



Realization of character recognition application on text images by convolutional neural network

Ebru Somuncu¹ , Nesrin Aydın Atasoy^{2*} 

¹Institute of Graduate Programs, Department of Computer Engineering, Karabük University, Karabük, 78050, Turkey

²Department of Computer Engineering, Faculty of Engineering, Karabük University, Karabük, 78050, Turkey

Highlights:

- Text based image analysis
- Character analysis using deep learning algorithm
- Text prediction enhancement

Keywords:

- Character recognition
- Convolutional recurrent neuralnetwork
- Deep learning
- Levenshtein distances
- Recurrent neural network

Article Info:

Research Article
Received: 22.01.2021
Accepted: 01.05.2021

DOI:

10.17341/gazimmfd.866552

Correspondence:

Author: Nesrin Aydın Atasoy
e-mail: nesrinaydin@karabuk.edu.tr
phone: +90 538 414 2745

Graphical/Tabular Abstract

In this study, a deep learning model that is able to recognize texts in signage images is used for indoor navigation. Convolutional neural network due to its success on images, recurrent neural network due to its success in sequencing problems and connectionist temporal classification method due to its success on elimination of mistakes on sequencing were used. With using Levenshtein distances, the estimation performance of the model was improved, and better results were obtained than previous studies.

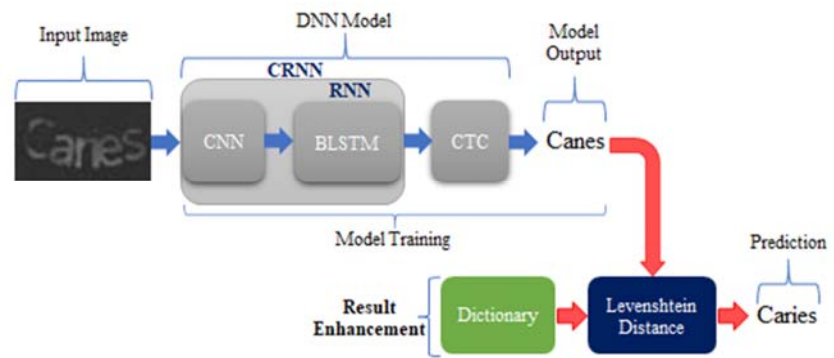


Figure A. Trained - tested model and enhancement of model outputs

Purpose: The aim of the study is to develop a robust deep learning model that can recognize the texts in signage images for location detection indoors. The performance of the proposed model has been tested in a large dataset.

Theory and Methods:

In this study, CNN and RNN architectures were used sequentially. Next, connectionist temporal classification method was used. Thus, character recognition-based words were estimated. However, the estimated words may have incorrect or missing characters as shown in Figure A. For solving this problem, Levenshtein distances were calculated between estimated word and words in dictionary. The estimation result was updated with nearest word. As a result, recognition performance was increased.

Results:

The recognition of the texts on signage images obtained under different lighting conditions was successfully achieved. The performance of the CRNN model was increased by using Levenshtein distances. With the improvements, success of the proposed approach reaches 96%.

Conclusion:

According to the test results, recognition of the indoor signage images has been improved with Levenshtein distance. Also, the CRNN model trained with English texts, is able to use for another language which has same character set by changing the dictionary which is used for Levenshtein distances.



Evrişimli tekrarlayan sinir ağı ile metin görüntüleri üzerinde karakter tanıma uygulaması gerçekleştirilmesi

Ebru Somuncu¹ , Nesrin Aydın Atasoy^{2*} 

¹Karabük Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Bölümü, Karabük, 78050, Türkiye

²Karabük Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Karabük, 78050, Türkiye

Ö N E Ç İ K A N L A R

- Metin tabanlı görüntü analizi
- Derin öğrenme algoritması kullanarak karakter analizi
- Metin tahmini iyileştirme

Makale Bilgileri

Araştırma Makalesi
Geliş: 22.01.2021
Kabul: 01.05.2021

DOI:

10.17341/gazimmfd.866552

Anahtar Kelimeler:

Karakter tanıma,
evrişimli tekrarlayan sinir
ağı,
derin öğrenme,
levenshtein mesafesi,
tekrarlayan sinir ağı

ÖZ

Literatürde iç mekan yönlendirmesi için kullanılan etkin ve yaygın teknolojiye rastlanmamıştır. Problemin çözümü için iç mekanlarda kullanılan metin görüntülerinin tanınması ve mevcut konunun görüntüleri göre tespit edilmesi en uygulanabilir çözüm olarak gözükmektedir. Bu çalışmada iç mekanlarda kullanılan tabela görüntülerinin tanınması için birçok alanda başarıyla kullanılan derin öğrenme yöntemleri kullanılmıştır. Oluşturulan model, Evrişimli Sinir Ağı ile görüntülerdeki köşe, kenar gibi yüksek seviyeli özellikleri çıkarmaktadır. Daha sonra kullanılan tekrarlayan sinir ağı tabanlı İki Yönlü Uzun-Kısa Süreli Bellek ile harflerin dizilim özellikleri korunmuştur. Dizilimdeki tekrarlar ve diğer problemlerin çözümü için Bağlantıcı Zamansal Sınıflandırma yaklaşımı kullanılmıştır. Eğitilen modelin ürettiği karakterlerle oluşturulan kelimelerdeki hatalar ise Levenshtein mesafesi kullanılarak iyileştirilmiştir. Modelin Synth90k veri seti ile yapılan testlerdeki başarısı %96, Synth90k ile eğitilmiş modelin aynı karakter setine sahip Türkçe görüntülerdeki başarısı ise %93 olmuştur. Elde edilen sonuçlar önerilen yaklaşımın başarısını ortaya koymaktadır.

Realization of character recognition application on text images by convolutional neural network

H I G H L I G H T S

- Text based image analysis
- Character analysis using deep learning algorithm
- Text prediction enhancement

Article Info

Research Article
Received: 22.01.2021
Accepted: 01.05.2021

DOI:

10.17341/gazimmfd.866552

Keywords:

Character recognition,
convolutional recurrent
neural network,
deep learning,
levenshtein distance,
recurrent neural network

ABSTRACT

In the literature, effective and common technology used for in door navigation has not been found. For the solution of the problem, recognition of the text images used indoors and determination of the current location according to the images seems to be the most feasible solution. In this study, deep learning methods, which are used successfully in many fields, were used to recognize signage images used in interior spaces. The created model extracts high-level features such as corners and edges in the images with the Convolutional Neural Network. Next, the sequencing properties of the letters are preserved with the Recurrent Neural Network-Based Bidirectional Long-Short-Term Memory. The Connectionist Temporal Classification approach is used to solve the repetitions and other problems in the sequence. Mistakes in words created with the characters generated by the trained model are improved using Levenshtein distance. The success of the model in the tests performed with the Synth90k data set was 96%, and the success in Turkish images with the same character set of the model trained with the Synth90k was 93%. The results obtained demonstrate the success of the proposed approach.

1. GİRİŞ (INTRODUCTION)

Gün geçtikçe artan insan nüfusu ve gelişen teknoloji ile onlarca katlı büyük binalar küçük birer şehir haline dönüşmüştür. İnsanlar, bu büyük ve karmaşık yapılarda hedefledikleri yere ulaşmaları için yönlendirilmeler yapacak bir rehber ihtiyacı duymaktadır. Açık alanda çalışan Küresel Konumlama Sistemi (Global Positioning System - GPS) teknolojik yapısı sebebiyle kapalı mekanlarda verimli olarak kullanılmamaktadır. İç mekan yönlendirme için donanımsal bazı çözümler olmasına karşın, günlük yaşantıda her bireyin bu donanımlara sahip olmaması, problemin çözümünün yaygınlaşmasının önündeki engeldir. İç mekan yönlendirmesinde konum tespiti için görüntülerin kullanılması daha yaygınlaştırılabilir bir yöntem olarak görülmektedir. Bu durumda bulunan konumun belirlenmesi için sahne tanıma veya iç mekandaki tabela yazılarının tanınması gerekmektedir. Görüntü verilerinin işlenmesi ve elde edilen bilgiye göre konumla ilişkilendirilmesi kapalı mekamlarda konumlandırma için kullanılabilir. Bu sebeple tabela görüntülerinin tanınması çalışmanın motivasyonunu oluşturmaktadır.

