**T.C.**

**SÜLEYMAN DEMİREL ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**

**OSMANLICA EL YAZMASI METİNLER İÇİN YAPAY ZEKÂ TABANLI KELİME İNDEKSLEME YÖNTEMİ**

**İbrahim ÖZŞEKER**

**Danışman**

**Dr. Öğr. Üyesi Ufuk ÖZKAYA**

**YÜKSEK LİSANS TEZİ**

**ELEKTRONİK VE HABERLEŞME MÜHENDİSLİĞİ ANABİLİM DALI**

**ISPARTA - 2022**

© 2022 [İbrahim ÖZŞEKER]

**TEZ ONAYI**

**İbrahim ÖZŞEKER** tarafından hazırlanan **"** **Osmanlıca El Yazması Metinler İçin Yapay Zeka Tabanlı Kelime İndeksleme Yöntemi"** adlı tez çalışması aşağıdaki jüri üyeleri önünde Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü **Elektronik ve Haberleşme Mühendisliği Anabilim Dalı**’nda **YÜKSEK LİSANS TEZİ** olarak başarı ile savunulmuştur.

**Danışman Dr. Öğr. Üyesi Ufuk ÖZKAYA** ..............................

Süleyman Demirel Üniversitesi

**Jüri Üyesi Prof. Dr. Adı Soyadı** ..............................

Süleyman Demirel Üniversitesi

**Jüri Üyesi Prof. Dr. Adı Soyadı** ..............................

Süleyman Demirel Üniversitesi

**Enstitü Müdürü Doç. Dr. Şule Sultan UĞUR** ..............................

**TAAHHÜTNAME**

Bu tezin akademik ve etik kurallara uygun olarak yazıldığını ve kullanılan tüm literatür bilgilerinin referans gösterilerek tezde yer aldığını beyan ederim.

**İbrahim ÖZŞEKER**

# İÇİNDEKİLER

**Sayfa**

[İÇİNDEKİLER i](#_Toc89686379)

[ÖZET ii](#_Toc89686380)

[ABSTRACT iii](#_Toc89686381)

[TEŞEKKÜR iv](#_Toc89686382)

[ÇİZELGELER DİZİNİ vi](#_Toc89686383)

[SİMGELER VE KISALTMALAR DİZİNİ vii](#_Toc89686384)

[1. GİRİŞ 1](#_Toc89686385)

[2. KAYNAK ÖZETLERİ 2](#_Toc89686386)

[3. EVRİŞİMSEL SİNİR AĞLARI 4](#_Toc89686387)

[3.1. Evrişim Katmanı 4](#_Toc89686388)

[3.2. Ortaklama 4](#_Toc89686389)

[3.3. Tam Bağlantı 5](#_Toc89686390)

[3.4. Aktivasyon Fonksiyonları 6](#_Toc89686391)

[4. ÇEKİŞMELİ ÜRETİCİ AĞLAR 7](#_Toc89686392)

[4.1. DCGAN 7](#_Toc89686393)

[4.2. Şartlı Çekişmeli Üretici Ağlar 8](#_Toc89686394)

[4.3. Şartlı Çekişmeli Üretici Ağlar ile Imgeden Imgeye Dönüşüm. 9](#_Toc89686395)

[5. VERİ SETLERİ 11](#_Toc89686396)

[5.1. Zorlayıcı Tarihi Veri Seti 11](#_Toc89686397)

[5.2. VML-MOC: Çok Yönlü ve Eğimli El Yazması Veri Seti 12](#_Toc89686398)

[5.3. VML-AHTE: Arapça El Yazısı Veri Seti 13](#_Toc89686399)

[6. GAN MODELİ İLE SATIR BÖLÜTLEME 14](#_Toc89686400)

[6.1. Eğitim ve Test Verilerinin Hazırlanması 14](#_Toc89686401)

[6.2. Model ve Eğitim 15](#_Toc89686402)

[6.3. İşleme Sonrası ve Test Verilerinin Uygulanması 15](#_Toc89686403)

[7. PERFORMANS METRİKLERİ 17](#_Toc89686404)

[8. SONUÇLAR 19](#_Toc89686405)

[9. TARTIŞMA 22](#_Toc89686406)

[KAYNAKLAR 23](#_Toc89686407)

[ÖZGEÇMİŞ 25](#_Toc89686408)

# ÖZET

**Yüksek Lisans Tezi**

**OSMANLICA EL YAZMASI METİNLER İÇİN YAPAY ZEKÂ TABANLI KELİME İNDEKSLEME YÖNTEMİ**

**İbrahim ÖZŞEKER**

**Süleyman Demirel Üniversitesi**

**Fen Bilimleri Enstitüsü**

**Elektronik ve Haberleşme Mühendisliği Anabilim Dalı**

**Danışman: Dr. Öğr. Üyesi Ufuk ÖZKAYA**

Tarihi belgeler araştırmacıların ve tarihçilerin çalışmalarında kullandığı önemli referanslardır. Fakat bu belgelerin incelenip doğru içeriğe ve bilgiye sahip olmak oldukça zahmetli bir iştir. Daha pratik ve hızlı bir araştırma için tarihi belgelerin dijitalleştirilmesi çok büyük bir öneme sahiptir. Bu işlem kendi içinde birçok zorluk barındırmaktadır.

Matbu belgeler için satır bölütleme ve tanımlama işlemi oldukça başarılı hale gelmiştir. Fakat el yazması belgelerde bu başarının düştüğü görülmektedir. Özellikle tarihi belgelerde satır düzensizliği, belge hasarı örtüşme gibi problemler daha çok görülmektedir.

Bu tezde satır bölütleme için derin öğrenme tabanlı çekişmeli üretici ağlar üzerine çalışılmıştır. Özellikle Osmanlıca gibi işlenmesi ve tanımlanması zor bir dil hedef alındığı için alfabe benzerliği sebebiyle Arapça veri setleri kullanılmıştır.

