

Derin Öğrenme ile İki Boyutlu Optik Karakter Tanıma

2D Optical Character Recognition Based on Deep Learning

Yrd.Doç.Dr. Arif Koyun
Süleyman Demirel Üniversitesi, Mühendislik
Fakültesi,
Bilgisayar Mühendisliği Bölümü, Isparta
arifkoyun@sdu.edu.tr

Emin Afşin
Süleyman Demirel Üniversitesi,
Fen Bilimleri Enstitüsü,
Bilgisayar Mühendisliği Bölümü
Yüksek Lisans öğrencisi, Isparta
eminafsin909@gmail.com

Öz

Görüntülerdeki optik karakterlerin tanımlanması işlemi, kağıt üzerindeki verilerin saklanması veya görüntü dosyalarındaki verilerde arama yapılması amacıyla önemlidir. Optik karakterlerin doğru bir şekilde tahmin edilebilmesi için makine öğrenmesi yöntemlerinden yararlanılmaktadır ve bu çalışmada görüntülerdeki optik karakterleri tanımak için makine öğrenmesi yöntemlerinden biri olan derin öğrenme ile iyi doğruluk oranı ve yüksek hız ile tanıma amaçlanmaktadır. Bu nedenle çalışmanın yapıldığı Matlab ortamındaki karakter tanıma aracı ile performans karşılaştırması yapılmıştır.

Anahtar Sözcükler: Derin öğrenme ,Optik Karakter tanıma,Derin öğrenme ile optik karakter tanıma, Yapay sinir ağı,Evrimsel Yapay Sinir Ağı.

Abstract

Identification of optical characters in images is important for storing data on paper or for searching data in image files. Machine learning methods are used to predict optical characters accurately and in this study, it is aimed to recognize with high accuracy and high speed by deep learning which is one of the machine learning methods to recognize the optical characters in the images. For this reason, performance comparison with character recognition tool in Matlab environment.

Keywords: Deep learning, Optical Character Recognition, Optical character recognition with deep learning, Artificial neural network

1. Giriş

Optik Karakter Tanıma (OCR), taranmış dosyalar, basılı materyaller ve ya elektronik görüntüler üzerindeki karakterlerin düzenlenebilir ve aranabilir verilere dönüştürmemize olanak sağlar[1,4,6]. OCR uygulamaları daha çok makineler tarafından yazılmış karakterleri tanırlar. El yazısı karakterlerinin tanınmasında da Intelligent character recognition (ICR) kullanılır [5,7].

Günümüzde elektronik belge yönetim sistemlerinin (EBYS) tümü yasalara göre OCR barındırması gerekir. Bu nedenle OCR, hem iş dünyasında hem de devletin bir çok kurumunda kullanılması zorunlu bir araç haline gelmiştir. OCR uygulamaları sıklıkla şu alanlarda kullanılmaktadır:

- Doküman İşleme
- Masatüsti Yayıncılık
- Sipariş İşleme
- Vergilendirme ve Tahsilat
- Mahkemelerle İlgili Alanlar
- Personel Kayıt Yönetimi
- Nüfus Sayımı Formlarının İşlenmesi
- Çek İşleme
- Ödeme İşleme
- Emekli Fonu İşleme
- CAPTCHA Aşma

OCR uygulamaları karakterleri tanımak için makine öğrenim yöntemlerinden biri olan yapay sinir ağlarını (YSA) kullanırlar. YSA, insanların öğrenme şeklini taklit ederek yeni bilgiler üretme ve oluşturma gibi yetenekleri gerçekleştirmek için insanın öğrenme sürecinin matematiksel olarak modelleyen [3,8] veya örneklerden olaylar arasındaki ilişkileri öğrenerek daha sonra hiç görmediği örnekler hakkında öğrendikleri bilgileri kullanarak karar veren sistemlerdir[11]. Makine öğreniminin amacı, bir

Gönderim ve kabul tarihi : 17.05.2017 - 22.09.2017

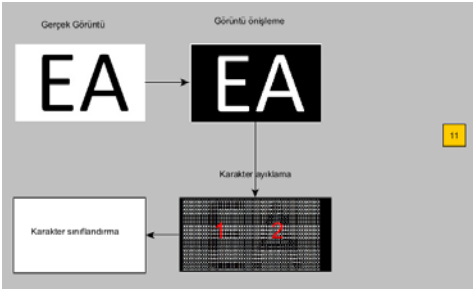
TÜRKİYE BİLİŞİM VAKFI BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ DERGİSİ (2017 Cilt: 10 - Sayı:1) - 11

insana benzer bir şey hissetmek, hatırlamak, öğrenmek ve tanımdır.