Dijital veri miktarının her geçen gün artmasıyla, verilerin analiz süreçleri önem kazanmaktadır. Makine öğrenmesi algoritmaları, verileri kullanarak, makinelerin insanlar gibi beceri kazanmasını ve analizler yapmasını hedeflemektedir. Makine öğrenmesi temel olarak denetimli, denimsiz ve takviyeli öğrenme yöntemleri kullanarak giriş veri seti için çıktılar tahmin eden algoritmik bir yaklaşımdır. Makine öğrenmesinin bir alt dalı olan derin öğrenme, verilerin temsillerinin öğrenilmesine dayalı yapıdan oluşan, çok katmanlı sinir ağı modelidir [1]. Derin öğrenme algoritmaları; yüz tanıma sistemleri, sağlık sektörü [2], ses tanıma çalışmaları [3], otonom sistemler, görüntü işleme [4], tahmin uygulamaları [5], doğal dil işleme, metin ve karakter tanıma çalışmaları, savunma ve güvenlik, nesne tanıma, sınıflandırma [6] gibi birçok alanda kullanılmaktadır. Derin Sinir Ağı (Deep Neural Network - DNN), çok fazla veri girişi ile ayırt edici özellikleri öğrenebilmektedir. Derin öğrenme ilk olarak 2012'de büyük ölçekli görsel tanıma (ImageNet) [7] yarışmasında nesne sınıflandırma alanında başarısını ortaya koymuştur. Öğrenme başarısı ağı eğitiminde kullanılan veri setinin büyüklüğüne bağlı olarak değişir. Verilerin birden çok katmandan geçmesi ile her bir adımda gereken bilgi öğrenilir. Yapılan çalışmalarda problemin yapısına ve kullanılan katman sayısına bağlı olarak farklı derin öğrenme mimarileri tercih edilmektedir.

En sık kullanılan derin öğrenme mimarilerinden Evrişimli Sinir Ağı (Convolutional Neural Network - CNN), çok katmanlı ileri yönlü bir yapay sinir ağıdır. En önemli avantajlarından biri büyük ağ modellerinde daha az sayıda eğitime ve daha az parametreye ihtiyaç duymasındır. CNN, görüntü üzerinde nesne tanıma, görüntü analizi [8] ve doğal dil işleme çalışmalarında kullanılmaktadır [9]. Doğal dil ifadelerinden metin analizi [10], cümlelerin analizi ve sınıflandırılması [11] gibi çalışmalar CNN'nin bu alandaki

başarısını ortaya koymaktadır. Derin öğrenme yaklaşımlarından Tekrarlayan Sinir Ağı (Recurrent Neural Network - RNN) diğer DNN mimarilerinden farklıdır. Zaman değişkenine bağlı olarak sıralı bilgileri kullanır. Hafıza yapısı ile bir dizinin her adımında aynı işlemleri önceki çıktılara bağlı olarak gerçekleştirir. RNN mimarisi ses verilerinin işlenmesi ve metne çevrilmesi [12], CNN mimarisi ile görüntülerden nesnelere ve konumlarının tespiti [13] gibi çalışmalarda başarılı sonuçlar ortaya koymuştur. El yazısı karakter tanıma ve görüntü sınıflandırma için RNN ve CNN mimarilerinin birlikte kullanıldığı çalışmalar [14, 15], bu mimarilerin farklı veri setleri üzerindeki kullanımlarını ve başarılarını göstermektedir. Görüntü üzerinde karakter tespiti için yapılan ilk çalışmalarda Optik Karakter Tanıma (Optic Character Recognition - OCR) için görüntü işleme yöntemi tercih edilmiştir [16]. Çalışmalarda kullanılan veri miktarının artması ve donanımsal altyapının gelişmesiyle DNN mimarilerinin kullanıldığı uygulamalar son yıllarda giderek artmıştır. CNN [17] ve RNN [15], metin analizi ve karakter tanıma problemlerinde dikkate değer bir başarı göstermektedir. Saxena vd. [18], bir belgenin içeriğinin belirlenmesi için CNN ve Uzamsal Dönüştürücü Ağı (Spatial Transformer Network - STN) kullanarak bir model oluşturmuşlardır. Model, metin görüntülerinin karakterlere dönüştürülmesi ve tanınması için gerçekleştirilmiştir. Oluşturulan model, görüntülerdeki metin satırlarını algılama ve aynı zamanda bu metinlerdeki kelimeleri tanıyabilme kabiliyetine sahiptir. He vd. [19], görüntü bileşenlerindeki metin ve ilgili alanların çıkarılmasında yeni bir uzman sistem önermiştir. Önerilen Metin Dikkatli Evrişimli Sinir Ağı (Text-Attentional Convolution Neural Network) sahne tespiti için yeni bir yöntem kullanmaktadır. Bu sayede karmaşık arka plana sahip metinlerin algılanmasının mümkün olduğunu ortaya koymuşlardır. Wang vd. [15] çalışmalarında yerel özelliklerin çıkarılmasında avantajlı olan CNN ve RNN mimarilerini kullanarak spesifik olarak güçlü bir model sunmaktadırlar. Önerilen tanımlama modeli, özelliklerin çıkarılmasında, çıkarılan özelliklerin bir araya getirilmesinde CNN ve RNN türevi olan Uzun Kısa Süreli Bellek (Long Short-Term Memory - LSTM) ağlarının avantajlarından yararlanmaktadır. Li vd. [20], plaka görüntüleri üzerinde otomatik karakter tanıma işlemini; plakanın tespiti ve karakter tanıma adımlarından oluşan iki adımdan oluşan mevcut yöntemlerden farklı olarak tek bir derin sinir ağı ile çözmeyi önermişlerdir. Yuan vd. [21], büyük ve küçük harf İngilizce karakterleri tanımak için eğitim aşamasında hatalı örneklerle dayalı takviyeli öğrenme stratejisi geliştirerek LeNet-5 mimarisini değiştirerek kullanmışlardır. LeNet-5 CNN diğer modellerden farklı olarak havuzlama katmanlarında en büyük değer yerine ortalama değeri kullanmaktadır. Tabela metinleri zaman zaman ilgi çekmek için karakterlerin farklı yüksekliklerde veya açılarda konumlandırılmasıyla oluşturulmaktadır. Bu yönü ile problem geleneksel görüntü işlemeye dayalı OCR probleminden ayrılmaktadır. Çin alfabesi gibi dillerde ise çizgilerin konumları kelime anlamlarını etkilediği için literatürde Çince üzerine yapılan çalışmalar incelenmiştir.

Çince el yazısı karakter tanıma için yapılmış çok sayıda çalışma vardır. Wu vd. [22], literatürden farklı olarak CNN mimarisinde tam bağlantılı katmanda veya uzamsal havuzlama katmanında yapılan düzenleme işlemi gevşemeye dayalı evrişim katmanında yapmaya odaklanmışlardır. Zhong vd. [23], ise doğrusal olmayan dönüşüm ile çoklu havuzlama ve veri artırma yaklaşımını tercih etmişlerdir. Yang vd. [24], derin CNN mimarisine verilecek resimlerdeki alana özgü bilgileri kaybetmemek için giriş katmanında deformasyon uygulayarak veri artırma yolunu tercih etmişlerdir. He vd. [25], tanıma sonuçlarının güvenilirliğinin nasıl ölçüleceği için CNN tabanlı güvenilirlik ölçümünün kullanışlı bir metrik olduğunu 827685 test örneğinden elde ettikleri sonuçlara dayanarak önermişlerdir. Yapılan çalışmalar derin öğrenme algoritmalarının harf konumu ve açısındaki farklılıklara rağmen hala gürbüz olduğunu göstermektedir.