**Anahtar Kelimeler:** GAN, Derin Öğrenme, Satır Bölütleme, El Yazısı Belge, Doküman İmgesi Analizi, Metin Satırı Çıkarımı

**2022, 25 sayfa**

# ABSTRACT

**M.Sc. Thesis**

**ARTIFICIAL INTELLIGENCE BASED WORD INDEXING METHOD FOR HISTORICAL OTTOMAN MANUSCRIPTS**

**İbrahim ÖZŞEKER**

**Süleyman Demirel University**

**Graduate School of Natural and Applied Sciences**

**Department of Electronics and Communication Engineering**

**Supervisor: Asst. Prof. Dr. Ufuk ÖZKAYA**

Historical documents are important references used by researchers and historians in their work. However, examining these documents and obtaining the right content and information is an exceedingly challenging task. The digitization of historical documents has immense importance for a more practical and rapid research. This process contains many challenges

Line segmentation and definition process for printed documents has become quite successful. However, this success has decreased in manuscript documents. Especially in historical documents, problems such as line irregularity, document damage and overlap are more common.

In this thesis, deep learning-based generative adversarial networks are studied for line segmentation. Arabic datasets were used due to alphabet similarity, especially since it was aimed at a language that is difficult to process and define, such as Ottoman Turkish.

**Keywords:** GAN, Deep Learning, Text Line Segmentation, Handwritten Document, Document Image Analysis, Text-line Extraction

**2022, 25 pages**

# TEŞEKKÜR

Tez çalışmasında bilgileri ve tecrübesiyle beni yönlendiren tez süresince çalışmama yön veren, planlanmasında ve gerçekleştirilmesinde her türlü desteği sunan danışmanın Dr. Öğr. Üyesi Ufuk ÖZKAYA’ya teşekkürlerimi sunarım.

Aynı zamanda OPTİMA Araştırma Grubu’nda benden desteklerini esirgemeyen Arş. Gör. Gizem DURSUN ve Arş. Gör. Ali Alper DEMİR’e yardımları ve yönlendirmeleri için teşekkür ederim.

Ayrıca bu tez çalışmasını FBG-2019-7324 numaralı proje ile destekleyen Süleyman Demirel Üniversitesi Bilimsel Araştırma Projeleri Koordinatörlüğüne destekleri için teşekkürlerimi sunarım.

İbrahim Özşeker

ISPARTA, 2022

ŞEKİLLER DİZİNİ

[Şekil 3.1. Evrişimsel sinir ağı katman diyagramı 4](#_Toc89691782)

[Şekil 3.2. Evrişim katmanı diyagramı 4](#_Toc89691783)

[Şekil 3.3. Maksimum ortaklama katmanı 5](#_Toc89691784)

[Şekil 3.4. Ortalama ortaklama katmanı 5](#_Toc89691785)

[Şekil 3.5. Tam bağlantı katmanı 5](#_Toc89691786)

[Şekil 3.6. Düzeltilmiş doğrusal birim 6](#_Toc89691787)

[Şekil 4.1. DCGAN akış diyagramı 7](#_Toc89691788)

[Şekil 4.2. DCGAN kullanılarak üretilmiş yüz imgeleri (Redford vd., 2016) 8](#_Toc89691789)

[Şekil 4.3. CGAN akış diyagramı 9](#_Toc89691790)

[Şekil 4.4. CGAN ile üretilmiş veri (Mirza vd., 2014). 9](#_Toc89691791)

[Şekil 4.5. U-net mimarisi (Ronneberger vd., 2015) 10](#_Toc89691792)

[Şekil 4.6. "pix2pix" mimarisi 10](#_Toc89691793)

[Şekil 5.1. Satır hattı içeren veri seti ve kesin referans 11](#_Toc89691794)

[Şekil 5.2. Eğimli veri seti ve etiketleri 12](#_Toc89691795)

[Şekil 5.3. Veri setinden bir sayfa ve etiketleri 13](#_Toc89691796)

[Şekil 6.1. Oluşturulan veri setinden örnekler 14](#_Toc89691797)

[Şekil 6.2. Bölütleme için kullanılan U-Net ve PatchGAN içeren GAN modeli 15](#_Toc89691798)

[Şekil 6.3. Kayan pencere uygulanmadan(a) ve uyguladıktan sonrası(b) 16](#_Toc89691799)

[Şekil 7.1. Örnek bir kesin referans ve çıktı 17](#_Toc89691800)

[Şekil 7.2. Karakter tabanlı metrikler için metrik hesaplama yönteminin piksel tabanlı gösterilmesi 18](#_Toc89691801)

[Şekil 8.1. Eğimsiz giriş sayfası ve işlemler sonrası elde edilen çıktı 19](#_Toc89691802)

[Şekil 8.2. Çok yönlü metin içeren bir sayfa ve sonuçları 20](#_Toc89691803)

[Şekil 8.3. Yatay bir satırın kesin referans ve çıktı örtüşmesi 21](#_Toc89691804)

[Şekil 8.4. Açılı bir satırın kesin referans ve çıktı kesişimi 21](#_Toc89691805)

[Şekil 9.1. Bir satır yapısındaki farklı yönelmelere örnek 22](#_Toc89691806)

[Şekil 9.2. Satır sonuna eklenen notlar 22](#_Toc89691807)

# ÇİZELGELER DİZİNİ

[Çizelge 8.1. Test verilerinin performans metrikleri 20](#_Toc89685939)

# SİMGELER VE KISALTMALAR DİZİNİ

CNN Evrişimsel Sinir Ağları

GAN Üretici Çekişmeli Ağlar

DCGAN Derin Evrişim Üretici Çekişmeli Ağlar

ReLU Düzeltilmiş Doğrusal Birim

XML Genişletilebilir İşaretleme Dili

P Keskinlik

R Duyarlılık

# GİRİŞ

Satır bölütleme dokuman analizi için temel bir adamdır. Temel amaç dokuman içerisindeki satırların tek tek yüksek doğrulukla ayrıştırmak ve diğer adımlara öncülük sağlamaktır. Dokuman analizi adımlarının başında gelen satır bölütlemenin analiz ve tanımlama gibi diğer adımların başarısını doğrudan etkilemektedir. Matbu belgeler için bölütleme işlemlerinde yüksek başarılar elde edilmiş olsa da el yazması belgelerde halen büyük problemler yaşanmaktadır. Özellikle tarihi dokümanlarda çok fazla zorlayıcı neden bulunduğu için bölütleme işleminin daha da zorlaşmasına neden olmaktadır. Satırların düzensiz, eğri, çok yakın ya da uzak, örtüşme olması gibi zorlayıcı problemler tarihi dokümanlarda satır bölütleme işlemi için başlıca zorlayıcı sebeplerdir. Arapça ve Osmanlıca gibi dillerde ise metin yönünün değişmesi, yazı içindeki noktalamalar ve örtüşmenin artmasından dolayı bu zorluklar artmaktadır. Bu çalışmada Osmanlıca tarihi metinlerde satır bölütleme üzerine çalışılmıştır. Osmanlıca dili Arap alfabesi kullandığı için eğitimi ve test işlemleri için Arapça veri setlerinden yararlanılmıştır. Bu veriler ile son yıllarda imgeden imgeye dönüşüm için kullanılan çekişmeli üretici ağlar eğitilerek maskelerin üretilmesi ve bu maskelerden satırların çıkarılması üzerine çalışılmıştır. Performansın belirlenmesi için çalışmanın durumuna uygun olarak bilinen performans metrikleri üzerinde geliştirmeler yapılmıştır.

