3’ten %83.6’ya ani bir geçiş ile görüntü sınıflandırmada büyük bir atılım sağlanmıştır.

Alexnet mimarisinin yapısı aşağıdaki görselde görülmektedir.



Şekil 4. Alexnet mimarisi

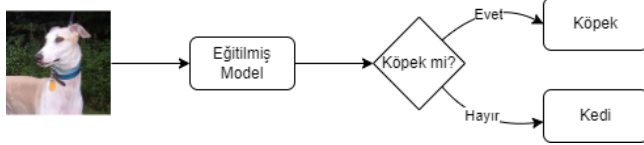
Alexnet’i farklı kılan ise ReLU Nonlinearity, Multiple GPUs ve Overlapping Pooling yaklaşımlarıydı. Alexnet, zamanında sürekli olarak tanh fonksiyonunun kullanılmasına karşı ReLU aktivasyon fonksiyonunu kullanıma geçirdi. Bu ise tanh fonksiyonunun kullanılmasından 6 kat hızlı bir eğitim süreci geçirmeye yaradı. Diğer bir farklı özellik, birden çok GPU kullanmalarıydı. Bu sayede yarışmada 1.2 milyon sayıda fotoğrafı eğitim sürecinde daha kısa sürelerde tamamlayabileceklerdi. Son durum ise, CNN’lerin, overlapping yapmadan geleneksel bir şekilde komşu nöron gruplarının çıktılarını pool ederler. Alexnet’te bu birleştirmenin yerini örtüşme aldı ve hatada yaklaşık %0.5’lik bir azalma görüldü.[7]



Şekil 5. Eğitim akış şeması

Şekilde eğitim için akış şeması gösterilmektedir. Veri seti ilk olarak boyutlandırılacak ve alexnet için hazır hale getirilecektir. Bu boyutlandırmada 227x227x3 haline

getirilmesim gerekmektedir. 227x227 kısmı resmin x ve y için piksel sayısını, 3 sayısı ise RGB'yi ifade etmektedir.



Şekil 6. Tahmin akış şeması

Eğitimden sonra bir fotoğraf verildiğinde eğitilmiş model bunun köpek veya kedi olduğunu tahmin edip 1 ve 0 sonuçlarını döndürür. 1 köpeği, 0 ise kediyi temsil eder. Bu çalışma kapsamında, model %88'lik bir doğruluk oranına ulaşmıştır.

IV.MATEMATİK

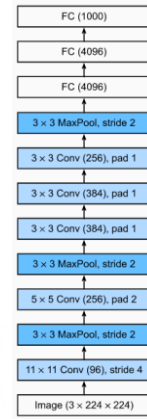
Kullandığımız batch size yani bölük boyutları 32 olarak belirlenmiştir ve eğitim süreci yaklaşık 3 saat sürmüştür. Bu eğitimde Alexnet mimarisi kullanılmıştır. Alexnet mimarisinde, görsel 227x227x3 şeklinde alınır. Buradaki ilk iki sayı görsel piksellerini, son sayı ise RGB, üç renkten oluşacağını gösterir. İlk katmanda convolutional katmanı 11x11 kernel'den oluşur ve stride sayısı ise 4'tür. Ayrıca tüm katmanlarda olduğu gibi burada da reLU aktivasyon fonksiyonu kullanılır. Her katman sonrasında kod içerisinde normalizasyon yapılması gerekir. İlk normalizasyon işleminden sonra max pooling işlemi uygulanır. Burada 3x3'lük bir max pooling işlemi vardır. Daha sonrasında 3 tane arka arkaya 3x3'lük convolutional katmanı uygulanır. Daha sonrasında tekrar bir 3x3'lük max pooling katmanı uygulanır. Stride değeri olarak 2 kullanılır. Daha sonrasında 3 tane tam bağlı katman ile düzleştirilip eğitilir. En son katmanda aktivasyon fonksiyonu olarak softmax kullanılır. Ayrıca kayıp fonksiyonu kategorik cross entropy'dir. Optimizer olarak 0.001 öğrenme oranı ile stokastik dereceli azalma kullanılır. Son katmanda ise softmax kullanılmasının sebebi; bir zar atıldığında 7 gelme olasılığının 0 olması gibi anlamsız bir değer çıkmasına karşın, CNN'de bu kontrolü bizim yapmamız gerekir. Belirlenen sınıflardan birine ait olduğu anlaşılmazsa negative bir değer alması olasılığa aykırıdır ve tüm sınıfların olasılık toplamı 1 olması gerekir. [9]

