

T.C.
İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR BİLİMİ VE MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR BİLİMİ VE MÜHENDİSLİĞİ BİLİM DALI

İNGİLİZCEDE ÇEMBER DOLGU ORANI YÖNTEMİ
İLE OPTİK KARAKTER TANIMA

YÜKSEK LİSANS TEZİ

Sahra TILKI

İstanbul
Haziran, 2020

T.C.
İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR BİLİMİ VE MÜHENDİSLİĞİ
YÜKSEK LİSANS PROGRAMI

İNGİLİZCEDE ÇEMBER DOLGU ORANI YÖNTEMİ İLE
OPTİK KARAKTER TANIMA

YÜKSEK LİSANS TEZİ

Sahra TILKI

Tez Danışmanı
Dr. Yahya ŞİRİN

İstanbul
Haziran, 2020

TEZ ONAYI

Lisansüstü Eğitim Enstitüsü Müdürlüğüne,

Bu çalışma, jürimiz tarafından Bilgisayar Bilimi ve Mühendisliği Anabilim Dalı, Bilgisayar Bilimi ve Mühendisliği Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman
Dr. Yahya ŞİRİN

Üye
Dr. Öğr. Üyesi Akhtar JAMIL

Üye
Dr. Öğr. Üyesi Abdulfetah Abdela SHOBOLE

Onay

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylıyorum.

Prof. Dr. Ali GÜNEŞ

Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİMİ

Yüksek lisans tezi olarak hazırladığım “**İngilizcede Çember Dolgu Oranı Yöntemi ile Optik Karakter Tanıma**” adlı çalışmanın öneri aşamasından sonuçlandığı aşamaya kadar geçen süreçte bilimsel etiğe ve akademik kurallara özenle uyduğumu, tez içindeki tüm bilgileri bilimsel ahlak ve gelenek çerçevesinde elde ettiğimi, tez yazım kurallarına uygun olarak hazırladığımı, bu çalışmamda doğrudan veya dolaylı olarak yaptığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu beyan ederim.



Sahra TILKI

ÖNSÖZ

Öncelikle çalışmalarım süresince kıymetli bilgi ve tecrübeleriyle bana yol gösteren değerli danışmanım Dr. Yahya Şirin'e teşekkürü bir borç bilirim. Ayrıca daima yanımda olan anneme, babama ve değerli arkadaşlarım Hasibe Büşra Doğru ve Büşranur Muslu'ya teşekkürlerimi sunarım.

Sahra TİLKİ

İstanbul - 2020



ÖZET

İNGİLİZCEDE ÇEMBER DOLGU ORANI YÖNTEMİ İLE OPTİK KARAKTER TANIMA

Sahra Tilki

Yüksel Lisans, Bilgisayar Bilimi ve Mühendisliği

Tez Danışmanı: Dr. Yahya Şirin

Haziran-2020, 83 Sayfa

Çeşitli alanlarda, mevcut olan verilerden bilgisayar ortamına bilgi depolaması yapabilmek, el yazısıyla yazılmış veya basılı dokümanlarda bulunan bilgileri, bilgisayarlar aracılığıyla tekrar kullanabilmek için fazla talep bulunmaktadır. Ancak bu bilgilerin olduğu belge veya dosyaları yeniden kullanabilmek, saklayabilmek için, bilgilerin okunması ve sorgulanması gerekmektedir. Bu nedenler doğrultusunda görüntü dosyalarından, özellikle de metinlerdeki bilgileri otomatik olarak alma ve depolama tekniklerine ihtiyaç vardır.

Optik Karakter Tanıma (OCR), görüntülerden otomatik olarak metin çıkarma ve işleme yeteneğine sahip bir bilgisayar sistemi geliştirmeye çalışan, son yıllarda araştırmacıların ilgisini çeken aktif bir araştırma alanıdır. OCR' nin amacı, herhangi bir metin biçiminde veya el yazısı içeren belgelerde değişiklik veya dönüşüm elde etmektir. Bu sayede daha derin ve daha ileri işlemler için yazdırılabilir veya taranmış metin görüntüleri, dijital formatta düzenlenebilir ve saklanabilir. Görüntüdeki karakterlerin doğru tanınması için görüntü üzerinde var olan çeşitli bozukluklar, karakterlerin font özellikleri, ışık miktarı, gibi birçok etkenin düzenlenmesi gerekmektedir. Bu nedenle kullanılacak görüntüler üzerinde gürültü azaltma, rotasyon düzeltme, ikili hale getirme ve normalizasyon gibi çeşitli görüntü işleme teknikleri uygulamak gerekir. Yapılan bu işlemler, sonraki aşamalarda karşılaşılabilecek zorluklara çözüm olup, doğru tanıma yapabilmek açısından oldukça önemli bir yere sahiptir.

Bu çalışmada, iskelet temsilini baz alan bir yöntem olan Çember Dolgu Oranı yönteminden elde edilen bilgiler ile derin öğrenme tekniklerinden biri olan CNN (Konvolüsyonel Sinir Ağı) modelinin birlikte kullanılmasının sonucunda daha az veri ile tanıma hızının artırılması ve yüksek doğruluk oranında tanıma amaçlanmaktadır.

Çalışmada kullanılan görüntüler, Matlab platformunda gerekli görüntü işleme teknikleri uygulanarak oluşturulan CNN modelinde eğitim için kullanılmak üzere hazır hale getirilmiştir. İlk aşamada CNN modelinde girdi olarak karakter görüntüleri kullanılmıştır. Daha sonra iskelet temsilleri oluşturulan görüntülerin, iskelet noktalarıyla ve Çember Dolgu Oranı yöntemi ile elde edilen noktalara PCA (Temel Bileşenler Analizi) işlemi uygulanarak görsel biçimde kullanılması sağlanmış ve bu işlemle birlikte yeni görüntüler elde edilmiştir. Bu görüntüler aynı CNN modelinde kullanılıp eğitim yapılmıştır. Gerçekleştirilen eğitimler sonucunda ulaşılan test sonuçlarının doğruluk oranları arasında performans kıyaslamaları yapılmış ve deneysel çalışmalarla sonuçları sunulmuştur.

Anahtar Kelimeler: OCR, Görüntü İşleme, Nesne Tanıma, Derin Öğrenme, CNN.

ABSTRACT

OPTICAL CHARACTER RECOGNITION WITH CIRCLE FILLING RATE METHOD FOR ENGLISH

Sahra Tilki

Supervisor: Dr. Yahya Şirin

June-2020, 83 Pages

There is a lot of demand to be able to store information from the data available in various fields to the computer environment, to reuse information which is handwritten or printed documents via computers. However, in order to reuse and store the documents or files containing this information, the information must be read and questioned. For these reasons, there is a need for automatic retrieval and storage of information from image files, especially text.

Optical Character Recognition (OCR) is an active research area that has been attracting researchers in recent years, trying to develop a computer system capable of automatically extracting and processing text from images. The purpose of OCR is to achieve changes or conversions in any text format or documents containing handwriting. Thus, printed or scanned text images can be edited and stored in digital format for deeper and further processing. In order for the characters in the image to be recognized correctly, various factors such as various defects on the image, the font properties of the characters, and the amount of light should be regulated. For this reason, it is necessary to apply various image processing techniques such as noise reduction, rotation correction, binarization and normalization on the images to be used. These procedures are a solution to the difficulties that may be encountered in the next stages and have a very important place in terms of making accurate recognition.

In this study, it is aimed to increase the recognition speed and recognition with less data as a result of using the information obtained from the Circle Fill Rate method, which is a method based on skeletal representation, and CNN (Convolutional Neural Network) which is one of the deep learning techniques, together with less data.

The images used in the study are made ready to be used for training in the CNN model created by applying the necessary image processing techniques in Matlab

platform. In the first stage, character images were used as inputs in the CNN model. Later, the images created with skeletal representations were used visually by applying PCA (Principle Components Analysis) to the points obtained with the skeleton points and the Circle Filling Rate method, and new images were obtained with this process. These images were used in the same CNN model and trained. Performance comparisons were made between the accuracy rates of the test results obtained as a result of the trainings conducted and the results were presented with experimental studies.

Keywords: OCR, Image Processing, Object Recognition, Deep Learning, CNN.



İÇİNDEKİLER

DIŞ KAPAK	
İÇ KAPAK	
TEZ ONAYI	i
BİLİMSEL ETİK BİLDİRİMİ.....	ii
ÖNSÖZ.....	iii
ÖZET.....	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
TABLolar LİSTESİ.....	xi
ŞEKİLLER LİSTESİ.....	xii
KISALTMALAR	xiv
BİRİNCİ BÖLÜM.....	1
GİRİŞ	1
1.1 Amaç	2
1.2 Kapsam.....	3
1.3 Katkılar	3
İKİNCİ BÖLÜM	5
LİTERATÜR TARAMASI	5
ÜÇÜNCÜ BÖLÜM	14
NESNE TANIMA.....	14
3.1 Veri Girişi.....	15
3.2 Ön İşleme:	15
3.3 Öznitelik Çıkarma.....	17
3.4 Tanımlama	20
3.5 Tanıma Stratejileri	21
DÖRDÜNCÜ BÖLÜM	23

OPTİK KARAKTER TANIMA	23
4.1 Optik Karakter Tanıma Tarihiçesi.....	24
4.2 Optik Karakter Tanıma Sistemlerinin Kullanım Alanları.....	25
4.3. OCR Adımları.....	27
4.3.1. Ön İşleme Aşaması	27
4.3.2. Özellik Çıkarımı	30
4.3.3. Sınıflandırma	31
4.4. OCR Doğruluğunu Etkileyen Faktörler.....	33
4.5. OCR Zorlukları	34
4.5.1. Görüntü Karmaşıklığı	35
4.5.2.Düzensiz Işıklandırma	35
4.5.3.Eğrilik (Rotasyon).....	35
4.5.4. En- Boy Oranları (Aspect Ratios).....	36
4.5.5.Eğim (Tilting, Perspective Distortion):	36
4.5.6.Odaklanma ve Görüntü Netliği.....	36
BEŞİNCİ BÖLÜM	37
MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME	37
5.1 Makine Öğrenmesi:	37
5.2 Derin Öğrenme	38
5.2.1 Konvolüsyon Sinir Ağları (CNN).....	39
5.3 CNN Modelleri :	47
5.3.1 Alex Net.....	48
5.3.2.ZF Net.....	49
5.3.3 GoogleNet.....	50
5.3.4.VGG Net.....	51
5.3.5.ReSNet.....	51
ALTINCI BÖLÜM	53

YAPILAN ÇALIŞMA.....	53
6.1. İskelet Temsili.....	53
6.2. Literatürdeki Çalışmalar.....	53
6.3. Çember Dolgu Oranı Yöntemi.....	55
6.3.1. İskelet Zenginleştirme Değeri e.....	57
YEDİNCİ BÖLÜM.....	59
DENEYSEL ÇALIŞMA.....	59
7.1 Veri Seti.....	59
7.2 Ön İşleme.....	60
7.3 x, y ve r Noktalarının Belirlenmesi.....	63
7.4 Boyut Azaltma:.....	63
7.5 CNN Modeli.....	65
7.5.1. Değerlendirme Kriterleri.....	66
SEKİZİNCİ BÖLÜM.....	73
SONUÇ VE ÖNERİLER.....	73
KAYNAKÇA.....	75

TABLolar LİSTESİ

Tablo 4.1: Yapay sinir ađına dayalı OCR uygulamaları.	33
Tablo 7.1: Görüntüleme ön işleme uygulanmadan yapılan eğitimin test sonuçları.....	68
Tablo 7.2: İskelet görüntüleriyle yapılan eğitimin test sonuçları.....	69
Tablo 7.3: X ve y noktaları ile yapılan eğitimin test sonuçları.	70
Tablo 7.4: X, y ve r noktaları ile yapılan eğitimin test sonuçları.	70
Tablo7.5: İç-dış sınır noktalarından elde edilen yeni görüntülerle yapılan eğitimin test sonuçları.....	72



ŞEKİLLER LİSTESİ

Şekil 3.1: Nesne Tanıma ve algılama aşamaları.....	15
Şekil 3.2: Gürültülü ve ön işlem uygulanmış görüntü.....	16
Şekil 3.3: SIFT vektörlerinin görüntü üzerinde gösterimi.....	18
Şekil 3.4: SURF vektörlerinin görüntü üzerinde gösterimi.....	19
Şekil 3.5: Orijinal görüntü ve HOG vektör görüntüsü.....	20
Şekil 3.6: Kümeleme ve Sınıflandırma işlemleri.....	21
Şekil 3.7: Farklı tanıma stratejileri.....	22
Şekil 4.1: OCR temel yapısı.....	24
Şekil 4.2: Gürültülü ve gürültüsü azaltılmış karakter görüntüsü.....	25
Şekil 5.1: AI, ML ve DL arasındaki ilişki.....	38
Şekil 5.2: CNN genel mimarisi.....	39
Şekil 5.3: Bir CNN ağında genişlik ve derinlik kavramları.....	40
Şekil 5.4: Konvolüsyon işlemi.....	42
Şekil 5.5: Sigmoid işlevi grafiği.....	43
Şekil 5.6: Çıkış verisi üzerindeki ReLu etkisi.....	44
Şekil 5.7: Maksimum havuzlama işlemi.....	46
Şekil 5.8: Ortalama havuzlama işlemi.....	46
Şekil 5.9: Orijinal ağ ve DropOut katmanı uygulanmış ağ yapısı.....	47
Şekil 5.10: AlexNet Mimarisi.....	48
Şekil 5.11: ReLu fonksiyonunun CIFAR-10 veri setindeki etkisi.....	49
Şekil 5.12: ZF Net Mimarisi.....	50
Şekil 5.13: GoogLeNet mimarisi.....	50
Şekil 5.14: VGGNet mimarisi.....	51
Şekil 5.15: Residual blokları.....	52
Şekil 6.1: Anahtar resmi üzerinde maksimum yarıçap gösterimi (Şirin ve Demirci, 2013).....	56
Şekil 6.2: A,B ve C iskelet noktaları olan yaprak şeklinin iskelet temsili. (Şirin, 2016).....	57
Şekil 7.1: Veri setindeki karakter ve harflerin örnek görüntüleri.....	59
Şekil 7.2: Arka plan ve karakter yazı renkleri değiştirilen görüntü örnekleri.....	60
Şekil 7.3: İskelet temsilleri oluşturulan görüntü örnekleri.....	61
Şekil 7.4: Yapılan düzeltmeler sonrasında elde edilen görüntü örnekleri.....	61

Şekil 7.5: Bounding box işlemi uygulanan örnek görüntüler.....	62
Şekil 7.6: Görüntülere uygulanan tüm işlemlerin adım adım gösterilmesi.....	63
Şekil 7.7: CNN mimarisi katman bilgileri.....	66
Şekil7.8: Veri setindeki görüntülere ön işlem uygulanmadan yapılan eğitim sonuçlarının doğruluk ve kayıp grafikleri.....	67
Şekil7.9: İskelet temsili oluşturulan görüntülerle yapılan eğitim sonuçlarının doğruluk ve kayıp grafikleri.....	68
Şekil 7.10: X ve y noktaları ile yapılan eğitimin doğruluk ve kayıp grafikleri.....	69
Şekil 7.11: X, y ve r noktaları ile yapılan eğitimin doğruluk ve kayıp grafikleri.....	70
Şekil 7.12: İç ve dış sınır noktalarından elde edilen görüntü örnekleri.....	71
Şekil 7.13: İç-dış sınır noktalarından elde edilen yeni görüntülerle yapılan eğitim sonuçlarının doğruluk ve kayıp grafikleri.....	71

KISALTMALAR

AI	: Yapay Zeka (Artificial Intelligence)
ANN	: Yapay Sinir Ağı (Artificial Neural Network)
ANPR	: Otomatik Plaka Tanıma (Automatic Number Plate Recognition)
CNN	: Konvolüsyonel Sinir Ağı (Convolutional Neural Network)
CV	: Bilgisayarlı Görü (Computer Vision)
DL	: Derin Öğrenme (Deep Learning)
GPU	: Grafik İşlemci Birimi (Graphics Processing Unit)
GT	: Temel Gerçek (Ground True)
HOG	: Yönlendirilmiş Gradyan Histogramı (Histogram of Oriented Gradients)
ILSVRC	: Büyük Ölçekli Görsel Tanıma Yarışması (Imagenet Large Scale Visual Recognition Challenge)
IOT	: Nesnelerin İnterneti (Internet of Things)
LSTM	: Uzun Kısa Süreli Hafıza (Long Short Term Memory)
ML	: Makine Öğrenmesi (Machine Learning)
NLP	: Doğal Dil İşleme (Naturel Language Processing)
OCR	: Optik Karakter Tanıma (Optical Character Recognition)
OR	: Nesne Tanıma (Object Recognition)
PCA	: Temel Bileşenler Analizi (Principal Components Analysis)
RELU	: Düzleştirilmiş Doğrusal Birim Katmanı (Rectified Linear Units Layer)
SIFT	: ÖlçekBağımsız Öznitelik Dönüşümü (Scale-Invariant Feature Transform)
SR	: Süper Çözünürlük (Super Resolution)
SURF	: Hızlandırılmış Dayanıklı Öznitelikleri (Speeded-Up Robust Features)

BİRİNCİ BÖLÜM

GİRİŞ

Günümüzde gelişmekte olan teknolojiyle birlikte insanlar, hayatlarında en fazla ihtiyaç duydukları önemli bir araç olan bilgiyi almakta, kullanmakta ve başkalarıyla paylaşmaktadır. Kullanılan bilgi insanın okuyabildiği ve anlayabileceği bir belge veya doküman biçimde olduğunda bu belgeleri elektronik bir ortamda saklamak ve bilgisayarın doğrudan anlamasını sağlamak mümkün olamamaktadır. Bu gibi problemlere çözüm olarak optik karakter tanıma teknolojisi kullanılmaktadır. Birçok tanıma teknolojileri ile karakterler bilgisayar ortamında kullanılarak tanıma işlemi gerçekleştirilmektedir.

OCR, "Optical Character Recognition" (optik karakter tanıma) teriminin kısaltmasıdır ve taranmış belgeleri, PDF dosyalarını veya dijital kamerayla çekilmiş görüntüler gibi farklı türdeki belgeleri düzenlenebilen ve aranılabilen verilere dönüştürmenize imkan sağlayan bir teknolojidir. Basılı bir sayfa üzerindeki doküman veya karakterin klavye aracılığıyla yazılmasına gerek duyulmadan, kısa bir zaman içinde elektronik bir metin dosyası formatına dönüştürülmesi anlamına gelmektedir.

1950'lerden bu yana çok aktif bir araştırma alanı olan OCR, otomatik örüntü tanımanın en başarılı uygulamalarından birisidir (Trier, Jain ve Taxt, 1996). Amacı, taranmış, kaydedilmiş veya resimdeki makine baskısı veya el yazısı metinlerin, rakamların, harflerin, sembollerin resimlerini sınıflandırmak ve istenilen formatta bir tanım oluşturmaktır. Optik karakter tanıma, sağlık, bankacılık, sipariş işlemleri, tıbbi istekler, hukuk, yüz tanıma ve parmak izi okuma gibi birçok uygulama alanında kullanılabilir. Ayrıca son zamanlarda kullanımı artan elektronik doküman yönetimi sistemleri, hem iş dünyasında hem de devlet yönetim merkezlerinde OCR teknolojilerinden yararlanmaktadır. Ancak uygulama alanlarından kaynaklanan birçok sorun bulunmaktadır. Örneğin, karakter kümesinin sabit bir konumda ve kameranın önünde tam olarak gösterileceği garantisizdir. Aynı zamanda, yazdırılan tüm karakterler aynı boyutta olmayabilir. Bu dezavantajların üstesinden gelmek için, OCR sisteminin konumlarından, yönlerinden ve boyutlarından bağımsız karakterleri tanıma yeteneğine sahip olması gerekmektedir.

OCR sisteminin genel yapısı ön işleme, özellik çıkarımı, tanıma ünitelerinden oluşur. Özellik çıkarma yönteminin seçimi muhtemelen yüksek tanıma performansının elde

edilmesinde en önemli faktördür (A.g.e.). Literatürde çok geniş bir özellik çıkarma yöntemleri yelpazesinin mevcut olduğu göz önüne alındığında, belirli bir uygulama için uygun bir yonteme karar vermek oldukça önemlidir. Bununla birlikte, hangi yontemin daha iyi performans gösterdiği, kullanılan veri setine bağlıdır. Ek olarak, farklı özellik tiplerinin farklı sınıflandırma tiplerine ihtiyaç duyabilmesinden dolayı OCR sistemi performansı kullanılan sınıflandırma türüyle de doğrudan ilişkilidir.

Literatürde açıklanan birçok özellik çıkarma tekniklerinden en yaygın olanları, Fourier tanımlayıcıları (Cai ve Lui, 2001; Chen ve Bui, 1999; Zahn ve Roskies, 1972), dalgacık tanımlayıcıları (Cai ve Lui, 2001; Cvetkovic, Ubeyli ve Cosic, 2008; Tang, vd., 1996), zernike momentleri (Khotanzad ve Hong, 1988), bağımsız bileşen analizi (Ozawa ve Kotani, 2000), ve bölgelendirme (zoning) (Freitas ve Oliveira, 2007) şeklinde sıralanabilir. Bununla birlikte, her özellik çıkarma yonteminin etkinliği, özel tanıma sorununa ve mevcut verilere bağlı olarak değişiklik gösterir. Ek olarak, literatürde bildirilen özellik çıkarma yöntemlerinin performansını karşılaştırmak kolay değildir, çünkü farklı veri setlerine dayanmaktadırlar (Trier, Jain ve Taxt, 1996). Bu nedenler, belirli bir uygulama için en iyi özellik çıkarma yonteminin zorlu görev seçimini yapar.

Bu tezde İngilizce karakterleri tanıma için bir çalışma ele alınmıştır. Kullanılan veri setinde bulunan karakter resimlerine, görüntü işleme tekniklerinin uygulanmasından sonra iskelet temsili ile bağlantılı olan Çember Dolgu Oranı yonteminin kullanılmasıyla öznitelik değerleri elde edilmiştir. Oluşturulan CNN (Convolutional Neural Network) modelinde ilk olarak sadece karakter görüntüleri ve iskelet temsillerinden oluşan görüntülerle eğitim yapılmıştır. Daha sonra bu öznitelik değerlerine uygulanan PCA işlemi sonucunda elde edilen noktaların görsel olarak kullanılması için tekrar oluşturulan görüntülerle eğitim yapılmış, kullanılan yöntem ve oluşturulan CNN modelinin, tanıma oranları üzerindeki etkisinin gözlemlenmesi önerilmiştir.

1.1 Amaç

Bu tezde gerçekleştirilen çalışmanın amacı, yukarıdaki bölümde anlatılan OCR sistemlerinde daha önceden yapılan uygulamalarda kullanılan yöntemlerden farklı bir yöntemle karakter tanıma yapabilmektir. OCR sistemlerinde kullanılan yöntemler üzerinde hem doğruluk oranları hem de güvenilirlik açısından gerekli literatür

taraması ve çeşitli araştırmalar yapılmış, elde edilen sonuçlar değerlendirilmiştir. El yazısı ve farklı yazı tiplerinde yazılmış karakterlerin tanınması, karakterin yazıldığı çok çeşitli seçenekler içerdiğinden, daha doğru tanıma oranlarına ulaşabilmek adına iskelet temsiline dayanan bir yöntem ile CNN modeli birlikte kullanılmıştır. Çalışma sonucunda OCR sistemleri için zenginleştirilmiş iskelet tabanlı bir yöntem olan Çember Dolgu Oranı yöntemi ile elde edilen noktalarla tekrar çizdirilen görüntülerin, oluşturulan CNN modelinde eğitilmesiyle bir doğruluk oranı hesaplanmaya çalışılmıştır.

Yapılan çalışmada ilk olarak kullanılan CNN modelinde yalnızca görüntülerle eğitim yapılarak elde edilen doğruluk oranı ile görüntüler üzerinde Çember Dolgu Oranı yöntemi sonucunda elde edilen noktalarla çizdirilen görüntülerin, aynı CNN modelinde eğitilmesiyle ulaşılan doğruluk oranları karşılaştırılmıştır.

1.2 Kapsam

Sunulan çalışma karakter tanıma kavramını ve karakter görüntüsünün sayısallaştırılmasından karakterin tanınmasına kadar tüm adımları kapsamaktadır. Karakter görüntülerinden elde edilen iskelet temsillerinin zenginleştirilmesini içeren bir yöntem ve CNN modeli kullanılarak tanıma işleminin gerçekleştirilmesi önerilmiştir. Kullanılan yöntem ve model doğruluk, dayanıklılık, uygulanabilirlik ve güvenilirlik açısından yapılan deneylerle gözlemlenmiştir.

1.3 Katkılar

Bu tezde karakter tanıma alanı için karakter görüntüleri, oluşturulan CNN modelinde eğitilerek ve görüntüler üzerinde iskelet temsillerinin zenginleştirilmesine dayanan yöntem uygulanması sonucunda elde edilen noktaların aynı modelde eğitilmesi sonucunda doğruluk oranlarına ulaşılmıştır.

Yapılan çalışmanın sağladığı katkılar aşağıdaki gibidir:

- Literatür tarama bölümünde anlatılan daha önceki çalışmalarda kullanılmış OCR sistemlerindeki yöntem ve modellere göre daha farklı algoritmalar kullanılmıştır.
- Çember Dolgu Oranı yönteminin veri seti içinde bulunan karakter görüntüleri üzerine uygulanması doğrultusunda elde edilecek olan noktaların görsel olarak kullanılabilmesi için PCA ile boyut azaltma işlemi uygulanmış ve bu

işlem sonrasında oluşan noktalar ile yeni görüntüler elde edilmiştir. Yeni görüntülerin, eğitilecek CNN modelinden sonuçlanacak doğruluk oranına olan etkisi açıklanmıştır.

- Farklı yazı tiplerinde ve el yazısı biçimde olan karakter görüntülerinden oluşan veri seti üzerinde uygulanan yöntem yapılan deneylerle birlikte değerlendirilmiştir.
- Yapılan deneyler sonucunda ortaya çıkan doğruluk oranları, performans, süre, güvenilirlik özellikleri açısından karşılaştırılarak, oluşturulan model ve kullanılan yöntemin etkileri değerlendirilmiştir.

Yukarıda ele alınanlara dayanarak, bu tez aşağıdaki bölümlerden oluşmaktadır:

Bölüm 2’de üzerine çalışılan konuyla ilgili literatürde yer alan benzer çalışmalar hakkında ayrıntılı bilgiler verilmiştir.

Bölüm 3’te OCR alanını içinde barındıran Nesne Tanıma (OR) alanı anlatılmış ve uygulanan aşamaları açıklanmıştır.

Bölüm 4’te OCR ile ilgili kapsamlı bilgiler verilir, tarihçesi, aşamaları, kullanım alanları ve zorlukları gibi bölümlerde detaylı olarak ele alınmıştır.

Bölüm 5’te Makine Öğrenmesi (ML) ve Derin Öğrenme (DL) konuları hakkında bilgiler verilmiş, CNN’lerin yapısı ve çeşitleri anlatılmıştır.

Bölüm 6’da yapılan çalışmanın içeriğinde bulunan ve deneysel çalışmalarda kullanılan yöntem tanıtılarak detaylı bilgiler verilmiştir.

Bölüm 7’de çalışmanın performansını gözlemleyebilmek için yapılan deneysel çalışmalar anlatılmış ve elde edilen bulgular aktarılmıştır.

Son olarak Bölüm 8’de ise ulaşılan sonuçlar açıklanarak gelecekte yapılabilecek çalışmalar önerilmiştir.

İKİNCİ BÖLÜM

LİTERATÜR TARAMASI

Bilindiği üzere, OCR alanında birçok farklı yöntemleri kullanan farklı çalışmalar yapılmıştır. Nitekim bazı çalışmalar bizim çalışmamız için yararlı olabilecek makine öğrenmesine ve diğer alanlara ait diğer algoritmaları ve yöntemleri kullanmaktadır. Bu bölümde, yapılan literatür taraması ile konumuzla ilgili daha önce yapılmış bazı çalışmalar anlatılmıştır.

Makine Öğrenmesinde el yazısı, rakam tanıma sorunu üzerine birçok çalışma ve algoritmalar geliştirilmiştir. Önerilen ML algoritmalarının çoğu, algoritmanın çalıştığını göstermek için problem üzerinde test edilir. MNIST (LeCun, Cortes ve Burges, 1998) veri seti literatürde birçok kez kullanılan ve iyi bilinen bir veri setidir. Derin öğrenme mimarilerin kullanımı ile veri seti üzerindeki test hata yüzdeleri % 1'in altına düşmüştür.

Düzenleme kullanarak ML algoritmalarının eğitimi aktif bir araştırma alanıdır. Düzenlemeye en temel yaklaşım ağırlık azaltılmasıdır (Krogh ve Hertz, 1992). Buradaki katmanlar arasındaki bağlantı parametreleri kendi normlarıyla düzenlenir. Bağlantı parametresinin ağırlığının amaç fonksiyonuna eklenmesi, optimizasyon sürecini, bağlantılardaki gürültünün azaltıldığı ve çözümü mümkün olduğu kadar küçük olan, daha genelleştirilebilir bir çözüme yönlendirmektedir (Bach, vd.,2012).

Literatürde Osmanlı-Türk dilleri üzerine karakter tanıma çalışmaları da yapılmıştır. Ataer ve Duygulu (2007) tarafından geliştirilen, Yapay Sinir Ağı (YSA) kullanarak Osmanlıca karakterleri tanıyan sistemdir ve % 85,5 sınıflandırma doğruluğuna ulaşmıştır.