El yazısı tanınmasında CNN mimarisinin başarısı farklı dillerde ortak yöntemlerin kullanılmasını sağlamıştır. Bai vd. [26], CNN mimarisinde paylaşımlı gizli katman kullanarak İngilizce ve Çince karakterleri tanımak için DNN geliştirmişlerdir. Sunulan çalışma farklı iki dili eş zamanlı olarak tanıyabilme kabiliyetine sahiptir. DNN mimarilerinin Arapça gibi harfin kullanıldığı sıraya göre değişkenlik gösteren farklı alfabeler üzerinde gelişimi hala aktif bir çalışma konusudur [27].

Önceki çalışmalarda CNN ve RNN derin öğrenme mimarilerinin karakter tanıma çalışmaları üzerindeki başarısı ortaya koyulmuştur. Bu çalışmada bu iki mimari birleştirilerek Evrişimli Tekrarlayan Sinir Ağı (Convolutional Recurrent Neural Network - CRNN) mimarisi oluşturulmuş ve tabela görüntüleri üzerinde karakter analizi ile metin tanıma uygulaması gerçekleştirilmiştir. Tanınan metindeki karakter hatalarının azaltılması için kullanılacağı yerdeki tabela metinlerinden oluşturulmuş sözlük içinde kendisine en yakın metinle eşleştirilerek model başarısı oldukça artırılmıştır. Kullanılan veri seti yeterince zengin olduğundan sadece sözlük değiştirilerek tekrar tekrar eğitim yapılmadan kullanılabilir bir yapıdadır. İngilizce kelimelerden oluşan veri seti ile eğitilen model Türkçe tabela tanımda %93'lük bir başarı sağlamıştır. Çalışmanın ikinci bölümünde kullanılan veri seti ve derin öğrenme modellerinin özellikleri, üçüncü bölümde uygulama, dördüncü bölümde deneysel sonuçlar ve son bölümde sonuçların değerlendirilmesi ve gelecek çalışmalar sunulmuştur.

2. MATERYAL - METOT (MATERIAL - METHOD)

2.1. Veri Seti (Dataset)

Uygulamada kullanılan Synth90k veri seti 90000 İngilizce kelimeyi kapsayan 9 milyon resimden oluşmaktadır [28]. Görüntülerdeki yazılar, boyut ve pozisyon çeşitliliği ile 1400 farklı yazı tipinden meydana gelmektedir. Şekil 1'de sentetik olarak oluşturulmuş veri setinden örnek metin görüntüleri görülmektedir. Veri seti görüntü sayısı bakımından literatürdeki en büyük veri setidir [17].



Şekil 1. Model eğitim ve test veri seti örnekleri (Model training and test data set examples)

2.2. Metot (Method)

Karmaşık hesaplamalar ve büyük veri işleme çalışmalarında başarılı sonuçlar ortaya koyan derin öğrenme yöntemleri, metin ve karakter tanıma çalışmalarında da başarı göstermektedir. Bunun sonucunda karakterlerin tespiti için OCR tekniği, karakter ayıklama yapan klasik görüntü işleme teknikleri [29], LibSVM, SVM, Naïve Bayes, K-NN [30] gibi makine öğrenmesi algoritmaları yerine, bu çalışmada CNN ve RNN' in olumlu yönleri birleştirilerek CRNN modeli oluşturulmuştur.

2.2.1. Evrişimli sinir ağı (Convolutional neural network)

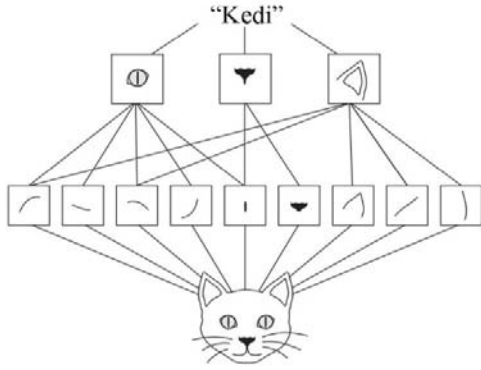
CNN eğitilebilen birçok katmandan ve parametreden oluşur. Katmanlı yapısıyla öz nitelikleri tespit ederek daha başarılı sonuçlar ortaya koymaktadır. Bu nedenle, çalışmada giriş görüntüleri üzerinde tanıma için, CNN tercih edilmiştir. CNN mimarisi, evrişim, havuzlama ve tam bağlı katman olarak üç kısımdan oluşur [31]. Evrişim katmanında giriş verileri filtrelenerek öz nitelik haritaları çıkartılır, böylece bölgesel örüntüler öğrenilmiş olur [32]. Şekil 2'de görüldüğü gibi hiyerarşik olarak öğrenilen yerel özelliklerden kedi gibi daha yüksek seviyedeki nesnelere öğrenilebilir. Havuzlama katmanında öz nitelikler örneklendirilir. Örnekleme ile boyut azaltılmış ve dolayısıyla parametre sayısında düşüş sağlanmış olur. Böylece, modelin daha hızlı öğrenmesi sağlanır. Tam bağlı katmanda her nöronun bir önceki katmandan iletilen girişlerine bağlı olarak çıktı değeri elde edilir. Tam bağlantılı katman hedeflerin optimizasyonu için kullanılabilir.

CNN mimarisi görüntüler üzerinde iyi çalışmasına rağmen, metin tabanlı çalışmalarda zaman zaman yetersiz kalabilmektedir. Özellikle sıralı işlem veya zaman serileri gibi hafıza gerektiren işlemlerde kabiliyetleri sınırlıdır.

2.2.2. Tekrarlayan sinir ağı (Recurrent neural network)

RNN mimarisinin en belirgin özelliği hafızaya sahip olmasıdır. Bu yapı birimler arasında bağlantılar ile döngü oluşturabilen ve bu şekilde zaman serilerine dayanan ve

sıralı işlemlerde başarılı sonuçlar ortaya koyan bir sinir ağı modelidir [33, 34]. RNN her bir adım işlemin bir önceki çıktıya bağlandığı çalışma mimarisine dayandırılmıştır. Her birime ait bilgilerin kullanılabilmesi özellikle görüntü üzerinde gerçekleştirilen işlemlerde yüksek doğruluk ortaya koymaktadır. Bu da ileri beslemeli ağlara göre en büyük avantajıdır. RNN mimarisinde katman sayısına bağlı olarak katlanarak değişen çarpımsal gradyan sebebiyle uzun süreli bağımlılıkları tespit etmek kolay değildir. Bu sorunun üstesinden gelmek için mimariye geçit eklenerek geliştirilmiş ve elde edilen bu mimariye ise Uzun Kısa Süreli Bellek (LSTM) adı verilmiştir.