# KAYNAK ÖZETLERİ

Satır bölütleme işlemi için pek çok yöntem ortaya atılmıştır. Yazı hatlarının bulunması için projeksiyon tabanlı (Arivazhagan vd., 2007), bağlı bileşen tabanlı (Koo vd., 2011 & Ryu vd., 2014) ve Hough dönüşümü tabanlı (Loulodis vd., 2008 & Likforman-Sulem vd., 1995) gibi uzamsal metotlar önerilmiştir. Projeksiyon tabanlı modeller matbu metinlerde genellikle başarılı çalışmasına rağmen el yazmalarında bu başarının çok düşük olduğu görülmektedir. Eğimli yazı hatlarının lokal projeksiyonları dikkate alınarak bölütleme işlemi yapılabilir fakat hatlarda çok yönlü eğim varsa bu başarı düşmektedir.

Bağlı bileşen tabanlı çalışmalarda genellikle her pikselin önemi eşittir. Bu varsayım herhangi bir piksel değişiminde sonuçlar değişilmektedir. Bu sebeple çok uygulanabilir değildir. Ayrıca bu tarz metotlar Arapça gibi çok sayıda ve küçük parçalar içeren metotlarda başarılı çalışmamaktadır. Hough dönüşümü tabanlı çalışmalar farklı eğimli satırlar içeren metinlerde başarılı çalışmaktadır fakat satır içeriğinde farklı yönelimi olan kelimeler var ise satırlar elde edilememektedir. Bunların dışında alternatif olarak süper piksel tabanlı doku veya piksel benzerliği kullanılarak yapılan sınıflandırma yapılabilmektedir. Fakat bu yöntemde oluşturulan kümeler aynı yazı hattını içerebilmektedir.

Çalışmalarda da görüldüğü gibi eğitimsiz metotlarda satır bölütleme için zorlayıcı metinlerde limitler bulunmaktadır. Özellikle el yazması arşivler içerikleri sebebiyle çok zorlayıcı olmaktadır. Bu zorlukları aşmak için eğitimsiz modeller çok başarılı değildir. Bu zorluklar:

* Eğimli ve yatık satırlar,
* Tek bir satır içeriğinde bulun eğimli ve dalgalı yapılar,
* Dokunan ve örtüşen karakterler,
* Başlangıç ve bitiş arasındaki konum farklılıklarıdır.

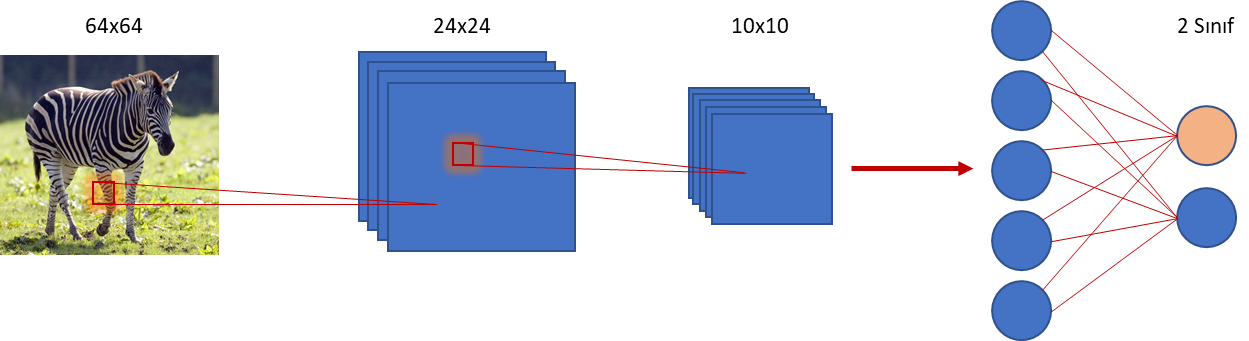
Eğitimsiz metotların yetersiz geldiği bu sorunların üstesinden gelebilmek için derin öğrenme gibi eğitimli yöntemler son yıllarda ortaya çıkmıştır. Derin öğrenme modellerinin temel motivasyonu satır bölütleme işlemindeki problemleri kompleks modellerin eğitimi sayesinde satırların uzamsal davranışlarını ve konumlarının tespitini sağlamaktır (Louloudis vd., 2006; Demir ve Özkaya, 2019). Barakat, vd. (2018) çalışmasında bir tam bağlantılı evrişimsel sinir ağı önerilmiştir. İşlem sonrası aşamasında bağlı olmayan satırlara genişletilme işlemi uygulanmıştır. Bölütleme performans metrikleri sayfa içeriğindeki eğimli ve çok yönelimli satırlara bağlı olarak hesaplanmıştır. Yazar tarafından tanımlanan keskinlik, duyarlılık ve F1 skoru bölütleme performansları Cohen vd. (2014) ile toplanan sonuçlarla karşılaştırılmıştır. Önerilen metotta eğimli satılarda bölütleme hataları görülmektedir.

El yazması metinlerde satır bölütleme için güçlü ve yeni bir metot olan çekişmeli üretici ağlar (GAN) önerilmiştir. GAN modelleri kompleks yapıları işlemekte çok başarılı sistemlerdir. Aynı zamanda GAN yapıları sentetik veri üretiminde de oldukça başarılıdır. El yazısı metinlerde satır bölütleme işlemi için GAN bir imgeden imgeye dönüşüm problemi olarak ele alınmıştır.

Kundu vd. (2020) çalışmasında satır bölütleme performansı GAN modellinin L1 ve L2 gibi farlı kayıp fonksiyonları için ölçülmüştür. Bu çalışma sonucunda en iyi el yazması metinlerde satır bölütleme performansının “U-Net” ’in üretici “PatchGAN” yapısının ayırıcı olarak kullanan bir GAN modeli tarafından sağlandığı görülmüştür. Önerilen modelde GAN yapısı kompleks ve örtüşme bulunan veri seti ile beslenmiştir ve kabul edilebilir satır maskeleri elde edilmiştir.