Osmanlıca-Latince transkripsiyon üzerine az sayıda çalışma bulunmaktadır. Andrews, vd.,(2010) Osmanlıca-Türkçe metinlerden Latin kökenli bir transkripsiyon sistemi geliştirmişlerdir.

Vural, vd., (2004) tarafından yapılan çalışmada, tablet bilgisayar arayüzü kullanılarak geliştirilen Türkçe için el yazısı metin tanıma sistemi sunulmaktadır. Her kaydedilen nokta için el yazısı verilerinden birkaç dinamik özellik çıkarılır. Tablet bilgisayar tarafından toplanan el yazısı verileri kelime ve harf tabanlı HMM modelleri (Hidden Markov Model) ile eğitilmiştir. HMM, özelliklerin mevcut

duruma veya geiře baėlı olarak bir olasılık daėılımı tarafından retildiėi varsayılan sınırlı sayıda durumla tanımlanmaktadır. Kelime modelleri kullanıldığında, veri setindeki her kelime iin bir kelime, o kelimenin birkaç eėitim rneėi kullanılarak eėitilir. Harf modelleri kullanıldığında, veri setindeki her kelime bař harf modellerine gre etiketlenmektedir. Daha sonra, o kelimenin harflerini temsil eden harf modellerini art arda sıralayarak bir kelime modeli oluřturulur. Gerekleřtirilen alıřma sonunda, 1000 kelimelik bir veri setindeki el yazısıyla yazılan szckler %94 oranında doėru tanınmıřtır.

OCR sistemlerinde optimum veya azaltılmıř hesaplama karmařıklıėı ile daha yksek tanıma oranlarına ulařılabilmektedir. Karakter tanıma srecine dahil olan hesaplama karmařıklıklarını bir font stili veya byklėnden baėımsız olan OCR sistemleri ile azaltmak olduka nemlidir. Bharath ve Rani (2017) tarafından karakter grntsnn sol, saė ve apraz ynlerine gre mesafe zellikleri kullanılarak, grntye gre font stilinin sınıflandırılması iin bir teknik nerilmiřtir. Bu alıřmanın temel amacı, genel OCR sistemlerinin karmařıklıėını font stili tanıma yntemiyle azaltmaktır. Karakter grntlerinin mesafe profili zellikleri, bir destek vektr makine sınıflandırıcısıyla eėitilmiř ve deneyler, eřitli belge kaynaklarından elde edilen karakter grntleri ile oluřturulmuř 10 byk harfli yazı tipi stilinin yanı sıra kk harflerden oluřan eėitim, 5 farklı font stilinden oluřan test veri setleri zerinde gerekleřtirilmiřtir. alıřmada nerilen algoritma ile % 80 doėruluk oranına ulařılmıřtır.

OCR, veri tabanı kullanarak taranan tm ėeleri veri tabanından karřılařtırıp tanıma iřlemini daha kolay hale getirir. Harneet ve Anmol'ın (2018) yapmıř olduėu alıřmada, bir OCR sisteminin farklı ařamalarda alıřması aıklanmıřtır. Bu alıřma, geleneksel sistemin eřitli dezavantajlarını bulmaya yardımcı olur ayrıca eksikliklerin nasıl giderilebileceėini ve gelecekte daha iyi OCR sistemlerinin nasıl saėlanabileceėini de anlatmaktadır.

Gnmzde, grme engelli insanlar mevcut teknolojileri kullanarak basılı metne eriřirken sayısız sorunlarla karřılařması, bu sorunlara zm olabilecek cihazların geliřtirilmesine duyulan ihtiyaı da arttırmaktadır. Kitapların dijitalleřtirilmesinden dolayı, endstri alanlarında ve arařtırma laboratuvarlarında gl bir belge analiz sistemi oluřturmak iin birok giriřim bulunmaktadır. Bunlardan biri olan Bhashitha, De Zoysa, vd.,(2018) tarafından geliřtirilmiř, mevcut sistemlerdeki sorunları zen

ve Sinhala, Tamil ve İngilizce dilleri için OCR ve yazılı metni ses çıkıtısına dönüştürme (Text to Speech) içeren bir Android tabanlı mobil uygulamadır. Önerilen sistemin yapılması için, kullanıcının cep telefon kamerası kullanarak basılı bir belgeyi optik görüntü olarak edinmesi gerekir. Belgenin açılı görünüşünden kaynaklanabilecek görüntü üzerindeki eğriltme faktörü, OCR doğruluğunu büyük ölçüde azaltmaktadır. Bu nedenle, görüntü eğriltme algılamasını yaptıktan sonra, görüntüyü tanınan kelimelerin harflerini temsil eden karakter akışlarına dönüştürmek için OCR sistemine geçilir. Son olarak, metin içeriğini bir ses çıkıtısına dönüştürmek amacıyla dönüştürülmüş metin çıkıtısına Text to Speech sistemi ile erişilir. Ek olarak, sayfalar arasında gezinmek için ses yardım sistemi de bulunmaktadır. Görme engelliler için yapılmış diğer mevcut sistemlere göre daha kolay, taşınabilir ve daha hızlı bir çözüm olmuştur.

Görme engelliler için yapılan bir diğer çalışmada Türkçe içindir. Bastan, Kandemir ve Cantürk (2016) tarafından gerçek zamanlı çalışma için hızlı gradyan tabanlı çok ölçekli metin algılama algoritması ve karakter tanıma için Tesseract OCR kullanarak Türkçe sahne ve kitap metni okuyabilen bir mobil sistem geliştirilmiştir. Yapılan sistem doğruluk, çalışma süresi, açısından değerlendirilerek diğer benzer yeni sistemlerle karşılaştırıldığında daha hızlı, bir mobil cihazda çalışabilir olduğu ve diğer dillere uygulanabilirliği gözlemlenmiştir.

OCR uygulamalarındaki doğruluk oranını artırmaya yönelik olan çalışmalardan bir diğerini ise Mor ve Wolf (2018) gerçekleştirmişlerdir. Bir OCR sisteminde çıkan hatalara göre sistem için bir güvenilirlik ölçümü amaçlamışlardır. Temelde çok sınıflı bir sınıflandırma problemi olan, sözlük (lexicon) tabanlı OCR sistemleri, doğruluğu arttırmak için genellikle sözlüğü dikkate alan yöntemler kullanır. Bu çalışmada, bir OCR sisteminin sözlükle ilgili herhangi bir bilgi gerektirmeyen ve sözlüğü kullanmayan bir durumda da yüksek doğruluğu yakalayabilmesi için iki teknik sunulmuştur. İlk teknik OCR modelinin ilk ve ikinci en muhtemel okumalara atadığı olasılıkları, bir girdi görüntüsü olarak kullanıp hata tahmin etmede etkili olmuştur. İkinci teknik ise OCR modeline girdi verilip verilmediğini veya numunenin reddedilmesi gerekip gerekmediğini doğrudan tahmin ederek hata tahmini yapmaktadır. Önerilen teknikler ile gerçekleştirilen deneyler sonucunda güvenilirlik ölçümünün OCR modellerindeki hataları tespit etmede oldukça etkili olduğu gösterilmiştir.

Hu, vd., (2018) tarafından yapılan çalışmada ise Wikipedia, Flickr30k gibi mevcut veri kümelerinin içerik çeşitliliğinden yoksun, sadece önceden tanımlanmış bazı sınıfları kapsayan, içerisinde bulunan metinlerin iyi düzenlenmiş bir dilde yazılmış olması nedeniyle gerçekçi uygulamalarla tutarsızlığa yol açması gibi dezavantajların üstesinden gelmek için Twitter100k veri kümesi oluşturulmuştur. Bu veri setinde Twitter'dan rastgele spor, mimarlık, gıda, hayvan, haber, bitki ve benzeri gibi çok çeşitli alanlardan alınan 100.000 görüntü-metin çifti bulunur. Tweetler, internet kullanıcıları tarafından gayri resmi dil ile yazıldığından dolayı Twitter100k içindeki metinler dilbilgisi ve sözcük dağarcığındaki diğer veri kümelerinden ayrılmaktadır. Oluşturulan veri setinin özelliklerini göz önüne alarak, görüntülerin OCR metinlerine dayanarak OCR sistemleri için performans iyileştirmesi önermişlerdir. Gelecekteki çalışmalarda, bu veri seti için daha iyi bir değerlendirme protokolü tasarlamayı ve görüntüler ile metinler arasındaki ilişkiyi, düşünce ve duygular gibi yüksek düzeyli anlamsal perspektiflerden ele almayı planlamaktadırlar.

Eski basım kitaplardaki metinler ile eğitilmiş OCRopus modellerinden elde edilen OCR metni için karakter hata oranlarını önemli ölçüde azaltan bir yöntem sunulan Reul, vd.,'nin (2018) yapmış oldukları çalışmada, çapraz eğitim ve güvene dayalı oylamanın kombinasyonunu kullanan bir yöntem önerilmiştir. Özellikle 15. ve 16. yüzyıldaki eski kitaplarda kullanılan farklı yazı tipleri arasındaki yüksek değişkenlik nedeniyle, model eğitimi için belirli miktarda temel gerçeğe (GT) ihtiyaç duyulur. Çalışmadaki amaç, farklı modelleri eğiterek ve bunları birleştirmek için belirli bir miktarda GT ile OCR doğruluğunu iyileştirmektir. Yedi adet eski basım kitap üzerinde yapılan deneylerde önerilen yöntem ile hata miktarını % 50 ve daha fazla oranda azaltılmıştır. Sonuçlar, önerilen yöntemin makul miktarda GT ile çalıştığını ve bu nedenle eski basım bir kitapta OCR işlemi sırasında tüm aşamalara uygulanabilir olduğunu göstermiştir. Bu çalışmada erken basılmış kitaplara odaklanmasına rağmen, sunulan fikirler herhangi bir zamanda basılmış diğer metinler için de uygulanabiliridir.

Video OCR, büyük dijital video haber arşivindeki ilgi alanlarını otomatik olarak altyazıların ve açıklamaların seçilip çıkarılması ve okunması yoluyla bulmaya yardımcı olabilecek bir tekniktir. Haber başlıkları genellikle sunulan videodaki insanların adları, yerlerin veya nesnelerin açıklamaları hakkında önemli bir arama bilgisi sağlar. Sato, vd.,(1998) videolar için düşük çözünürlüklü karakterler ve son

derece karmaşık arka planlar olmak üzere iki farklı karakter tanıma sorunu ele almışlardır. Bu problemleri çözmek için çoklu çerçeve entegrasyonu ve dört filtre kombinasyonu kullanmışlardır. Karakterleri bölümlere ayırma, tanıma tabanlı bir bölümlendirme yöntemiyle yapılır ve bölümlenmeyi iyileştirmek için ortalama karakter tanıma sonuçları kullanılır. Haber videosunda OCR yapmak ve sonuçlarını diğer video anlama teknikleriyle birleştirmek, haber videolarının genel olarak anlaşılmasını geliştirecektir. Gerçek haber verilerinde, aynı veriler üzerinde geleneksel OCR teknikleri kullanılarak elde edilenlerin neredeyse iki katı olan% 83,5 doğru karakter tanıma ve% 65,2 doğru kelime tanıma oranlarına ulaşmışlardır.

Metin belgelerinin yazdırılması ve taranması, modellenebilecek karakterlerde bozulmalara neden olur. Karakterleri yazdırma ve tarama işlemleriyle ortaya çıkan bozulmaları yöneten parametrelerin belirli kombinasyonları, bozulmuş karakterlerin benzer bir görünüme sahip olduğu şekilde etkilerken, diğer bozulmalar karakterleri çok farklı bir görünümde bırakır. Bir eğitim setiyle daha yakından eşleşen bir test setinin eğitim setiyle daha az uyumlu olandan daha yüksek hassasiyetle tanınacağı bilinmektedir. Aynı şekilde, sınıflayıcılar daha düşük varyansa sahip veri setlerinde daha iyi performans gösterme eğilimindedir. Smith ve Andersen (2005) tarafından yapılan çalışmada, bozulmuş karakter grupları arasındaki benzerliği tanımlamak için resmi bir yazıcı / tarayıcı bozulma modeli kullanan analitik bir yöntem araştırılmıştır. Bu benzerliğin, model odaklı bir eğitim seti verisi seçimi yoluyla bir sınıflayıcının tanınma doğruluğunu arttırdığı gösterilmiştir. Çalışmanın devamı olarak bozulma alanını bölmenin faydalarını daha derinlemesine araştırıp genişletmeyi amaçlamaktadırlar.

LSTM (Long Short Term Memory) ağları, literatürdeki el yazısı tanıma çalışmalarında önemli sonuçlar elde etmiş ve Breuel, vd., (2013) yaptıkları çalışmada makineyle basılan Latince ve Fraktur tanıma sorunu uygulanmasını çift yönlü LSTM ağlarını kullanarak gerçekleştirmişlerdir. OCR için, yeni bir algoritma ile birleştirilmiş tek boyutlu bir tekrarlayan ağ kullanmışlardır. Eğitim ve test için Washington Üniversitesi (UW3) veri seti ile bilim dergilerinden ve diğer ortak kaynaklardan 1.600 sayfa doküman görüntüsü elde etmişlerdir. Eğitim setinde 95.338 metin satırı ve test setinde 1.020 metin satırı içeren rastgele veri setleri oluşturulmuştur. LSTM mimarisi, İngilizce metin üzerinde karakter düzeyinde % 0.6 hata vermiştir. Temel olarak bozulmuş Fraktur veri seti eğitim ve test setlerine

bölündüğünde, sistem % 1.64 hata oranına ulaşmıştır. Fraktur'da basılan eğitim setinin bir parçası olmayan belirli kitaplarda, sistem % 0.15 (Fontane) ve % 1.47 (Ersch-Gruber) hata oranlarına ulaşmıştır. Bu tanıma doğrulukları, herhangi bir dil modellemesi veya herhangi bir işlem sonrası teknik kullanılmadan bulunmuştur.

OCR, Otomatik Plaka Numara Tanıma Sisteminde (ANPR) son aşamadır ve bu aşamada, plaka görüntüsü üzerindeki plaka karakterleri kodlanmış metinlere dönüştürülmektedir. Zhai, Bensaali ve Sotudeh (2012) tarafından gerçekleştirilen çalışmada, ANPR uygulaması için iki katmanlı ileri beslemeli bir sinir ağı tabanlı OCR algoritması sunulmuştur. Sunulan algoritmanın performansını test etmek için 3700 tane ikili karakter görüntülerinden oluşan bir veri tabanı kullanılmıştır. Elde edilen sonuçlar, önerilen algoritmanın bir ANPR sisteminin gerçek zamanlı gereksinimlerini karşılayabildiğini ve ortalama olarak % 97.3 başarılı tanıma oranı ile 8.4ms'de bir karakter görüntüsünü işleyebildiğini göstermiştir.

Plaka Tanıma Sistemlerine bir başka örnek çalışma ise Patel C ve Patel D (2012) tarafından yapılmıştır. Bu çalışmada OCR araçları olan Tesseract ve Transym kullanılarak plakalar üzerindeki sayılar çıkarılmaya çalışılmış ve bu iki araç arasında karşılaştırılma yapılmıştır. Çeşitli araç türlerinden 20 farklı tipte plaka görüntüleri 5 megapiksel kamera ile çekilerek elde edilmiştir. Tesseract renkli görüntüleri gri tonlamalı görüntülere dönüştürerek çalışmaktadır. Transym ise renkli görüntüleri gri tonlamalı görüntülere kaplayarak çalışır. Bu yüzden Transym'i kullanırken renkli resmi gri skalaya dönüştürülmesine gerek duyulmamaktadır. Tesseract ile Transym karşılaştırılmasında, Tesseract'ın veri setinde sadece % 47 hassasiyet sağlayan Transym'e kıyasla, renkli görüntü için % 61 ve gri tonlamalı görüntüler için % 70 doğruluk oranlarıyla daha iyi bir tanıma sağladığı gözlemlenmiştir. Sonuçlar doğrultusunda Tesseract'ın gri tonlamalı görüntülerde renkli görüntülere kıyasla daha iyi hassasiyet sağladığını gösterir. Renkli görüntülerin gri skalaya dönüştürülmesi ve Tesseract'a girdi olarak verilmesi durumunda, metin çıkarımının doğruluğunun arttığı gözlemlenmiştir. Görüntü işleme süreleri açısından yapılan karşılaştırmalar sonucunda ise Tesseract'ın renkli ve gri tonlu olan bir görüntüyü işlemek için sırasıyla ortalama 1 saniye ve 0,82 saniye, Transym'in ise bir görüntüyü işlemesi ortalama 6,75 saniye olarak ölçülmüştür. Bu sürelerle bakılarak Tesseract'ın, Transym'den daha hızlı olduğu açıklanmıştır.

Suzuki, vd., (2003) tarafından INFTY adlı matematiksel dokümanlar için bir OCR sistemi sunulmuştur. Eşzamanlı karakter tanıma ve iki tamamlayıcı tanıma sistemine dayalı matematik metin ayırma gibi bazı yeni teknikler daha iyi tanıma performansı için kullanılmıştır. Yaklaşık 500 sayfa matematiksel belge üzerinde yapılan çalışmalarda normal metinlerde % 99.44, matematiksel ifadelerde % 95.18 ve toplamda % 98.51 karakter tanıma oranları sonuçlarına ulaşılmıştır.

Sınıflandırıcı seçimi ve özellik çıkarımı için kullanılan teknik, OCR sisteminin tanıma doğruluğu ve yeteneğini değerlendirmek için önemli ana faktörlerdendir. Choudharya, Rishib ve Ahlawat (2013) tarafından ortaya konulan çalışmada İngilizce dilindeki el yazısı karakterlerinin tanınmasında ikilileştirme tekniği kullanılarak özellik çıkarımı işlemi gerçekleştirilmiştir. El yazısı karakter görüntülerinin tanınması, inceltme, ön ve arka plan gürültüsünü giderme, kırpma ve boyut normalleştirme gibi bazı ön işleme teknikleri kullanılıp, çok katmanlı ileri beslemeli yapay sinir ağı ile sınıflandırılma yapılmıştır. Yapılan deneyler sonucunda % 85.62'lik sınıflandırma doğruluğu elde edilmiştir.

Düşük çözünürlüklü görüntüler OCR performansını düşürebildiğinden, metin görüntüsü üzerindeki süper çözünürlük (SR), bilgisayarlı görme alanında giderek daha önemli bir sorun haline gelmiştir. Zhang, Liu ve Xiong (2017) tarafından yapılan çalışmada, genel görüntü SR ve metin görüntüsü SR için CNN önerilmiştir. Çalışmada, OCR'ı kolaylaştırmak için metin görüntüsü SR için CNN'i eğitirken yeni bir kayıp işlevi önerilmiş ve performansı daha da geliştirmek için model birleşimi yapılmıştır. Ayrıca, SR sırasında görüntü sınırlarını daraltmak için basit ama etkili bir görüntü dolgu yöntemi sunulmuştur. Deneysel sonuçlar ICDAR 2015 TextSR veri seti üzerinde uygulanmış ve % 78.10 oranında doğruluk elde edilmiştir.

Dijitalleşmiş belgelerin oluşturulması ve depolanması bilgi ve iletişim teknolojilerinde çok önemli bir yere sahiptir. Kişisel bilgi içeren dokümanların kötü niyetli davranışlarla sızdırabilir olmasından dolayı bireyler ve şirketler için bilgi kaybı sorunu oluşmaktadır. Lee, Song ve Won (2018) tarafından yapılan çalışmada, kişisel ve kurumsal bilgisayarlardaki bilgi sızıntılarını önceden önlemek, görüntü dosyalarında saklanan dokümanlar arasında kişisel bilgilerin varlığını daha etkili ve hızlı bir şekilde tespit etmek için bir yöntem önerilmektedir. OCR özelliklerini tanıma oranıyla sınıflandırarak ve algılama hızını artırmak için gereksiz olanları silerek kişisel bilgi algılama etkinliğini artırılmış ve sınıflandırılmış OCR

özelliklerinin referans frekansı kullanılarak tespit süresi azaltılmıştır. Kişisel bilgileri saptamak için mevcut yaklaşım, eğitim verilerini bir klasörde tutup, ayrı ayrı sınıflandırarak tek tek kontrol eden tanıma oranını temsil eden bir yöntem kullanılmıştır. Tanıma oranına göre, eğitim verilerinde önerilen yöntem, bunları tamamlanan, kullanılmış ve varsayılan olmak üzere üç klasöre ayırarak sınıflandırılmıştır. Eğitim verilerini sınıflandırma özelliklerini belirlemek için üç tanıma oranı kriterinin %70, %80 ve %90 olduğunu varsayarak en güvenilir kriteri belirlemek için bir test yapılmış ve sonuçlar %80'in en güvenilir tanıma oranı olduğunu göstermiştir. Ayrıca, tanıma oranının ve veri işleme hızının, bir görüntüye bağlı olarak eğitim verilerinin referans frekansını hesaplayarak ve frekansa dayalı eğitim verilerinin önceliğini belirlemek için bir yöntem ekleyerek mevcut yönteme göre daha iyi olduğu belirtilmiştir.

Avadesh ve Goyal (2018) tarafından yapılan çalışmada basılı antik el yazılarının tanınması için CNN tabanlı bir OCR önerilmiştir. Görüntüdeki harfleri tanımlamak için piksel yoğunlukları hesaplanmış ve kullanılan CNN modelinde % 93.32 doğruluk elde edilmiştir.

OCR için CNN'e dayanan bir yaklaşım önerilen Sharm, vd., (2018) tarafından yapılan çalışmada, öğrenciler tarafından özel bir formda doldurulmuş, İngiliz alfabesi ve rakamlarını içeren bir veri seti kullanılmıştır. Temel olarak bu OCR sisteminden, bir öğrencinin bilgilerini çıkarmak için yararlanılmaktadır. Çalışma sonucunda sayısal verilerin % 96.87'si ve alfabetik verilerin % 94.36'sı doğru bir şekilde tahmin edilmiştir.

Farklı metin türlerini algılamak için Khedkar, vd., (2020) tarafından yapılan çalışmada, görüntüdeki metnin dijital forma çevrilebilmesi için tek bir sözcüğü sınıflandırma amaçlanmaktadır. Kelimeleri doğrudan sınıflandırma ve karakter segmentasyonu bu görevi yerine getirmek için kullanılan iki temel yaklaşımdır. Kelimeleri doğru bir şekilde sınıflandırmak için CNN önerilmiştir. Farklı stil ve yazı tiplerine sahip görüntülerden oluşan veri setine segmentasyon işlemi uygulandıktan sonra CNN modelinde gerekli testler yapılarak % 83.3 doğruluk oranına ulaşılmıştır.

Ashiquzzaman ve Tushar (2017) tarafından yapılan çalışmada, Arapça rakam tanıma yöntemlerine kıyasla önemli ölçüde geliştirilmiş doğruluk gösteren, uygun aktivasyon fonksiyonu ve düzenleme katmanı kullanan derinlemesine sinir

ağları temel alan CNN modeli kullanılmıştır. Modelde gerçekleştirilen deneyler sonucunda % 97.4 oranında doğruluğa ulaşılmıştır.

Yin, vd.,'nin (2019) yapmış oldukları çalışmada, Nesnelerin İnterneti (IoT) uygulamalarında tanıma doğruluğunu geliştirmek için derin öğrenme destekli bir OCR tekniği önerilmiştir. Bu tekniği gerçekleştirmek için, Çince büyük harfli karakterlerin bir veri kümesi üzerinde CNN, VGG, ResNet ve CapsNet eğitilip test edilmiştir. Ayrıca, CNN'deki aktif olmayan nöronlar ve yoğun katmanlar silinmiştir. Yapılan deneyler sonucunda CNN'in % 97.70, VGG'in % 97.92, ResNet ve CapsNet'in sırasıyla % 99,17 ve % 99,38 doğruluğa ulaştığı gözlemlenmiştir.

Yapılan literatür taramasının sonucunda derin öğrenme ve makine öğrenimi teknikleri ile sinir ağlarının OCR ve bilgisayarlı görü uygulamalarında oldukça yaygın bir biçimde kullanıldıkları görülmüştür. Ayrıca yapılan çalışmalarda kullanılan yöntemin önemiyle birlikte, derin öğrenme tekniklerinin büyük veri setleri, performans ve elde edilen doğruluk oranlarına olan pozitif etkisi gözlemlenmiştir.

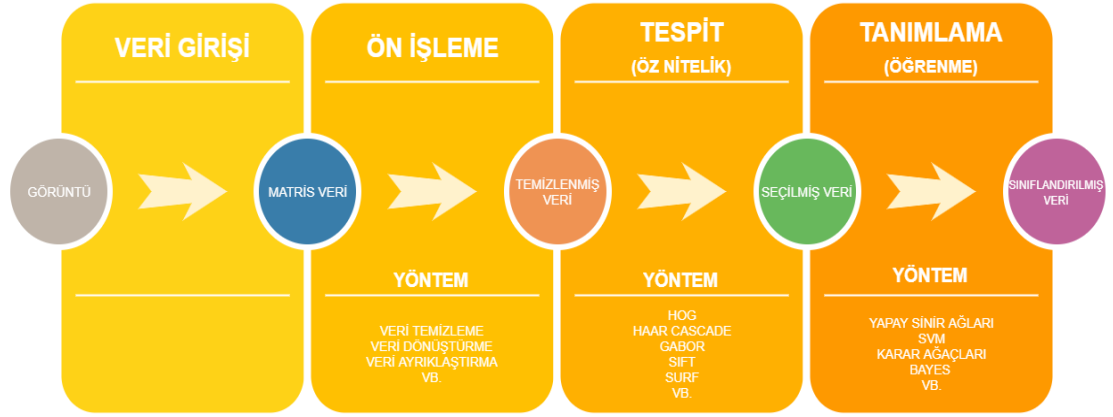
ÜÇÜNCÜ BÖLÜM

NESNE TANIMA

Üzerinde ölçüleme yapılabilen veya gözlenebilen bilgi nesne olarak tanımlanmaktadır. Nesne kavramı içerisine ses, görüntü, sinyal gibi değişkenler girmektedir. Günümüzde sayısız görüntü farklı formatlarda depolanmaktadır. Arama, tanımlama ve benzerlerini bulma işlemleri depolanan görüntüler üzerinde uygulanmaktadır. Ayrıca OR alanında gerçekleştirilen çalışmalara olan ilgi günden güne artmaya devam etmektedir.

Nesneyi algılayıp, belirlenen kriterler doğrultusunda tanımlama veya sınıflandırma işlemine OR denmektedir. Başka bir deyişle OR, dijital görüntülerde veya videolarda belirli insanlar, binalar veya arabalar gibi anlamsal nesnelere algılamaya yönelik, bilgisayarlı görme ve görüntü işleme ile ilgili bir bilgisayar teknolojisi. Bilgisayar vizyonundaki en eski alanlardan biri olup, hala birçok araştırmacı üzerinde çalışmalar yapmakta ve farklı algoritmaların geliştirilmesiyle dikkat çekmektedir. Paul Viola ve Michael Jones tarafından 2001 yılında sunulan Viola Jones algoritması OR ile ilgili devrim niteliğinde sayılabilecek bir çalışma olmuştur. Bu algoritma “Rapid Object Detection using a Boosted Cascade of Simple Features” başlıklı makale ile duyurulmuştur (Viola ve Jones, 2001). Devam eden süreçte birçok algoritma geliştirilmiş, yakın zamanda ise kullanılmaya başlayan GPU (Graphics Processing Unit) teknolojisi ile hız kazanan derin öğrenme sayesinde çok daha fazla doğruluk oranı ile tanımlama yapabilen yöntemler ortaya çıkarılmıştır.

Nesne tespit ve tanıma süreçlerini 5 ana başlık altında toplanarak anlatılmaktadır ve Şekil 3.1’de gösterilmiştir. Bunlar, veri girişi, veri ön işleme ve aşamaları, öznelik çıkarımı ve öz nitelik seçimi son olarak ise tanımlama olarak adlandırdığımız veri sınıflandırma aşamaları şeklindedir.



Şekil 3.1: Nesne Tanıma ve algılama aşamaları.

OR, görüntü algılama ve bilgisayarlı görme alanında birçok uygulamalara sahiptir. Bu uygulamalar tıp biliminde, savunma sanayisinde, trafikte, sismik hareket incelenmelerinde ve deprem tahminlerinde, ayrıca ses, el yazısı, parmak izi, yüz tanıma uygulamalarında, uydu aracılığıyla elde edilen resimlerin yorumlanması gibi birçok farklı alanlar üzerinde gerçekleştirilmektedir.

3.1 Veri Girişi

Bu aşamada, verinin sisteme girdi olarak verilmesi için gürültülerinden ayrıştırılması, istenilen formata getirilmesi gibi işlemlerin uygulanarak hazırlanması gerçekleştirilir. Nesne tespit veya sınıflandırma işlemi gerçekleştireceğimiz için kullanacağımız veriler doğal olarak görüntü olmaktadır. Dolayısıyla bir insan, otomobil veya ev istediğimiz veriye örnekken, daha fazla ve daha karmaşık nesnelerin birlikte bulunduğu görüntüler girdi olarak tanımlanabilir. Bu veri türlerine fotoğraf makineleri ile çekilen fotoğraflar, dijital ortamlarda oluşturulan resimler, videolar ve taranmış metinler örnek olarak verilebilir. Veri formatı, kullanılan teknoloji, dil, platform gibi birçok etkene bağlıdır fakat kullanılacak olan algoritmalar için geçerli format, matris haline getirilmiş görüntü pikselleridir.