$$\hat{y}_j = \frac{e^{x_j}}{\sum_k e^{x_k}}$$

Kural 1
olasılığı negatiften kurtardık

Kural 2
olasılıklar toplamının [0,1] arasında olmasını sağladık (normalizasyon işlemi)

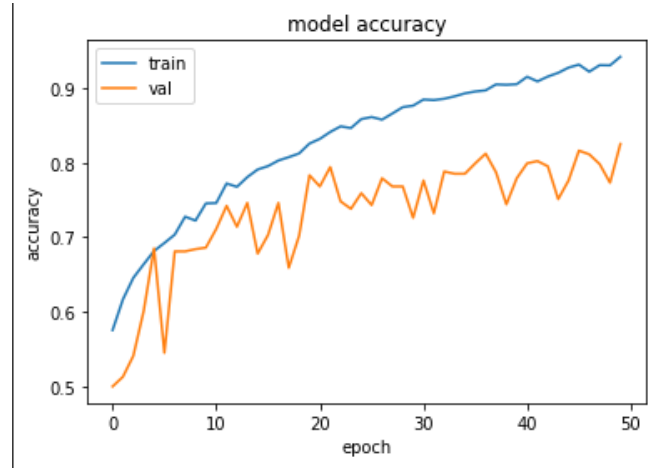
Şekil 7. Softmax



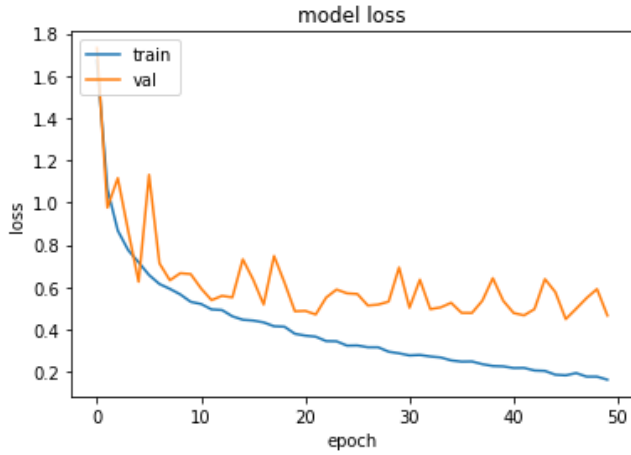
Şekil 8. AlexNet Mimarisi

V. ÇALIŞMA ÇIKTILARI VE SONUÇLARI

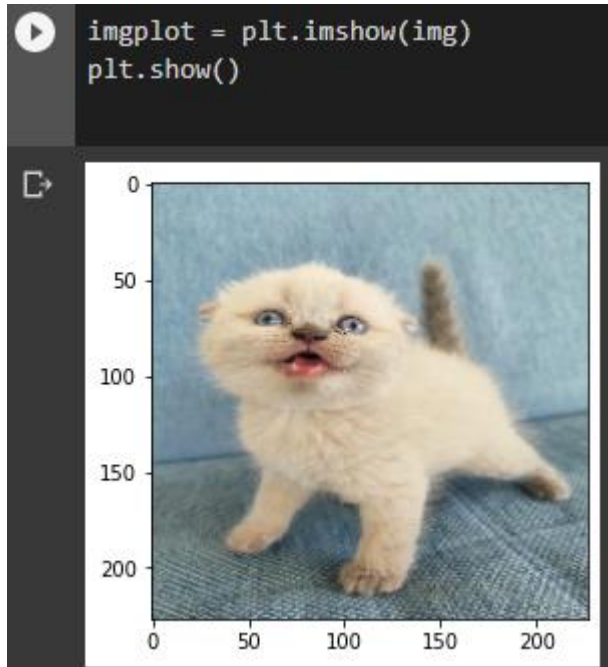
Çalışma sonucunda alınan doğruluk değeri %90'dır. Bunun anlamı yaklaşık olarak 100 tane daha önce modelin tanımadığı kedi köpek fotoğraflarından 90 tanesini doğru olarak sınıflandırmayı başarabildiğidir. Eğitim sırasında doğruluk değerinin zamana bağlı değişimi şekilde görüldüğü gibidir.