Şekil 2. Öğrenilen yerel örüntülerden yüksek seviyeli nesne tespiti (High level object detection from learned local patterns) [32]

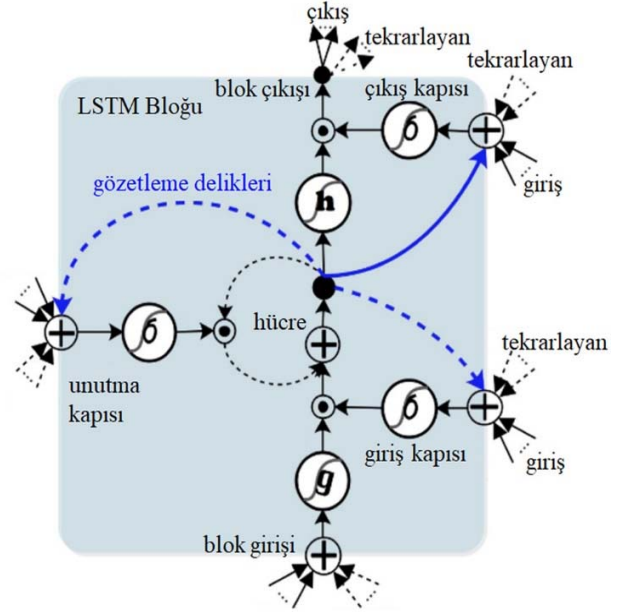
LSTM rastgele aralıklarla değerleri hatırlayan bir RNN mimarisine sahiptir. LSTM görüntü, ses, video gibi veri dizilerini geri bildirim bağlantıları ile işleyebilmektedir [35, 36]. Aynı zamanda yazı tanıma, duygu sınıflandırma, tıbbi veri sınıflandırma [5], konuşma [37] ve dil tanıma problemlerinde de zaman serisi veriler üzerinde işlem yaparak analiz gerçekleştirebilmektedir. Zaman serisinde önemli olaylar arasında gecikme olabileceğinden gradyan problemine karşılık Şekil 3'te görülen LSTM mimarisi [38] kullanılmaktadır.

Şekil 3'te görülen LSTM temel olarak RNN'den farklı bir mimariye sahip değildir. Ancak ara durumu hesaplamak için farklı bir fonksiyon kullanır. LSTM'in hafızası bir önceki durumu ve o andaki giriş X değerini girdi olarak alan bir mimaridir. Bu yapı önceki durumu, mevcut belleği ve girdiyi birleştirerek uzun vadeli bağımlılıkları yakalamada etkili sonuçlar ortaya koymaktadır.

İki Yönlü Tekrarlayan Sinir Ağları ise (Bidirectional Recurrent Neural Network-BRNN) zaman serileri üzerinde hem geçmiş hem gelecek bilgileri kullanabilmeyi sağlamaktadır. Klasik RNN mimarisinden farklı olarak hem geri hem ileri katmanlardan elde edilen sonuçlar ile mevcut andaki değeri Eş. 1'de görüldüğü gibi hesaplanmaktadır. T zaman dilimindeki $y^{<t>}$ çıkışı:

$$\begin{aligned} a^{<t>} &= g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a) \quad (1) \\ y^{<t>} &= g_2(W_{ya}a^{<t>} + b_y) \end{aligned}$$

Eş. 1'de W ve b geçici paylaşılan katsayıları, g aktivasyon fonksiyonlarını, a ise bir önceki çıkışa uygulanan aktivasyon fonksiyonunu temsil etmektedir. İkili indisler; hangi katmanın çıkışının sonraki katmanda nereye bağlandığını, tekli indisler ise aktivasyon fonksiyonunun veya katsayısının bulunduğu katmanı temsil etmektedir. Üst indisler de zamanı göstermektedir.

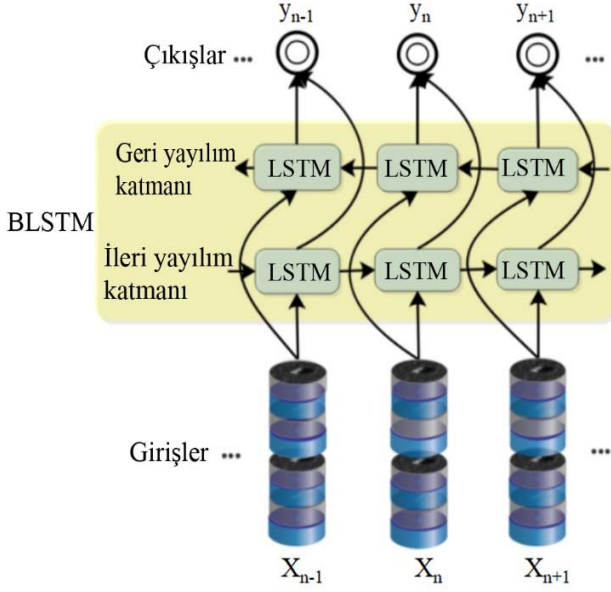


Şekil 3. Uzun Kısa Süreli Bellek blok diyagramı (LSTM block diagram) [38]

BRNN üzerindeki tekrar eden hücrelerde LSTM mimarisi kullanılarak, İki Yönlü Uzun-Kısa Süreli Bellek (Bidirectional Long Short-Term Memory-BLSTM) mimarisi oluşturulmuştur. Bir RNN mimarisi olan BLSTM, ileri ve geri yönlü hesaplamalar ile metin ve ses tanıma gibi sıralı veriler üzerinde başarılı şekilde kullanılmaktadır [39]. Bu çalışmada metin bazlı giriş görüntüleri üzerinde her karakterin geçmiş ve gelecek bağlamları hesaplanarak karakterlerin analizi ve yazının tahmininde BLSTM mimarisinin sıralı veriler üzerindeki avantajlarından yararlanılmıştır. Şekil 4'te görüldüğü gibi BLSTM ağları ileri ve geri yönde gelen bilgileri dikkate alır [38].

2.2.3. Bağlantıcı zamansal sınıflandırma (Connectionist temporal classification -CTC)

Danışmanlı öğrenme algoritmaları kullanılırken dizilim problemlerinde hizalama ile ilgili sorunlarla karşılaşılabilir. Örneğin, 128 piksel genişliğindeki görüntü RNN'e giriş dizisi olarak verildiğinde her bir zaman adımının etiketlenmesi gerekmektedir. Bu şekilde yapılacak etiketleme çok uygulanabilir bir yöntem değildir. Bu durumda CTC yöntemi RNN mimarisindeki her bir zaman diliminin etiketlenmesi gerekliliğini ortadan kaldırmak için kullanışlı bir yöntemdir. Bunun için CTC sadece harf olarak değil, kelime hatta satır olarak verilen metin ve metnin gerçek etiketini alarak bu sorunu çözer.



Şekil 4. İki yönlü Uzun Kısa Süreli Bellek blok diyagramı (BLSTM block diagram)

Ayrıca harf tekrarı olması durumunda da araya boşluk karakteri ekleyerek kelimenin doğru bir şekilde oluşturulmasına yardımcı olur. Örneğin “Serra” ismi RNN tarafından tahmin edildiğinde [“S”, “S”, “e”, “e”, “e”, “r”, “r”, “a”, “a”, “a”, “a”] olarak üretilir. Bu durumda metinde tekrarlayan karakterlerin birleştirilmesi, sorunu çözmek için iyi bir yaklaşım olsa da birleştirme sonucunda elde edilecek “Sera” kelimesi yanlış tahmindir. CTC dizilimde tekrarlı harflerin arasına eklediği boş karakter ile doğru çıktı “Serra” kelimesinin üretilmesini sağlar.

2.2.4. Levenshtein mesafesi (Levenshtein distance)

Levenshtein Mesafesi (LD) iki dizge arasındaki benzerliği derecelendirmek için kullanılmaktadır [40]. Gerçek dünya problemlerinde genellikle karakter dizgeleri üzerinde arama sonuçlarının benzerliğini ölçmek ve bu sayede hatalı olan dizgeler için önerilerde bulunmak için tercih edilmektedir. Modelin tahmin ettiği dizilim ile gerçek dizilim arasındaki hatanın minimize edilmesi için kullanılmıştır. Böylece İngilizce karakterler ile eğitilen modelin Türkçe metinlerden oluşan veri setinde kabul edilebilir bir başarı göstermesi sağlanmıştır.