3.2 Ön İşleme:

Ön işleme aşamasında veri, gürültülerden, tutarsız veya tam olmayan verilerden arındırılarak yapılacak olan işlemin amacına uygun hale getirilmesi için hazırlanır. Şekil 3.2’de (a) gürültülü bir görüntü ile (b) ön işlem yapılmış bir görüntü verilmiştir.



Şekil 3.2: Gürültülü ve ön işlem uygulanmış görüntü.

Bir nesne, birden çok varlık içeren bir sahnenin görüntülerinden tanınması gerektiği için, OR'nın karmaşıklığı çeşitli faktörlere bağlıdır. Bu faktörler aşağıdaki gibi sıralanabilir.

Sahne sabitliği: Kullanılan görüntüler genellikle aydınlatma, arka plan, model olarak kamera parametreleri ve bakış açısı gibi benzer koşullarda elde edilir. Sahne koşulları aynı nesnenin görüntülerini çarpıcı biçimde etkiler. Farklı sahne koşulları altında, farklı özellik dedektörlerinin performansı önemli ölçüde farklı olacaktır. Arka plan, diğer nesnelere ve aydınlatma, ne tür özelliklerin etkin ve güvenilir bir şekilde algılanabileceğini belirlemek için göz önünde bulundurulmalıdır.

Görüntü-model uzayları: Bazı uygulamalarda görüntü elde edilir ve üç boyutlu nesnelere iki boyutlu olarak kabul edilebilir. Modeller üç boyutlu ise ve perspektif efektleri göz ardı edilemezse, durum daha karmaşık hale gelir. Bu durumda özellikler iki boyutlu görüntü alanında algılanırken, nesnelerin modelleri üç boyutlu alanda olabilir. Dolayısıyla aynı üç boyutlu özellik, bir görüntüdeki farklı bir özellik olarak görünebilir. Bu durum nesnelerin hareketi nedeniyle dinamik görüntülerde de meydana gelebilmektedir.

Model veri tabanındaki nesne sayısı: Nesne sayısı çok küçük olduğunda hipotez oluşturma aşamasına ihtiyaç duyulmayabilir. Bir sıralı ayrıntılı eşleştirme kabul edilebilir. Hipotez oluşumu çok sayıda nesne için önemli hale gelir. OR için uygun özelliklerin seçilmesinde harcanan çaba miktarı, nesne sayısındaki artışla birlikte hızla artar.

Görüntüdeki nesne sayısı ve oklüzyon (occlusion) olasılığı: Eğer görüntü üzerinde yalnızca tek bir nesne varsa, nesne tamamen, anlaşılır bir biçimde görünebilir. Görüntüdeki nesne sayısı ile, oklüzyon olasılığı arasında doğru orantılı bir ilişki vardır. Oklüzyon, beklenen özelliklerin olmaması ve beklenmeyen özelliklerin üretilmesi ile sonuçlanır ve birçok temel görüntü hesaplamasında ciddi bir sorundur. Genel olarak tanıma görevindeki zorluk, görüntüdeki nesne sayısı ile artmaktadır. Görüntü bölümlendirmedeki zorluklar, görüntülerde çok sayıda cisim olmasından kaynaklanmaktadır.

3.3 Öznitelik Çıkarma

Öznitelik çıkarma aşamasında ön işlem uygulanan veri üzerinde daha önce belirlenmiş olan nesnenin elde edilme işlemi gerçekleştirilir. Başka bir deyişle, bir nesnenin önceden tanımlanan kriterleri ve özelliklerinin kullanılmasıyla görüntüler üzerinden otomatik olarak tespit edilerek, detaylarının elde edilmesi işlemidir.

OR için birçok özellik türü kullanılır. Çoğu özellik, görüntüdeki bölgelere veya sınırlara dayanır. Bir bölgenin veya kapalı bir sınırın, bir nesnenin parçası olan varlığa karşılık geldiği varsayılmaktadır. Yaygın olarak kullanılan özelliklerden bazıları aşağıdaki gibidir.

Genel özellikler: Genellikle görüntülerdeki bölgelerin, alan boyut, çevre ve tanımlayıcılar gibi bazı özellikleridir. Küresel özellikler, bir bölge içindeki tüm noktaları göz önüne alarak bir bölge için veya yalnızca bir bölge sınırındaki noktalar için edinilebilir. Her durumda amaç tüm noktaların, yerlerini, yoğunluk özelliklerini ve mekansal ilişkileri göz önünde bulundurarak elde edilen tanımlayıcıları bulmaktır.

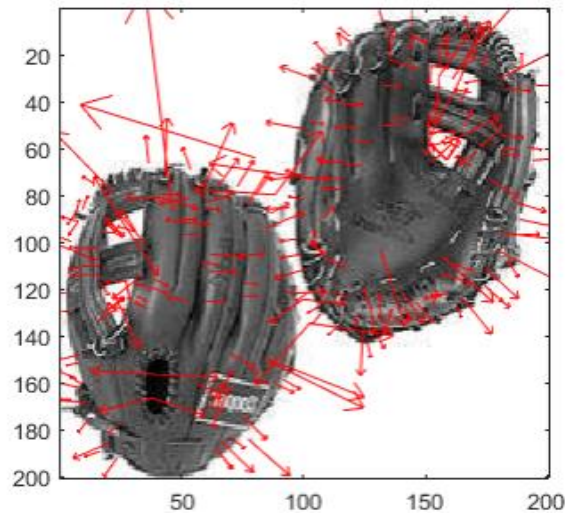
Yerel özellikler: Yoğunlukla bir nesnenin sınırındadır veya bir bölgenin ayırt edici küçük bir alanını temsil ederler. Yaygın olarak kullanılan bazı yerel özellikler eğrilik, sınır bölümleri ve köşelerdir. Eğrilik, bir sınırdaki eğrilik olabilir veya bir yüzey üzerinde hesaplanabilir. Yüzey, yoğunluklu bir yüzey veya 2, 3 boyutlu bir yüzey olabilir. Yüksek eğrilik noktalarına genellikle köşe denir ve OR'da önemli bir rol oynar. Yerel özellikler, belirli bir küçük sınır segmenti şekli veya bir yüzey yaması içerebilir.

İlişkisel özellikler: Farklı varlıkların, bölgelerin, kapalı hatların veya yerel özelliklerin göreceli konumlarına dayanır. Bu özellikler genellikle özellikler ile bağlı oryantasyon ölçümleri arasındaki mesafeyi içerir ve birçok bölge veya görüntüdeki

yerel özellikleri kullanarak bileşik nesnelerin tanımlanmasında çok yararlıdır. Çoğu durumda, varlıkların göreceli konumu nesnelere tanımlayan şeydir. Tam olarak aynı özellik, biraz farklı ilişkilerde, tamamen farklı nesnelere temsil edebilmektedir.

Öznitelik çıkarımı için kullanabileceğiniz çok fazla algoritma bulunmaktadır. Bunlardan birkaç tanesini söylemek gerekirse; SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), HOG (Histogram of Oriented Gradients) şeklinde sıralayabiliriz.

SIFT: Lowe (2004) tarafından önerilmiş olan bu algoritma, ilgi noktalarını tespit etmek ve bunlardan tanımlayıcılar oluşturmak için iyi bilinen bir yöntemdir. Bu yöntem nesne tespiti ve tanınması için ölçek, rotasyon ve çeviri özellik vektörleri sağlar. SIFT tanımlayıcılarını elde etmek için öncelikle potansiyel anahtar noktalar, görüntünün farklı ölçeklerini kullanarak tespit edilir. Bu adımda Gaussian (DoG) filtrelerin farkı kullanılmıştır. Daha sonra, kesin anahtar noktaları, düşük kontrastlı anahtar noktaları ve zayıf lokalize kenar, anahtar noktalarının çıkarılmasıyla bir eşikleme sistemi tarafından belirlenir. Bundan sonra anahtar noktaların oryantasyonları belirlenir. Son olarak tanımlayıcılar, 16x16 boyutundaki gradyan yönlerini ve büyüklüklerini kullanarak her bir anahtar noktası için hesaplanır. Böylece, temel noktanın bu boyuttaki örtüşmeyen her 4x4 penceresi için gradyan oryantasyonları ve büyüklükleri kullanılarak üretilen histogram birleştirilerek 128 boyutlu özellik vektörleri elde edilir (Keser, 2018). SIFT vektörleri ile bir örnek Şekil 3.3'te verilmiştir. Görüntü üzerindeki kırmızı renkli vektörler, ana noktaların konumunu, ölçeğini ve yönünü simgeleyen SIFT vektörleridir.



Şekil 3.3: SIFT vektörlerinin görüntü üzerinde gösterimi.

SURF: SURF yöntemi, literatürde OR için yaygın olarak kullanılan özellik algılama ve açıklama yöntemlerinden biridir (Bay, TuytelaarsveVan Gool, 2006). SURF özellik vektörlerinin elde edilmesi, SIFT hesaplamasındaki gibi adımlardan oluşur. Fakat özellikleri daha hızlı elde edebilmek için bu adımlar geliştirilir. SURF özelliklerinin hesaplanması için, SIFT işleminden farklı olarak kutu filtreleri kullanılır. Ölçek boşluğu, farklı boyutlardaki kutu filtreleri kullanılarak elde edilir. Hessian matrisine dayanan Blob dedektörleri, ilgilenilen noktaları bulmak için kullanılır. Hessian matrisinin determinantının maksimum olduğu noktalar seçilir ve temel noktalar belirlendikten sonra temel nokta yönelimleri hesaplanır. Temel noktaların yönünü belirlemek için yatay ve dikey ekseninde dalgalık tepkileri kullanılır. Daha sonra tanımlayıcı hesaplama işlemi gerçekleştirilir.

SURF vektörleri için ölçek ve rotasyon özellikleri değişmezdir. Bununla birlikte, işlemi hızlandırmak için U-SURF olarak adlandırılan rotasyon varyantı SURF vektörleri kullanılabilir. Şekil3.4'te Caltech-256 veri setinden (Griffin, Holub ve Perona, 2007) üretilmiş bir alt küme olan Örnek 3'ün bir örnek görüntüsü üzerinde en güçlü 20 SURF vektörü gösterilmiştir.



Şekil 3.4: SURF vektörlerinin görüntü üzerinde gösterimi.

HOG: Literatürde OR için yaygın olarak kullanılan özellik vektörlerinden biri, Yönlendirilmiş Gradyan Histogramıdır (HOG). Ancak HOG özelliklerinin insan tespiti için kullanılabileceği gösterilinceye kadar popüler değildi (Dalal ve Triggs, 2005). HOG özelliklerini hesaplamak için, öncelikle görüntünün dikey ve yatay gradyanları elde edilir. Bu gradyan değerleri, büyüklük ve yön değerleri olan kutupsal koordinat değerlerine dönüştürülür. Görüntünün üç renkli kanalı varsa, her

piksel için renk kanalları ile ilgili büyüklük ve yön değerleri arasından maksimum büyüklük ve yön değerleri seçilir. Bu gradyan büyüklüğü ve yön değerlerine göre, her bir (tipik olarak 8x8) görüntü bloğu için gradyan vektörlerinin histogramı elde edilir. Daha sonra oluşturulan histogram vektörleri, 16x16 blok normalizasyonu adı verilen bir teknikle normalleştirilir. Uygulamaya göre parametrelerde değişiklik olabilmektedir. Son olarak bu vektörler, tek bir HOG vektörü oluşturmak için birleştirilir.

HOG özellikleri, normalizasyon adımı sayesinde aydınlatma değişikliklerine karşı dayanıklıdır. Ancak rotasyon ve ölçek değişmez değildir. Ayrıca HOG özelliğinin boyutunun, görüntü bloklarından elde edilen histogram vektörlerinin miktarı nedeniyle görüntü boyutuna göre değiştiğine dikkat edilmelidir. Şekil3.5'te orijinal bir görüntü ve bu görüntünün HOG vektörlerinin çıkarılmış hali gösterilmiştir.

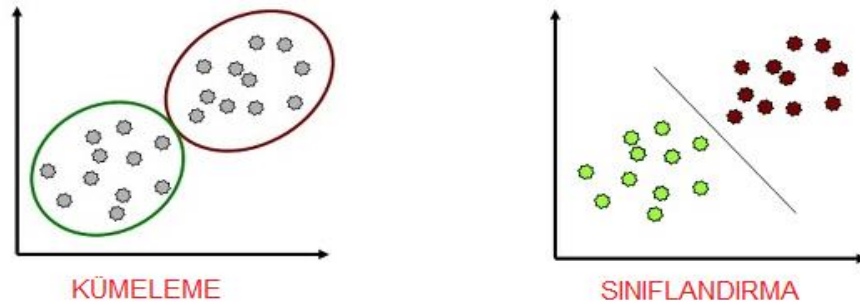


Şekil 3.5: Orijinal görüntü ve HOG vektör görüntüsü.

3.4 Tanımlama

Tanımlama, tespit edilen görüntü üzerinde çıkarım yapabilmek ve bu sayede görüntüyü tanıyabilmek, yani bu görüntünün ne olduğunu anlama aşaması olarak açıklanabilir. Bu bölümde temel iki kavram olan Sınıflandırma ve Kümeleme işlemlerine değinilmiştir. Sınıflandırma, öğrenme, kestirim ve örüntü tanıma yöntemleri ile veriyi önceden belirlenmiş sınıflardan birine dahil etme işlemidir. Sınıflandırmadaki temel fikir, özelliklere dayalı nesnelere tanımdır. OR yaklaşımları bu kategoriye girer ve örnekleri birçok uygulamada gösterilmiştir. Sinirsel ağ tabanlı yaklaşımlar da bu sınıfa girer. Bu sınıftaki tüm teknikler, görüntülerde N özelliklerinin tespit edildiğini ve bu özelliklerin aynı metrik alanda temsil edilmeleri için normalleştirildiğini varsaymaktadır. Sık kullanılan bazı

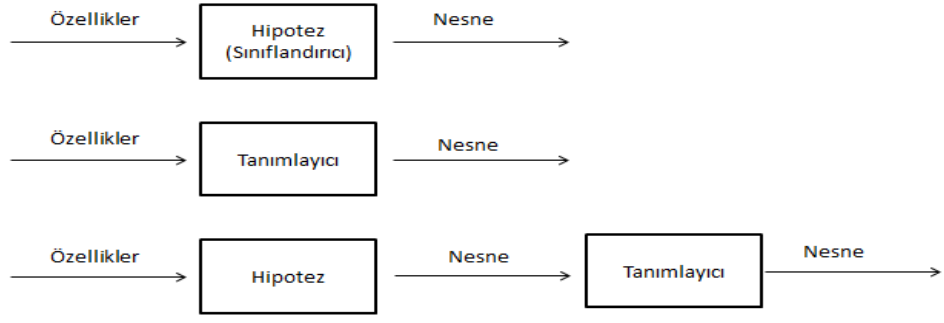
sınıflandırma tekniklerine Sinir Ağları, Bayes Ağları, En Yakın Komşu gibi sınıflandırıcılar örnek olarak verilebilir. Kümeleme ise birbirleriyle benzer olan verileri, benzer özellikleri bulunanlar ile aynı grupta toplama işlemidir. Sınıflandırma ve Kümeleme işlemleri Şekil 3.6’da gösterilmiştir.



Şekil 3.6: Kümeleme ve Sınıflandırma işlemleri.

3.5 Tanıma Stratejileri

OR, uygun özelliklerin tespit edilmesinden sonra yapılması gereken adımların dizisidir. Bir görüntüdeki algılanan özelliklere dayanarak, görüntüdeki olası nesnelere hakkında hipotezler oluşturulmalıdır. Bu hipotezler nesne modelleri kullanılarak doğrulanmalıdır. OR tekniklerinin tümü güçlü hipotez oluşturma ve doğrulama adımları gerektirmez. Çoğu tanıma stratejisi bu iki adımı farklı miktarlarda birleştirmek için geliştirilmiştir. Şekil 3.7’de gösterildiği gibi, bu iki adımın üç farklı olası kombinasyonu gösterilmiştir. Buradaki bir veya daha fazla adımın nasıl uygulanacağı, daha önce bahsettiğimiz faktörlerle karakterize edilebilir. Sorunun karmaşıklığına bağlı olarak, bir tanıma stratejisinin hipotez oluşturma ve doğrulama adımlarından birini veya ikisini birden kullanması gerekebilmektedir.



Şekil 3.7: Farklı tanıma stratejileri.

DÖRDÜNCÜ BÖLÜM

OPTİK KARAKTER TANIMA

Belgeler, insanların okuyabildiği ve anlayabileceği kağıtlar şeklindedir. Ancak bilgisayarın bu belgeleri doğrudan anlaması mümkün değildir. Optik karakter tanıma, herhangi bir insan müdahalesine gerek kalmadan, makine tarafından basılan veya el ile yazılmış metinlerin, sayıların, harflerin ve sembollerin bilgisayarla işlenebilir biçimde sınıflandırılması problemi ile ilgilidir. Tanıma işlemi, verilen karakter görüntüsünden elde edilen özellikler ile görüntü modelleri kütüphanesi arasında bir eşleşme aranarak gerçekleştirilir (Chen, 2003).

Karakter tanıma sistemleri veri toplama süreci (çevrimiçi veya çevrimdışı) ve metin türü (makine baskılı veya el yazısı) olmak üzere iki ana kritere göre sınıflandırılabilir (Arıca ve Yarman-Vural, 2001). Bununla birlikte, çevrimiçi karakter tanıma bazen Optik Karakter Tanıma ile karıştırılmaktadır. Optik karakter tanıma, sistemin karakter görüntüsünün sabit, statik şeklini tanıdığı, çevrimiçi karakter tanıma ise el yazısı sırasındaki dinamik hareketi tanıdığı açık karakter tanıma anlamına gelir.

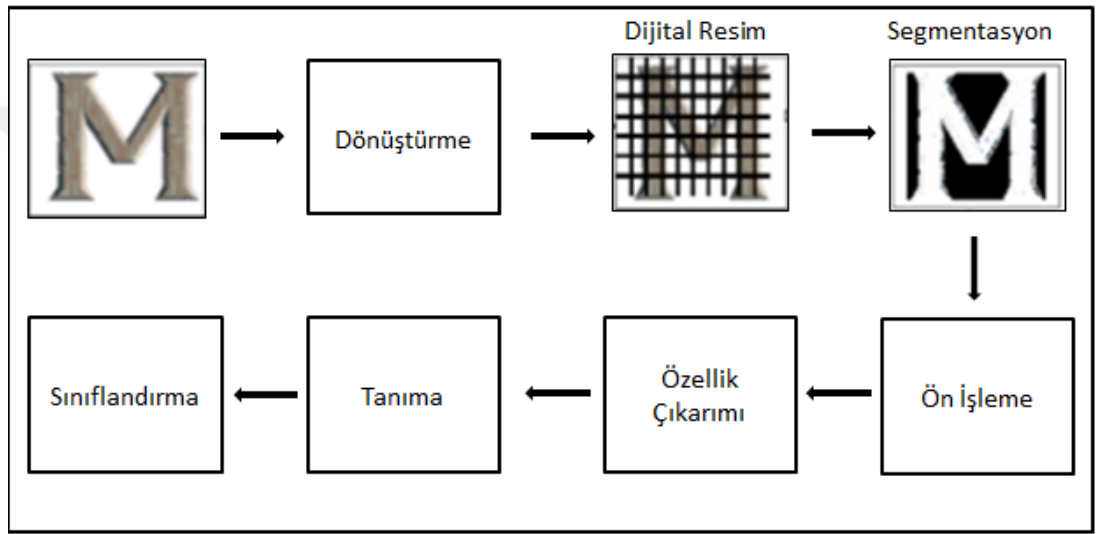
Çevrimiçi Karakter Tanıma: Karakterlerin gerçek zamanlı olarak tanıma işlevini yerine getirir. Çevrimiçi sistemler, zamanlama bilgisine sahip olduklarından tanıma gerçekleştirmek için daha iyi bilgiye sahiptir ve karakterlerin çevrimdışı olan benzerlerini bulmak için ilk arama adımından kaçınırlar. Örnek verilecek olursa, çevrimiçi sistemler kalemin konumunu doğrudan, yazıyı yazıldığı gibi tutan dönüştürücüden alır. Bu cihazlardan en yaygın olanı elektronik tablet veya dijitalleştiricilerdir (digitizer). Karakterlerin çevrimiçi olarak tanınması, karmaşık karakter şekilleri ve farklı modlarda yazılmış karakter sembollerinin büyük çeşitliliği nedeniyle zorlu bir problem olarak bilinir (Arıca, 1998; Arıca ve Yarman-Vural, 2001).

Çevrimdışı Karakter Tanıma: Herhangi bir belge üzerinde bulunan yazıların tanınabilmesi için kullanılan yöntemlerdir. Çevrimdışı karakter tanımada daktilo veya el yazısı ile yazılmış karakterler, optik tarayıcı veya kamera tarafından taranıp, ikili veya gri tonlamalı bir görüntüye dönüştürüldüğünden dolayı Optik Karakter Tanıma olarak da bilinmektedir. Karakterlerin taranması, ikili veya gri tonlamalı görüntülere dönüştürülmesi işlemleri arasındaki karmaşık etkileşimler, çevrimdışı karakter tanımada ek zorluklara neden olmaktadır. Bu nedenle çevrimdışı karakter

tanıma, çevrimiçi karakter tanıma göre daha zor bir görev olarak kabul edilir (A.g.e.).

Bu iki tanıma arasındaki en büyük fark, çevrimiçi karakter tanımanın gerçek zamanlı bağlamsal bilgilere sahip olması ancak çevrimdışı verilerin olmamasıdır. Bu yüzden işleme mimarileri ve yöntemlerinde önemli bir farklılık olmaktadır.

Hangi metin türünde yazılmış olursa olsun OCR sistemleri görüntü ön işleme, ikili hale dönüştürme, segmentasyon, öznelik çıkarma ve sınıflandırma aşamalarından oluşmaktadır. Şekil 4.1’de OCR temel yapısı gösterilmektedir.



Şekil 4.1: OCR temel yapısı.

4.1 Optik Karakter Tanıma Tarihi

OCR teknolojisiyle ilgili ilk patenti 1929 yılında Alman bilim adamı Gustav Tauschek almıştır. Tauschek'in ilkesi o zamandan teknolojiyi ayıran şablon/maske eşleşmesiydi. Onun makinesi, fotoğraf dedektörü ve mekanik şablon eşleştirmesi kullanan mekanik bir cihazdı. Sistemde tanınacak olan karakter ile şablon üst üste geldiğinde ışık geçmeyeceğinden dolayı fotoğraf dedektör ile algılanamazdı. Günümüzde de Tauschek'in bu ilkesi hala kullanılmaktadır (Öz, Çit ve Yazgan, 2010). 1933'te ise Handel, OCR üzerine ABD patentini aldı. (ABD Patenti 1.915.993).

Başlangıçta bir OCR geliştirmenin kolay olacağı düşünülüyordu. 1950 ve 1960 yılları, araştırmacıların ideal bir OCR hayal ettikleri fakat sorunların da farkında

oldukları yıllardı. İlk modern karakter tanıyıcılar, 1940'larda dijital bilgisayarların gelişmesiyle birlikte ortaya çıktı. Karakterlerin otomatik olarak tanınmasıyla ilgili yapılan ilk çalışmalar, iyi basılmış metinlere veya küçük, iyi ayırt edilmiş el yazısı metinlerine yoğunlaşmıştı. 1951 yılında 23 harfi tanıyabilen ve ilk modern bilgisayar desteğinin bulunduğu OCR sistemi olan GİSMO, Amerika'da David Shepard tarafından geliştirilmiştir (Musayev, 2004). 1965'te İngiltere'deki fatura ödeme sistemlerinde devrim yaratan bir süreç olan, OCR teknolojisini kullanan bir bankacılık sistemi olan Ulusal Giro planlanmaya başlandı. 1980'li yıllara kadar olan çalışmalarda eksiklikler, gelişmiş algoritmaların, güçlü donanımların ve optik cihazların olmamasından kaynaklandı. Bilgisayar donanım ve yazılım teknolojisinde gerçekleşen patlama ile önceden geliştirilen metodolojiler, birçok uygulama alanında ve OCR sistemlerinin geliştirmesindeki hızlı büyüme için oldukça verimli bir ortam oluşturmuştur (Arıca, 1998). 1994 ve 1997 yıllarında ise Arap alfabesindeki harfleri tanıyan bir OCR sistemi oluşturdu (Ahmed, 1994). Daha sonra Inoue, Miyazaki ve Suzuki (1998) tarafından matematiksel formüllerin bulunduğu Japonca dökümanlara ait olan taranmış sayfadaki matematiksel formülleri ve Japonca terimleri ayrı ayrı tanıyan bir OCR sistemi geliştirildi. Bunlara benzer sistemleri takiben Leung ve Sze (1997) ile Zhang, Zhao ve Wang (2002) el yazısı ile yazılmış olan Çin karakterlerini tanıyan çalışmalar gerçekleştirdiler. Günümüzde kullanılan gelişmiş teknolojiler ve hızlı mikroişlemcilerin ortaya çıkmasıyla birlikte OCR'nın popülaritesi her yıl artmakta ve kullanılan araçlara son derece gelişmiş tanıma teknikleri sağlayan daha iyi tarama teknolojileri bulunmaktadır.

4.2 Optik Karakter Tanıma Sistemlerinin Kullanım Alanları

OCR sistemlerinin temel amacı, fotoğrafı çekilmiş veya taranmış herhangi bir resim formundaki belgenin, bilgisayar tarafından işlenebileceği ve tekrar kullanılabilir bir gösterime dönüştürmektir. OCR sistemleri, günümüzde birçok farklı alanda araştırma konusu olmakta, kullanılmakta ve gün geçtikçe bu kullanım alanları artış göstermektedir. Bu artışın en önemli sebebi ise çok fazla miktarda bulunan verinin yalnızca birkaç dakika içinde veri tabanlarına ya da bilgisayarlara kolayca ulaştırılabilmesidir.

OCR sistemlerinin sıklıkla uygulandığı ve önemli uygulama alanları aşağıda verilmiştir.

- Adres Okuma: Adres okuyucuları, hedef bloğu ve posta kodunu bu adres bloğuna göre sıralayan ve konumlandıran posta servisleri tarafından kullanılır, daha sonra postalar teslimat amaçlı olarak hedeflere bağlı olarak gruplanır.
- Form Okuma: Bir form okuma sistemi önceden basılmış form talimatları arasında ayırım yapar ve verilerle doldurulur. Sistem ilk önce boş bir formla eğitilir ve alanları verinin yazdırılacağı forma kaydeder. Form tanıma aşamasında sistem, verilerle doldurulması gereken bölgeyi taramak için eğitimden elde edilen bilgileri kullanır.
- Fatura Görüntüleme: Makbuzları, faturaları ve envanter belgelerini okumak, finansal kayıtları izlemek ve ödemelerin birikmesini engellemek için sayısız kuruluş uygulamasının bir parçası olarak yaygın şekilde kullanılmaktadır. Devlet dairelerinde ve özerk kuruluşlarda, OCR farklı prosedürler arasında bilgi toplama ve analizini kolaylaştırır (Ganis, Wilson ve Blue, 1998).
- Plaka Tanıma: Araçların otomatik olarak plakalarını okuyup tanımlamak için görüntü işleme ve karakter tanıma teknolojisini kullanır. Tipik plaka tanıma uygulamaları arasında özel otopark yönetimi, trafik takibi, otomatik trafik bileti verilmesi, otomatik ücret ödemesi, gözetim ve güvenlik uygulamaları bulunmaktadır.
- El Yazısı Tanıma: Bir bilgisayarın kaynaklardan, örneğin kağıt kayıtlarından, fotoğraflardan, dokunmatik ekranlardan ve farklı aygıtlardan anlaşılır el yazısı verilerini alma ve tercüme etme kapasitesidir. Yazılan içeriğin resmi, optik tarama (optik karakter tanıma) veya zekice bir kelime tanıma ile bir miktar kağıttan "çevrimdışı" algılanabilir. Öte yandan, kalem ucundaki gelişmeler, örneğin kalem tabanlı bir bilgisayar ekran yüzeyi ile "çevrimiçi" olarak tespit edilebilir.
- Hukuk Sektörü: Hukuk endüstrisi OCR yeniliğinin alıcılarından biridir. OCR, belgeleri dijitalleştirmek ve özel olarak bilgisayar veritabanına girmek için kullanılır. Birkaç anahtar kelime yazarak çok fazla veri içeren veri tabanlarından istenen belgeler bulunabilir.
- Bankacılık: OCR'nın bir başka zorunlu kullanımı da insan müdahalesi olmadan çeklerin işlenmesinde kullanıldığı bankacılık alanıdır. Sistem

filtreleme yapabilen ve doğru miktarda nakit alışverişi yapılan bir makinaya gömülebilir. Yapılan bu çalışmalar, bankalarda gerçekleştirilen işlemlerin bekleme süresini azaltmaya yardımcı olmaktadır.