Bu grafikte görüldüğü üzere 50 epoch gerçekleşmiştir. Batch size olarak 32 belirlenmiş ve eğitim işlemi başladığında doğruluk değeri 0.57 olarak gözükmemektedir. Validation ise yaklaşık 0.5 ile başlamıştır. Daha sonrasında epoch arttıkça doğruluk oranının arttığını görebiliriz. Eğitim tamamlandığında doğruluk oranı 0.9417 olarak tamamlanmıştır. Aynı şekilde ikinci grafiğimizde de kayıp fonksiyonunu görebiliriz. O da beklenildiği şekilde giderek azalan bir yapıdadır. Başlangıçta 1.8 değeri ile başlayan kayıbmız, eğitim süreci sonunda 0.16 değerine kadar gerilemiştir. Onunla ilgili grafiği de aşağıda görebilirsiniz.



Çalışma sonucunda bulunan yeni kedi görselimiz aşağıdadır ve bunu tek olarak tahmin etmesi için modelimize vermemiz gerekir. Tahmin yapıldığında verilen görselin kedi olduğu sonucuna vararak doğru tahminde bulunmuştur.



```
[27] print(prediction)

kedi
```

REFERANSLAR

- [1] Ergin T. (2018, Ekim). Convolutional Neural Network (ConvNet yada CNN) nedir, nasıl çalışır?. Available: <https://medium.com/@tuncerergin/convolutional-neural-network-convnet-yada-cnn-nedir-nasil-calisir-97a0f5d34cad>
- [2] Image Classification. (n.d.). Papers With Code. Retrieved June 16, 2022, from <https://paperswithcode.com/task/image-classification>
- [3] Boesch G. (2020, Ocak). A Complete Guide to Image Classification in 2022. Available: <https://viso.ai/computer-vision/image-classification/#~:text=Image%20Classification%20is%20the%20Basis%20of%20Computer%20Vision,-The%20field%20of&text=It%20forms%20the%20basis%20for,%2C%20machine%20vision%2C%20and%20more.>
- [4] AlexNet. (2022, Temmuz 24). In asgag. <https://en.wikipedia.org/wiki/AlexNet>
- [5] Kızrak A. (2018, Mayıs). DERİNE DAHA DERİNE: Evrişimli Sinir Ağları. Available: <https://ayyucekizrak.medium.com/deri-CC%87ne-daha-deri-CC%87ne-evri-C5%9Fimli-sinir-a-C4%9Flar%C4%B1-2813a2c8b2a9>
- [6] Evrişimli Sinir Ağları, (2022, Kasım 3) In asgag. https://tr.wikipedia.org/wiki/Evri-C5%9Fimli_sinir_a-C4%9Flar%C4%B1
- [7] Wei J. (2019, Temmuz) AlexNet: The Architecture that Challenged CNNs. Available: <https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951>
- [8] Deep Convolutional Neural Networks (AlexNet). (n.d.). Dive Into Deep Learning. Retrieved June 16, 2022, from https://d2l.ai/chapter_convolutional-modern/alexnet.html
- [9] Gülüm S. (2021, Mart) Softmax: Bir Aktivasyon Fonksiyonu. Available: <https://medium.com/deeper-deep-learning-tr/softmax-bir-aktivasyon-fonksiyonu-da8382d8a281>