3. UYGULAMA (APPLICATION)

3.1. Geliştirme Ortamı (Development Environment)

Uygulama, Anaconda Tümlşik Geliştirme Ortamı (Integrated Development Environment - IDE) ve Jupyter Notebook yazılımı kullanılarak geliştirilmiştir. Geliştirme dili olarak Python, derin öğrenme kütüphaneleri olarak Keras ve TensorFlow kullanılmıştır. Tablo 1’de, uygulamanın gerçekleştirildiği bilgisayarın yazılım ve donanım özellikleri görülmektedir.

Tablo 1. Modelin tasarlandığı ortam özellikleri (Environment features for which the model is designed)

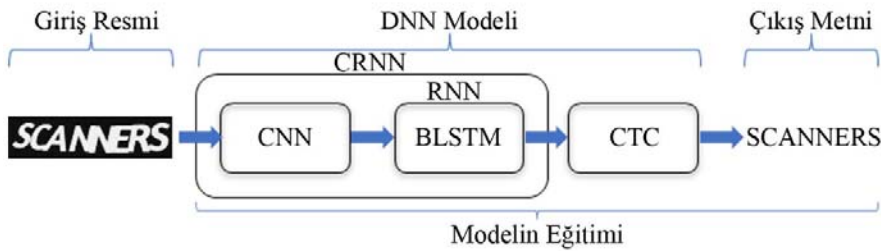
CPU	Intel Core i7-4700HQ 2.40 Ghz
GPU ¹	Nvidia GeForce GTX 850M 4GB
(Model Eğitimi İçin)	
GPU ²	Intel HD Graphics 4600
Bellek	16GB
İşletim Sistemi	Windows 10 Pro 64 bit
Anaconda Navigator	1.10
Keras	2.3
TensorFlow	1.13

3.1. Veri Setinin Hazırlanması (Preparation of The Dataset)

Veri üzerinde ön işleme adımları çoğu zaman veri ile gerçekleştirilecek işlemler için bir analiz niteliği taşır. Bu analiz veriler üzerinden anlamlı bilgilerin çıkarılması esasına dayanır. Veri temizleme, veri indirgeme, veri dönüştürme, veri birleştirme gibi birçok ön işleme tekniği mevcuttur [41]. Bu teknikler, veriyi anlama, eksik verileri tespit etme ve düzenleme, dönüştürme, bütünleştirme, boyutlandırma, normalleştirme adımları ile verinin kullanım amacına uygun hale getirilmesini sağlamaktadır.

Çalışmada, Synth90k veri setinden 50000 adet 3 kanal renk bilgisine sahip giriş görüntüsü 0-255 aralığında gri seviyeli resme dönüştürülmüştür. Giriş veri boyutu 128×32 olarak belirlenmiştir. Daha sonra Eş. 2’de verilen normalizasyon işlemi uygulanarak 0-255 aralığındaki piksel gri seviyeleri 0-1 aralığına dönüştürülmüştür. Böylece 0: siyah, 1: beyaz ve aradaki değerlerin gri seviyeleri 0 ve 1’e yakınlıklarına göre temsil edilmektedir.

$$x' = (x - \min(x)) \div (\max(x) - \min(x)) \quad (2)$$



Şekil 5. Oluşturulan modele genel bakış (Overview of the created model)

3.2. Derin Öğrenme Modelinin Oluşturulması (Creation of The Deep Learning Model)

Bu çalışmada, RNN mimarisinin avantajları kullanılarak CNN mimarisinin metinler üzerindeki performansını iyileştiren birleştirilmiş CRNN modeli önerilmektedir. Şekil 5'te çalışmada kullanılan model mimarisinin genel adımları görülmektedir.

128×32 boyutlarında modele verilen giriş görüntüsünün özellikleri CNN mimarisi ile oluşturulan ağ tarafından öğrenilmektedir. Çıkarılan özelliklerden görüntü üzerindeki harf barındıran kısımlara odaklanılmaktadır. Ardından RNN türevi BLSTM mimarisi ise bilgiyi daha uzun mesafelere yaydığı için kullanılmıştır. RNN katmanından alınan matris formundaki çıktı için şartlı olasılık hesaplanır. Bu olasılığa göre karakterler ve dizilim oluşturulur. CTC yöntemi sayesinde tekrarlı harfler veya benzer örüntüler içeren metinlerde başarılı sonuçlar elde edilir. Çalışmada, karakter tanıma için oluşturulan CRNN modelinin yapısı Şekil 6'da gösterilmektedir.

Çalışmada kullanılan mimaride görüntü boyutları (128,32,1) olarak düzenlenmiştir. CNN katmanında ele alınan giriş görüntüleri üzerinde özellik çıkarımı için katmanlar 64'ten 512'ye artarak sekiz evrişim katmanı ile ele alınır. Kullanılan 2x2, 3x3 boyutlarında filtreler ile katmanlı mimari üzerinde dönüşüm işlemine dayalı çıkış verileri elde edilir. Dönüşüm işlemi filtrelerin tüm görüntüler üzerinde gezinme işlemine dayanmaktadır. Evrişim işlemi sonucunda aktivasyon haritası (özellik haritası) oluşur. Filtrelere bağlı özelliklerin çıkarımı sağlanır ve çıktı değerlerine Eş. 3'te ki Düzleştirilmiş Doğrusal Birim (Rectified Linear Unit - RELU) aktivasyon fonksiyonu uygulanarak ağına daha hızlı öğrenmesi sağlanır.

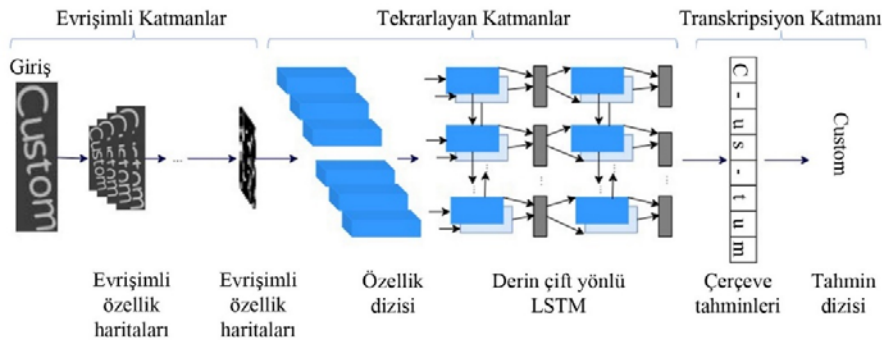
$$f(x) = \begin{cases} 0 & \text{eğer } x < 0 \\ x & \text{eğer } x \geq 0 \end{cases} \quad (3)$$

Derin öğrenme çalışmalarında veriden anlamlı bilgilerin analizi için aktivasyon fonksiyonları kullanılmaktadır. Çalışmada aktivasyon fonksiyonu olarak çıkış katmanında Softmax diğer katmanlarda doğrusal ve hesaplaması daha hızlı olduğu için RELU katmanı uygulanmıştır. Evrişim katmanına 2x2 ve 2x1 boyutunda maksimum havuz katmanı

eklenerek geniş ölçekli özelliklerin de çıkarımı sağlanmış ve model karmaşıklığı, performansı düşürmeden parametre sayısını düşürerek azaltılmıştır. Çalışmada aşırı öğrenmeyi engelleyerek eğitim performansını artırmak için belirlenen oranda bağlantıları rastgele kaldıran Seyreltme (Dropout) katmanı kullanılmıştır. Ağın eğitim süresini kısaltması ve daha iyi performans göstermesi için ağına içinde gerçekleşen veri dağılımının değişmesine bağlı olarak katmanlar arası geçişlerde Yığın Normalleştirme (Batch Normalization) kullanılmıştır. Elde edilen CNN çıktısı lambda işlevi ile sıkıştırılmış (squeeze) BLSTM ile uyumlu hale getirilmiştir. Daha sonra her biri 128 birim olan iki BLSTM katmanı ile sonuç olarak boş karakter dahil toplam çıktı sayısı kadar her etiketleme için olasılıklar belirlenmiştir. Son katmanda CTC kullanılarak yüksek doğrulukta karakterlerin analizi sağlanmıştır. Tablo 2'de ağı oluşturan mimari tasarım görülmektedir.