- Sağlık: Basılı materyalleri işlemek için tıbbi hizmetlerde OCR sistemlerinin kullanımında bir genişleme görülmüştür. Tıbbi hizmet uzmanlarının sürekli olarak koruma çerçeveleri ve ayrıca genel sağlık formları dahil olmak üzere her hasta için geniş hacimli dokümanları yönetmesi gerekir. Bu verilerin her birinin farkında olmak için ilgili bilgilerin elektronik bir veri tabanına girilmesi önemlidir. OCR işleme araçlarıyla, yapılardan veri çıkarabilir ve veri tabanlarına eklemeler yapılabilir. Böylece her hastanın bilgisi hızlı bir şekilde kaydedilir ve gerektiğinde daha sonra da kullanılabilir.
- Captcha: Web sitelerinin güvenliğini sağlamak amacıyla bir önlem niteliğinde kullanılan Captcha, botlara karşı etkili bir çözüm olmaktadır. Bilgisayarların çözemeyeceği fakat insanların çözebileceği testlerden meydana gelen bir sistemdir. Örnek verilecek olursa, testte bulunan herhangi bir görüntüyü, bir insan inceleyip yorumlayabilirken bir bilgisayar bunu yorumlayarak doğru kararı verememektedir. “Completely Automated Public Turing test to tell Computers and Humans Apart” ifadesinin kısaltması olan Captca, Manuel Blum, Luis von Ahn, John Langford ve Nicholas Hopper tarafından 2000 senesinde tasarlanmıştır. Bu uygulamayı bireyler internet ortamından indirerek kendi sistemlerinde bot önleyici olarak kullanabilmektedirler (Gossweiler, Kamvar ve Baluja, 2009).

4.3. OCR Adımları

Bu bölümde, optik karakter tanımanın temel ve önemli aşamalarına odaklanmaktadır. Bu aşamalar ön işleme, segmentasyon, normalizasyon, özellik çıkarımı ve sınıflandırma olarak gruplandırılabilir. Çeşitli uygulamalarda bazı aşamalar birleştirilmiş veya eksiltişmiş olabilmektedir. OCR ile ilgili etkili bir uygulama tasarlamak ve yüksek karakter tanıma oranı elde etmek için her aşamada ortaya çıkabilecek zorlukları göz önünde bulundurmak gerekmektedir.

4.3.1. Ön İşleme Aşaması

Ön işlemin amacı, görüntüdeki istenmeyen özellikleri veya gürültüyü, hiçbir önemli bilgiyi kaçırmadan ortadan kaldırmaktır. Metin veya grafik içeren renkli, gri seviyeli

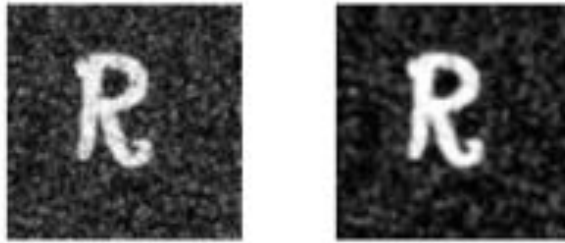
veya ikili belge görüntülerinde ön işleme tekniklerini uygulamak gerekmektedir. Renkli görüntülerin işlenmesi hesaplama açısından daha maliyetli olduğundan, karakter tanıma sistemlerindeki uygulamaların çoğu ikili veya gri görüntüler kullanmaktadır.

Ön işleme, tutarsız veri ve gürültüyü azaltır. Görüntüyü geliştirir ve sonraki OCR aşamaları için kullanmaya hazır hale getirir. Görüntüyü, ilk aşama olan ön işleme aşamasında uygun formata dönüştürerek, sonraki aşamalarda işlenecek görüntünün etkinliğini ve kolaylığını artırabiliriz. Bu nedenle karakter tanıma oranındaki azalmaya neden olan gürültüyü azaltmak, ön işleme aşamasında ana önemli konudur. Dolayısıyla ön işleme, verinin ardışık fazlarla uygunluğunu kontrol ettiğinden, özellik çıkarımı aşamasından önceki bir aşama ön işleme aşamasıdır. OCR Zorlukları bölümünde anlatacağımız çoğu zorlukların ön işleme aşamasında ele alınması gerekmektedir.

Ön işleme aşamasını daha detaylı bir biçimde inceleyecek olursak; gürültü azaltma, ikilileştirme, segmentasyon ve normalizasyon gibi alt başlıklara ayırabiliriz.

4.3.1.1. Gürültü Azaltma

Gürültü azaltmanın amacı, istenmeyen herhangi bir bit örüntüsünü (bit-pattern) ortadan kaldırmak ve genellikle görüntü elde etme cihazının yüzeyini veya düşük örnekleme oranını yazarak ortaya çıkan saf noktaları kullanmaktır. Temel fikir, komşu piksellerin gri değerlerinin bir fonksiyonu olarak bir piksele değer atamak için görüntüyü önceden tanımlanmış bir maske ile sarmaktır. Gürültülü ve gürültüsü azaltılmış görüntüler Şekil 4.2’de verilmiştir.



Şekil 4.2: Gürültülü ve gürültüsü azaltılmış karakter görüntüsü.

a) Segmentasyon

Segmentasyon, görüntünün analizini kolaylaştırmak için, görüntüyü bileşenlerine ayırma işlemini ifade etmektedir. Başka bir deyişle, bir resmin içindeki metin

bileşenini görüntünün arka planından izole etme işlemidir. Tanıma oranlarını doğrudan etkileyen görüntü analizinde önemli bir aşamadır. Karakter segmentasyonu, kelimelerden sadece karakterleri çıkardığımız süreçtir. Segmentasyon, iç ve dış olmak üzere iki gruba ayrılabilir. Dış segmentasyon, görüntüyü mantıksal birimlerine ayırır. Örneğin, doküman analizinde, sayfa düzeni paragraflara, cümlelere ve kelimelere ayırmaktır. İç segmentasyon ise görüntüyü ayrı ayrı semboller, karakterler, şekiller vb. biçimde ayırma yapmayı amaçlayan bir işlemdir (Jain, 1989). Düzenlenebilir metin satırlarının tanınan karakterlerden uygun şekilde yeniden düzenlenmesi için önce metin satırını bölümlere ayırır, sonra kelimeler bölümlenmiş satırdan bölümlenir ve son olarak karakterlerin bölümlendirilmesi gerçekleştirilir. Belge segmentasyonu, bir OCR sisteminin uygulanmasında büyük bir ön işleme aşamasıdır. Bir belge görüntüsünü homojen bölgelere sınıflandırmak, yani her bölgenin metin, şekil, tablo veya yarı tonlu görüntü gibi yalnızca bir tür bilgi içermesini sağlayan bir işlemdir. Çoğu durumda, OCR ile ilgili sistemlerin doğruluk oranı, büyük ölçüde kullanılan sayfa bölümlenme algoritmasının doğruluğuna bağlıdır.

Segmentasyonun üç farklı algoritması vardır;

- Yukarıdan aşağıya doğru yöntemler (Top-down methods): Bu yaklaşımda bir belgedeki büyük bölgeler tekrarlı olarak daha küçük alt bölgelere ayrılır. Kriter karşılandığında, belge bölümlendirme işlemi durur ve bu aşamada elde edilen aralıklar son bölümlenmenin sonuçlarını oluşturur.
- Aşağıdan yukarıya doğru yöntemler (Bottom-up methods): Bu yaklaşımda ise ilgili pikseller aranarak başlanır ve bu ilgili pikseller gruplandırılır. Daha sonra bu ilgili pikselleri, kelimelerle birleştirilen karakterleri, satırları veya metin bloklarını birleştiren bağlı bileşenlerde yönetirler.
- Hibrit yöntemler (Hybrid methods): Aşağıdan yukarıya ve yukarıdan aşağıya yöntemlerinin entegrasyonu ile oluşturulan yöntemlerdir.

Son yıllarda OCR sisteminin farklı yönleriyle ilgili olarak, segmentasyon için birçok yaklaşım önerilmeye devam edilmiştir.

b) Normalizasyon

Normalizasyon işleminde, segmentasyon işleminin bir sonucu olarak, özellik çıkarma aşaması boyunca işlenmeye hazır olan izole edilmiş karakterler elde edilir.

Bu nedenle izole edilen karakterler kullanılan algoritmalara bağılı olarak belirli bir boyuta indirgenir. Segmentasyon işlemi, görüntüyü $m * n$ matris formuna dönüştürdüğü için çok önemlidir. Bu matrisler daha sonra, boyutu en aza indirerek ve herhangi bir etkili bilgiyi kaçırmadan, gereksiz bilgileri görüntüden kaldırarak bu işlemler sonucunda normalize edilmiş olur (Trier, Jain ve Taxt, 1996).

c) İkileştirme

Görüntü ikileştirme, gri ölçekli bir görüntünün ikili görüntüye dönüştürülmesi anlamına gelir. Bir görüntüdeki her piksel bir bite dönüştürülür, pikselin değerine ve eşik değerine bağılı olarak "1" veya "0" değeri atanır. İkileştirme süreci küresel ve yerel olmak üzere iki gruba ayrılmaktadır. Küresel ikili hale getirme, görüntünün yoğunluk histogramından arka plan düzeyinin tahminine dayanan, görüntünün tamamına bir eşik değeri seçer. Yerel ikileştirme ise her piksel için yerel alan bilgisine göre farklı eşik değerleri kullanır.

4.3.2. Özellik Çıkarımı

Bu aşama, özellik vektörlerini oluşturmak için ilgili özellikleri nesnelere veya alfabelere çıkarma işlemidir. Tanıma problemini çözülebilir hale getirmek için modeli özellik vektörüne dönüştürmek gerektiğinden, özellik çıkarımı aşaması model tanıma uygulamalarında en önemli adımlardan biridir. Bu özellik vektörleri daha sonra sınıflandırıcılar tarafından giriş birimini objektif çıkış ünitesi ile tanımlamak için kullanılır. Sınıflandırıcının bu özellikleri dikkate alarak sınıfları ayırt etmesi zahmetsiz hale gelir, çünkü belirlenmesi oldukça kolaydır (Pradeep, Srinivasan ve Himavathi, 2011).

Literatürdeki karakterlerden özelliklerin çıkarılması ile ilgili birkaç teknik önerilmiştir. Dinesh Acharya ve Subbareddy (2007) el yazısıyla yazılan Kannada rakamları için son puanları tanıma, yatay / dikey vuruşları özelliklerini kullanarak %90,50 tanıma doğruluğu elde etmişlerdir. Ancak, bu yöntemde bazı özelliklerin kaybına neden olan inceltme işlemi kullanılmıştır (Pradeep, Srinivasan ve Himavathi, 2011). Sharma, Ghose ve Shah (2012) el yazısı alfabe tanıma için Euler sayısını kullanan gelişmiş bir bölge bazlı hibrit özellik çıkarma modeli önermişlerdir. Suen'nin (1986) çalışmasına göre, istatistiksel özellikler ve yapısal özellikler olmak üzere iki ana özellik sınıfı vardır. Bir karakter matrisinde, her noktanın istatistiksel

dağılımından bölgeleme, açı, kesit, fourier dönüşümü ve izdüşüm histogramları gibi istatistiksel özellikler elde edilebilir (Rehman ve Saba, 2014).

Başlangıçta makine tarafından basılan karakterleri tanımak için istatistiksel özellikler sunulur. Öte yandan, yapısal veya topolojik özellikler, tasarlanacak karakter setinin geometrisi ile ilgilidir. Karakterlerde dış büyüklükler ve sıkışıklıklar, delik sayısı, bitiş noktası gibi özellikler örnek olarak verilebilir.

4.3.3. Sınıflandırma

OCR sistemleri, geniş bir şekilde her bir örneği, daha önceden tanımlanmış bir sınıfa atayan örüntü tanıma metodolojilerini kullanır. Özellik çıkarma ve seçme aşaması tamamlandıktan sonra birkaç mevcut yaklaşım kullanılarak bir sınıflandırıcı tasarlanabilir. Sınıflandırıcı, bir örnekteki çeşitli özellikleri alan ve örneğin ait olduğu sınıfı öngören bir işlemdir. Bu aşamada farklı girdileri farklı sınıflara ayırırken, homojen niteliklere sahip gruplar oluşturmak için, algılanan bilgilere göre girdilerin karşılaştırma sınıflarına dağıtılması işlemi gerçekleştirilir.

Sınıflandırmayı seçmek kolay değildir, çünkü sınıflandırma için tek bir optimal yaklaşım yoktur. Sınıflandırıcının seçilmesi, serbest parametre sayısı, mevcut eğitim seti gibi birkaç özelliğe bağlıdır. Her biri özellik alanının belirli bölümlerinde desenleri ayırt edebilen birçok kabul edilebilir sınıfsal tasarım yaklaşımı vardır. Denetimli öğrenmede, hangi sınıflandırma yaklaşımı kullanılırsa kullanılsın sınıflandırma süreci, eğitim ve sınıflandırma olmak üzere iki bölümde çalışır (Jain, Duin ve Mao, 2000). Eğitim bölümünde, sınıflandırıcı bilinen her sınıf modelinin ölçülen özelliklerine dayanarak eğitilir. Öğrenilen sınıf, temel olarak eğitim seti içindeki özellikler ve sınıf etiketi arasındaki ilişkinin bir modelidir. Sınıflandırma bölümünde ise eğitilmiş sınıflandırıcı, giriş desenini ölçülen özellikleri göz önüne alarak desen sınıflarından birine tahsis eder.

Sonuç olarak sınıflandırıcının performansı hem mevcut eğitim örneklerine hem de örneklerin spesifik değerlerine bağlıdır. Buradaki düşünce, eğer sınıf özellikleri ve sınıflar arasındaki ilişkiyi gerçekten yakalarsa, örneklerin sınıflarını tahmin edebilmesi gerektiği şeklindedir.

Temel sınıflandırma yaklaşımlarını aşağıdaki gibi kategorize edebiliriz.

- Şablon Eşleştirme: Depolanan modellerin algılanacak kelime veya karakter ile eşleştirilmesi açısından karakter tanıma için en az karmaşık yöntemdir.

Şekillerin, piksellerin, eğriliğin ve benzerlerinin bir araya getirilmesiyle yapılan eşleştirme işlemi, iki vektör arasındaki benzerlik seviyesini belirler. Gri seviyeli veya ikili giriş karakterleri, depolanan modellerin standart bir düzenlemesiyle karşılaştırılır. Başka bir deyişle, tanınacak model, izin verilen tüm değişiklikleri dikkate alarak depolanan modellerle eşleştirilir. Şablon eşleştirmesi hesaplama açısından zordur ve görüntüleme işlemi, bakış açısı değişikliği veya desenler arasındaki büyük sınıf içi değişimler nedeniyle modellerin bozulması durumunda başarısız olur (A.g.e.). Bu stratejinin tanınma oranı, gürültü ve girdi bozulmalarına karşı son derece hassastır.

- İstatistiksel Yaklaşım: Bu yaklaşımda, gözlem yoluyla elde edilen ve bir d boyutlu örüntü vektörü x olarak gösterilen bir ölçüm kümesi verildiğinde, örüntü, olası C sınıflarından birine atanır ($\omega_i = 1, \dots, c$). Özelliklerin örüntü sınıfında koşullandırılmış bir olasılık yoğunluğu işlevine sahip olduğu varsayılır ve özellik vektörü x , koşullu olasılık işlevi $p(x | \omega_i)$ sınıfından rasgele çekilmiş bir gözlem olarak görülür ve ölçüm alanını c bölgelerine ayırır ($\Omega_i = 1, \dots, c$). Eğer bir gözlem vektörü Ω_i 'de ise, ω_i sınıfına ait olduğu varsayılır. Sınıf yoğunlukları hakkında mevcut bilgilere bağlı olarak, sınıflandırıcı tasarlamak için çeşitli stratejiler kullanılmaktadır. Sınıf yoğunlukları biliniyorsa Bayes karar kuralı kullanılabilir fakat bu pratikte yaygın bir durum değildir. Böyle bir durumda parametrik ve parametrik olmayan iki farklı strateji izlenebilir. Parametrik yaklaşım, sınıf koşullu yoğunlukların biçiminin bilindiğini varsayar, ancak bazı yoğunluk parametrelerinin eğitim verileri kullanılarak tahmin edilmesi gerektiğini varsayar. Bununla birlikte, sınıf koşullu yoğunlukların şekli bilinmiyorsa, yoğunluk fonksiyonunu tahmin etmek veya doğrudan karar sınırını oluşturmak gerekir. İstatistiksel tanıma yöntemleri hakkında ayrıntılı bilgi Duda, Hart ve Stork'un (2000) çalışmasında bulunabilir.
- Nöral Ağlar: Nöral Ağlar: Karakter sınıflandırma sorunu, insanların karakterleri öğrenmelerine ve deneyimlerine göre algılayabildikleri için sezgisel mantıkla tanımlanır. Bu nedenle, doğada fazlasıyla sezgisel olan sinir ağları bu tür sorunlar için oldukça uygundur (A.g.e.). Bir sinir ağı, nöron

olarak adlandırılabilen, düğümlere bağlı sayısal değerler olarak gösterilen ve graf geçişi algoritmaları ile elde edilen modeller arasındaki dönüşümler şeklinde temsil edilmektedir. Graftaki düğümler, giriş düğümleri veya çıkış düğümleri olarak bölünmüştür ve birbirlerine paralel olarak bağlanmışlardır. Tüm bağlantılar giriş ve çıkış fonksiyonunu belirleyen ağırlıklara sahiptir ve öğrenme algoritması ile ayarlanırlar. Paralel bir mimariye sahip olduğu için klasik tekniklere göre daha yüksek oranda hesaplamalar yapılabilir. Ek olarak, sinir ağları verilerdeki değişikliklere uyum sağlayabilir ve girdinin karakteristiğini öğrenebilir. Tablo 3.2’de yapay sinir ağına dayalı önerilen bazı OCR uygulamalarının sonuçları verilmiştir.

Tablo 4.1:Yapay sinir ağına dayalı OCR uygulamaları.

Yazarlar	OCR Uygulaması	Doğruluk %
Shah, vd., (2009)	Numara Tanıma	95.49
Zhai, Bensaali ve Sotudeh (2012)	Otomatik Plaka Tanıma	97.3
Shamsher, vd., (2007)	Urdu dili için OCR	98.3
Yetirajam, Nayak ve Chattopadhyay (2012)	Bozuk Karakterleri Tanınma ve Sınıflandırılma	68.33

4.4. OCR Doğruluğunu Etkileyen Faktörler

OCR teknolojilerinde var olan gerekli işlemler yapıldıktan sonra çıkacak olan sonuçların doğruluğu değerlendirilirken göz önünde bulundurulması gereken çok sayıda önemli faktör vardır. Bu önemli faktörlerden bazıları aşağıda listelenmiştir:

Tarama yöntemleri: Eğer görüntü gri tonlama veya daha iyi olarak gösteriliyorsa OCR doğruluğunu artırma olasılığı, diğer tüm tarama mekanizmalarından daha fazladır. Gri tonlamalı görüntüler açıkça okunabilir durumdadır, bu nedenle taranan

görüntünün standardı düşerse, OCR doğruluğu da azalabilir. Taranan görüntünün standardını etkileyecek başka faktörler varsa bunlar da dikkate alınmalıdır.

Orijinal belgenin niteliği: Metnin görüldüğü orijinal kağıt, taranan görüntü için çok önemlidir. Çok koyu bir arka plana karşı görünen metnin, kelimelerin veya grafiklerin üzerine basılmış olanları okumak zor bir işlemdir. Bir sistemi sadece ilgili verileri yorumlayacak ve geri kalanını göz ardı edecek şekilde programlamak da karmaşık bir iştir. OCR sistemi karakter ile kağıt arka plan gürültüsü arasında ayırım yapamıyorsa, karakteri yanlış tanıyabilir. Siyah beyazın aksine gri tonlamalı görüntüler, OCR yazılımına metin ve gürültü arasında daha iyi bir ayırım yapma şansı verir ve böylece doğruluk artırılmış olur.

Baskı Türü: Metnin asıl olan niteliği OCR doğruluğunda önemli bir fark yaratabilir. Açıkçası metin zayıf yazdırılırsa veya yazılırsa ve karakterler kırılırsa, soluksa veya belirgin kenarlara sahip değilse, bu bir OCR sisteminin desenleri tanıma ve benzer şekilli karakterleri ayırt etme yeteneğini etkiler. Bu nedenle baskının netliği göz önünde bulundurulması gereken bir faktördür. Örneğin, harfleri ve sayıları birleştiren bir karakter dizisini incelerken, büyük bir harf olan "O" ve "0" rakamları arasında çok az görünür bir fark vardır. İnsanlar doğru anlamı belirlemek için cümleyi veya paragrafın tamamını tekrar okuyabilir. Ancak bu prosedür, bir makine için çok daha zordur. Bazı fontlarda diğerlerine göre daha büyük punto boyutları kullanılarak baskı netliğini de arttırabilir.

Karmaşıklıkların Biçimlendirilmesi: Yazı tipi boyutu ve yazı yüzeyindeki değişiklikler, karakterlerin yanlış anlaşılmasına neden olabilir. Aşırı mürekkep veya kağıt bozulmalarından kaynaklanabilecek durumlarda tanıma işlemi gerçekleştirilemeyebilir.

4.5. OCR Zorlukları

Yüksek doğrulukta karakter tanıma için OCR teknikleri, farklılaştırıcı metin ve arka plan gibi bazı temel yapısal özelliklere sahip yüksek kaliteli veya yüksek çözünürlüklü görüntüler bekler. Görüntülerin üretilme şeklinin görüntü kalitesi üzerinde büyük bir etkisi bulunmaktadır. Bundan dolayı görüntülerin üretilme şekli OCR'nın doğruluğu ve başarısı açısından oldukça önemli ve belirleyici bir faktördür. Tarayıcılar tarafından üretilen görüntülerle çalışan bir OCR, genellikle yüksek

doğruluk ve performans sağlar. Kameralar tarafından üretilen görüntüler ise genellikle çevresel veya kamerayla ilgili oluşabilecek faktörler nedeniyle OCR için kullanılacak taranmış görüntüler kadar iyi bir girdi değildir. Bu ve benzeri durumlar nedeniyle aşağıdaki gibi çok sayıda hata ortaya çıkabilmektedir.

4.5.1. Görüntü Karmaşıklığı

Bir OCR sisteminde kullanılacak olan görüntüdeki metin veya karakter yanında çok sayıda nesne ve semboller görebiliriz. Bu nesne ve semboller, karşılaştırmalı yapılara ve işlenen görüntüdeki metni tanımayı çok zorlaştıran görünümlere sahip olabileceklerinden dolayı görüntü karmaşıklığı oluşumuna neden olurlar. Yaşanan bu sorunun nedeni çevredeki sahnenin, metnin bulunmadığı bölümü ayırmakta zorlanmasıdır (Ye ve Doermann, 2015).

4.5.2. Düzensiz Işıklandırma

Doğal ortamlarda görüntü çekmek istenildiğinde çoğu zaman eşit olmayan aydınlatma ve gölge problemleri ile karşılaşılır. Bu durum görüntünün özelliği ve kalitesini düşürmeye neden olduğundan OCR için bir zorluk teşkil etmektedir. Dolayısıyla daha az doğru algılama, segmentasyon ve tanıma sonuçları ortaya çıkar (A.g.e.). Düzensiz veya eşit olmayan aydınlatma durumu, taranan görüntü ve kamera ile üretilmiş görüntü arasındaki farkı ayırt etmeyi sağlar. Bir fotoğraf makinesinde flaş kullanılması, eşit olmayan ışıklandırma gibi bu tür sorunları ortadan kaldırılabileceği de, yeni zorluklar ortaya çıkarmaktadır. Kamera kullanılarak elde edilen görüntülerde aydınlatma veya gölge gibi problem oluşabileceğinden OCR sistemlerinin başarısı ve doğruluğu açısından taranan görüntülerin kullanılması tercih edilmektedir.

4.5.3. Eğrilik (Rotasyon)

Optik karakter tanıma sistemleri için, kameradan alınan görüntü veya kullanılan diğer aygıtların bakış açısı, tarayıcı girişi gibi sabit olmadığından metin satırlarının kaymasına neden olabilmektedir. Bunun sonucunda görüntü üzerinde oluşabilecek bir eğri, OCR sınıflandırıcısı tarafından eğitildiğinde büyük ölçüde zayıf sonuçlar görülecektir. Bu tür problemlerin çözümü için RAST algoritması, Hough dönüşümü, Fourier dönüşümü gibi görüntü önemseme amaçlı birçok teknik bulunmaktadır.

4.5.4. En- Boy Oranları (Aspect Ratios)

Metin farklı boyut oranlarına sahiptir. Metin, trafik işaretleri gibi kısa olabilirken diğer metinler ise video başlıkları gibi daha uzun olabilir. Metnin yeri, ölçeği ve uzunluğu, yüksek işlemsel karmaşıklığı ortaya çıkaran metni saptamak için arama prosedürüyle dikkate alınmalıdır.

4.5.5.Eğim (Tilting, Perspective Distortion):

Tarayıcılar tarafından elde edilen metin görüntüleri, sensör düzlemine paralel konumdadırlar. Ancak taşınabilir bir kamera tarafından elde edilen metin görüntüleri, genellikle düzleme paralel olmadıklarından dolayı kameradan uzak olan metin satırları, kameraya yakın olanlardan daha küçük boyutta görünürler. Bu durum fotoğrafların eğilmesine neden olur. Sonuç olarak düşük tanıma ve düşük doğruluk oranı gözlemlenmektedir.

4.5.6.Odaklanma ve Görüntü Netliği

Farklı mesafe ve açılarda çalışmak için birden fazla sayıda kamera ihtiyacı olduğundan, kullanılan kameranın odaklanma özelliği OCR sistemleri için önemli bir faktördür. Karakter tanıma, karakter segmentasyonu ve en iyi doğruluk elde edebilmek için kullanılacak kameranın odaklanma özelliğine dikkat edilmeli ve net görüntülerin kullanılması gerekmektedir.

BEŞİNCİ BÖLÜM

MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME

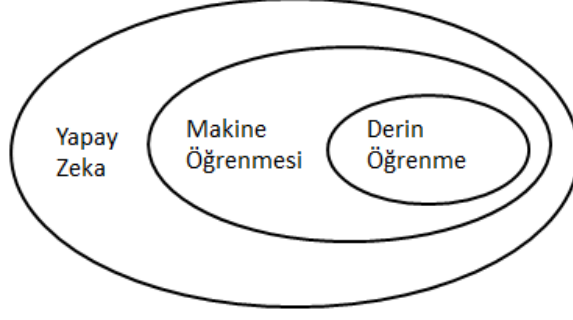
5.1 Makine Öğrenmesi:

Yapay Zeka (AI), 1950'lerde John McCarthy tarafından tanıtıldı. AI'nın amacı insan zihninin normalde yaptıklarını taklit edebilecek makinelere sahip olabilmektir. AI ile ilgili olarak iki farklı kavram vardır. Bunlardan ilki, insan seviyesindeki AI olarak da bilinen Genel AI'dır. Dünyayı algılayan ve çevresini bir insan gibi düşünen bir makine anlamına gelmektedir. İkinci kavram olan Dar AI ise makinelerin yalnızca bir insanın yaptığı belirli odaklanmış alanları taklit edebildiği anlamına gelmektedir. Bu kavramların gerçekleşmesi için ML algoritmaları tanıtılmıştır.

“Öğrenme” terimini açıklamamız gerekirse, çalışarak edinilen bir tür beceri olarak tanımlanabilir. Ayrıca bilmenin veya deneyime sahip olmanın anlamına da gelebilir. Tecrübe veya bilgi sahibi olmak insanın iki özelliğidir. Öğrenme terimini tanımlayabildiğimiz için makine öğrenmesinin, makinelerin insanlar gibi davranması amacıyla tasarlandığı veya programlandığı bir bilgisayar bilimi alanını ifade ettiğini söyleyebiliriz. Başka bir deyişle ML, geçmişte yaşanmış olan olaylara dayanarak şimdi veya gelecek zamanda daha iyisini yapmayı öğrenmekle ilgilidir.

Bilgisayara insan öğrenmesini simüle etmeyi öğreten makine öğrenmesi, Bayes Ağları, Karar Ağaçları, Kümeleme, DL gibi çok çeşitli yaklaşımları içerir. Hedefin denetimsiz öğrenmeye açık olduğu ve son zamanlarda daha fazla araştırma ve geliştirme alan, öğrenme algoritmasına etiket verilmediği için denetlenen öğrenme arasında geliştirilen makine öğrenmesi, böylece geniş çapta mevcut ve elde edilebilecek etiketler olmadan büyük verilerden yararlanmanın yolunu sunmaktadır. Şekil 5.1'de DL, ML ve AI'nın birbirleriyle olan ilişki gösterilmektedir.

Sonuç olarak, DL'in birkaç AI kavramına ulaşmak için makine öğreniminin bir dalı olduğu belirtilmektedir. Bu tezde, DL yaklaşımı çeşitli nedenlerle seçilmiştir. Kullanılacak veriler çok büyük olduğunda DL performansı artırmaya yardımcı olabilir.



Şekil 5.1: AI, ML ve DL arasındaki ilişki.

5.2 Derin Öğrenme

Bir ML dalı olan DL, verilerin kendisini kullanıp öğrenmeyi gerçekleştirerek, verinin gizli gösterimini keşfetmeyi amaçlar. Bu tezde OCR çalışması yapıldığı için DL tanımımızı görüntü tanıma konusuna dayandırmak istersek; bir makinenin bir görüntü veya nesnenin içeriğini tanımasına olanak sağlayan eğitim tekniğidir diyebiliriz.