# EVRİŞİMSEL SİNİR AĞLARI

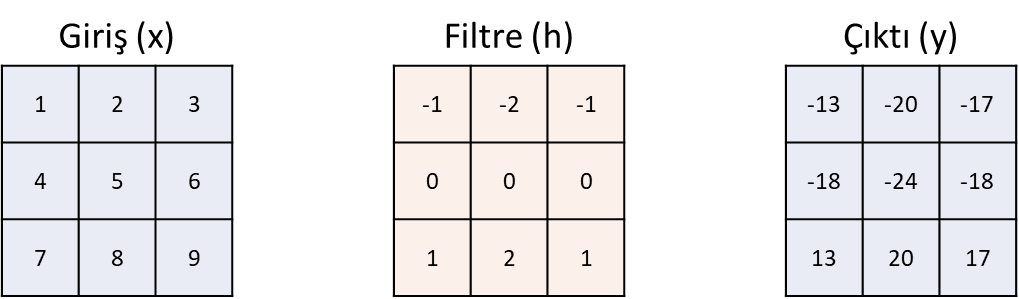
Evrişimsel sinir ağları (CNN) girişine 3 veya daha fazla boyutlu veri alan yapılardır. Giriş verisi bir takım eğitilebilir ve statik katmanlardan geçirilerek sonuç elde edilir.



Şekil 3.1. Evrişimsel sinir ağı katman diyagramı

## Evrişim

Konvülüsyon çok önemli ve temel bir sinyal işleme aracıdır. Evrişim olarakta bilinen bu işlem kayan pencere mantığı ile çalışır. Görüntü işlemede ise yaygın olarak 3 boyutlu yada 2 boyutlu evrişim işlemi kullanılmaktadır.



Şekil 3.2. h filtresi için evrişim örneği

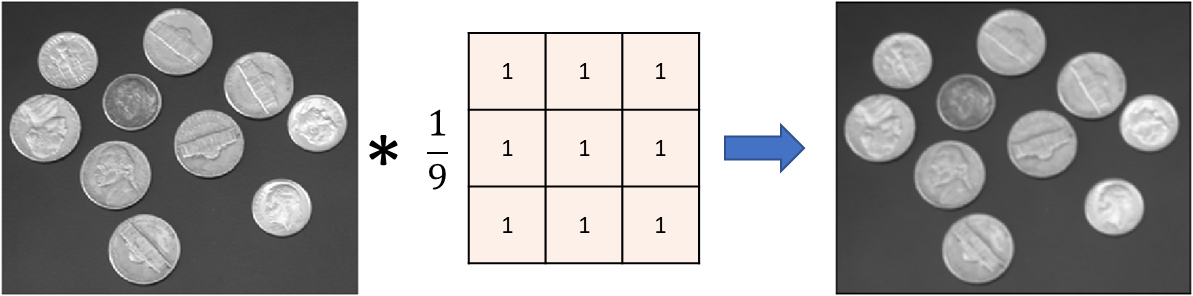
Şekil 3.2’de x giriş matrisi, h filtre ve y bu filtrenin giriş ile 2 boyutlu evrişim sonucudur.



Şekil 3.3. Evrişim h filtresi indisleri

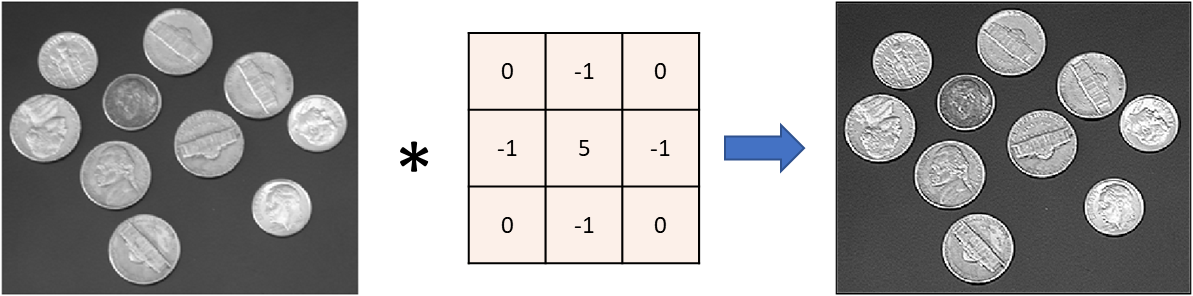
Şekil 3.3 filtreniz indislerini temsil etmektedir. Bu durumda bir matrise 2 boyutlu evrişim (no) formülü ile hesaplanır.

Görüntü işleme yöntemleri arasında yaygın olarak bilinen evrişim filtreleri vardır.



Şekil 3.4. Bulanıklaştırma filtresi

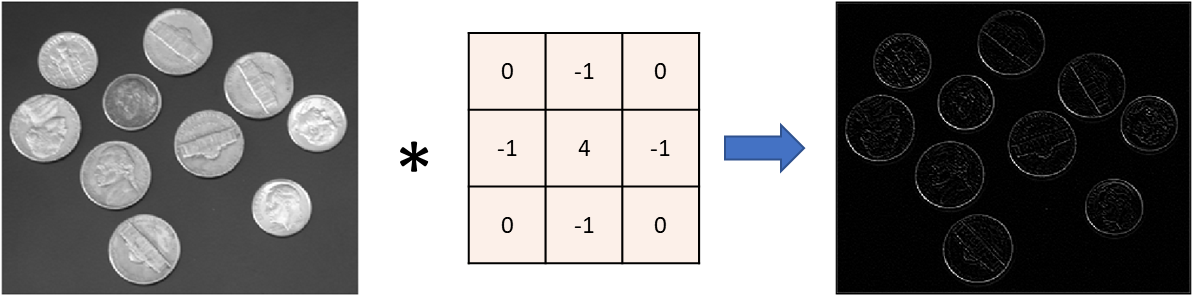
Şekil 3.4. bulanıklaştırma filtresinin sonuçları görülmektedir. Bu filtre imge ile konvolüsyon işlemi uygulayarak kenar geçiş pikselleri birbirine yakınlaştırılır.