3.3. Modelin Eğitimi (Training of The Model)

Çalışmada görüntüler üzerinde özelliklerin doğru tespiti ve önemli bilgilerin analizini gerçekleştirmek için normalizasyon işlemi uygulanmıştır. Her bir giriş görüntüsü CNN'e girdi olarak verilerek evrişim katmanında filtre parametreleri ile metin görüntüleri üzerinde özelliklerin çıkarımı gerçekleştirilmiştir. Elde edilen çıkış bilgileri RNN katmanında BLSTM mimarisi ile ele alınmıştır. Sıralı kelime bilgilerinin zamansal özellikleri korunarak hem geçmişten geleceğe hem gelecekte geçmişe analiz yapılmıştır. BLSTM mimarisinin çıktıları ile metin görüntülerindeki içeriğin her bir karakter için olasılığı hesaplanmıştır. Son katmanda CTC kullanılmıştır. CTC, zamanlamanın göz önüne alındığı çalışmalarda RNN mimari ile dizi sorunlarının çözümünü sağlayan sinir ağı çıktısıdır. CTC ile modellenen RNN yazı ve konuşma tanıma çalışmalarında yüksek performans göstermektedir [42]. Google'ın konuşma tanıma çalışmalarında da RNN mimarisi ile kullanılan CTC ile %49'luk bir başarı artışı sağlanmıştır. Çalışmalarda belirlenen boyuttaki görüntüler RNN'e girdi olarak verilmeden önce her bir zaman adımı etiketlenmektedir. Bu çalışmada CTC direkt olarak görüntü üzerindeki metni bir görüntüye etiketleyerek ve RNN çıktısında elde edilen aynı karakterleri birleştirerek en yüksek olasılığa sahip karakterleri olarak tahmin doğruluğunu arttırmaktadır.

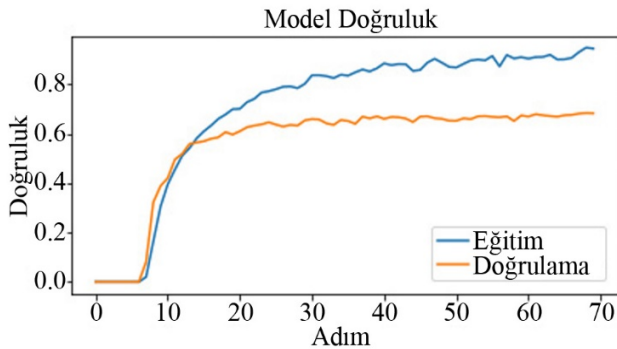


Şekil 6. Karakter tanıma için oluşturulan CRNN model yapısı (Created CRNN model structure for character recognition)

Tablo 2. Oluşturulan ağın mimari yapısı (Architectural structure of the designed network)

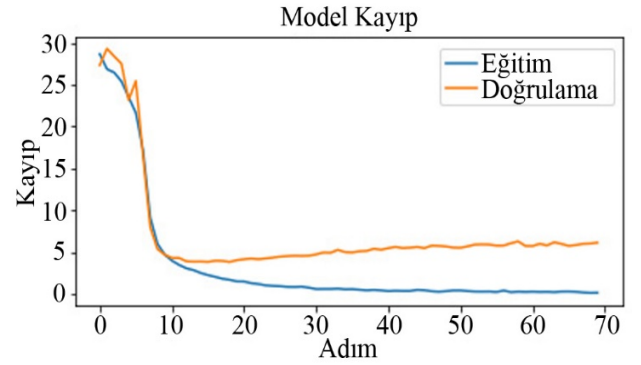
Katmanlar	Katman türü	Filtre boyutu	Filtre sayısı	Çıkış boyutu (Düğüm sayısı)	Eğitilecek parametre sayısı
E1	Giriş	-	-	$32 \times 128 \times 1$	0
	Evrişim	3×3	64	$32 \times 128 \times 64$	640
	ReLU	-	-	$32 \times 128 \times 64$	0
H2	Havuzlama	2×2	-	$16 \times 64 \times 64$	0
E3	Evrişim	3×3	64	$16 \times 64 \times 64$	36928
	ReLU	-	-	$16 \times 64 \times 64$	0
E4	Evrişim	3×3	128	$16 \times 64 \times 64$	73856
	ReLU	-	-	$16 \times 64 \times 64$	0
H5	Havuzlama	2×2	-	$8 \times 32 \times 128$	0
E6	Evrişim	3×3	256	$8 \times 32 \times 256$	295168
	ReLU	-	-	$8 \times 32 \times 256$	0
E7	Evrişim	3×3	256	$8 \times 32 \times 256$	590080
	ReLU	-	-	$8 \times 32 \times 256$	0
H8	Havuzlama	2×1	-	$4 \times 32 \times 256$	0
E9	Evrişim	3×3	512	$4 \times 32 \times 512$	1180160
	ReLU	-	-	$4 \times 32 \times 512$	0
TN10	Toplu	-	-	$4 \times 32 \times 512$	2048
	Normalleştirme	-	-	-	-
E11	Evrişim	3×3	512	$4 \times 32 \times 512$	2359808
	ReLU	-	-	$4 \times 32 \times 512$	0
TN12	Toplu	-	-	$4 \times 32 \times 512$	2048
	Normalleştirme	-	-	-	-
H13	Havuzlama	2×1	-	$2 \times 32 \times 256$	0
E14	Evrişim	2×2	512	$1 \times 31 \times 512$	1049088
	ReLU	-	-	$1 \times 31 \times 512$	0
L15	Lambda	-	-	31×512	0
İY16	İki-Yönlü	-	128	31×256	656384
İY17	İki-Yönlü	-	128	31×256	394240
T18	Tam bağlı	-	-	31×63	16191

Synth90k veri setindeki 50000 görüntü üzerinden %90'ı eğitim, %10'u doğrulama veri seti olarak belirlenmiştir. Tasarlanan model yapısına göre, 45000 eğitim, 5000 doğrulama görüntüsü için Şekil 7 ve Şekil 8'de modelin doğruluk ve kayıp (loss) değerlerinin grafikleri görülmektedir.



Şekil 7. Eğitim ve doğrulama için modelin doğruluk grafiği (Accuracy graphic of the model for training and validation)

Tablo 3'te, çalışmanın veri seti büyüklüğü, adım (epoch) değeri ve yığın (batch) büyüklüğü değiştirildiğinde elde edilen doğruluk oranlarındaki değişim görülmektedir.



Şekil 8. Eğitim ve doğrulama için modelin kayıp grafiği (Loss value graphic of the model for training and validation)

Tablo 3. Modelin farklı boyutta Synth90k verileri için eğitim başarımları (Training performance values for different size Synth90k data of the model)

Model	Veri boyutu	Adım sayısı	Yığın büyüklüğü	Doğruluk değeri (%)
CRNN	11000	50	100	90,0
CRNN	22000	50	150	81,0
CRNN	45000	40	256	83,0
CRNN	50000	70	256	94,0

4. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

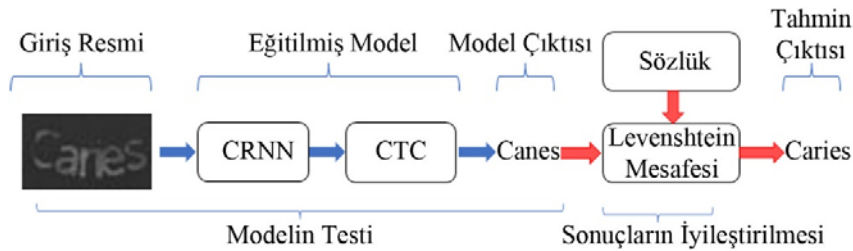
Eğitimi 50000 görüntü ve %94'lük doğrulukla tamamlanan model, aynı veri setinden daha önce modele hiç gösterilmemiş 7395 adet görüntü ile test edilmiştir. Test için kullanılan veri seti ön işlemlerden geçirilerek eğitilmiş modele girdi olarak verilmiştir. Modelin test verileri için ürettiği tahmin sonuçlarında karakter hatası olan dizilimlerin de yanlış tahmin olarak sınıflandırılması sonucunda elde edilen tahmin sonuçlarının iyileştirilmesine karar verilmiştir. Bu sebeple modelin ürettiği tahmin sonuçları üzerine LD kullanarak yapılan iyileştirme sonucunda test verilerindeki başarının %96'luk bir doğruluğa ulaştığı gözlenmiştir. Tahmin yapan modelin genel yapısı Şekil 9'da verilmiştir. Şekil 10'da modelin test görüntüleri üzerindeki tahminlerinden örnek bir kısmı görülmektedir.