DL'i anlamak için AI'da standart bir teknik olan ve makinelerin öğrenmesini sağlayan denetimli eğitimi yeniden gözden geçirmek gerekir. Somut olarak, bir programın bir nesneyi nasıl tanıyacağını öğrenmesi için, programı farklı sınıflara göre etiketlediğimiz birçok nesne görüntüsü ile eğitmeniz gerekmektedir. Eğitim, saatler hatta günler ve bazı durumlarda haftalar sürebilir. Eğitimi elma görüntülerini kullanarak gerçekleştirdiğimizi düşünürsek, eğitim sonunda makine, başka yeni bir elma görüntüsüyle karşılaştığı zaman elmayı tanıyabilecektir.

Bir görüntü üzerinde herhangi bir nesne veya karakter tanımaya çalışırken, nesne veya karakter üzerinde çok fazla özellik olduğu için benzersiz özellikleri bulabilmek zaman alıcı ve zor bir işlemdir. Dolayısıyla geleneksel kodlama yöntemleri bu tür problemleri etkili bir şekilde çözememektedir. Bu noktada DL, özelliklerin manuel olarak çıkarıldığı klasik makine öğreniminin aksine, ilgili özellikleri otomatik olarak ham verilerden çıkarmaktadır.

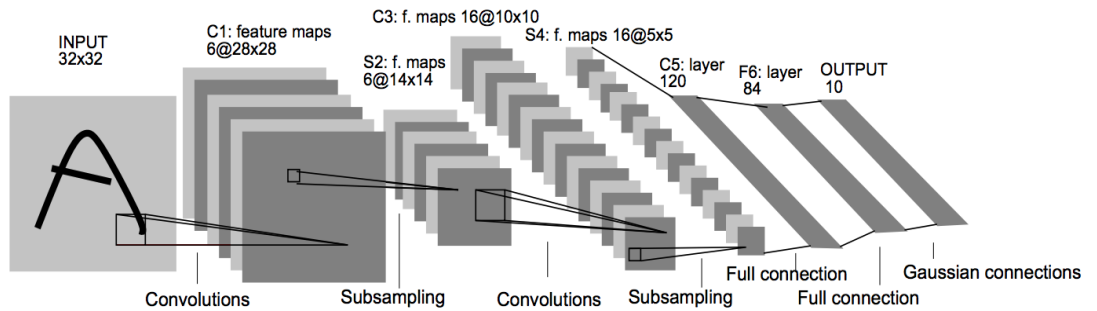
DL, birçok gizli katmana sahip sinir ağıdır ve bu gizli katmanlar binlerce veya milyonlarca olabilirler. Ağ üzerinden bir görüntü eğitildikten sonra basit kavramlardan karmaşık kavramlar oluşturabilirler. Bir görüntü DL ağında eğitildiğinde, şekil, kenarlar ve köşeler gibi basit özellikleri birleştirerek karakterler,

yüzler, arabalar, tablolar vb. nesnelere öğrenebilir. Görüntü, katmanlardan geçerken her katman birer birer giriş olarak bir sonraki katmana geçmeden önce basit bir özellik öğrenir. Katmanlar arttıkça ağ daha da karmaşık özellikleri öğrenebilir ve sonunda bunları birleştirerek görüntünün ne olduğu tahmin edilebilmektedir. DL ile veri arasında bütünleştirici bir ilişki vardır. Bu ilişkiye bağlı olarak daha fazla veri genellikle sonuçların etkinliğini artırır.

DL yöntemlerinin dikkate değer bir başarı elde ettiği en belirgin alanlar Doğal Dil İşleme (NLP), Tıbbi Araştırma, Konuşma Tanıma, Dil Çevirisi, OCR ve OR olarak belirtilebilir. Bu alanlarda alınan en gelişmiş doğruluk sonuçları ile DL yöntemlerinin önemi kanıtlanmıştır.

5.2.1 Konvolüsyon Sinir Ağları (CNN)

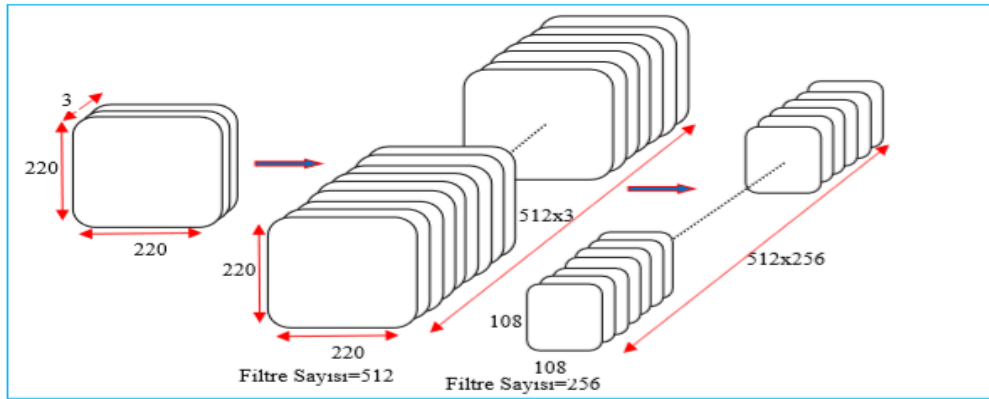
Konvolüsyon Sinir Ağları (CNN) bir derin öğrenme tekniği ve güçlü bir tür sinir ağıdır. Genellikle görüntülerdeki nesnelere tanımak, bilgisayarlı görü (CV) alanındaki sorunları çözmek için kullanılırlar. CNN'lerin kullanılmasıyla gerçekleştirilen CV ve görüntü analizi ile ilgili pratik uygulamalarda büyük başarılar elde edilmiştir. CNN mimarisinde, görüntüdeki bir nesneyi tanımak için birlikte çalışan birçok farklı katman bulunmaktadır. Şekil 5.2'de ilk CNN modeli olarak bilinen, LeNet'in (LeCun, vd.,1998) yapısı gösterilmektedir. Giriş şekli $32 \times 32 \times 1$ olarak gösterilir. Bu gösterimde; İlk değer bir filtre genişliğini, ikinci değer filtre yüksekliğini gösterir ve üçüncü değer filtre numarası anlamına gelen derinliği ifade etmektedir.



Şekil 5.2: CNN genel mimarisi.

Derinlik, toplam katman sayısı, genişlik ise konvolüsyon katmanlarında bulunan filtreler ile ilişki kavramlarıdır. Filtre sayılarının genişlik kavramı üzerinde gösterdiği

etki Şekil 5.3'te gösterilmiştir. Her bir filtre ile görüntünün bir özelliği keşfedilir. Dolayısıyla ne kadar fazla filtre kullanılırsa görüntünün o kadar fazla özelliği öğrenilir ve buna bağlı olarak da ağın genişliğinde artma meydana gelir. Şekil 5.3'te bir CNN ağının ilk üç katmanı örnek olarak gösterilmiştir. Giriş katmanında renkli görüntü boyutu $220 \times 220 \times 3$, ikinci katmandaki filtre sayısı 512 ve bir sonraki katmandaki filtre sayısı 256 olarak belirlenmiştir. Filtrelerin uygulanması sonucunda, 512×3 görüntü ikinci katmanda, 512×256 adet görüntü ise üçüncü katmanda oluşmaktadır. Bu işlem konvolüsyon katmanlarında kullanılan filtre sayısı ile orantılı olarak artmasıyla ilişkili olarak ağın genişliğini de artırır. Şekil 5.3'te gösterildiği üzere ağın derinliği 3 katman olduğundan dolayı 3'tür. Daha sonra detaylı olarak belirtilecek olan havuzlama katmanı boyut azaltma etkisi yaptığından, oluşturulacak ağda her havuzlama katmanı kullanılırsa, ağ derinliği giriş katmanındaki görüntü boyutu ile ilişkili hale gelir. Bu nedenle giriş görüntüsü boyutunun yüksek tutulması ile ağ katmanlarının sayısı artırılabilir. Bu durumun tam tersi şekilde giriş görüntü boyutu düşük olduğunda, ağdaki havuzlama katmanında uygulanan boyut azaltma işleminin sonucunda sıkı bir ağ tasarlanmış olur.



Şekil 5.3: Bir CNN ağında genişlik ve derinlik kavramları.

CNN yapısını yüz, nesne veya karakter tanıma uygulamalarına göre örneklerle açıklayacak olursak, ilk gizli katmanda nöronlar kenarlar, kulakların şekli veya burun gibi basit kalıpları öğrenirler ve daha sonra görüntü, daha belirgin özellikleri öğrenen bir sonraki katmana iletilir. Katmanlar arttıkça nöronlar, tüm yüz veya araba gibi üst düzey özellikleri öğrenmeye başlarlar. Bir katmandaki her nöron görüntünün sadece küçük bir kısmına yoğunlaşır ve bu kısımdaki aynı özelliği bir nöron arar.

Görüntünün başka bir bölümünde farklı özelliklerin aranmasını geri kalan nöronlar gerçekleştirir.

Bir CNN modelinin iki önemli avantajı vardır. Birincisi, kullanılan parametrelerin azaltılmasıdır. Birçok parametre olduğunda, çok fazla nöronun eğitilmesine ihtiyaç duyulur. Bunun sonucunda birçok sorun ortaya çıkabilmektedir. Giriş olarak yüksek çözünürlüğe sahip bir görüntü varsa, aşırı öğrenmeye (overfitting) ve çok fazla hesaplama gücü gerektirebilecek çok derin ve tam olarak bağlı katmanlara ihtiyaç olacaktır. Bununla birlikte, CNN'ler, özellikle bilgisayarla görü ile ilgili görevlerde aşırı öğrenme olasılığını azaltabilirler.

Toplam parametreler standart sinir ağları kullanılarak ve farklılıkları gözlemlemek için sırasıyla CNN tabanlı model kullanılarak hesaplanmıştır. Her ikisi de sadece bir katman için hesaplanır. Denklem 5.2.1'de CNN kullanmak yerine klasik bir sinir ağı kullanıldığında çıktıyı göstermektedir.

- $$\text{Toplam Parametreler} = (\text{filtre_yüksekliđi} \times \text{filtre_genişliđi} \times \text{filtre_derinliđi} + 1) \times (\text{yeni_yükseklik} \times \text{yeni_genişlik} \times \text{filtre_numarası}) \quad (5.2.1)$$

Ancak denklem 5.2.2'de CNN kullanıldığındaki çıktının denklem 2.2'deki çıktı ya göre parametre sayısını önemli ölçüde azalttığı görülmektedir.

- $$\text{Toplam Parametre} = (\text{filtre_yüksekliđi} \times \text{filtre_genişliđi} \times \text{filtre_derinliđi} + 1) \times \text{filtre_numarası} \quad (5.2.2)$$

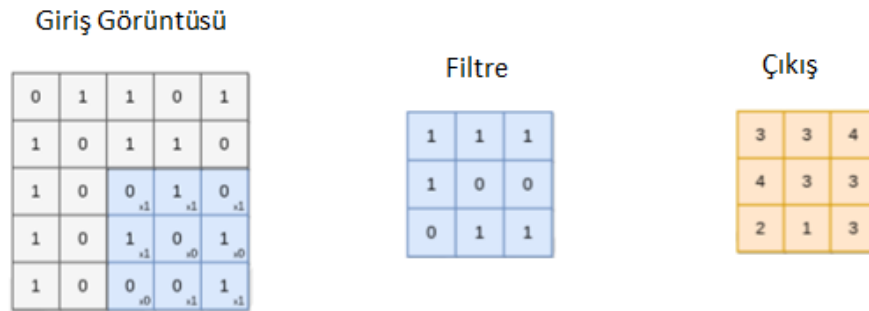
İkinci ana avantaj ise parametre paylaşımıdır. CNN ile öğrenilen parametreler bir sonraki katmana girdi olarak paylaşılır. Bu nedenle katmandaki aynı ağırlıklar tekrar kullanılır, bu da katmanlarda tekrar öğrenmeye gerek olmadığı anlamına gelmektedir. Böylece, daha sonraki katmanlar daha karmaşık özellik ve kalıpları öğrenebilir. CNN'lerin çalışmasını daha iyi anlamak için, bu ağın ana katmanları aşağıdaki bölümde detaylı olarak açıklanmıştır.

a) Konvolüsyon Katmanı

Dönüşüm katmanı olarak da bilinen bu katman, CNN temelini oluşturmaktadır. Tüm görüntü üzerinde belirli bir filtrenin dolaştırılmasıyla dönüşüm işlemi gerçekleştirilir.

Bu yüzden filtreler CNN için önemli bir bileşendir. Kullanılan filtreler 3*3, 2*2 gibi farklı boyutlarda olabilir. Filtreler, çıkış verisini oluşturabilmek için bir önceki katmandan gelen görüntüler üzerinde konvolüsyon işlemi gerçekleştirirler ve bu işlem sonucunda özellik haritası (feature map) oluşur. Özellik haritasını her bir filtrenin kendine ait olan özelliklerinin belirlendiği bölgeler olarak tanımlayabiliriz. CNN modelinde yapılan eğitim süresince filtrelerin katsayıları, eğitim kümesindeki her öğrenme yinelenmesiyle güncellenir. Burada özelliklerin belirlenmesi için, verinin hangi bölgelerinin önemli olduğu keşfedilir.

Şekil 5.4'te konvolüsyon işleminin nasıl gerçekleştiğini gösterilmektedir. Bu örnekte, giriş 5x5 değer matrisi olarak temsil edilir ve 3 × 3 boyutunda bir filtre vardır. Filtre, her pozisyonda görüntünün etrafında kaydırılır. Dolayısıyla, filtrenin değerleri, o pencere boyutundaki görüntünün değerleri ile çarpılır. Görüntünün tüm değerlerini, bu pencere boyutunda bir sayı temsil eder. Filtreler, genellikle kare şekindedirler. Verilen örnekte, pencere bir piksel kadar kayar.



Şekil 5.4: Konvolüsyon işlemi.

Filtreler sürekli güncellenir ve model tarafından öğrenilir. Standart sinir ağlarında w gibi parametreler öğrenilmiş parametrelerdir. Çeşitli filtreler modelin, bir görüntünün içeriğini oluşturmasına yardımcı olur. Konvolüsyon işleminin çıktısı genellikle orijinal görüntüden daha küçük bir boyuta sahiptir. Denklem 5.2.3, konvolüsyon işleminden sonra bir özelliğin yeni boyutunun nasıl hesaplandığı gösterilmiştir. Denklemdeki P değeri dolgu oranını (padding), S değeri de adım miktarını (stride) temsil etmektedir.

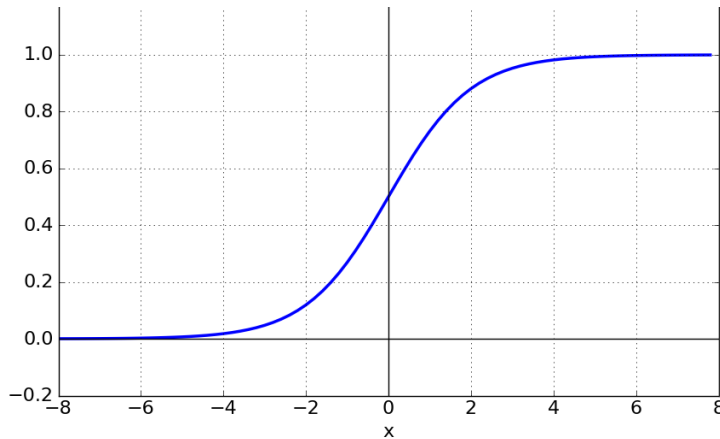
- $\text{new_height} = (\text{input_height} - \text{filter_height} + 2 \times P) / S + 1$
- $\text{new_width} = (\text{input_width} - \text{filter_width} + 2 \times P) / S + 1$ (5.2.3)

b) Aktivasyon Katmanı

Aktivasyon katmanından önce kullanılan konvolüsyon katmanında, belli başlı matematiksel işlemler uygulandığından dolayı ağ doğrusal bir yapıda olmaktadır. Bu ağı doğrusal olmayan bir yapıya dönüştürmek için aktivasyon katmanında kullanılan aktivasyon fonksiyonları uygulanır. Böylece ağ daha hızlı öğrenir. Bir sinir ağı mimarisinde aktivasyon fonksiyonlarını seçmek oldukça önemlidir. Değeri belirli bir aralıkta haritalama yapan bir fonksiyon kullanılmalıdır ve en yaygın olarak kullanılan fonksiyonlar aşağıda anlatılmaktadır.

Bunlardan ilki, Sigmoid fonksiyonudur. Giriş değerlerini 1 ile 0 arasında alan sigmoid fonksiyonu, sınıflandırma problemleri için yaygın olarak kullanılır. Denklem 5.2.4 ve Şekil 5.5'te Sigmoid işlevi gösterilmiştir.

$$A(x) = \frac{1}{1 + e^{-x}} \quad (5.2.4)$$



Şekil 5.5: Sigmoid işlevi grafiği.

Verilen grafiğe baktığımızda fonksiyonun uçlarına doğru, y değerleri x'teki değişikliklere çok az tepki verdiğini görebiliriz. Dolayısıyla bu bölgelerde türev değerleri 0'a yakınsar ve çok küçük olur. Bu olaya gradyanların kaybolması veya ölmesi (vanishing gradient) denilmektedir. Ayrıca bu durumda öğrenme olayı

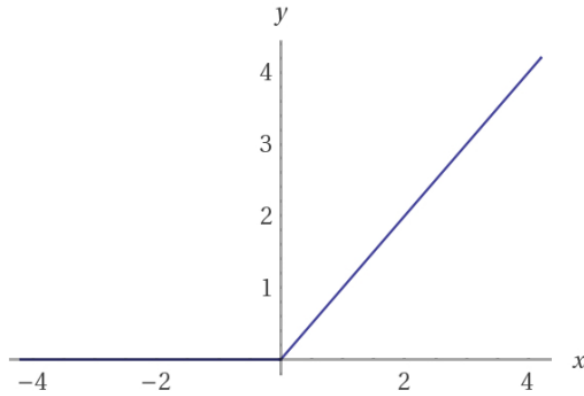
minimum düzeyde gerçekleşir. Türev değeri 0 olduğunda ise öğrenme gerçekleşmez. Yavaş bir öğrenme olayı olduğunda hatayı minimuma indirgeyen optimizasyon algoritması yerel minimum değerlere takılabilir ve modelden alınabilecek maksimum performans elde edilmeyebilmektedir.

Softmax, çoklu kategoriler için Sigmoid aktivasyon fonksiyonunun bir genellemesidir. Tamamen bağlanmış son katmanın çıktısını, girdi olarak alır ve toplamı bire eşit olacak şekilde olasılıklar ile eşleştirir. Böylece, Softmax çıktıları normalleştirir ve toplamı 1'e eşitler. Denklem 5.2.5'te Softmax fonksiyonunun formülü verilmiştir. K, toplam kategori sayısını belirtirken, j ise kategoriler arasındaki tek bir sayıyı ifade eder.

$$A(x_j) = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}} \quad (5.2.5)$$

Bir diğer aktivasyon fonksiyonunda Düzleştirilmiş Doğrusal Birim Katmanı (Rectified Linear Units Layer - ReLu)'dur. ReLu, negatif değerleri sıfıra çekerek giriş verisini etkilemektedir. Çoğu araştırmada aktivasyon fonksiyonu olarak tercih edilmektedir. Şekil 5.6 ve denklem 5.2.6'da ReLU'nun işlevi gösterilmektedir.

$$A(x) = \max(0, x) \quad (5.2.6)$$



Şekil 5.6: Çıkış verisi üzerindeki ReLu etkisi.

ReLU $[0, +\infty)$ aralığında değer almaktadır. Çok fazla nöronlu büyük bir sinir düşündüğümüzde, ağdaki bazı nöronların aktif olmasını sağlayarak verimli bir

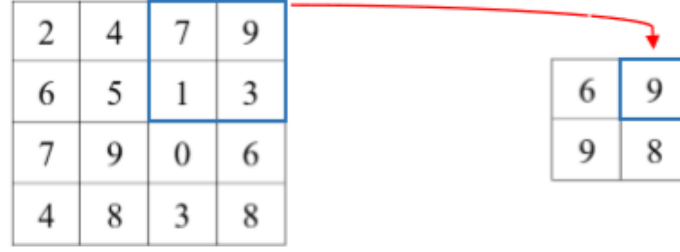
hesaplama yükü oluşturulması ReLU ile sağlanmış olur. Sayı negatif olursa, çıkış sıfır olur fakat sayı negatif olmazsa çıktıdaki değer korunur. Bu da ağıın daha hızlı çalışacağı anlamına gelmektedir. ReLU'nun hesaplama yükünün diğer aktivasyon fonksiyonlarına göre az olması ve gradyan kaybolmasını engellemesi, çok katmanlı ağlarda daha çok tercih edilmesine neden olmuştur.

c) Havuzlama Katmanı (Pooling Layer):

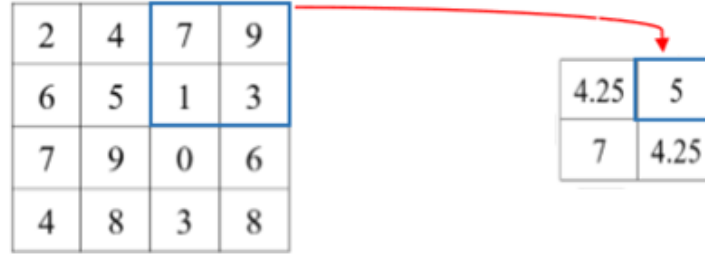
Amacı bir sonraki katmanda kullanılacak olan giriş boyutunu (genişlik*yükseklik) azaltmak olan havuzlama katmanı, konvolüsyon katmanı gibi çalışmaktadır. Burada belirtilen boyut azaltma işlemi, bilgi kayıplarının yaşanmasına neden olmaktadır. Ancak oluşacak bu kayıplar, sonraki ağ katmanları için daha az hesaplama yükü oluşturma ve modelin aşırı öğrenme veya ezberlemesini önleme gibi faydalar sağlamaktadır.

Konvolüsyon işleminde uygulanan filtreleri havuzlama katmanında da kullanılmaktadır. Bu filtreler, belli bir adım atma değerine göre tüm görüntü üzerinde gezdirilir ve görüntüdeki piksellerin değerleri alınarak işlem yapılır. Piksel değerleri, maksimum veya ortalama değerlerinin alınmasıyla iki farklı şekilde hesaplanabilir. Maksimum piksel değerleri alındığında maksimum havuzlama, ortalama piksel değerleri alındığında ortalama havuzlama isimleri kullanılır. Daha iyi performans gösterdiği ve dolayısıyla genellikle maksimum havuzlama işlemi tercih edilmektedir. Ayrıca maksimum havuzlamada, bir görüntünün en önemli özellikleri korunurken parametreler azaltılabilmektedir.

Konvolüsyon katmanı sonucunda oluşan filtre sayısının görüntülerin her biri için havuzlama işlemi gerçekleştirilir. CNN'lerde havuzlama katmanı isteğe bağlıdır ve bazı modellerde bu işlem gerçekleştirilmeyebilir. Şekil 5.7'de, 2 adımla yapılan maksimum havuzlama işlemi ve Şekil 5.8'de ortalama havuzlama işlemi gösterilmiştir.



Şekil 5.7: Maksimum havuzlama işlemi.



Şekil 5.8: Ortalama havuzlama işlemi.

d) Tam Bağlantılı Katman (Fully Connected Layer)

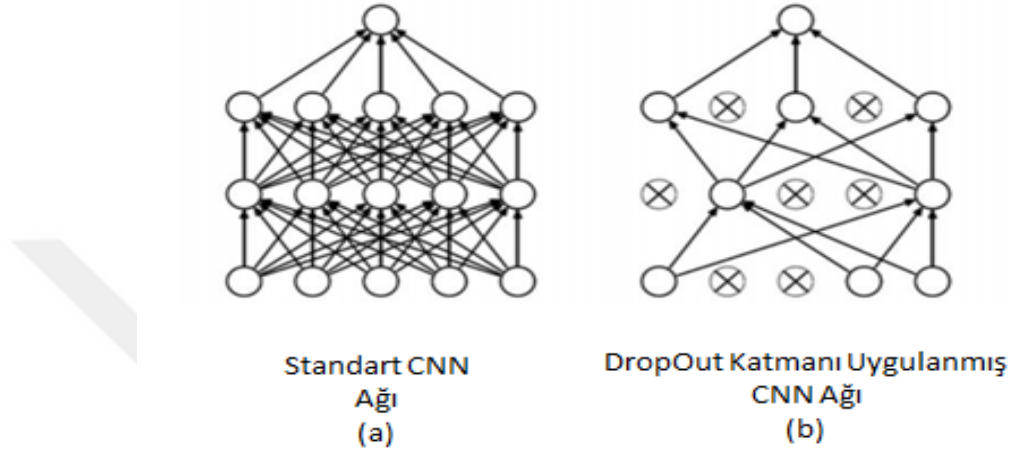
CNN mimarisini oluşturan ve art arda gelen konvolüsyon, ReLu ve havuzlama katmanlarından sonra gelen tam bağlantılı katman, kendinden önceki katmanın tüm alanlarına bağlı şekildedir. Bu sebepten dolayı bu katmana tam bağlantılı katman denilmektedir. Bu mimarinin sonunda, çıktının eğitim ve tahmin için kullanılacak tam bağlantılı bir katmana düzleştiği söylenebilir. Farklı mimarilerde bu katmanın sayısı değişiklik gösterebilmektedir.

e) Dropout

Dropout, derin öğrenmede en çok kullanılan ağ düzenleme tekniklerinden biridir (Srivastava, vd.,2014). CNN'de büyük veriler kullanılarak eğitim yapıldığında bazı durumlarda ağ ezberleme yapar. Bu durumu önlemek için Dropout katmanı kullanılmaktadır.

Temel olarak kullanılan mantık, ağın bazı düğümlerinin kaldırılmasıdır. Düğümler kaldırılırken rastgele eleme yöntemi veya bir eşik değeri belirlenerek eleme yapılır. Eşik değeri kullanıldığında genellikle 0,1 aralığında bir değer tanımlanmaktadır. Tüm katmanlarda aynı dropout eşik değerinin kullanılması zorunlu değildir. Probleme ve veri setine bağlı olarak bu değer değişiklik gösterebilir. Dropout,

özellikleri öğrenerek bir modeli, rastgele durumlardan daha sağlam hale getirmeyi sağlamaya çalışır. Burada en önemli nokta, eğitim sırasında nöronlara bağımsızlık kazandırarak sistemin daha iyi verim elde etmesini sağlayabilmektir. Şekil 5.9(a)'da ağın orijinal yapısı, (b)'de Dropout katmanının uygulandığından sonraki hali gösterilmiştir.



Şekil 5.9: Orijinal ağ ve DropOut katmanı uygulanmış ağ yapısı.

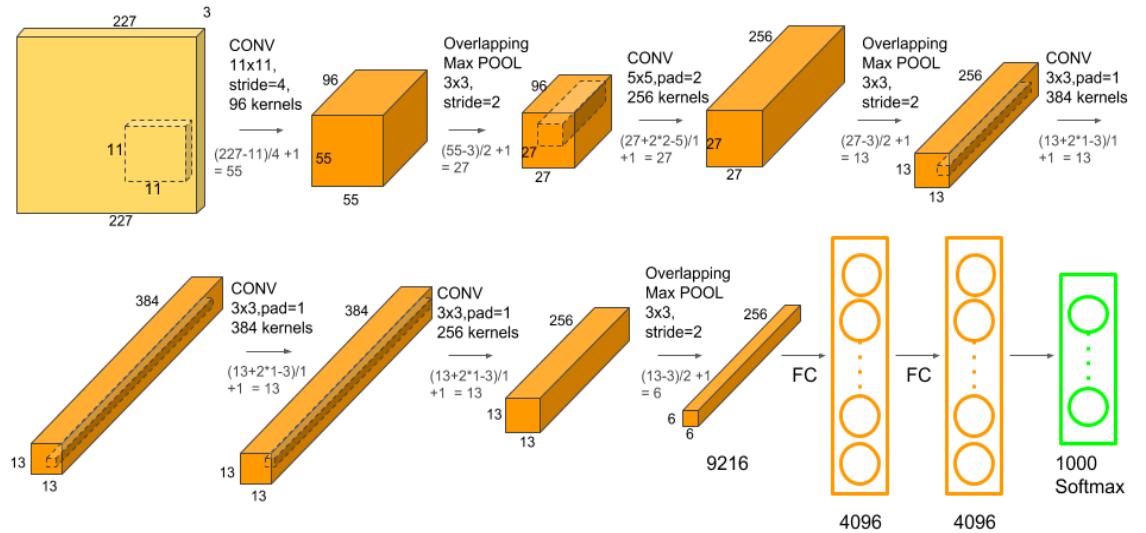
5.3 CNN Modelleri :

Derin öğrenmenin 1998 yılında Yann LeCun'nun yayınlamış olduğu makale (Lecun, vd.,1998) ile ilk uygulama yapıldığı söylene de, bilim dünyasındaki ilk büyük etki, 2012 yılında nesne tanımlama alanında en büyük yarışma olan, Büyük Ölçekli Görsel Tanıma Yarışması (ILSVRC 2012) ile meydana gelmiştir. Aynı yıl yarışma, %26,1 olan Top-5 hata oranı %15,3 gibi bir orana düşüren ve derin öğrenmede temel mimari kabul edilen CNN, Krizhevsky, Sutskever ve Hinton (2012) ile kazanılmıştır. Elde edilen bu sonuçlar sayesinde DL alanında yükseliş yaşanmış ve çalışmalar popüler hale gelmiştir. Devam eden yıllardaki yarışmalarda katılımcıların çoğu modellerini derin öğrenme mimarisine göre tasarlamayı tercih etmişlerdir. Modellerin her biri derin öğrenmenin temel taşları olarak kabul edilmektedir. ILSVRC yarışmasını kazanan modellere ek olarak, görüntü içindeki nesne tanımlama için geliştirilmiş olan modeller ile ilgili bilgiler bu kısımda anlatılmaktadır.