Şekil 3.5. Keskinleştirme filtresi

Keskinleştirme filtresi ise Şekil 3.5’te görüldüğü gibi imge üzerindeki kenar bilgilerini yoğunlaştırır.

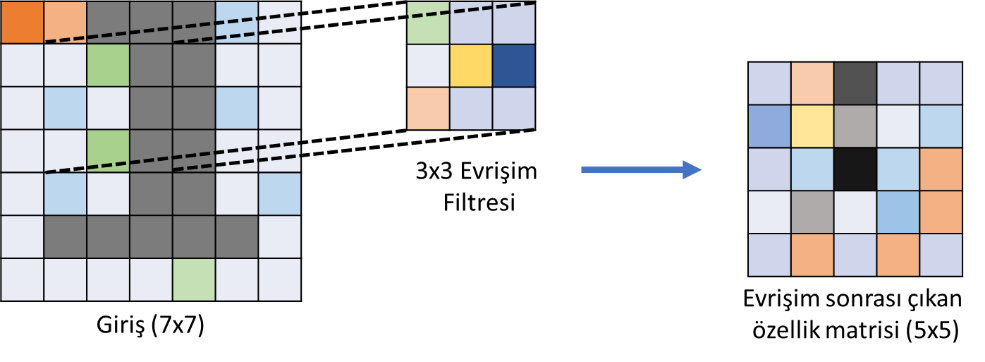
Diğer bir filtre ise kenar bulmaya yarayan bir yapıdadır. 2 eksendeki piksellerin negatifleri merkez pikselin 4 katı ile toplanır ve çevre pikseller arası bir fark varsa bir köşe bilgisi ortaya çıkarır.



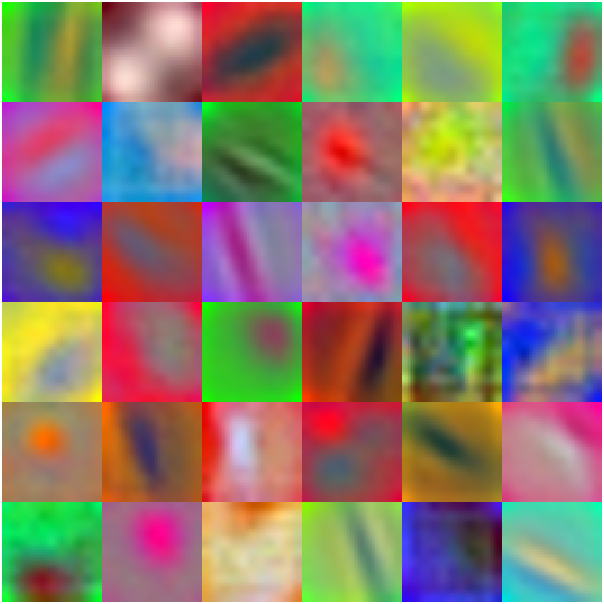
Şekil 3.6. Kenar bulma filtresi

## Evrişim Katmanı

Evrişim katmanı giriş verisinin üzerinde belirlenen SxS boyutunda bir konvolüsyon işlemi uygulanarak bir öznitelik haritası elde edilmesini sağlar.



Şekil 3.7. Evrişim katmanı diyagramı



Şekil 3.8. ImageNet (Krizhevsky vd., 2012) ağının ilk evrişim katmanı özellikleri

metin, döşeli, renkli, boyanmış içeren bir resim

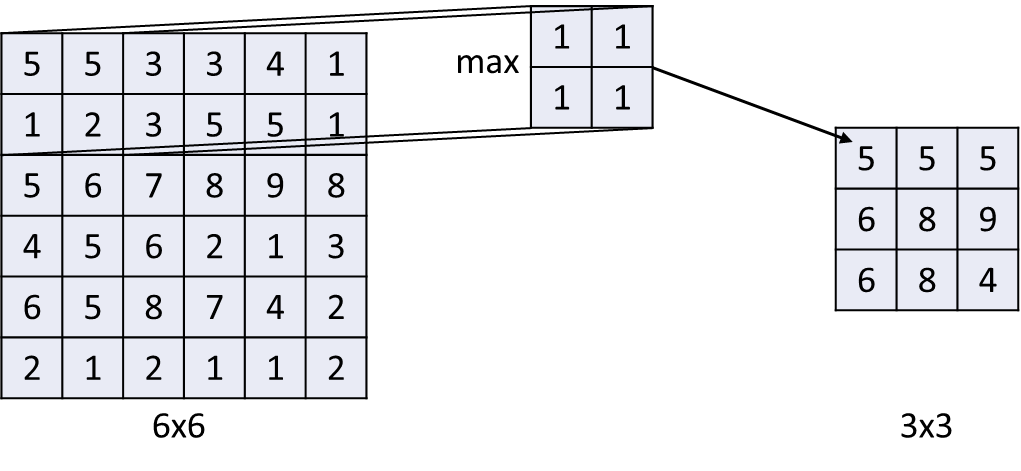
Açıklama otomatik olarak oluşturuldu

Şekil 3.9. ImageNet (Krizhevsky vd., 2012) ağının 5. evrişim katmanı özellikleri

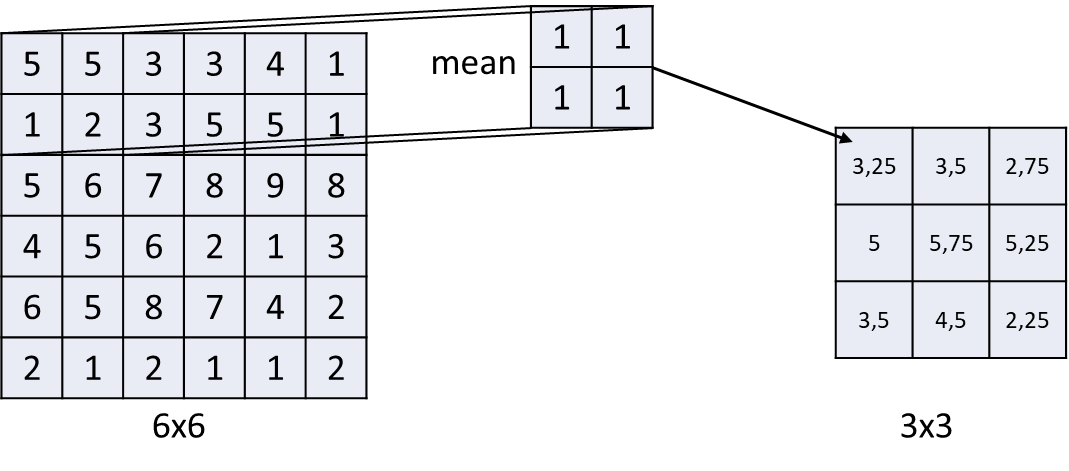
Şekil 3.8 ve Şekil 3.9 incelendiğinde evrişimsel sinir ağında derin katmanlara inildikçe evrişim katmanının incelediği özellik kompleksliği artmaktadır.