Model İngilizce dilinde yeterince zengin bir veri seti ile eğitilmiş ve tahmin doğruluğu başarısı LD ile artırılmıştır. LD için kullanılan sözlüğün değiştirilmesi ile İngilizce ile aynı veya benzer karakter setini kullanan dillerde de başarının yüksek olacağı düşünülmüştür. Bu öngörünün doğruluğunu test etmek üzere 100 resimden oluşan Türkçe tabele görüntülerindeki metinler için İngilizce metinlerle eğitilmiş CRNN modeli ve modelin ürettiği sonuçların iyileştirilmesi için Türkçe sözlük kullanan LD ile %93 doğrulukla tahmin başarısı elde edildiği gözlemlenmiştir. Elde edilen sonuçlar tezin başarısını doğrulamaktadır.

Tablo 4'te literatürde farklı veri setleri ile gerçekleştirilen çalışmalarda ve bu çalışmada kullanılan veri seti ile elde edilen CRNN başarımlarının karşılaştırması verilmiştir.

Tablo 4. Farklı çalışmalardaki CRNN mimarilerinin başarı oranının karşılaştırması
(Success rate comparison of CRNN architectures in different studies)

Çalışmalar	Model	Veri seti	Doğruluk (%)
Saxena vd. [17]	CRNN_STN	Synth90k	90,7
Gordo [43]	CRNN	SVT	91,8
Jaderberg vd. [44]	CRNN	IIIT5k	89,6
Bissacco vd. [45]	CRNN	SVT	90,4
Jaderberg vd. [46]	CRNN	IC03	91,5
Oluşturulan	CRNN	Synth90k	96,0
Oluşturulan*	CRNN	Türkçe Görüntüler	93,0



Şekil 9. Model çıktısının LD ile iyileştirilmesi (Enhancement of model output with LD)

orjinal_metin = Frogmarching orjinal_metin = Renaud
tahmini_metin = Frogmarching tahmini_metin = Renaud

orjinal_metin = fudges orjinal_metin = lucbrates
tahmini_metin = fudges tahmini_metin = lucbrates

orjinal_metin = gauze orjinal_metin = Mealybug
tahmini_metin = gauze tahmini_metin = Mealybug

Şekil 10. Test verileri için model tarafından oluşturulan tahminler (Generated estimations by the model for test data)

Bu çalışmada önerilen yöntemin performansı, literatürdeki çalışmalarla karşılaştırıldığında elde edilen başarı açıkça görülmektedir. Hem benzer model-farklı veri seti, hem de farklı model-benzer veri seti kullanılarak yapılan çalışmalara göre iyi bir eğitim ve test başarısı elde edilmiştir.

5. SONUÇLAR (CONCLUSIONS)

Bu çalışmada iç mekan yönlendirmesinde kullanılmak üzere metin tabanlı görüntülerde karakter bazlı kelime tespiti için tabele yazılarını tespit edebilen CNN ve RNN mimarilerinin birleştirildiği ve sonuçlarının iyileştirildiği bir yaklaşım geliştirilmiştir. Çalışmada CNN'in görüntülerdeki öznitelik tespiti ve özellik çıkarımındaki başarısı, RNN katmanında ele alınan BLSTM mimarisinin, sıralı harf bilgilerinin zamansal özellikleri korunarak hem geçmişten geleceğe hem gelecekte geçmişe analiz başarısı birleştirilmiştir. Tasarlanan CRNN mimarisinde CNN katmanlarıyla görüntülerdeki köşe, kenar vb. yüksek seviyeli özellikler elde edilmiştir. Elde edilen özellik vektörleri RNN tabanlı BLSTM mimarisine giriş olarak verilmiş ve dizilim olasılıkları hesaplanmıştır. Elde edilen dizilimlerdeki harf tekrarı gibi sorunların üstesinden gelmek için ağırlık son katmanında CTC yöntemi kullanılmıştır. Böylece modelin eğitim mimarisi tamamlanmıştır. Ancak CTC algoritmasının da zaman zaman yetersiz kalabildiği durumlar olduğu için modelin tahmin ettiği kelimelerle haricen oluşturulmuş bir sözlükteki benzer kelimeler arasında LD hesaplanarak karakter bazlı hataların önüne geçilerek doğruluk başarısı %96'ya kadar iyileştirilmiştir. Son adımda yapılan iyileştirmenin katkısını daha net ortaya koymak adına, modelin eğitiminde kullanılan karakter setine sahip farklı bir dildeki tahmin başarısı test edilmiş ve daha önce harflerin dizilimleri hakkında dahi bilgiye sahip olmayan önerilen

model %93' lük doğrulama başarısı göstermiştir. LD için kullanılan sözlüğün modelden bağımsız olması aynı zamanda çalışma zamanında hızlanmayı da beraberinde getirmektedir. Çünkü bir dildeki bütün kelimeler yerine, arama uzayının sınırlandırılması toplam hesapsal maliyeti azaltacaktır.

Gelecekte yapılacak çalışmalarda, CRNN mimarisi ile görüntülerden tespit edilen bu metinlere göre iç mekanda konum ilişkilendirilmesi gerçekleştirilecek ve böylece yönlendirme için mobil cihazlarda kullanılacak bir uygulama geliştirilecektir.

KAYNAKLAR (REFERENCES)