5.3.1 Alex Net

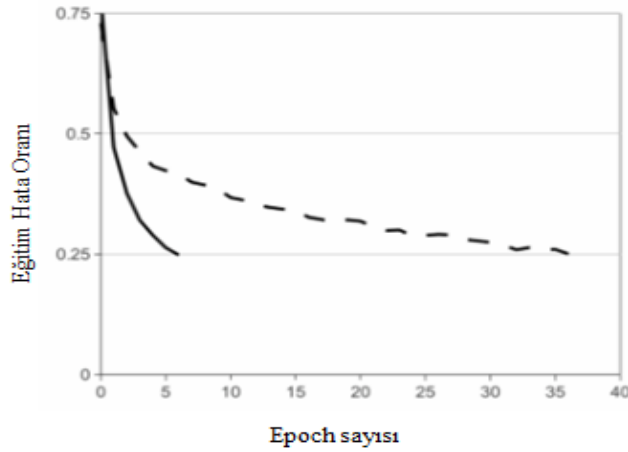
2012 yılında gerçekleştirilen ILSVRC yarışmasının kazananı olan AlexNet modeli, "ImageNet Classification with Deep Convolutional Networks" isimli makale ile yayınlanmıştır(A.g.e.).Mayıs 2019 tarihi itibari ile 40022 alıntı yapılan mimari ile birlikte bilgisayarlı OR hata oranı %26,2'den %15,4'de düşmüştür. Şekil 5.10'da yapılan çalışmanın mimarisi gösterilmektedir. 1000 nesneyi sınıflandıracak şekilde tasarlanan mimari, 5 konvolüsyon, havuzlama ve 3 tam bağlantılı katmandan oluşmaktadır. Adım kayma sayısı (stride) 4 olacak biçimde belirlenen 11x11 boyutunda filtreler kullanılmıştır.

AlexNet'in tek bir konvolüsyon katmanında, genellikle aynı büyüklükte birçok kernel (çekirdek) vardır. Örneğin, AlexNet'in ilk konvolüsyon katmanında 11x11x3 boyutunda 96 tane kernel bulunmaktadır. Kernel genişliği ve yüksekliği genellikle derinliklerin kanal sayısı ile aynı olmaktadır. İlk iki konvolüsyon katmanından sonra maksimum havuzlama katmanları gelmektedir. Bu katmanlar, üçüncü, dördüncü ve beşinci konvolüsyonlu katmanlar ile doğrudan bağlanır. Beşinci konvolüsyon katmanını, çıktısı tamamen birbirine bağlı iki katmanın içine giren ve üst üste gelen bir maksimum havuzlama katmanı izler. İkinci tamamen bağlı katmanında, 1000 sınıf etiketli softmax sınıflandırıcısı ile eğitim yapılır.



Şekil 5.10: AlexNet Mimarisi.

AlexNet'in önemli bir özelliği de ReLU aktivasyon fonksiyonunun kullanılmasıdır. Bu modelde derin CNN'lerin, tanh veya sigmoid gibi aktivasyon fonksiyonlarına göre çok daha hızlı bir şekilde eğitilebileceğini ortaya konulmuştur. Şekil 5.11'de ReLu aktivasyon fonksiyonunun, CIFAR-10 veri seti üzerinde test edilerek epoch sayısı ile eğitim hata oranı arasındaki ilişki gösterilmektedir.

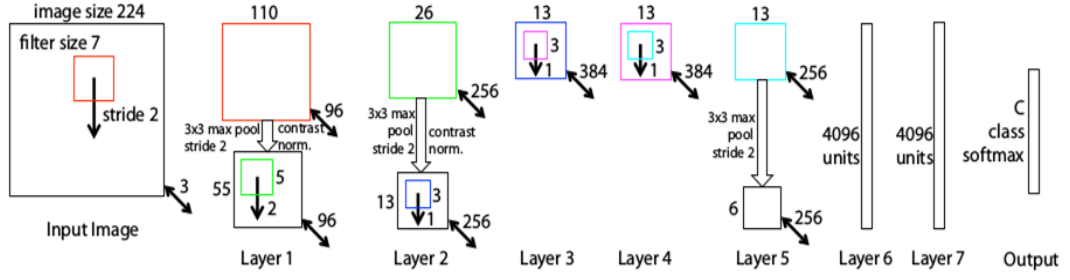


Şekil 5.11: ReLu fonksiyonunun CIFAR-10 veri setindeki etkisi.

5.3.2.ZF Net

Bu mimari AlexNet mimarisinin geliştirilmiş hali olan ZF Net, Matthew Zeiler ve Rob Fergus tarafından tasarlanmış ve 2013 yılında ILSVRC yarışmasının kazananı olmuştur (Zeiler ve Fergus,2014). Bu model ile OR'da hata oranı %11,2'ye indirilmiştir ve Şekil 5.12'de model mimarisi gösterilmektedir.

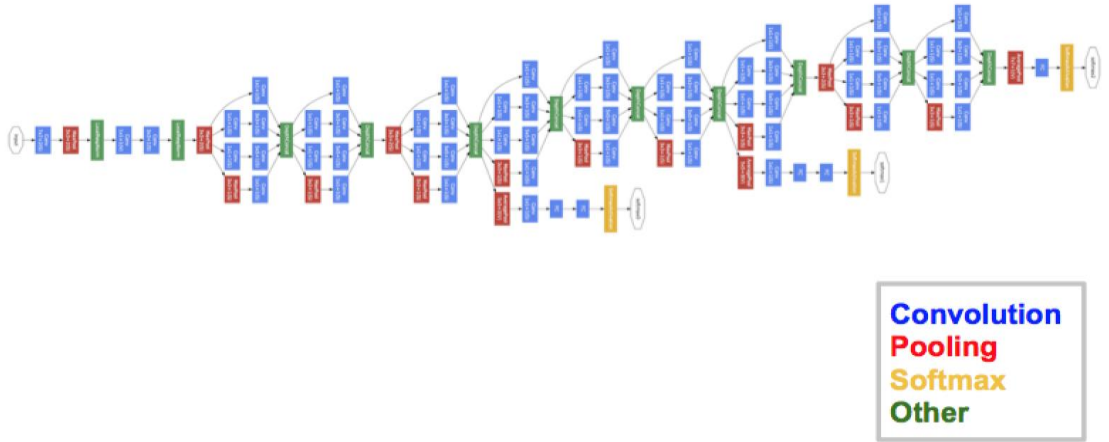
ZF Net modeli, ilk katman için AlexNet'in uyguladığı 11x11 boyutlu filtrelerden farklı olarak 7x7 boyutundaki filtreleri ve havuzlama katmanında uygulanan 2 adım kayma miktarını kullanmıştır. Burada yapılan değişikliğin nedeni, ilk konvolüsyon katmanına kullanılacak daha küçük bir filtre boyutunun, giriş boyutundaki birçok orijinal piksel bilgisinin korunmasında yardımcı olacak olmasıdır. Aktivasyon fonksiyonu için ReLu, hata fonksiyonu için Cross-Entropy Loss ve eğitim için Stochastic Gradient Descent kullanılmıştır. Bir GTX 580 GPU üzerinde on iki gün süren bir eğitim yapılmıştır.



Şekil 5.12: ZF Net Mimarisi.

5.3.3 GoogLeNet

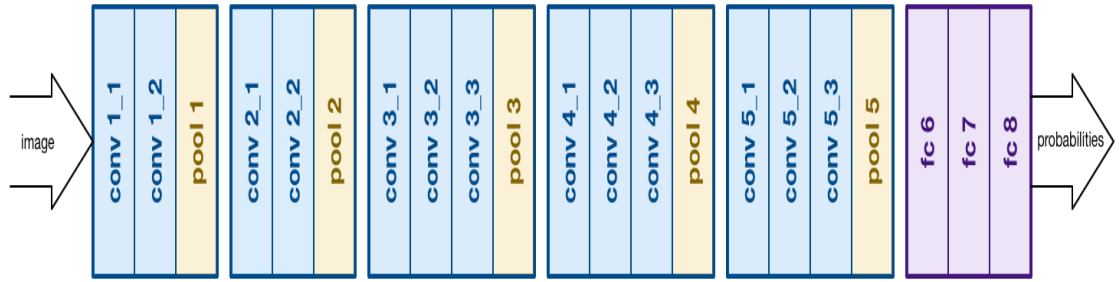
2014 ILSVRC yarışmasının kazananı GoogLeNet olmuştur (Szegedy, vd., 2015). Bu modelde 22 katman bulunur ve %5.7 oranında top-5 hata oranı elde edilmiştir. LeNet mimarisinden etkilenecek ona benzer bir CNN kullanılmış, ancak farklı olarak da başlangıç modülü olarak isimlendirilen yeni bir öge uygulanmıştır. Bu mimari genel olarak, ardışık bir yapıda konvolüsyon ve havuzlama katmanlarını art arda getirmeyen ilk CNN mimarilerinden birisidir. Katmanların her birini yığnlamak ve çok sayıda filtre eklemek, hesaplama ve bellek maliyeti oluşturur. Ayrıca ezberleme olasılığını da artırdığından dolayı bu yeni model bellek ve güç kullanımı üzerinde önemli bir yere sahiptir. GoogLeNet’de, bu gibi problemlerin çözümü için paralel olarak birbirine bağlı modüller kullanılmıştır. Bu model, parametre sayısını büyük ölçüde azaltmak için çok küçük çeşitli konvolüsyonlara dayanmaktadır. Şekil 5.13’te GoogLeNet’in mimarisi gösterilmektedir.



Şekil 5.13: GoogLeNet mimarisi.

5.3.4.VGG Net

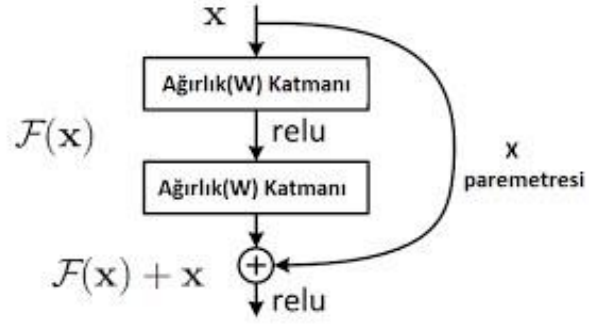
ILSVRC 2014 yarışmasında ikinci olan ağ mimarisine VGGNet adı verilmiş ve Simonyan ve Zisserman (2014) tarafından geliştirilmiştir. VGGNet, 16 evrimli katmandan oluşur ve oldukça düzgün bir mimariye sahiptir. AlexNet'e benzemektedir ve sadece 3x3 konvolüsyon içermesine rağmen çok sayıda filtre bulunur. 2-3 hafta boyunca 4 GPU'de eğitimi sürdürülmüştür. Resimlerden özellik çıkarmak için toplulukta en çok tercih edilen mimaridir. VGGNet'in ağırlık yapılandırması halka açıktır ve temel özellik çıkarıcı olarak birçok diğer uygulama ve zorluklarda kullanılmıştır. Bununla birlikte, VGGNet, ele alması biraz zor olabilen 138 milyon parametreden oluşur. Diğer modellere benzer şekilde matrislerin yükseklik ve genişlik boyutları girişten çıkışa doğru azalırken, derinlik değeri artmaktadır. VGGNet'in mimarisi Şekil 5.14'te verilmiştir.



Şekil 5.14: VGGNet mimarisi.

5.3.5.ReSNet

Diğer tüm mimarilerden farkı, daha derin olarak tasarlanması olan ResNet, 152 katmandan meydana gelmektedir (He, vd.,2016). Aynı zamanda %3,6 hata oranı ile 2015 ILSVRC yarışmasının kazananı olmuştur. ResNet mimarisi Residual bloklardan oluşmaktadır. Residual bloğunda x girişi, konvolüsyon -ReLU-konvolüsyon serisinden sonra bir $F(x)$ sonucuna ulaşmaktadır. Daha sonra elde edilen bu sonuç, orijinal x girişine eklenir ve $H(x) = F(x) + x$ denklemi şeklinde ifade edilir. Residual blokları Şekil 5.15'te gösterilmiştir.



Şekil 5.15: Residual blokları.



ALTINCI BÖLÜM

YAPILAN ÇALIŞMA

Bu bölümde iskelet temsili ile ilgili detaylı bilgi verilir, tez çalışmasında kullanılan yöntem olan Çember Dolgu Oranı yöntemi açıklanarak literatürde yer alan diğer örnek iskelet yöntemi çalışmalarından bahsedilmiştir.

6.1. İskelet Temsili

İskelet, görüntüdeki istenmeyen pikselleri silerek iskelet çıkarma işlemidir. Başka bir deyişle siyah arka plan piksellerini bir piksel genişlik iskeleti elde edilinceye kadar katman katman silen morfolojik işlemidir. Genellikle siyah (arka plan) ve beyaz (ön plan) piksellerden oluşan ikili görüntülere uygulanır. Girdi ikili bir resim olarak alınır ve çıktı olarak başka bir ikili resim üretilir. Bir iskelet algoritmasının etkili olması için, görüntüleri ince nesnelere indirgemeli, topolojik ve geometrik özellikleri de korumalıdır. Bununla birlikte, iyi bir iskelet algoritması aşağıdaki özelliklere sahip olmalıdır:

- Sonuçta ortaya çıkan iskeletler bağlantıyı sağlamalıdır.
- Elde edilen iskeletler birim piksel genişliğinde olmalıdır.
- Aşırı piksel silme işlemi yapılmamalıdır.
- Uygulama süresi açısından daha iyi performans göstermelidir.

İskelet temsili, son yıllarda birçok görüntü işleme uygulamasında önemli bir adım olarak sayılmaktadır. İşlenmesi gereken veri miktarını, önemsiz özellikleri, istenmeyen gürültüyü ve işlem süresini azaltmak, birleşme noktaları ve bileşenler arasında bağlantı gibi önemli özellikleri çıkarmak için birçok uygulamada yardımcı olabileceğinden dolayı dijital görüntüler üzerinde iskelet işlemi uygulamak gerekmektedir. OCR, görüntü tanıma, parmak izi sınıflandırması, kimlik doğrulama, imza doğrulama, tıbbi görüntüleme gibi birçok çeşitli görüntü işleme uygulamalarında kullanılmaktadır.

6.2. Literatürdeki Çalışmalar

Daha önceki yapılan OR veya OCR çalışmalarında iskelet temsili uygulamaları bulunmaktadır. Blum ve Roger (1978) tarafından tanımlanan medyan eksen yöntemi, eşleştirme, sınıflandırma gibi şekil analizi görevlerinde kullanılmıştır. İskelet temsili,

başka bir deęişle şeklin medyal eksenini olarak isimlendirebilir ve şekli tamamen içine alabilen en büyük yarıçaplı dairelerin merkez noktalarını birleştiren eğriden oluşur. Eğri noktaları, şeklin kenarlarına en az iki tane farklı noktada değen ve tamamen şeklin içinde kalan maksimum dairelerin merkezlerinden oluşan iskelet noktaları şeklinde de ifade edilebilir.

Nesne silüetlerine dayalı şekil tanıma uygulanan ve iskelet bitiş noktaları arasındaki yollarının karşılaştırılmasıyla yeni bir iskelet grafiklerini eşleştirebilen algoritma Bai ve Latecki (2008) tarafından ortaya konulmuştur. Algoritma, benzer görsel yapıdaki iskelet grafiklerinin tamamen farklı topolojik yapıya sahip olmasına dayandırılmıştır. Deneysel sonuçlar, önerilen algoritmanın, eklemelerin ve kontur deformasyonlarının varlığı ile birlikte doğru sonuçlar verebildiğini göstermiştir.

Quehl, Sack ve Yang (2012) tarafından yapılan çalışmada, video verilerindeki metinlerin farklı çözünürlüklerde ve karmaşık bir görüntü halinde olabilmelerinden dolayı OCR sonuçlarında zorlu kontrast oranlarına neden olma problemi için yeni bir ikilileştirme yöntemi önerilmiştir. Bu yöntemde metnin eğim yönü, görüntü iskeletlerinin içerik dağılımını analiz edilerek yaklaşırlır. Ardından metnin temel pikselleri, iskelet piksellerinin ortalama gri skala değeri hesaplanarak seçilir. Bu işlem sonucunda video metni görüntüleri standart bir OCR sistemi tarafından okunabilir formata dönüştürülür. Son olarak, metin piksellerini elde etmek için otomatik bir temel bölge büyütme algoritması uygulanır. Deneysel sonuçlar önerilen yaklaşımın video metin görüntülerini tanımak için diğer referans yöntemlerini gerçekleştirdiğini göstermektedir.

İkili görüntülerin bağlantısını, topolojisini ve hassasiyetini korumak için gerekli olan iskeletleştirme algoritmalarında piksel silme kriterlerinin özelliklerini içeren Goyal ve Luthra (2016) tarafından yapılan çalışmada, farklı iskelet temsili algoritmalarının performansı, inceltme hızı, bağlı bileşenlerin sayısı, uygulama süresi gibi farklı parametrelere göre karşılaştırmalar gerçekleştirilmiştir. Yapılan bu karşılaştırmalar, bağlı bileşenlerin sayısı, uygulama süresi ile birlikte Zhang ve Suen (1984) algoritmasının Ortalama Kare Hatası (Mean Square Error) ile Guo ve Hall (1989) algoritması üzerine odaklanmıştır. Deneysel karşılaştırmalarla Zhang ve Suen algoritmasının bağlantı, ortalama kare hata oranı ve uygulama süresi açısından Guo ve Hall algoritmasından daha iyi olduğu sonucuna varılmıştır.

Chatbri ve Kameyama (2014) inceltme algoritmalarını gürültüye karşı dayanıklı hale getirmek için ikili çizim resimlerinin inceltilmesinde ölçek alanı filtrelemesi uygulamışlardır. Bu çalışmada, bir girdi görüntüsünün birden fazla filtreleme ölçeğinde birden fazla temsili türetilir. Daha sonra gürültü giderme ve şekil bozulma arasında en iyi dengeyi sağlayan filtreleme ölçeği seçilir. Ölçek seçimi, yüksek miktarda filtrelemenin ve şekil bozulmalarının neden olduğu fazladan kusurları (yedek dallar ve çizgiler) tespit eden bir performans ölçüsü kullanılarak yapılır. Ek olarak bu çalışmanın, eşleştirme performanslarını iyileştiren bir ön işleme ve normalizasyon adımı olarak fayda gösterdiği gözlemlenmiştir.

6.3. Çember Dolgu Oranı Yöntemi

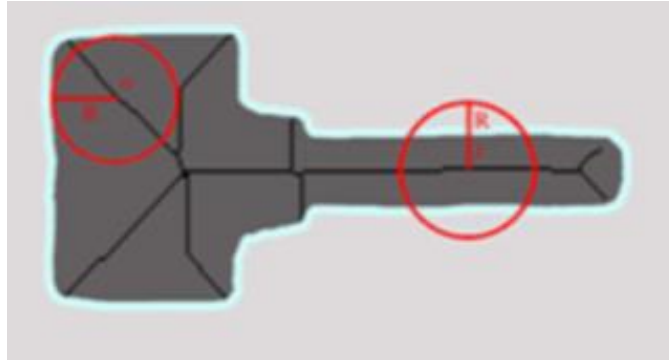
Şekil tanımlayıcıları genellikle şeklin sınırları veya iç özellikleri üzerinden tanımlanır. Tanımlayıcılar bir veri tabanı içerisindeki benzer şekilleri belirleme potansiyellerine göre değerlendirmektedirler. Önerilen tekniklerin etkinliğini şekiller üzerindeki temsil gücüne göre değerlendirmek yeterli değildir. MPEG-7, şekil tanımlayıcılarının etkinliğini değerlendirmek üzere bulma doğruluğu, genellenebilirlik, kompakt uygulanabilirlik, sağlamlık, düşük hesaplama karmaşıklığı, ve kabadan inceye doğru hiyerarşik detay temsili gibi kriterler belirlemiştir (Kim H ve Kim J, 2000).

Tanımlayıcılar genellikle iki grupta incelenirler (Zhang ve Lu, 2004). Bunlardan ilki olan sınır tabanlı tanımlayıcılar, isminden de kolay anlaşılacağı gibi şeklin sınırlarını kullanarak işlem yaparlar ve sürekli (global) yaklaşım, ayrık (structural) yaklaşım olmak üzere iki temel gruba ayrılırlar. Sürekli yaklaşım şekil kenarlarını bölmeden işlem yaparak şekli tek bir bütün halinde değerlendirirken, ayrık yaklaşım ise şekil kenarlarını parçalı (primitive) olarak kullanmaktadır. İkinci tanımlayıcı grubu olan alan tabanlı tanımlayıcılar, sınır tabanlı tanımlayıcıların aksine şeklin tüm noktalarını kullanırlar.

Literatürdeki birçok şekil tanıma uygulamaları ve kullanılan tanıma algoritmaları incelendiğinde, bu uygulama ve algoritmaların şekil yön değişimlerine karşı duyarlı olduğu görülmektedir. Şirin ve Demirci (2013) tarafından yapılan çalışmada sunulan Çember Dolgu Oranı yöntemi, iskelet koordinatları ve bu noktalar kullanılarak oluşturulmuş, yerel şekil doldurma oranı değerleri kullanarak çeşitli nesnelere tanınması işlemini gerçekleştirmektedir. Bu yöntem, iskelet noktalarını çember

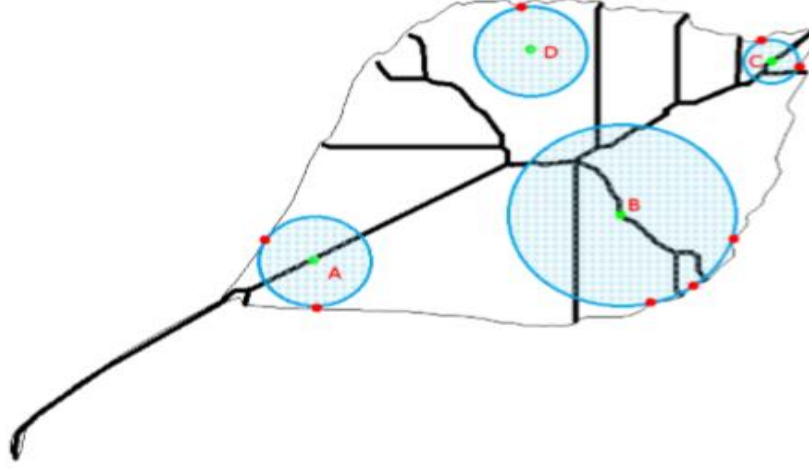
merkezleri olarak alıp, bu noktalardan çizilen tarama çemberleri ile şeklin örtüşme oranlarını nokta tanımlayıcı olarak alan bir yöntem olduğundan dolayı şekil yarıçaplarına göre oranlar belirlenmektedir. Buna bağlı olarak şekillerin büyüklüklerinden ve pozisyon değişikliklerinden bağımsız bir yöntem oluşturulmuştur. Ayrıca MPEG7 tarafından şekil tanımlayıcılarının etkinliğini değerlendirmek için belirlenen kriterlere uygun bir yöntemdir.

Sunulan çalışmada ilk olarak farklı boyutlardaki resimlerin normalizasyon işlemi ile belirli boyuta getirilmesinden sonra iskelet noktaları belirlenmiştir. Belirlenen her bir iskelet noktası için x ve y koordinat bilgisi kullanılarak o iskelet noktasının etrafına çizilebilecek, şeklin içerisinde kalan en büyük dairenin r yarıçap değeri hesaplanmıştır. Bu hesaplama ile veri tabanında bulunan şekillerin tümü içinden maksimum yarıçap R değeri elde edilmiştir. İskelet noktasının tanımına göre, bu noktayı merkez kabul eden r yarıçap değerindeki çemberin tüm noktaları şeklin içinde kalmaktadır. Aşağıda verilen Şekil 6.1’de görüldüğü üzere şekildeki iskelet noktaları için hesaplanan maksimum yarıçap değeri R ve şekildeki anahtar resminin uç bölgesinde çizilen r yarıçaplı çemberin r değeri, R değerine kadar artırılıp şekil içinde kalan noktalar hesaplanmıştır. Hesaplanan bu noktalar doğrultusunda şekil üzerindeki değişimler gözlemlenmiştir.



Şekil 6.1: Anahtar resmi üzerinde maksimum yarıçap gösterimi (Şirin ve Demirci, 2013).

Bir şekil üzerindeki her nokta iskelet noktası olarak kabul edilememektedir. İskelet noktası olarak kabul edilebilmesi için şeklin kenarlarını en az iki farklı noktada dokunması gerekir. Şekil de gösterilen A B ve C noktaları en az iki farklı noktada şekle dokunduğu için iskelet noktası olarak kabul edilmiş, D noktası ise sadece bir tane noktada şekle dokunduğu için iskelet noktası olarak kabul edilmemiştir.



Şekil 6.2: A,B ve C iskelet noktaları olan yaprak şeklinin iskelet temsili. (Şirin, 2016).

Şekil6.2’de kabul edilen A,B ve C iskelet noktalarının eşitlik üzerinde gösterimi aşağıdaki gibidir;

$$p_i = \{x_i, y_i, r_i\} \quad (6.3.1)$$

Eşitlikteki p_i değeri iskelet noktasını, (x_i, y_i) koordinat noktalarını, r_i değeri ise o noktadaki yarıçap değerini ifade etmektedir.

6.3.1. İskelet Zenginleştirme Değeri e

Hesaplamalarda kullanılması için iskelet temsili oluşturulan şekil veya karakterlerin üzerine çizilebilecek dairelerin en büyük yarıçapa sahip olanı saklanmaktadır. Bu işlemin gerçekleştirilmesiyle yarıçap, her iskelet noktası için kendi yarıçapından, saklanan maksimum yarıçap değerine kadar artırılır ve artırılan bu yarıçap değerlerine göre daireler çizilir. Çizilen bu daireler şeklin içerisinde kalabilir veya şeklin sınırlarından taşabilir. İskelet üzerindeki o noktaya çizilen dairenin içerisinde kalan toplam piksel sayısı ile şekil içerisinde kalan alandaki piksel sayılarının oranlanmasıyla zenginleştirilmiş iskelet değeri e hesaplanmaktadır. Bu değer formülasyonu aşağıda gösterilmektedir:

$$e^i = \sum_{k=r_i}^{R+\varepsilon} \frac{n_i^k}{N_i^k} \quad (6.3.2)$$

Buradaki k değeri yarıçapı, N_i^k yarıçapındaki daireyi oluşturan toplam piksel sayısı, n_i^k şekil içindeki piksel sayısı, R değeri ise şekle ait olan iskelet içinde bulunan maksimum yarıçap değerini ve ε değeri küçük pozitif bir sabiti ifade etmektedir.

Hesaplanan e değeri ile şekil iskelet noktaları içerisine yeni değerler eklenir ve şeklin iskeleti zenginleştirilir. Bu sayede orijinal iskelet temsilinden daha farklı ve ayırt edici bir özelliğe sahip olmaktadır. Zenginleştirilen iskeletler (x, y, r, e) değerlerinden oluşan 4 boyutlu vektörler şeklinde temsil edilebilirler. Temsil edilen bu gösterim sayesinde iskeletler ve şekiller arasındaki ilişki, aynı yarıçap değerlerine sahip olan alanlara karşılık gelen farklı e değerleri ile ayırt edilebilmektedir.

Bu tezde, oluşturulan CNN modelinde ilk olarak veri setindeki karakter görüntüleri kullanılmış, ikinci olarak görüntülerin iskelet temsilleri oluşturulmuş ve bu görüntülerin kullanılmasıyla yapılan eğitim sonuçlarının tanıma oranlarına ulaşılmıştır. Daha sonra önerilen Çember Dolgu Oranı yönteminin veri setindeki karakter görüntüleri üzerine uygulanmasıyla elde edilen iskelet noktaları hesaplanmıştır. Veri setindeki her görüntü için hesaplanan bu değerlere PCA uygulanarak boyut azaltma işlemi gerçekleştirilmiştir. Bu işlem sonunda elde edilen vektör değerlerinin görüntü olarak kullanılabilmesi sağlanmış ve tekrar görüntü çizimleri yapılmıştır. Oluşan görüntüler, aynı CNN modelinde eğitilerek tanıma oranı belirlenmiştir. Ulaşılan tanıma oranlarına göre, hem CNN modelinin hem de önerilen yöntemin tanıma oranı üzerindeki etkileri gözlemlenmeye çalışılmıştır.

YEDİNCİ BÖLÜM

DENEYSEL ÇALIŞMA

Bu bölüm, çalışmada kullanılan veri seti üzerinde yapılan işlemleri ve uygulama süreçlerini açıklamaktadır.

El yazısıyla ve farklı yazı tiplerinde yazılan rakam ve karakter tanıma problemini çözmek için veri seti, ön işleme, öznitelik çıkarma, iskelet noktalarının belirlenmesi, PCA ile boyut azaltma ve oluşturulan CNN modelinde uygulanan eğitim aşamalarından oluşan bir yaklaşım izlenmiştir. Bu aşamalar aşağıdaki bölümlerde adım adım detaylı olarak anlatılacaktır.