Makine öğrenmesinin yeni bir alanı olan Derin Öğrenme, son dönemde geliştirilen yapay zekâ uygulamalarının başarısını oldukça yüksek düzeylere çıkarmıştır[2]. Bu çalışmada makine öğreniminde en çok talep gören yöntemlerden biri olan derin öğrenme kullanılmıştır. Derin öğrenme de aslında çok katmanlı bir yapay sinir ağıdır. Veri miktarının artması ve ekran kartlarının (GPU) işlem gücünün artması [9,10] ile 2000'li yılların ikinci yarısından itibaren derin öğrenme en çok talep gören yöntemlerden biri haline gelmiştir.

2. Yöntemler

Optik karakter tanıma işlemi birkaç aşamadan meydana gelmektedir. Yapılan çalışmadaki karakter tanıma sistemi Şekil 1'deki gibi görüntü dosyasının önışleme tabi tutulması ,satırları bulma ve satırlardaki karakterleri ayıklama ve karakterleri sınıflandırmaya olarak üç ana bölümden meydana gelir.



Şekil 1: Karakter Tanıma aşamaları

Gerçek görüntü; genellikle taranmış kağıt evraklar ve ya dijital kamerayla çekilen görüntülerdir. Optik karakter tanıma işlemi metinlerin yazılmış olduğu bir çok belge türü içerdiğinden bu çalışma da gerçek görüntülerde tablo benzeri öğeleri içermeyen basit içerikleri tanımak olacaktır.

2.1. Görüntü Önışleme

RGB formatındaki resimler kırmızı, gri ve yeşil olmak üzere üç katmandan oluşurlar.

| | | | | | | |
|----|-----|-----|-----|-----|-----|-----|
| 54 | 58 | 255 | 8 | 0 | | |
| 45 | 0 | 78 | 51 | 100 | 74 | |
| 85 | 47 | 34 | 185 | 207 | 21 | 36 |
| 22 | 20 | 148 | 52 | 24 | 147 | 123 |
| 52 | 36 | 250 | 74 | 214 | 278 | 41 |
| | 158 | 0 | 78 | 51 | 247 | 255 |
| | | 72 | 74 | 136 | 251 | 74 |

Şekil 2: RGB formatındaki resmin 3 katmanlı matris görünümü

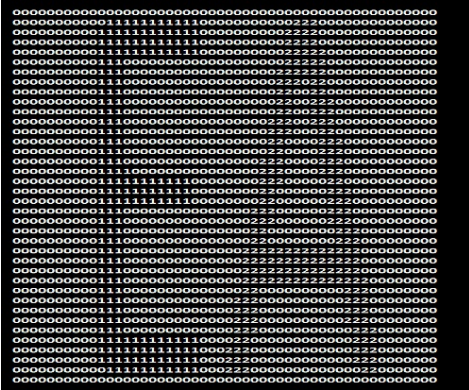
Her bir katman Şekil 2'deki gibi 0 ile 255 arasında sayı matrislerinden oluşmaktadır.

Görüntünün önışleme tabi tutulması işlemi; görüntünün öncelikle RGB formatından gri formata çevrilmesi yani tek bir katmanın kullanılması daha sonra da 0 ve birlerden oluşan binary formata çevrilmesi işlemidir. Binary formata çevirme işlemi basit bir şekilde; belli bir değerin üzerindeki sayıların 1 aşağısındaki sayıların ise 0'a çevrilmesi ile elde edilebilir. Bu şekilde 1 sayısı olan pikseller beyaz, 0 olanlar ise siyah görünür. Bu işlem kırmızıya mavimsi gri renkteki katmanlarda yakın sonuçlar verir. Bu şekilde siyah ve beyaz görüntü elde edilmiş olur.

2.2. Karakter Ayıklama

Karakter ayıklama işlemi; birbirine bağlı olan piksel gruplarını bulma işlemidir yani görüntünün son hali binary olduğundan, görüntüde birbirine bağlı 1 sayıların ayrı ayrı gruplandırılması işlemidir. Bu şekilde, her bir karaktere farklı bir gerçek sayı verilir ve gruplandırılır. Şekil 3'de gösterildiği gibi ilk harf olan E harfi 1 sayısıyla, ikinci harf olan A harfi ise 2 sayısıyla ifade edilmiştir.

Bu şekilde karakterler ayıklandığında birbirinden bağımsız iki ve ya üç komponente sahip olan i,İ ,ü,Ü,ö,Ö,ğ ve Ğ karakterlerinin tespiti için ayrıca aynı dikey eksende olup olmadıkları kontrol edilmek zorunda.