1. Şeker A., Diri B., Balık H.H., Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme, Gazi Mühendislik Bilimleri Dergisi, 3 (3), 47-64, 2017.
2. Kaya A., Keçeli A.S., Can A.B., Examination of various classification strategies in classification of lung nodule characteristics, Journal of the Faculty of Engineering and Architecture of Gazi University, 34 (2), 709-725, 2019.
3. Atasoy F., Eltanashi S., A Proposed Speaker Recognition Model Using Optimized Feed Forward Neural Network and Hybrid Time-Mel Speech Feature, International Conference on Advanced Technologies Computer Engineering and Science (ICATCES), Karabuk-Turkey, 130-140, 3-5 June, 2020.
4. Yıldız O., Melanoma detection from dermoscopy images with deep learning methods: A comprehensive study, Journal of the Faculty of Engineering and Architecture of Gazi University, 34 (4), 2241-2260, 2019.
5. Atasoy F., Hajaj E., Thyroid Disorder Prediction Using Advance Deep Learning Paradigms, 1st International Conference on Computer, Electrical and Electronic Sciences, Konya-Turkey, 149-154, 8-10 October, 2020.
6. Metin İ.A., Karasulu B., A novel dataset of human daily activities: Its benchmarking results for classification performance via using deep learning techniques, Journal of the Faculty of Engineering and Architecture of Gazi University, 36 (2), 759-777, 2020.
7. Deng J., Dong W., Socher R., Li L., Li K., Fei-Fei L., ImageNet: A large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition, Miami-FL-USA, 248-255, 20-25 June, 2009.
8. Hanbay K., Hyperspectral image classification using convolutional neural network and two-dimensional complex Gabor transform, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (1), 443-456, 2020.
9. Grefenstette E., Blunsom P., de Freitas N., Hermann K.M., A Deep Architecture for Semantic Parsing, Proceedings of the ACL 2014 Workshop on Semantic Parsing, Baltimore-MD, 22-27, 26 June, 2014.
10. Kim Y., Convolutional neural networks for sentence classification, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha-Qatar, 1746-1751, 25-29 October, 2014.
11. Graves A., Jaitly N., Towards end-to-end speech recognition with recurrent neural networks, Proceedings of the 31st International Conference on Machine Learning (ICML), 32 (2), 1764-1772, 2014.
12. Karpathy A., Fei-Fei L., Deep Visual-Semantic Alignments for Generating Image Descriptions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (4), 664-676, 2017.
13. Salouhou A., El yazısı karakter tanıma ve resim sınıflandırmada derin öğrenme yaklaşımları, Yüksek Lisans Tezi, Fatih Sultan Mehmet Vakıf Üniversitesi, Lisansüstü Eğitim Enstitüsü, İstanbul, 2019.
14. Dutta K., Krishnan P., Mathew M., Jawahar C.V., Improving CNN-RNN Hybrid Networks for Handwriting Recognition, 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), USA, 80-85, 5-8 August, 2018.
15. Wang R., Li Z., Cao J., Chen T., Wang L., Convolutional Recurrent Neural Networks for Text Classification, International Joint Conference on Neural Networks, Budapest-Hungary, 1-6, 14-19 July, 2019.
16. Hamad K., Kaya M., A Detailed Analysis of Optical Character Recognition Technology, International Journal of Applied Mathematics, Electronics And Computers, 4 (1), 244-244, 2016.
17. Jaderberg M., Simonyan K., Vedaldi A., Zisserman A., Reading Text in the Wild with Convolutional Neural Networks, International Journal of Computer Vision, 116 (1), 1-20, 2016.
18. Saxena S.S., Saranya G., Aggarwal D., A Convolutional Recurrent Neural Network (CRNN) Based Approach for Text Recognition and Conversion of Text to Speech in Various Indian Languages, International Journal of Advanced Science and Technology, 29 (06), 2270-2276, 2020.
19. He T., Huang W., Qiao Y., Yao J., Text-Attentional Convolutional Neural Network for Scene Text Detection, IEEE Transactions on Image Processing, 25 (6), 2529-2541, 2016.
20. Li H., Wang P., Shen C., Toward End-to-End Car License Plate Detection and Recognition with Deep Neural Networks, IEEE Transactions on Intelligent Transportation Systems, 20 (3), 1126-1136, 2019.
21. Yuan A., Bai G., Jiao L., Liu Y., Offline handwritten English character recognition based on convolutional neural network, 10th IAPR International Workshop on Document Analysis Systems (DAS), Gold Coast-QLD-Australia, 125-129, 27-29 March, 2012.
22. Wu C., Fan W., He Y., Sun J., Naoi S., Handwritten Character Recognition by Alternately Trained Relaxation Convolutional Neural Network, 14th International Conference on Frontiers in Handwriting Recognition, Hersonissos-Greece, 291-296, 1-4 September, 2014.
23. Zhong Z., Jin L., Feng Z., Multi-font printed Chinese character recognition using multi-pooling convolutional

- neural network, 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis-Tunisia, 96–100, 23-29 August, 2015.
24. Yang W., Jin L., Xie Z., Feng Z., Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge, 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis-Tunisia, 551-555, 23-26 August, 2015.
 25. He M., Zhang S., Mao H., Jin L., Recognition confidence analysis of handwritten Chinese character with CNN, 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis-Tunisia, 61–65, 23-26 August, 2015.
 26. Bai J., Chen Z., Feng B., Xu B., Image character recognition using deep convolutional neural network learned from different languages, IEEE International Conference on Image Processing (ICIP), Paris-France, 2560-2564, 27-30 October, 2014.
 27. Rizvi M.A.I., Deb K., Khan M.I., Kowsar M.M.S., Khanam T., A comparative study on handwritten Bangla character recognition, Turkish Journal of Electrical Engineering and Computer Sciences, 27 (4), 3195–3207, 2019.
 28. Jaderberg M., Simonyan K., Vedaldi A., Zisserman A., Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition, NIPS Deep Learning Workshop, Montreal, 1-10, 12 December, 2014.
 29. Çetiner H., Çetişli B., Çetiner İ., Gerçek Zamanlı T.C. Kimlik Numarası Tanıma, SAÜ. Fen Bilimleri Dergisi, 16 (2), 123-129, 2012.
 30. Akbani O., Gokrani A., Quresh M., Khan F.M., Behlim S.I., Syed T.Q., Character recognition in natural scene images, International Conference on Information and Communication Technologies (ICICT), Karachi-Pakistan, 1-6, 12-13 December, 2015.
 31. Lecun Y., Bengio Y., Hinton G., Deep learning, Nature, 521 (7553), 436–444, 2015.
 32. Chollet F., Deep Learning with Python, Manning Publications, 2018.
 33. Liang M., Hu X., Recurrent convolutional neural network for object recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston-MA-USA, 3367–3375, 07-12 June, 2015.
 34. Dogan Y., Improvement of recurrent deep neural networks algorithm by feature selection methods and its usage of automatic identification system data evaluated as time series, Journal of the Faculty of Engineering and Architecture of Gazi University, 35 (4), 1897–1911, 2020.
 35. Graves A., Schmidhuber J., Framewise phoneme classification with bidirectional LSTM networks, Neural Networks, 18 (5-6), 602-610, 2005.
 36. Fernández S., Graves A., Schmidhuber J., An application of recurrent neural networks to discriminative keyword spotting, International Conference on Artificial Neural Networks, Porto-Portugal, 220–229, 9-13 September, 2007.
 37. Graves A., Mohamed A., Hinton G., Speech Recognition with Deep Recurrent Neural Networks, IEEE International Conference on Acoustics Speech and Signal Processing, Vancouver-BC-Canada, 6645-6649, 26-31 May, 2013.
 38. Fei G., Teng H., Wang J., Sun J., Hussain A., Zhou H., A Novel Multi-Input Bidirectional LSTM and HMM Based Approach for Target Recognition from Multi-Domain Radar Range Profiles, Electronics, 8 (5), 535, 2019.
 39. Graves A., Jaitly N., Mohamed A., Hybrid speech recognition with Deep Bidirectional LSTM, IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc-Czech Republic, 273-278, 8-12 December, 2013.
 40. Yujian L., Bo L., A Normalized Levenshtein Distance Metric, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (6), 1091–1095, 2007.
 41. Oğuzlar A., Veri Ön İşleme, Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 0 (21), 67–76, 2003.
 42. Bluche T., Ney H., Louradour J., Kermorvant C., Framewise and CTC training of Neural Networks for handwriting recognition, 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis-Tunisia, 81-85, 23-26 August, 2015.
 43. Gordo A., Supervised mid-level features for word image representation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston-MA-USA, 2956–2964, 07-12 June, 2015.
 44. Jaderberg M., Simonyan K., Vedaldi A., Zisserman A., Deep structured output learning for unconstrained text recognition, International Conference on Learning Representations (ICLR), California-US, 1–10, 7-9 May, 2015.
 45. Bissacco A., Cummins M., Netzer Y., Neven H., PhotoOCR: Reading text in uncontrolled conditions, IEEE International Conference on Computer Vision, Sydney-NSW-Australia, 785–792, 1-8 December, 2013.
 46. Jaderberg M., Vedaldi A., Zisserman A., Deep Features for Text Spotting, 13th European Conference, Zurich-Switzerland, 512-528, 6-12 September, 2014.