7.1 Veri Seti

Bu tezde kullanılan Çember dolgu oranı yöntemi ve CNN modelinin geçerliliği için karakter ve rakam görüntülerinden oluşan bir veri seti gerekmektedir. Çalışmada kullanılan veri seti, el yazısıyla ve farklı yazı tiplerinde yazılmış olan rakam ve İngilizce karakter görüntülerinden oluşmaktadır. Veri seti içinde beyaz arka plan üzerine siyah ve renkli biçimde olan karakter ve rakam görüntülerinden toplam 40.657 tane örnek bulunmaktadır. Bu görüntüler üzerinde Görüntü İşleme teknikleri kullanılarak ön işlem ve segmentasyon aşamaları uygulanır. Segmentasyon aşamasından sonra elde edilen sonuçlar ile Çember Dolgu Yöntemi için hazır görüntüler elde edilir. Yapılan bu işlemler sonucunda veri setinde tutarsızlığı önlemek ve daha doğru sonuçlara ulaşabilmek için veri setindeki düzeltilemeyecek bozuk ve anlamsız olan görüntüler çıkarılarak, 30.000 tane karakter ve rakam görüntüleri, oluşturulan CNN modeli için kullanılabilir halde kaydedilir.

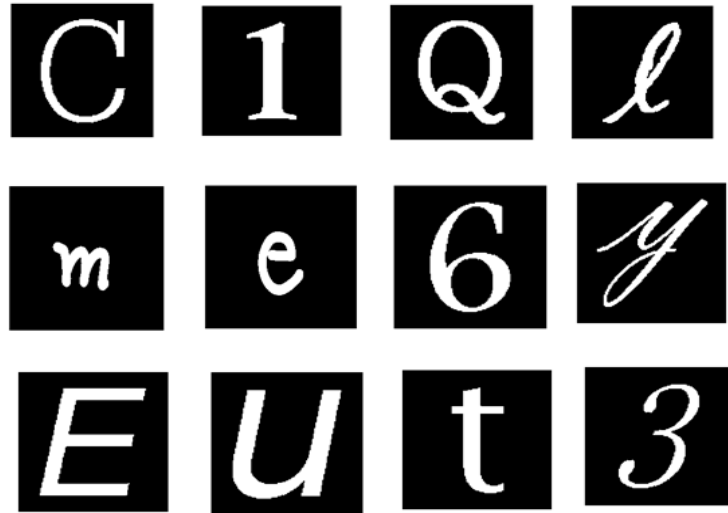


Şekil 7.1: Veri setindeki karakter ve harflerin örnek görüntüleri.

Veri seti içinde bulunan bazı rakam ve karakterlere ait görseller Şekil 7.1’de verilmiştir.

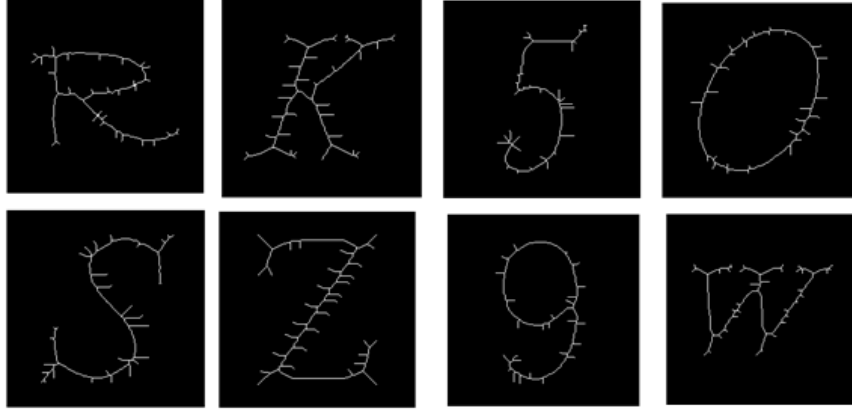
7.2 Ön İşleme

Ön işleme aşamasında, görüntü üzerindeki gürültü ve bozulma gibi sorunları, herhangi önemli bir bilgi eksikliğine neden olmadan ortadan kaldırması amaçlanmaktadır. Kullanılan veri setine ait örnek karakter görüntülerinden de anlaşılacağı gibi görüntüler, beyaz arka plan üzerine siyah karakterler veya renkli olarak yazılmış haldedir. Tüm veri setindeki görüntülerin bu şekilde kullanılması sonuçlar için elverişli olmayacağından ve görüntüler üzerinde ikilileştirme işlemi uygulayabilmek için tüm karakterler siyah arka plan üzerine beyaz yazılmış karakterler şeklinde değiştirilerek kaydedilmiştir. Bu işlemler MATLAB platformu kullanılarak gerçekleştirilmiştir. Yapılan işlemlerin sonucunda kaydedilen yeni görüntülerden örnekler Şekil 7.2’de gösterilmektedir.



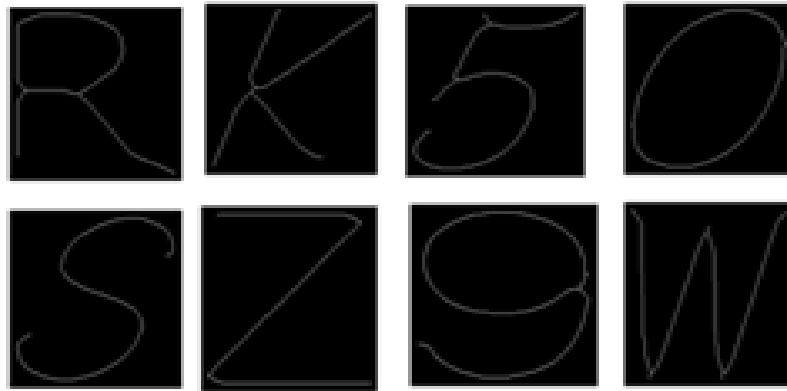
Şekil 7.2: Arka plan ve karakter yazı renkleri değiştirilen görüntü örnekleri.

Arka plan renkleri düzenlenen veri setindeki görüntüler üzerine morfolojik işlem uygulanmıştır. Veri setindeki her bir görüntünün iskelet temsili oluşturulmuştur. Bu işlemden sonra elde edilen görüntü iskeletleri, daha sonraki aşamalar olan CNN modelinde eğitme ve Çember Dolgu Oranı yönteminde kullanılmak üzere kaydedilmiştir. Veri seti içindeki iskelet temsilleri oluşturulan görüntülerden örnekler Şekil 7.3’te gösterilmektedir.



Şekil 7.3: İskelet temsilleri oluşturulan görüntü örnekleri.

Yapılan bu işlem sonucunda iskelet temsili oluşan karakter görüntüsü üzerindeki belirsizliklerin, silinmelerin ve dallanmaların meydana geldiği gözlemlenmiştir. Bu durum tanıma ve kullanılan yöntem kısmında yapılacak hesaplamalar için problem oluşturabileceğinden dolayı gerekli morfolojik düzeltme işlemleri uygulanmıştır. Gerçekleştirilen işlemler sonucunda daha düzgün, gürültüleri giderilen, anlaşılabilirhalde karakter görüntüleri elde edilmiş, tanıma aşaması ve yöntemde hesaplanacak değerler için daha doğru görüntülerin kullanılması sağlanmıştır. Şekil 7.4'te uygulanan düzeltme işlemleri sonucunda elde edilen görüntü örnekleri verilmiştir.



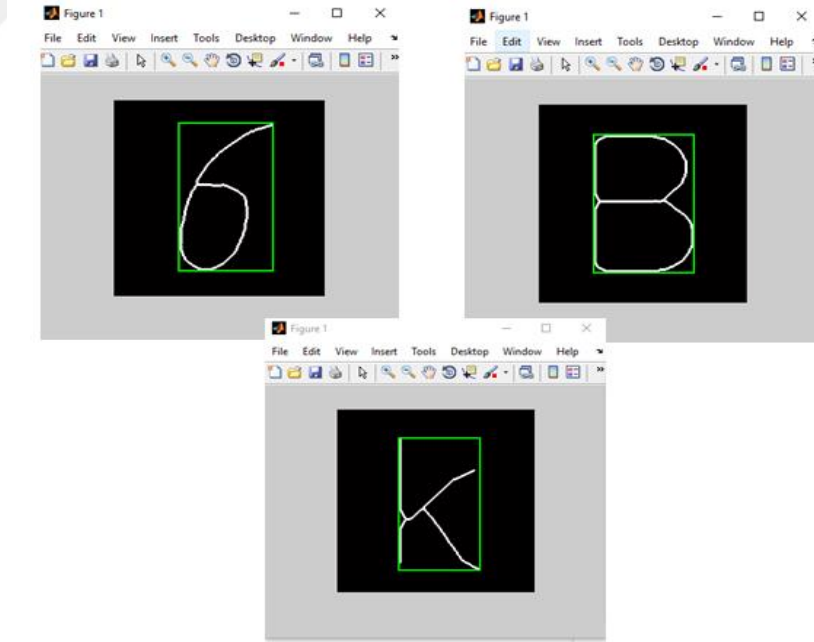
Şekil 7.4: Yapılan düzeltmeler sonrasında elde edilen görüntü örnekleri.

Görüntüler üzerinde var olan problemlerin giderilmesinden sonra OR, OCR alanlarında kullanılan ve önemli bir yere sahip olan segmentasyon işlemi uygulanmıştır. Veri seti içindeki karakter görüntüleri için segmentasyon aşamasında

Matlab platformu üzerinde Bounding Box işlevi kullanılmıştır. Bu işlev, karakterlerin başlangıç ve bitiş noktalarını içeren bölgeyi bir sınırlayıcı kutu içine alarak geri kalan kısımları yok etmektedir. Başka bir deyişle, sadece karakterin yazılmış olduğu sınırlara odaklanıp karakterin daha doğru ve düzgün şekilde algılanmasını sağlamaktadır. Bounding Box ile kelime veya bir cümle içinde bulunan tüm karakterler, tek tek sınırlandırıcı kutular içine alınarak ayrılabilirler.

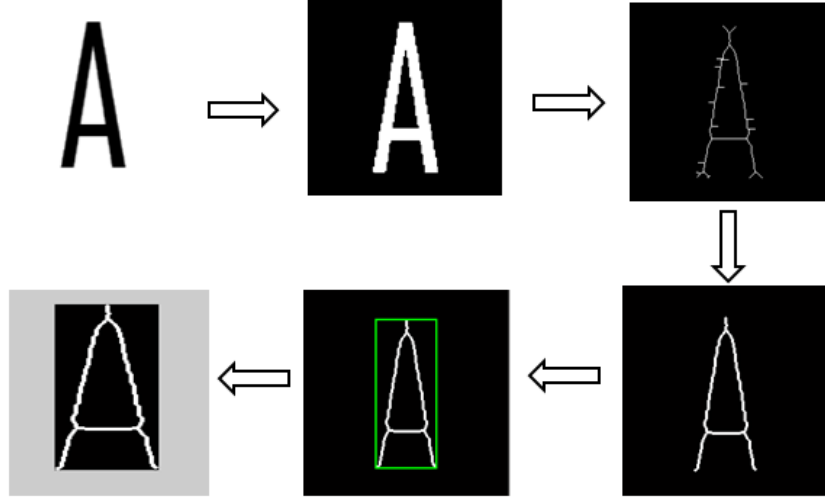
Yapılan bu işlemden sonra her bir karakterin yazılış biçimleri farklı olduğundan çizilen sınırlayıcı kutular da farklı boyutlarda oluşmuştur. Bundan dolayı tüm karakter görüntüleri farklı boyutlarda kaydedilmiştir. Her biri farklı boyutlarda olan görüntüler, oluşturulacak olan CNN modelindeki eğitimin daha doğru ve düzenli gerçekleşebilmesi için 224*224 genişlik ve yükseklik değerlerine sahip olacak şekilde düzenlenerek aynı büyüklükte boyutlandırılmıştır. Bu durumda CNN modelinde yapılacak eğitim sonunda daha yüksek doğruluk oranları elde edilerek daha güvenilir tanıma işlemi yapılması için katkı sağlanmaktadır.

Veri seti içindeki örnek üç karakter görüntüsüne uygulanan Bounding Box işlemi, yeşil sınırlandırıcı kutu şeklinde kullanılmış ve Şekil 7.5'te gösterilmiştir.



Şekil 7.5: Bounding box işlemi uygulanan örnek görüntüler.

Şekil 7.6'da, veri seti içinde bulunan tüm karakter ve sayı görüntülerine uygulanan işlemler tek bir görüntü örneği üzerinde detaylı olarak adım adım gösterilmiştir.



Şekil 7.6: Görüntülere uygulanan tüm işlemlerin adım adım gösterilmesi.

7.3 x, y ve r Noktalarının Belirlenmesi

Bu aşamada CNN ağ modelinde eğitim yapılmak üzere iskelet noktalarından tekrar elde edilecek görüntülerin girdi olarak kullanılabilmesi için, veri setindeki görüntülere Bölüm 7.2’de belirtildiği şekilde işlemler uygulandıktan sonra Çember Dolgu Oranı yöntemiyle beraber hesaplamalar yapılmıştır. Bu hesaplamalar Bölüm 6.3’ te detaylı olarak açıklanan formüllerin kullanılmasıyla gerçekleştirilmiştir.

Veri setine uygulanan ön işleme aşamasından sonra elde edilen görüntülerin, Java programlama dilinde yazılan kod betiği ile x, y, r noktaları hesaplanmıştır. Her harf ve rakam için 400 tane iskelet noktası hesaplanmış ve ayrı ayrı oluşturulan dosyalara kaydedilmiştir.

7.4 Boyut Azaltma:

PCA, Görüntü İşleme, Örüntü Tanıma ve Bilgisayar Görme alanlarında yaygın olarak kullanılmaktadır. Temel olarak amacı boyut azaltma, verilerin yorumlanması ve özelliklerin çıkarılması olan istatistiksel bir yöntemdir. Yüksek boyutlu girdi verilerinde önemli örnekleri bulmak için uygulanmıştır. Bir görüntü, düşük boyutluluğa sahip bir özellik vektörü olarak verimli bir şekilde temsil edilebilir. Bu alt alandaki özellikler, tanıma için ham görüntüden daha belirgin ve daha zengin bilgiler sağlayabilmektedir. Kısacası PCA, bir veri matrisi veya veri kümesinin

kovaryans matrisi ile daha basit bir hale getirilmesidir. Ayrıca verilerin boyutsallığını azaltarak, önemli bilgileri kaybetmeden hesaplama hızının artmasını sağlamaktadır.

PCA tekniği (Turk ve Pentland, 1991), görüntünün satırlarını birleştirerek ve her pikselin yoğunluğunu tek bir özellik vektörü olarak kullanıp her görüntüyü yüksek boyutlu bir alanda bir özellik vektörü olarak kabul eder.

Matematiksel olarak N tane örnek ele alalım.

$$\{A_1, A_2, \dots, A_N\} \quad (7.4.1)$$

$\{A_1, A_2, \dots, A_N\}$ n boyutlu görüntü vektör uzayında değerler olarak her birinin c sınıfına $\{B_1, B_2, \dots, B_C\}$ ait olduğunu varsayalım. Tüm eğitim örneklerinin ortalama matrisi olan \bar{A} hesaplanır ve A_i orijinal değerlerinden çıkarılır ve sonuç ϕ_i de saklanır.

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i \quad (7.4.2)$$

$$\phi_i = A_i - \bar{A} \quad (7.4.3)$$

Bir sonraki adımda, kovaryans matrisi C aşağıda belirtilen formül ile hesaplanır;

$$C = \frac{1}{N} \sum_{i=1}^N \phi_i \phi_i^T \quad (7.4.4)$$

U_i öz vektörleri U_i ($i = 1, \dots, N$) ve karşılık gelen λ_i özdeğerleri λ_i ($i = 1, \dots, N$) hesaplanır. Yukarıdaki N özvektörlerinden, en büyük özdeğerlere karşılık gelen k seçilmelidir ve diğer özdeğerler büyükten küçüğe azalacak şekilde sıralanır. Böylece ilk dönüştürülen özellik, verilerdeki en fazla değişkenliği oluşturur. K özvektörleri U_k kullanarak, PCA tarafından yapılan özellik çıkarma aşağıdaki gibidir.

$$F_i = U_k^T (A_i - \bar{A}) \quad (i = 1, \dots, N) \quad (7.4.5)$$

Veri setindeki tüm görüntülerin x-y ve x-y-r noktaları hesaplanarak PCA işlemi için x-y ve x-y-r şeklinde gruplandırılmıştır. İlk olarak x-y grubu için 2*400 boyutlu noktalara PCA işlemi gerçekleştirilmiş ve görsel olarak kullanılabilmesi için bu noktalarla 20*20 boyutunda yeni görüntüler oluşturulmuştur. Aynı işlem diğer x-y-r grubuna da uygulanarak 20*20 boyutunda görüntü çizimleri yapılmıştır.

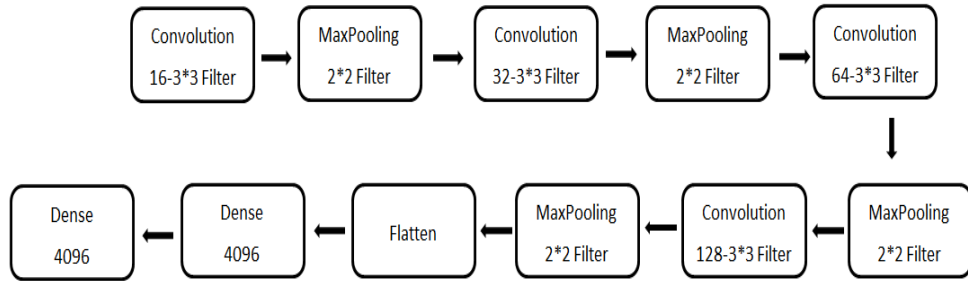
CNN modelinde eğitim yapılması için her harf ve rakamların tekrar elde edilen görüntüleri ayrı ayrı dosyalara kaydedilmiştir. Bu işlemlerde Python programlama dili kullanılmıştır.

Oluşturulan CNN modeli ve bu modelde yapılan eğitimlerin sonucunda elde edilen doğruluk oranları bir sonraki bölümde detaylı olarak belirtilmiştir.

7.5 CNN Modeli

CNN modeli, PyCharm platformunda Python programlama dili kullanılarak oluşturulmuştur. DL modellerinin uygulanması ve geliştirilmesini oldukça hızlı ve kolay hale getirebilmek için geliştirilen, Theano ve TensorFlow kütüphaneleri üzerinde çalışabilen bir Python kütüphanesi olan Keras kullanılmıştır. Bölüm 7.1’de açıklanan ve örnek görüntüleri verilen veri setinde bulunan karakter ve harf görüntüleri, Matlab platformunda gerekli olan görüntü işleme tekniklerinin uygulanmasıyla, oluşturulan CNN modelinde eğitime hazır hale getirilmiştir. Kullanılan bu görüntüler 224*224 boyutlarına sahiptir. Veri setindeki 30.000 görüntü %80 eğitim, %20 test olacak şekilde ayrılmıştır. İlk olarak veri setindeki görüntüler herhangi bir işlem uygulanmadan CNN modelinde girdi olarak kullanılmış ve tanıma oranı elde edilmiştir. Daha sonra bu görüntüler ön işlem aşamasından geçirilerek, Bölüm 7.3 ve 7.4’te belirtilen işlemler uygulanmış ve elde edilen bilgilerle tekrar eğitim yapılmıştır. Gerçekleştirilen bu eğitimler sonucunda varılan tanıma oranları kıyaslanmış ve grafiksel olarak görselleştirilmiştir. Oluşturulan CNN modelinin katman bilgileri aşağıda detaylı olarak verilmiştir.

4 farklı 2D konvolüsyon katmanından oluşan mimaride, ilk katman 16 çıkış boyutlu 3*3filtreye sahiptir. Sonraki katmanlar 32, 64 ve 128 çıkış boyutunda olup, her birinde 3*3 filtreler kullanılmıştır. Tüm Conv2D katmanlarından sonra 2*2 filtre boyutunda maksimum havuzlama katmanı bulunmaktadır. Oluşturulan bu katmanlardan sonra tam bağlı katmanlara geçiş için düzleştirici (flatten) katmanı ile elde edilen çıktılar düzleştirilmiştir. Daha sonra ReLu aktivasyon fonksiyonu ile etkinleştirilen 4096 çıkış boyutlu yoğunluk katmanları bulunmaktadır. Veri setindeki sınıf sayımız kadar çıkışlı ve softmax fonksiyonuyla son bir yoğun katman kullanılmıştır. Şekil 7.7’de CNN mimarisinin katman bilgileri verilmiştir.



Şekil 7.7: CNN mimarisi katman bilgileri.

7.5.1. Değerlendirme Kriterleri

Gerçekleştirilen eğitimler sonrasında yapılan tahminlerin performans analizi için bazı ölçütler kullanılmıştır.

- **Kesinlik/Hassasiyet (Precision):** Yapılan ölçümdeki aslen pozitif değerlerin sınıflandırıcı tarafından pozitif olarak tespit edilen sonuçlarının, tüm pozitif olarak tespit edilenlerin sayısına oranıdır. Denklem 7.5.1'deki şekilde hesaplanmaktadır.

$$Precision = \frac{TruePositives}{TruePositives+FalsePositives} \quad (7.5.1)$$

- **Hatırlatma (Recall):**Aslen pozitif olan değerlerin sınıflandırıcı tarafından bulunabilmesi yeteneğidir.Denklem 7.5.2'deki gibi hesaplanabilir.

$$Recall = \frac{TruePositives}{TruePositives+FalseNegatives} \quad (7.5.2)$$

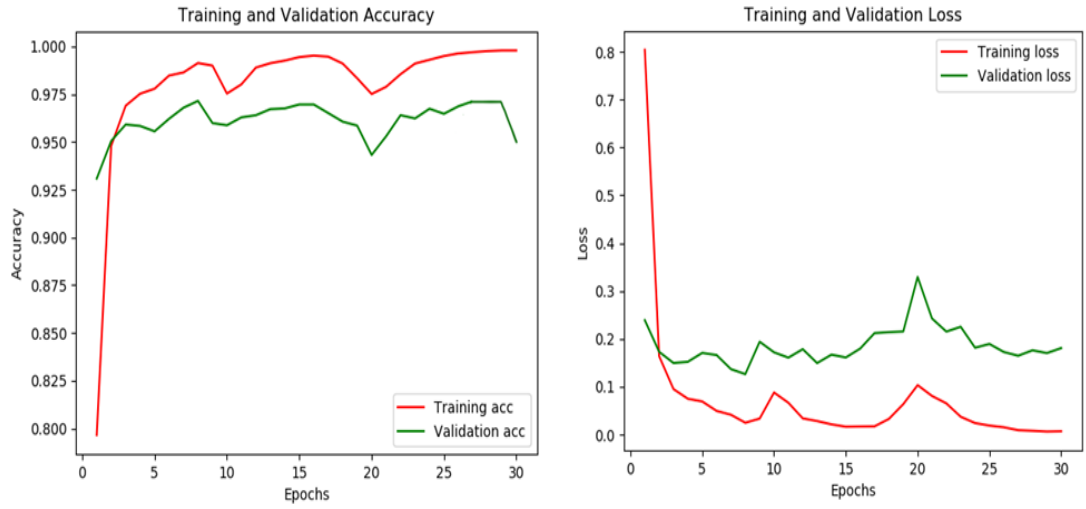
- **F1 Skoru (F-Measure/F-Score):** Kesinlik ve hassasiyet ölçütlerinin harmonik ortalamasıyla model doğruluğunun genel bir ölçüsüdür. Hesaplanması Denklem 7.5.3'te verilmiştir.

$$F1 = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (7.5.3)$$

- **Doğruluk (Accuracy):** Bir modelin doğru bir şekilde eğitilip eğitilmediğini ve genel olarak nasıl performans gösterdiği hakkında bilgi vermektedir. Matematiksel olarak, doğru tahminlerin tüm tahminlere oranı şeklinde açıklanmaktadır. Denklem 7.5.4' te verilen formül ile hesaplanmaktadır.

$$Accuracy = \frac{TrueNegatives+TruePositives}{TruePositives+TrueNegatives+FalsePositives+FalseNegatives} \quad (7.5.4)$$

İlk aşamada herhangi bir görüntü işleme tekniği uygulanmadan 30.000 görüntünün kullanılmasıyla gerçekleştirilen eğitimin sonucunda %95 oranında test doğruluğu elde edilmiştir. Epoch sayısı (eğitim turu) ile eğitim ve test sonuçlarının doğruluk ve kayıp oranlarına ait olan grafik Şekil 7.8'de gösterilmiştir. Aşağıda verilen Tablo 7.1'de epoch sayısının değişimine göre Doğruluk, Precision (Kesinlik), Recall (Hatırlatma) ve F1 skor değerleri verilmiştir.

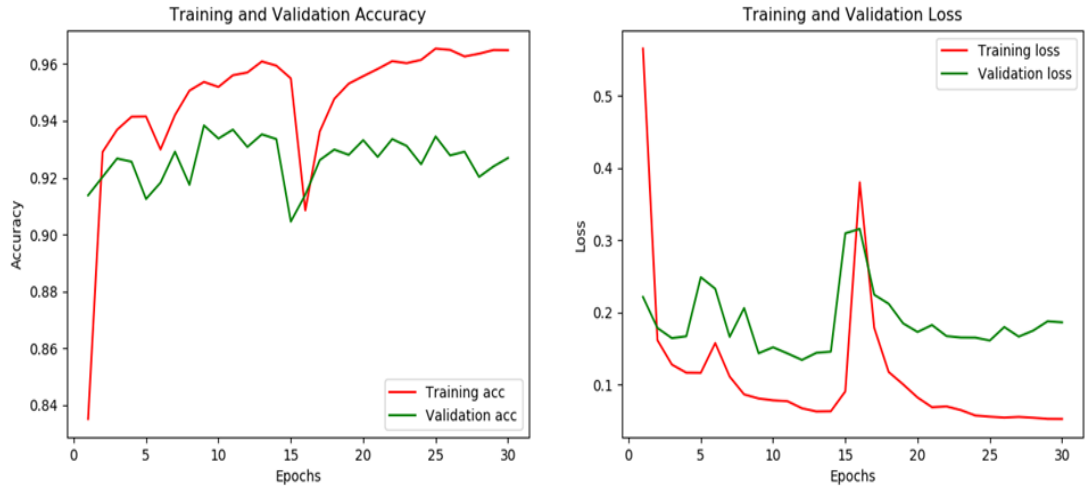


Şekil 7.8: Veri setindeki görüntülere ön işlem uygulanmadan yapılan eğitim sonuçlarının doğruluk ve kayıp grafikleri.

Tablo 7.1: Veri setindeki görüntülere ön işlem uygulanmadan yapılan eğitimin test sonuçları.

Eğitim Turu	Doğruluk (%)	Precision (%)	Recall (%)	F1 Score (%)
10	0.9588	0.9425	0.9492	0.9352
15	0.9696	0.9612	0.9520	0.9630
20	0.9492	0.9300	0.9326	0.9358
25	0.9648	0.9612	0.9652	0.9535
30	0.9502	0.9501	0.9456	0.9542

İkinci aşamada veri setindeki 30.000 görüntünün iskelet temsilleri oluşturulduktan sonra bu görüntülerle yapılan eğitim sonucunda %92 oranında test doğruluğu elde edilmiştir. Epoch sayısı ile eğitim ve test doğruluk oranlarının grafiği Şekil 7.9'da, kayıp oranları ise Şekil 'de verilmiştir. Tablo 7.2'de epoch sayısının değişimine göre Doğruluk, Precision, Recall ve F1 skor değerleri verilmiştir.



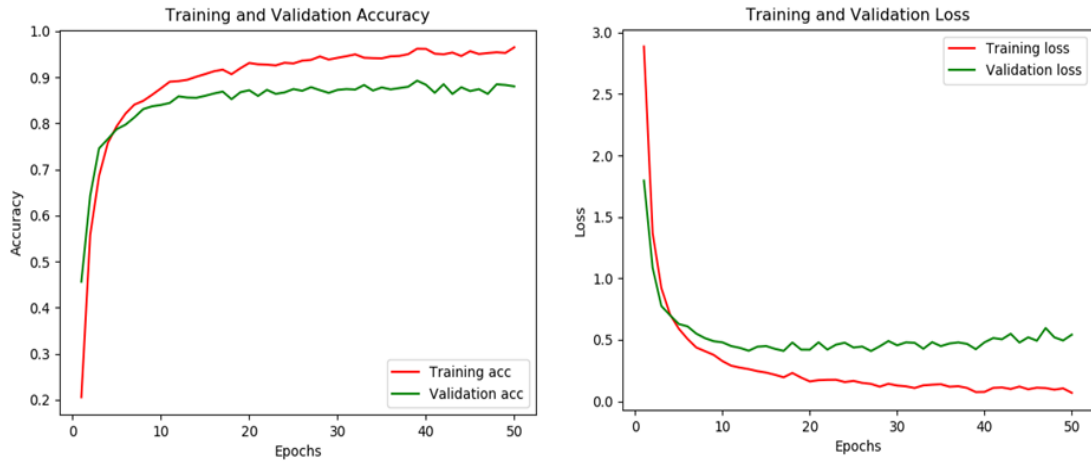
Şekil 7.9: İskelet temsili oluşturulan görüntülerle yapılan eğitim sonuçlarının doğruluk ve kayıp grafikleri.

Tablo 7.2: İskelet görüntüleriyle yapılan eğitimin test sonuçları.