Şekil 3: Karakter ayıklama işlemi

Aksi takdirde i ve benzeri karakterler iki, ü ve benzeri karakterler ise üç ayrı karaktermiş gibi algılanır. Karakter ayıklama işlemi için matlab ortamında bwlabel fonksiyonu kullanılmıştır. Bu fonksiyonun matlab ortamında kullanım biçimi basitçe ;

[komponentler,harfSayisi]=bwlabel(binaryResim);

şeklinde. Bu fonksiyon için kullanılan binaryResim belgedeki tek satırı ifade etmektedir. Çünkü görüntü dosyasının içindeki tüm karakterleri bwlabel fonksiyonu ile bulmaya çalışırsak hangi karakterin hangi satıra ait olduğunu bulma işlemi komplike bir hal alacaktır. Bu nedenle belgedeki satırlar bulduktan sonra her bir satır için ayrı bwlabel fonksiyonu kullanılarak ilgili karakterler sırası bir şekilde bulunur. Karakterler bulduktan sonraki işlem ise karakterler arasındaki boşluk hesabı ile kelimeler arasındaki boşluğu bulma işlemidir. Bu işlem için basit olarak her karakter arasındaki piksel sayısını hesaplayıp ortalama karakterler arasındaki piksel sayısını bulmak olacaktır. Karakterler arasındaki boşluk ortalama piksel sayısından fazla ise bu boşluğun, karakterler arasındaki boşluk değil de kullanıcı tarafından bırakılan boşluk olduğu varsayılmıştır. Bu şekilde kelimeler arasındaki boşluk bulunmuş olur.

2.3. Karakter sınıflandırma

Karakter sınıflandırma işlemi; ayıklanan karakterlerin hangi ASCII kodu olduğunu bulma işlemidir. Bu işlem derin öğrenmenin özelleşmiş yapısı olan Evrişimsel sinir ağı (Convolutional neural network) ile yapılmıştır. Bu sınıflandırma işlemi için öncelikle örnek uzay kümesinin evrişimsel sinir ağına

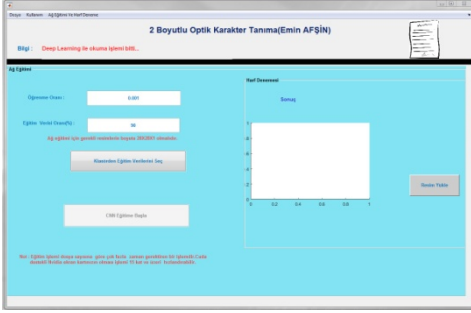
gösterilip ağı eğitimi işleminin yapılması gerekmektedir. Sinir ağlarının doğruluk oranı eğitim işlemindeki veriseti sayısı ile doğru orantılıdır. Bu nedenle Evrişimsel sinir ağına karakterleri tanınmasındaki başarıyı artırmak için yaklaşık olarak 2.160.000 karakter örneği ile eğitilmiştir. Bu eğitim işlemi için kullanılan karakter kümesi, C# yazılım dilinde olan Graphics sınıfının DrawString fonksiyonu ile her font ailesine ait karakterler Şekil 4'teki gibi oluşturuldu. C# ortamında oluşturulan karakterler RGB formatında ve her font ailesine göre farklı boyutta oluşmuştur. Oluşturulan bu karakterler genişliği 28 ve yüksekliği 28 olan gri görüntü dosyası haline getirilmiştir.



Şekil 4: Örnek eğitim verisi

Çünkü Evrişimsel sinir ağına eğitmek için kullanılan eğitim verilerinin hepsinin aynı boyutta ve aynı derinliğe sahip olması gerekmektedir. Örneğin RGB formatında görüntü dosyaları ile ağı eğitmek istersek her dosyanın RGB formatında olması ve aynı boyutta olması gerekmektedir genel anlamda karakterlerin tümü aynı boyutta ve aynı katman sayısına sahip olmalıdır. Bu nedenle oluşturulan karakterler 28X28 boyutunda tek katmanlı gri görüntü dosyasına çevrildikten sonra ağı eğitilmiştir.

Ağ eğitimi zahmetli ve zaman alan bir işlemdir. Şekil 5'te yapılan çalışmadaki ağ eğitimi ekranı bulunmaktadır. Bu işlem zamanın zamanın kısaltılması için kullanılan bilgisayarın ekran kartı(GPU) ile eğitim işlemi yapılmıştır. GPU ile işlem yapılması eğitim süresini 15 kata kadar azaltmaktadır. Buna rağmen eğitim işlemi 24 saat sürmüştür.