## Ortaklama

Ortaklama katmanı uzamsal bir örnekleme işlemi yapan bir katmandır. Bilinen iki türü maksimum veya ortalama ortaklama katmanlarıdır. Maksimum ortaklama filtre içindeki değerin maksimumunu alarak ortalama ortaklama ise filtre içinde kalan değerlerin ortalamasını birleştirerek boyut azaltır.



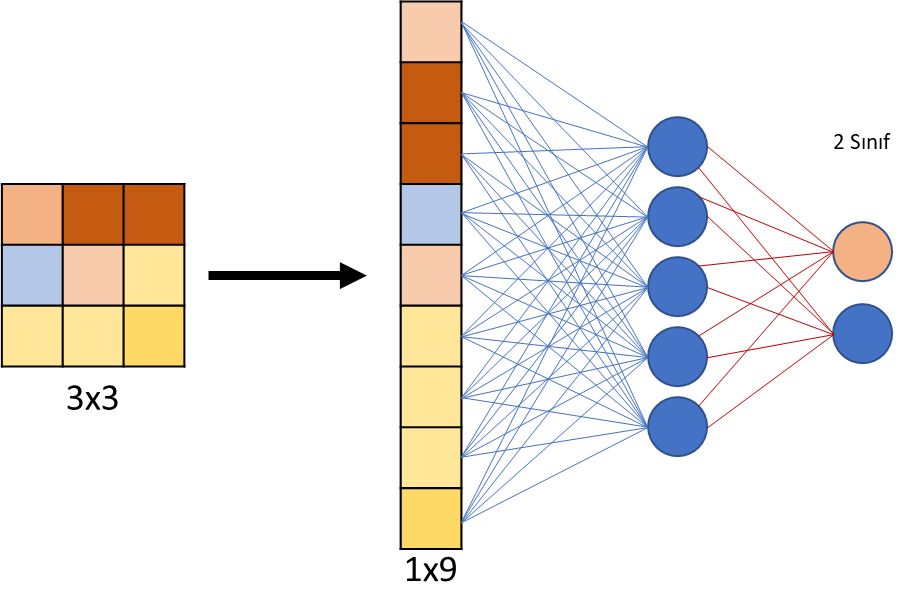
Şekil 3.10. Maksimum ortaklama katmanı



Şekil 3.11. Ortalama ortaklama katmanı

## Tam Bağlantı

Tam bağlı katmanda her giriş bütün nöronlara bağlı olan ve genelde CNN mimarisinin sonunda uygulanan bir katmandır. Sınıf veya özellik ağırlıkları hedeflerini optimizasyon için kullanılır.



Şekil 3.12. Tam bağlantı katmanı

yeşil, selentere içeren bir resim

Açıklama otomatik olarak oluşturuldu

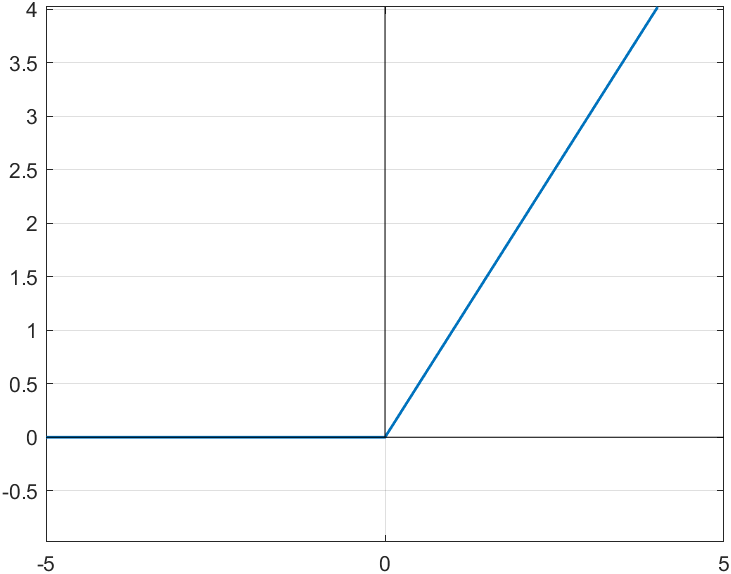
Şekil 3.13. ImageNet (Krizhevsky vd., 2012) ağının en son tam bağlantı katmanının bazı sınıfların çıktıları

## Aktivasyon Fonksiyonları

Aktivasyon fonksiyonları katman çıkışlarına eklenerek nöronların aktivasyonu kontrol edilir. Bu sayede eğitimde ezberleme gibi performans düşüren durumlar düzeltilir.

Düzeltilmiş doğrusal birim (ReLU) evrişimsel sinir ağlarında kullanılan temel bir aktivasyon fonksiyonudur.

Bu fonksiyonun (No)’da belirtilen tanımı gereği bilgisayar tarafından hesaplanması kolaydır ve biyolojik nöron yapısına yakındır.



Şekil 3.14. Düzeltilmiş doğrusal birim

Şekil 3.6’daki grafik incelendiğinde negatif değerlerde fonksiyon türevi sıfırdır. Bu sebeple geri yayılım sırasında parametrelerde bir güncellenme yaşanmaması bir problemdir. Sızıntı düzeltilmiş doğrusal birim aktivasyon fonksiyonu bu problemi çözmek için ortaya çıkmıştır negatif değerler sıfıra çok yakındır fakat sıfır değildir.