Eğitim Turu	Doğruluk (%)	Precision (%)	Recall (%)	F1 Score (%)
10	0.9338	0.9234	0.9365	0.9200
15	0.9046	0.9017	0.9014	0.9000
20	0.9332	0.9245	0.9324	0.9136
25	0.9345	0.9300	0.9312	0.9236
30	0.9269	0.9238	0.9304	0.9145

Bir diğer aşamada iskelet temsilleri oluşturulan görüntülerin önce x ve y noktaları için daha sonra da x, y ve r noktaları için PCA işlemi ile boyut azaltma yapılmış ve bu işlemler sonucunda elde edilen noktalarla yeni resimler oluşturulmuştur. X ve y noktaları için gerçekleştirilen eğitim sonucunda %88 oranında test doğruluğu elde edilmiştir. Epoch sayısı ile eğitim ve test doğruluk, kayıp oranlarının grafiği Şekil 7.10'da gösterilmiştir.

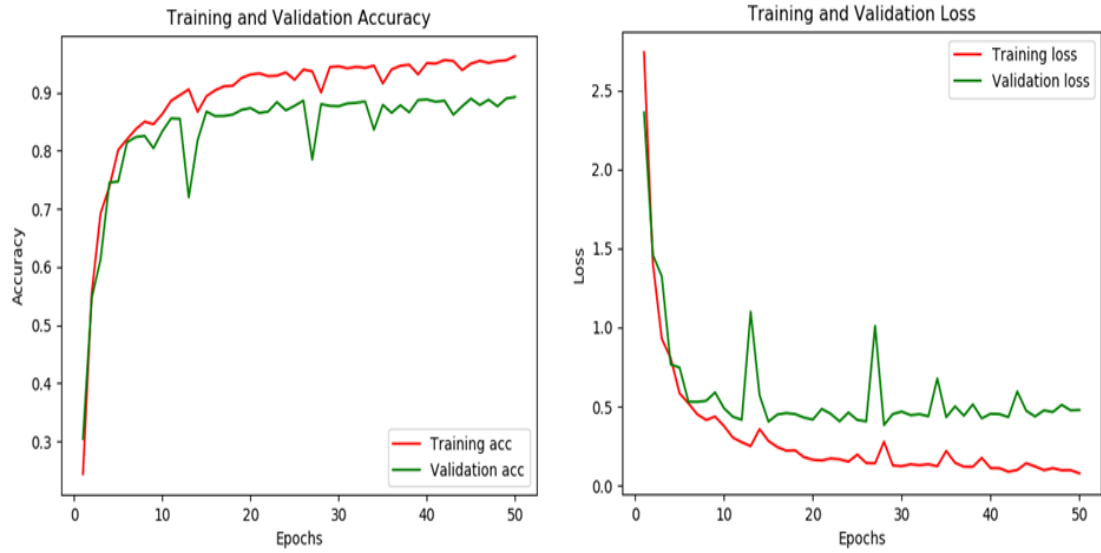
X, y ve r noktaları için yapılan eğitim sonucunda %89 test doğruluğu elde edilmiştir. Epoch sayısı ile eğitim ve test doğruluk, kayıp oranı arasındaki ilişkiyi gösteren grafik Şekil 7.11'de gösterilmiştir. Tablo 7.3'te x ve y noktaları için, Tablo 7.4'te ise x, y ve r noktaları için epoch sayısının değişimine göre Doğruluk, Precision (Kesinlik), Recall (Hatırlatma) ve F1 skor değerleri verilmiştir.



Şekil 7.10: X ve y noktaları ile yapılan eğitimin doğruluk ve kayıp grafikleri.

Tablo 7.3: X ve y noktaları ile yapılan eğitimin test sonuçları.

Eğitim Turu	Doğruluk (%)	Precision (%)	Recall (%)	F1 Score (%)
10	0.8400	0.8325	0.8292	0.8245
20	0.8720	0.8696	0.8632	0.8625
30	0.8728	0.8712	0.8652	0.8692
40	0.8839	0.8829	0.8727	0.8719
50	0.8804	0.8612	0.8856	0.8834

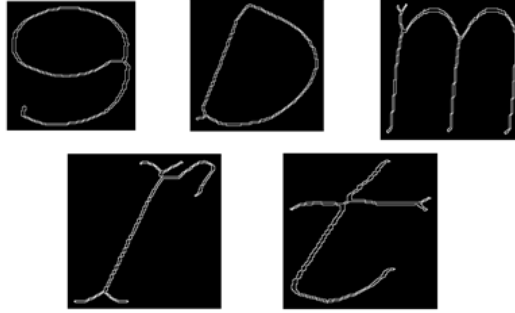


Şekil 7.11:X, y ve r noktaları ile yapılan eğitimin doğruluk ve kayıp grafikleri.

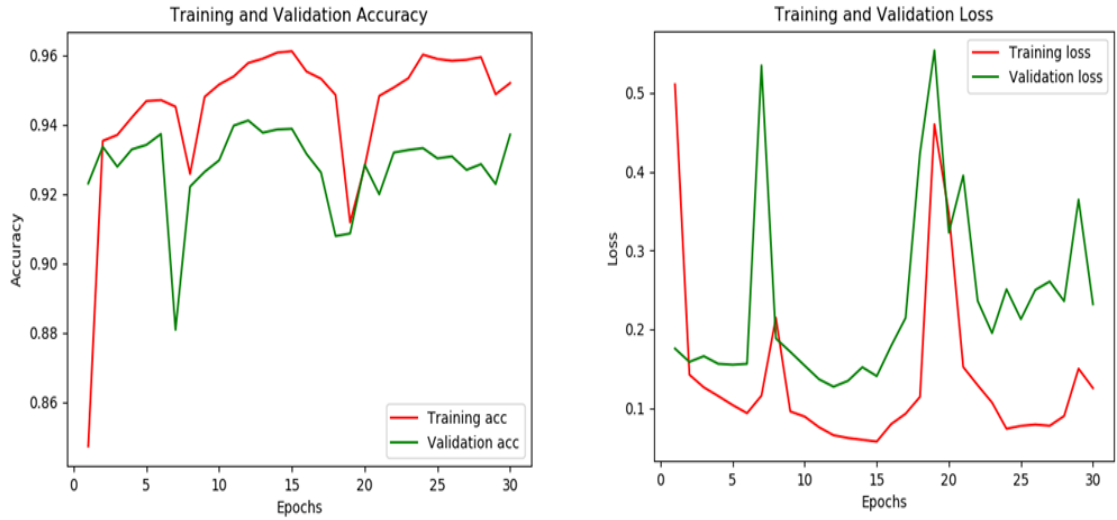
Tablo 7.4: X, y ve r noktaları ile yapılan eğitimin test sonuçları.

Eğitim Turu	Doğruluk (%)	Precision (%)	Recall (%)	F1 Score (%)
10	0.8334	0.8274	0.8312	0.8330
20	0.8736	0.8707	0.8712	0.8728
30	0.8767	0.8685	0.8765	0.8749
40	0.8883	0.8759	0.8820	0.8863
50	0.8924	0.8902	0.8822	0.8765

Son olarak iskelet temsilleri oluşturulan görüntülerin iç ve dış sınır noktalarından elde edilen görüntülerle yapılan eğitimin sonucunda %94 oranında test doğruluğuna ulaşılmıştır. Bu görüntülerin örnekleri Şekil 7.12’de verilmiştir. Eğitim ve test doğruluk oranlarının grafiği Şekil 7.13’te gösterilmiştir. Tablo 7.5’te epoch sayısının değişimine göre Doğruluk, Precision, Recall ve F1 skor değerleri verilmiştir.



Şekil 7.12:İç ve dış sınır noktalarından elde edilen görüntü örnekleri.



Şekil 7.13:İç-dış sınır noktalarından elde edilen yeni görüntülerle yapılan eğitim sonuçlarının doğruluk ve kayıp grafikleri.

Tablo 7.5:İç-dış sınır noktalarından elde edilen yeni görüntülerle yapılan eğitimin test sonuçları.

Eğitim Turu	Doğruluk (%)	Precision (%)	Recall (%)	F1 Score (%)
10	0.9297	0.9209	0.9206	0.9256
15	0.9338	0.9300	0.9305	0.9312
20	0.9284	0.9156	0.9189	0.9026
25	0.9303	0.9278	0.9250	0.9100
30	0.9460	0.9410	0.9303	0.9324

SEKİZİNCİ BÖLÜM

SONUÇ VE ÖNERİLER

OCR uygulamaları, diğer alternatiflere kıyasla elde edilen doğru sonuçların yanı sıra, çeşitli sistemlerde kullanım kolaylığı ve düşük maliyetleri nedeniyle, birçok alanda giderek daha fazla kullanılmaktadır.

OCR alanındaki araştırmalar onlarca yıldır devam ediyor olsa da, makine öğrenmesinin gelişimi ve özellikle derin öğrenme teknikleri ile gelecek vaat eden sonuçların elde edilmesiyle birlikte bu alanda yapılan çalışmalar ön plana çıkmıştır. Hem derin öğrenme yöntemleri hem de OCR işlemleri halen araştırılmaktadır. Birçok işe uygulanan derin öğrenme yöntemlerinin pratik sonuçları, bu yöntemlerin geleneksel yöntemleri aşan sonuçları elde etme yeteneğini göstermektedir.

Bu tez, OCR için yapılması gereken görüntü işleme tekniklerinden sonra Şirin Y. (2016) tarafından geliştirilen Çember Dolgu Oranı yöntemi ile elde edilen özniteliklerin, oluşturulan CNN modeliyle birlikte kullanımını araştırmayı amaçlamaktadır. Kullanılan veri seti içinde yer alan el yazısı ve farklı yazı tiplerinde yazılmış karakter ve sayı görüntüleri, gerekli ön işlem tekniklerinden geçirilerek gürültü, bozulma gibi problemlerinin giderilmesi sağlanmıştır. Oluşturulan CNN modelinde ilk olarak herhangi bir ön işlemden geçmeyen görüntülerin kullanılmasıyla eğitim yapılmış ve doğruluk oranı elde edilmiştir. Daha sonra aynı veri setindeki görüntülerin iskelet temsillerinden, Çember Dolgu Oranı yöntemi ile yapılan hesaplamalar sonucunda iskelet noktaları elde edilmiştir. Bu iskelet noktaları üzerinde uygulanan PCA işlemi ile boyutu düşürülen noktalarla tekrar çizdirilen görüntülerin, aynı CNN modelinde eğitilmesiyle tanıma işlemi gerçekleştirilmiş ve doğruluk sonucuna ulaşılmıştır. Yapılan bu deneysel çalışmalar sonucunda elde edilen doğruluk oranları karşılaştırılmıştır.

Tez kapsamında yapılan deneysel çalışmalar sonucunda sağlanan katkılar aşağıda sunulmuştur.

- Veri setindeki görüntülerin iskeletleştirilmesiyle ve maksimum yarı çaplarda çizilen çemberleri kullanan Çember Dolgu Oranı yöntemiyle birlikte tanıma aşaması için gerekli olacak bilgiler zenginleştirilmiştir.

- Birçok farklı yazı tipinde ve el yazısıyla yazılmış olan karakter ve sayı görüntülerinden oluşan veri seti üzerinde yapılan deneyler kapsamlı olarak değerlendirilmiştir.
- Kullanılan yöntem ile elde edilen bilgiler doğrultusunda veri sayısı düşürülmüş ve tanıma hızı artırılmıştır. Ayrıca veri setindeki görüntülerle ulaşılan doğruluk oranları ile Çember Dolgu Oranı yöntemi kullanılarak elde edilen noktalar üzerinde yapılan işlemler sonucunda ulaşılan doğruluk oranları karşılaştırılmış, daha az veri kullanılmasıyla yaklaşık olarak aynı doğruluk oranları elde edildiği görülmüştür.

Gelecekteki çalışmalarda, x , y ve r değerlerine yapılan işlemlere ek olarak iskelet zenginleştirme değeri olan e değerinin de kullanılmasıyla birlikte gerekli karşılaştırmaların yapılması planlanmaktadır. Ayrıca OCR alanındaki problemlere çözüm ve daha iyi doğruluk oranları elde edilebilmesi için kullanılan yöntem ve oluşturulan CNN modeli çeşitli veri setleri üzerinde kullanılabilir. Diğer tanıma algoritmalarının deneyimlenmesiyle elde edilen sonuçların mevcut oranlarla karşılaştırılması yapılabilir. Yazılı metinlerde birçok yazım hatalarının bulunduğu bilinmektedir. Bu gibi problemlere çözüm olabilmek için de ayrıca yazım denetimi ve gelişmiş hata düzeltme mekanizmaları kullanılarak doğruluk oranları arttırılabilir.

KAYNAKÇA

Ahmed, O.A.W. (1994). *Application Of Artificial Neural Networks to Optical Character Recognition* (Master's Thesis). King Fahd University Of Petroleum & Minerals, Saudi Arabia.

Andrews, W. G., Ğnan, M., Kebeli, S., & Waters, S. (2010). Rethinking The Transcription Of Ottoman Texts The Case For Reversible Transcription. *Turkish Studies International Periodical for the Languages, Literature and History of Turkish or Turkic Volume 5/2spring 2010*.

Arıca, N. (1998). *An Offline Character Recognition System for Free Style Handwriting* (Ms.thesis). The Graduate School of Natural and Applied Science of the Middle East Technical University, Ankara.

Arıca, N., & Yarman, F.T. (2001). An Overview of Character Recognition Focused on Off-line Hadwriting. *IEEE Transaction on Systems, Man and Cybernetics, Part C: Applications and Reviews* (Vol. 31, No. 2, pp. 216-232).

Ashiquzzaman, A., & Tushar, A.K. (2017). Handwritten Arabic Numeral Recognition using Deep Learning Neural Networks. *In Proceedings of the 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition, Dhaka, Bangladesh, 13–14 February 2017*(pp. 3–6).

Ataer, E., & Duygulu, P. (2007). Matching Ottoman Words: An image retrieval approach to historical document indexing. *In Proc. of ACM Int. Conf. on Image and Video Retrieval, ACM, 2007, New York* (pp. 341–347).

Avadesh, M., & Goyal, N. (2018). Optical character recognition for sanskrit using convolution neural networks. *In: Proceedings of the 13th IAPR international workshop on document analysis systems (DAS), Vienna* (pp: 447–452).

Bach, F., Jenatton, R., Mairal, J., & Obozinski, G. (2012). Optimization with Sparsity-Inducing Penalties. *Found. TrendsMach. Learn.*, 4(1), 1–106.

Bai, X., & Latecki, L. J. (2008). Path similarity skeleton graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7), 1282–1292.

- Bastan, M., Kandemir, H., & Canturk, B. (2016). MT3s: Mobile Turkish Scene Text-to-Speech System for the Visually Impaired. *arXiv:1608.05054 [cs]*.
- Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.
- Bharath, V., & Shobha Rani, N. (2017). A font style classification system for English OCR. *2017 International Conference on Intelligent Computing and Control (I2C2)*. IEEE, Coimbatore, India. (doi: 10.1109/I2C2.2017.8321962).
- Blum, H., Roger, N. N. (1978). *Pattern recognition*,10(3), 167–180.
- Breuel, T. B., Hasan, A.Ul., Azawi, M. A. A., Shafait, F. (2013). High-Performance OCR for Printed English and Fraktur using LSTM Networks. *2013 12th International Conference on Document Analysis and Recognition*. Washington, DC, USA. (doi: 10.1109/ICDAR.2013.140).
- Cai, J., & Lui, Z.Q. (December, 2001). Hidden Markov Model with Spectral Features for 2D Shape Recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence* (Vol.23, No.12).
- Chatbri, H., & Kameyama, K. (2014). Using Scale Space Filtering to Make Thinning Algorithms Robust Against Noise in Sketch Images. *International Conference on Pattern Recognition letters Elsevier* (vol. 42, pp. 1-10).
- Chen, G., & Bui, T.D. (1999). Invariant Fourier-wavelet descriptor for pattern recognition, *Elsevier Science Ltd., Pattern Recognition* (Vol. 32, pp. 1083-1088).
- Chen, Q. (2003). *Evaluation of OCR Algorithms for Images with Different Spatial Resolutions and Noises* (Ms. Thesis). School of Information Technology and Engineering Faculty of Engineering University of Ottawa, Ottawa.
- Choudharya, A., Rishib, R., and Ahlawat, S. (2013). Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique. *2013 AASRI Conference on Intelligent Systems and Control AASRI Procedia 4*, 306 – 312. (doi: 10.1016/j.aasri.2013.10.045).
- Cvetkovic, D., Ubeyli, E. D. and Cosic, T. (2008). Wavelet transform feature extraction from human PPG, ECG and EEG signal responses to ELF PEMF exposures: A pilot study. *Digital Signal Processing*, 18(5): 861-874.

- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893). IEEE.
- De Zoysa, D. S. S., Sampath, J. M., De Seram, E. M. P., Dissanayake, D. M. I. D., Wijerathna, L., & Thelijjagoda, S. (2018). Project Bhashitha - Mobile Based Optical Character Recognition and Text-to-Speech System. *2018 13th International Conference on Computer Science & Education (ICCSE)*. IEEE. Colombo, Sri Lanka. (doi: 10.1109/ICCSE.2018.8468858).
- Dinesh Acharya, U., & Subbareddy N. V. (2007). Krishnamoorthy: Isolated Kannada Numeral Recognition Using Structural Features and K-Means Cluster. *Proc. of IISN-2007*, 125-9.
- Duda, R. O., Hart, P. E. and Stork, D. G. (2000). Pattern Classification 2nd Edition. *Wiley Interscience, (October, 2000)*, ISBN-10: 0471056693.
- Freitas, C.O. de A., Oliveira, L.S., & Bortolozzi, F. (Seoul, Korea 2007). Zoning and metaclasses for character recognition. *Symposium on Applied Computing Proceedings of the 2007 ACM symposium on Applied computing* (pp. 632- 636). ISBN:1-59593-480-4.
- Ganis, M. D., Wilson C. L. and Blue, J. L. (1998, Aug). Neural Network-Based Systems For Handprint OCR Applications. *IEEE Transactions on Image Processing*, 7(8), 1097-112.
- Gossweiler, R., Kamvar, M. and Baluja, S., (2009, April). What's up CAPTCHA?: a CAPTCHA based on image orientation. In *Proceedings Of The 18th International Conference On World Wide Web* (pp. 841-850). ACM.
- Goyal. G., Luthra, R. (2016). Skeleton Generation for Digital Images Based on Performance Evaluation Parameters. *International Journal of Signal Processing, Image Processing and Pattern Recognition* (Vol. 9, No. 2, pp. 47-58). (doi: <http://dx.doi.org/10.14257/ijcip.2016.9.2.05>).
- Griffin, G., Holub, A., and Perona, P. (2007). Caltech-256 object category dataset. *Technical Report 7694*, Caltech.

- Guo Z., & Hall, R.W. (1989). Parallel Thinning with Two-Sub Iteration Algorithms. *Communications of the Association of Computer Machinery (ACM) Image Processing and Computer Vision*(vol. 32, no. 3, pp. 359-373).
- He, K. M., Zhang, X. Y., Ren, S. Q., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 Ieee Conference on Computer Vision and Pattern Recognition (Cpvr)*: 770-778.
- Hu, Y., Zheng, L., Yang, Y., & Huang, Y. (2018). Twitter100k: A Real-world Dataset for Weakly Supervised Cross-Media Retrieval. *IEEE Transactions on Multimedia* (Volume:20, Issue:40, April 2018). (doi: 10.1109/TMM.2017.2760101).
- Inoue, K., Miyazaki, R. and Suzuki, M. (1998). Optical Recognition Of Printed Mathematical Documents. *Proceedings of the 3th Asian Technology Conference in Mathematics* (pp. 280–289). Springer-Verlag.
- Jain, A. K. (1989). Fundamentals of Image Processing. *Prentice Hall, Englewood Cliffs*. NJ 07632, ISBN: 0-13-336165-9.
- Jain, A.K., Duin, R. P. W. and Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 22,1, pp. 4-37).
- Keser, R. K. (2018, Temmuz). *Reduced Dimensional Features For Object Recognition*, (Yüksek Lisans Tezi), İstanbul Teknik Üniversitesi, Bilişim Enstitüsü, Uygulamalı Bilişim Anabilim Dalı, Uygulamalı Bilişim Programı, İstanbul.
- Khedkar, A., Shukla, A., Hardas, A., Nigam, N., & Namdev, K. (April-2020). Review Of Text Recognizer Developed Using Cnn. *International Research Journal of Modernization in Engineering Technology and Science* (Volume:2/4).
- Khotanzad, A., & Hong, Y. H. (1988). Rotation Invariant Pattern Reconition Using Zernike Moments. *9th International Conference on Pattern Recognition, Rome, Italy* (Vol.1, pp. 326 - 328). ISBN: 0-8186-0878-1.
- Kim, H. K.,& Kim, J. D. (2000). Region-based shape descriptor invariant to rotation, scale and translation. *Sig. Proc.: Image Comm*,16(1-2), 87–93.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. In *NIPS*.

- Krogh, A., & Hertz, J.A. (1992). Simple Weight Decay Can Improve Generalization. *Advances In Neural Information Processing Systems 4, Morgan Kaufmann* (pp.950–957).
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*86(11), 2278–2324.
- LeCun, Y., Cortes, C., and Burges, C.J. (1998). The MNIST database of hand written digits.
- Lee, Y. K., Song, J., Won, Y. (2018). Improving personal information detection using OCR feature recognition rate. *The Journal of Supercomputing*2018 (Volume 75, Issue 4, pp. 1941–1952). (doi: <https://doi.org/10.1007/s11227-018-2444-0>).
- Leung, C.H., & Sze, L. (1997). Feature Selection In The Recognition Of Handwritten Chinese Characters. *Engineering Applications of Artificial Intelligence*, 10, 5, pp. 495-502.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- Mor, N., & Wolf, L. (2018). Confidence Prediction for Lexicon-Free OCR. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. Lake Tahoe, NV, USA. (doi: 10.1109/WACV.2018.00030).
- Musayev, E. (2004). *Bilgisayar Destekli Karakter Tanıma Sistemi Tasarımı* (Yüksek Lisans Tezi). İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Ozawa, S., & Kotani, M. (2000). A Study of Feature Extraction and Selection Using Independent Component Analysis. *Proc. of Int. Conf. on Neural Information Processing (ICONIP 2000-Taejon, Korea)*, (Vol. 1, pp. 369-374).
- Öz, C., Çit, G. ve Yazgan, H.R. (2010). Cep Telefonu İçin Bir Optik Karakter Tanıma Uygulaması. *International Science & Technology Conference 2010*. Kuzey Kıbrıs Türk Cumhuriyeti.
- Patel, C., Patel, A., & Patel, D. (2012, October). Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study. *International Journal of Computer Applications* (Volume 55– No.10, 0975 – 8887). (doi: 10.5120/8794-2784).

Pradeep, J., Srinivasan, E. and Himavathi, S. (2011, April). Diagonal Based Feature Extraction For Handwritten Character Recognition System Using Neural Network. In *Electronics Computer Technology (ICECT), 2011 3rd International Conference*, (Vol. 4, pp. 364-368). IEEE.

Quehl, B., Sack, H., Yang, H. (2012, May). A skeleton based binarization approach for video text recognition. *13.th Int. Workshop on Image Analysis for Multimedia Interactive Services*. Dublin Ireland. (doi: 10.1109/WIAMIS.2012.6226754).

Rehman, A., & Saba, T. (2014, August). Neural Networks For Document Image Preprocessing: State Of The Art. *Artificial Intelligence Review*,42(2), 253-273. (doi: 10.1007/s10462-012-9337-z).

Reul, C., Springmann, U., Wick, C., & Puppe, F. (2018). Improving OCR Accuracy on Early Printed Books by Utilizing Cross Fold Training and Voting. *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*.IEEE. Vienna, Austria. (doi: 10.1109/DAS.2018.30).

Sato, T., Kanade, T., Hughes, E., & Smith, M. (1998). Video OCR for Digital News Archives. *IEEE Workshop on Content-Based Access of Image and Video Databases (CAIVD'98)*. Bombay, India (pp. 52-60).

Shah, P., Karamchandani, S., Nadkar, T., Gulechha, N., Koli, K., & Lad K. (2009, Nov). OCR-based chassis-number recognition using artificial neural networks. In *Vehicular Electronics and Safety (ICVES), 2009 IEEE International Conference on 2009 Nov 11*(pp. 31-34). IEEE.

Shamsher, I., Ahmad, Z., Orakzai, J. K., & Adnan A. (2007). OCR for printed urdu script using feed forward neural network. In *Proceedings of World Academy of Science, Engineering and Technology 2007 Aug* (Vol. 23, pp. 172-175).

Sharm, A.S., Mridul, M.A., Jannat, M.E., & Islam, M.S. (2018). A Deep CNN Model for Student Learning Pedagogy Detection Data Collection Using OCR. In *Proceedings of the 2018 International Conference on Bangla Speech and Language Processing, Sylhet, Banglades*.

Sharma, O. P., Ghose, M. K. and Shah K. B. (2012, May). An Improved Zone Based Hybrid Feature Extraction Model For Handwritten Alphabets Recognition Using

Euler Number. *International Journal of Soft Computing and Engineering*, 2(2), ISSN: 2231-2307.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.

Singh, H., & Sachan, A. (2018). A Proposed Approach for Character Recognition Using Document Analysis with OCR. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE. Madurai, India. (doi: 10.1109/ICCONS.2018.8663011).

Smith, E. H. B., & Andersen, T. (2005). Text Degradations and OCR Training. *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. Seoul, South Korea. (doi: 10.1109/ICDAR.2005.226).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research (JMLR)*, 15, 1929–1958.

Suen, C. Y. (1986). Character Recognition By Computer And Applications. *Handbook Of Pattern Recognition And Image Processing*, 569-586.

Suzuki, M., Tamari, F., Fukuda, R., Uchida, S., & Kanahori, T. (2003). INFTY: An Integrated OCR System for Mathematical Documents. *Proceedings of the 2003 ACM Symposium on Document Engineering*. Grenoble, France. (doi: 10.1145/958220.958239).

Szegedy, C., Liu, W., Jia, Y. Q., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going Deeper with Convolutions. *2015 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*: 1-9. IEEE.

Şirin, Y. (2016). *Zenginleştirilmiş İskelet Noktaları ile Verimli 2B ve 3B Şekil Geri Getirme*. (Doktora Tezi). Tobb Ekonomi ve Teknoloji Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Şirin, Y., Demirci, M. F. (2013). Circle Filling Rate Descriptor for Object Recognition. *2013 21st Signal Processing and Communications Applications Conference (SIU)*. IEEE. Haspolat, Turkey. (doi: 10.1109/SIU.2013.6531162).

- Tang, Y.Y., Li, B. F., Ma, H., Liu, J., Leung, C.H., & Suen, C.Y. (Washington DC, USA 1996). A Novel Approach to Optical Character Recognition Based on Ring-Projection-Wavelet- Fractal Signatures. *IEEE Computer Society* (Vol. 2, pp. 325-329).
- Trier, Ø. D., Jain, A. K. and Taxt, T. (1996). Feature Extraction Methods for Character Recognition - A Survey. *Pattern Recognition* (Vol. 29,4, pp. 641-662).
- Turk, M., Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3 (1), 71–86.
- Viola, P., & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Accepted Conference on Computer Vision and Pattern Recognition*.
- Vural, E., Erdogan, H., Oflazer, K., & Yanikoglu, B. (2004). *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference, 2004*.(doi: 10.1109/SIU.2004.1338602).
- Webb, A. R. (2002). Statistical Pattern Recognition. *John Wiley & Sons, Ltd. Second Edition*, 2002, ISBN: 0-470-84513-9.
- Ye, Q., & Doermann, D. (2015). Text detection and recognition in imagery: A survey. *IEEE transactions on pattern analysis and machine intelligence* 2015 Jul 1, 37(7), 1480-1500.
- Yetirajam, M., Nayak M. R., and Chattopadhyay S. (2012). Recognition and classification of broken characters using feed forward neural network to enhance an OCR solution. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* Volume. 2012 Oct 28;1.
- Yin, Y., Zhang, W., Hong, S., Yang, J., Xiong, J., Gui, G. (2019, 04 April). Deep Learning-Aided OCR Techniques for Chinese Uppercase Characters in the Application of Internet of Things. *IEEE Access* (Volume: 7, Page(s): 47043 – 47049). (doi: 10.1109/ACCESS.2019.2909401).
- Zahn, C.T., & Roskies, R.Z. (March 1972). Fourier Descriptors for Plane Closed Curves. *IEEE Transaction on Computers* (Vol. C-21, No. 3, pp. 269-281).
- Zeiler, M. D. & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. *Computer Vision - Eccv 2014*, Pt I 8689: 818-833.

Zhai, X., Bensaali, F. and Sotudeh R. (2012). OCR-based neural network for ANPR. In *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings 2012 Jul 16*(pp. 393-397). IEEE.

Zhai, X., Bensaali, F., and Sotudeh, R. (2012). OCR-Based Neural Network for ANPR. *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings*. Manchester, UK. (doi: 10.1109/IST.2012.6295581).

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1), 1-19.

Zhang, H., Liu, D., Xiong, Z. (2017). CNN-based text image super-resolution tailored for OCR. *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE. St. Petersburg, FL, USA. (doi: 10.1109/VCIP.2017.8305127).

Zhang, L. X., Zhao, Y. N. and Wang, J. X. (2002). Feature Selection In Recognition Of Handwritten Chinese Characters. *Proceedings of the 1st International Conference on Machine Learning and Cybernetics*, 3, (pp. 1158-1162).

Zhang, Y. T., & Suen, C. Y. (1984). A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the Association of Computer Machinery (ACM)* (vol. 27, no. 3, pp. 236-239).