Şekil 5:Yapılan çalışmadaki ağ eğitimi ekranı

4.Sonuçlar ve Öneriler

GPU'ların paralel işlem yeteneği sayesinde ağ eğitimi için özellikle GPU kullanılması önerilmektedir.Aksi taktirde günlerce hatta aylarca veri setinin büyüklüğüne göre ağ eğitimi işlemi sürebilmektedir.

Karakter tanıma işlemi; karakterlerin birbirlerine olan benzerliklerinden dolayı sınıflandırma işlemini olumsuz etkilemektedir.Bu nedenle karakter tanıma işleminin doğruluğunu artırmak için kullanılan verinin fazla olması önerilmektedir.Ayrıca karakter tanıma işlemi için kullanılan veri setinin gri renkte yani tek katmanlı olması; ağıın doğruluk oranını artırmakla birlikte daha az veri seti ile ağı daha iyi doğruluk oranlarına ulaştırmıştır.

Çizelge 1'de gösterilen iki adet testte, derin öğrenme ile yapılan çalışma ve Matlab ortamında gömülü olarak gelen OCR aracının doğruluk oranı ve hız performansı karşılaştırılmıştır. Yapılan çalışmada satır sayısı arttıkça hızın daha da arttığı gözlemlenmiştir. Bunun sebebi derin öğrenme ile yapılan çalışmada her bir satırın ayrı ayrı paralel olarak bulunmaya çalışılmasıdır. Bu nedenle 10 satır ile yapılan testte yapılan çalışma diğer araca göre daha yavaş çalışmıştır. Ayrıca derin öğrenme ile yapılan çalışmada test yapılan bilgisayarın GPU'su test sonuçlarında büyük değişikliğe sebep olabilmektedir. Bu nedenle test sonuçlarındaki süre, farklı bilgisayarlar da farklı sonuçlar verebilir.

Çizelge 1: Sınıflandırma Performans Sonuçları

| Uygulama | Doğruluk (Hatalı karakter sayısı) | Süre (Sn) | Satır Sayısı | Harf Sayısı |
|-----------------|-----------------------------------|-----------|--------------|-------------|
| Yapılan Çalışma | % 100 (0) | 1.09 | 10 | 297 |
| | %99.51 (3) | 2.09 | 17 | 619 |
| Matlab OCR | %99.66 (1) | 0.50 | 10 | 297 |
| | %99.35 (4) | 2.48 | 17 | 619 |

5.Kaynakça

- [http://bidb.itu.edu.tr/sevirdefteri/blog/2013/09/08/ocr-\(optical-character-recognition---optik-karakter-tan%C4%B1ma\)](http://bidb.itu.edu.tr/sevirdefteri/blog/2013/09/08/ocr-(optical-character-recognition---optik-karakter-tan%C4%B1ma)), İnternet Erişim Tarihi: 22.04.2017.
- Rende, F. Ş., Bütün, G., Karahan, Ş., Enstitüsü, B. T., TÜBİTAK BİLGEM, G. Derin Öğrenme Algoritmalarında Model Testleri: Derin Testler.
- <http://www.ibrahimcaviroglu.com/dokumanlar/ilerialgoritmaanalizi/ilerialgoritmaanalizi-5.hafta-yapaysiniraglari.pdf>, İnternet Erişim Tarihi: 28.04.2017
- [http://bidb.itu.edu.tr/sevirdefteri/blog/2013/09/08/ocr-\(optical-character-recognition---optik-karakter-tan%C4%B1ma\)](http://bidb.itu.edu.tr/sevirdefteri/blog/2013/09/08/ocr-(optical-character-recognition---optik-karakter-tan%C4%B1ma)), İnternet Erişim Tarihi: 08.05.2017.
- <http://www.grafikerler.org/forum/konu/optik-karakter-tanima-ocr-nedir.5099/>, İnternet Erişim Tarihi: 07.05.2017.
- <https://www.abbyy.com/tr-tr/finereader/>, İnternet Erişim Tarihi: 23.12.2016.
- <http://www.grafikerler.org/forum/konu/optik-karakter-tanima-ocr-nedir.5099/>, İnternet Erişim Tarihi: 25.12.2016.
- <https://ekblc.files.wordpress.com/2013/09/vsa.pdf>, İnternet Erişim Tarihi: 25.12.2016.
- Simonyan, K., Zisserman, A. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR), 2015.
- Krizhevsky, A., Sutskever, I., Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 2012, pp. 1097-1105.
- Öztemel, E. Yapay Sinir Ağları. Papatya Yayıncılık, 2003, İstanbul.