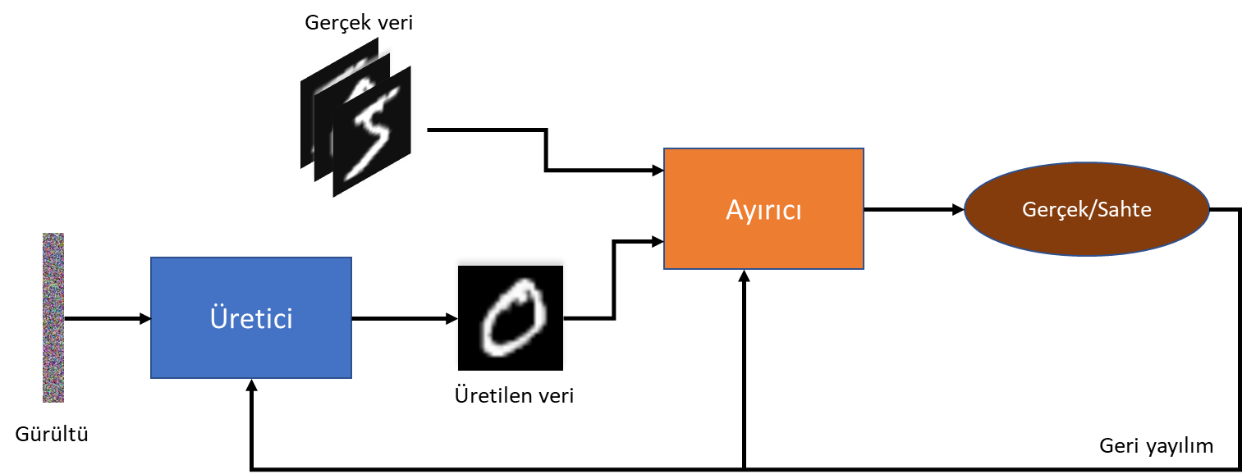
# ÇEKİŞMELİ ÜRETİCİ AĞLAR

Çekişmeli üretici ağları (GAN) 2 farklı ve birbirine zıt ağ yapıları birleştirilerek üretilmiş ağlardır. Ağlar arasında bir oyun teorisi kurularak Nash dengesine ulaştırılır (Goodfellow vd., 2014).

Çekişmeli üretici ağlar üretici ve ayırıcı olmak üzere iki birimden oluşmaktadır. Üretici giriş verisi üzerinde anlamlı bir sonuç üretmeye çalışırken ayırıcı üretilen verinin ne kadar gerçek olduğunu yorumlamaktır. Bu işlem sırasında ayırıcı ve üretici arasında bir minmax oyunu oynanır.

## DCGAN

Derin evrişimsel üretici sinir ağları (DCGAN) temel bir GAN çalışmasıdır. DCGAN ile tek boyutlu bir gürültüden bir imge üretilmektedir.



Şekil 4.1. DCGAN akış diyagramı

Şekil 4.1’de görüldüğü gibi DCGAN bir adet evrişimsel sinir ağı ve buna eklenen tam tersi yönlü bir evrişimsel sinir ağının birleşimidir. DCGAN üretilirken temel olarak yapılacak değişiklikler:

* Bütün havuzlama katmanları atlamalı konvolüsyon (ayırıcı) ve fraksiyonel konvolüsyon ile değiştirilir.
* Üretici ve Ayırıcı katmanlar arasında yığın normalizasyonu kullanılır.
* Tam bağlantılı gizli katmanlar ağdan çıkartılır.
* Üretici ağında ReLU aktivasyon fonksiyonu, üretici çıkış katmanında ise Tanh aktivasyon fonksiyonu kullanılır.
* Ayırıcı ağında LeakyReLU aktivasyon fonksiyonu kullanılır (Redford vd., 2016)

DCGAN kullanılarak z boyutlu bir veriden yapay imgeler oluşturulur.

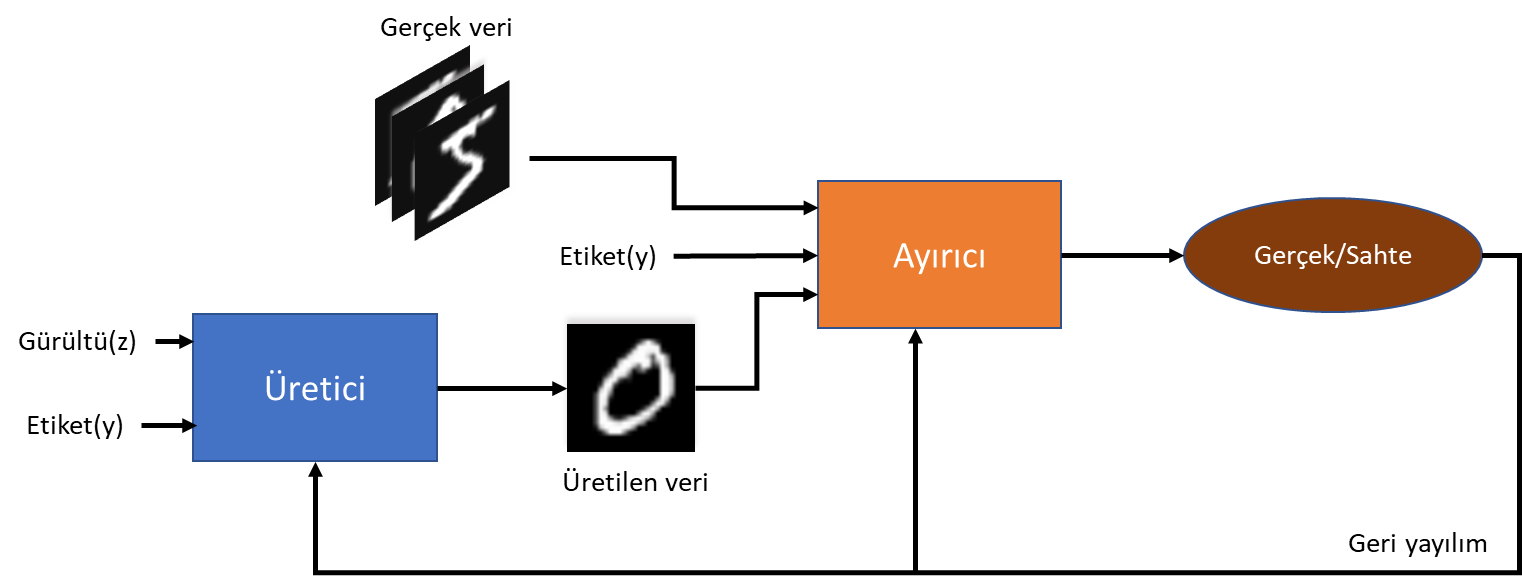
poz, kişi, iç mekan, grup içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 4.2. DCGAN kullanılarak üretilmiş yüz imgeleri (Redford vd., 2016)

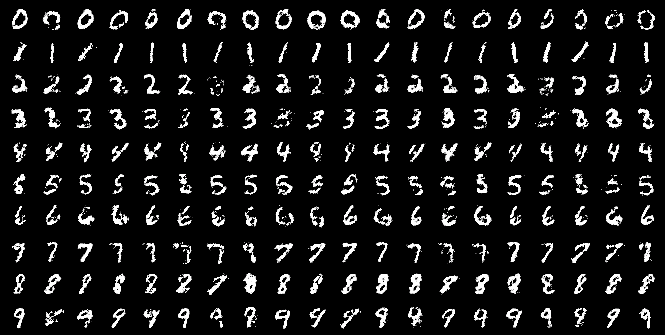
## Şartlı Çekişmeli Üretici Ağlar

Çekişmeli üretici ağlar üretici ve ayırıcı katmalarına bir y şartı eklenerek genişletilebilir ve yönlendirilebilir. y herhangi bir sınıf veya yardımcı veri olabilir. Üretici ve ayırıcı y verisi ile beslenerek gerçekleyebilir (Mirza vd., 2014).



Şekil 4.3. CGAN akış diyagramı

Yapılan çalışmada rastgele üretilmiş bir z verisi üzerine MNIST veri seti etiketini işaretleyen bir y dizisi eklenerek GAN yapısının istenen türde bir veri oluşturması sağlanmıştır.

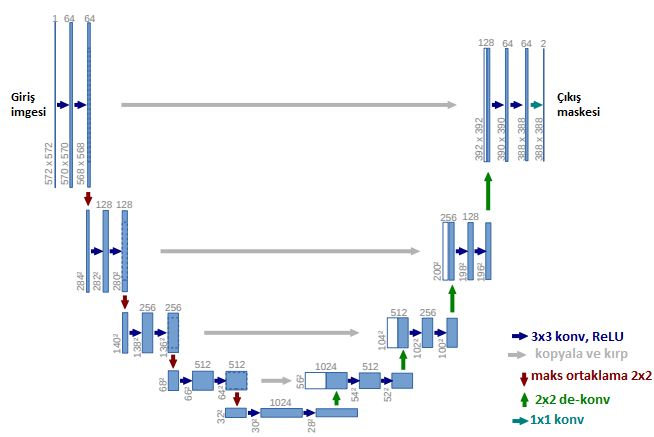


Şekil 4.4. CGAN ile üretilmiş veri (Mirza vd., 2014).

## Şartlı Çekişmeli Üretici Ağlar ile İmgeden İmgeye Dönüşüm.

Çalışmada rastgele bir giriş yerine bir referans imge verilerek buna bağlı bir görüntü üretmesi sağlanmıştır. “pix2pix” olarak bilinen bu ağ yapısında üretici bloğunda U-Net (Ronneberger vd., 2015), ayırıcı yapasında ise PathGAN (Isola vd., 2017) kullanılmıştır.

U-Net imgeden imgeye dönüşüm sağlayan bir evrişimsel sinir ağıdır. Genelde segmentasyon ve maskeleme için kullanılmaktadır.



Şekil 4.5. U-net mimarisi (Ronneberger vd., 2015)

Şekil 4.5’te belirtilen mimari ile üretilen imgeler bir ayırıcı ile birleştirilerek sistemin gerçekçi imgeler üretilmesi sağlanır.



Şekil 4.6. "pix2pix" mimarisi

# VERİ SETLERİ

El yazısı analizi için ICDAR ve benzeri yarışmalarda oluşturulmuş birçok veri seti bulunmaktadır fakat yaptığımız Osmanlıca el yazmaları için işaretlenmemiş kadı siciller gibi belgelere olmasına rağmen bunlar üzerinde işaretleme bulunmamaktadır. Osmanlıcada Arap alfabesi kullanılması sebebiyle Arapça veri setleri üzerinde durulmuştur.

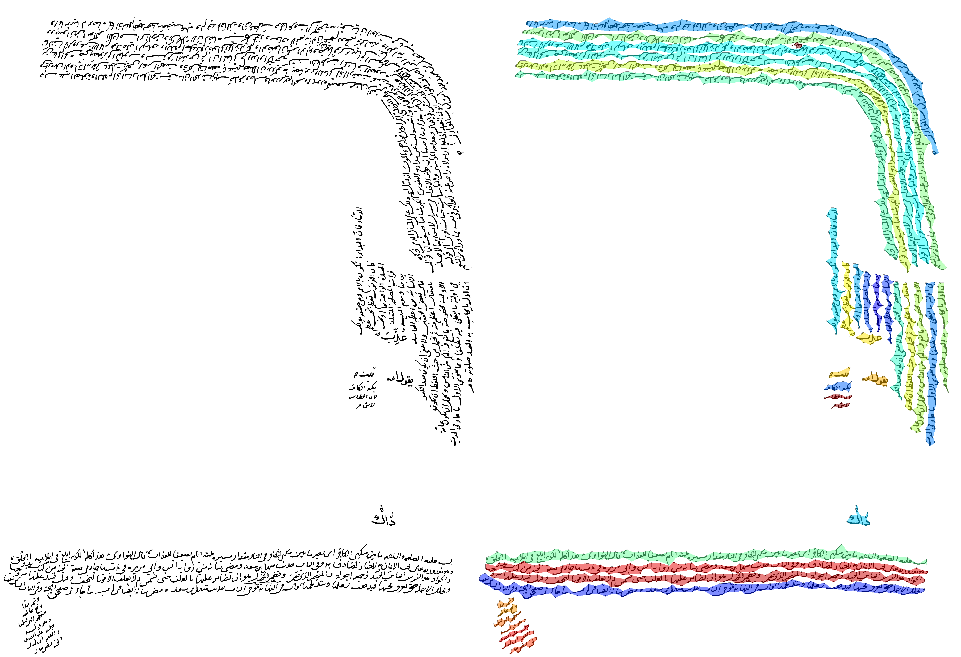
## Zorlayıcı Tarihi Veri Seti



Şekil 5.1. Satır hattı içeren veri seti ve kesin referans

Veri seti 30 sayfadan oluşan bir Arapça yazıdır. 2732 adet satırdan oluşmaktadır. Çok yönlü ve yamuk veriler barındırmaktadır. Veri seti içerinde sayfalar ve Şekil 5.1. gibi satır hatları kesin referansları bulunmaktadır

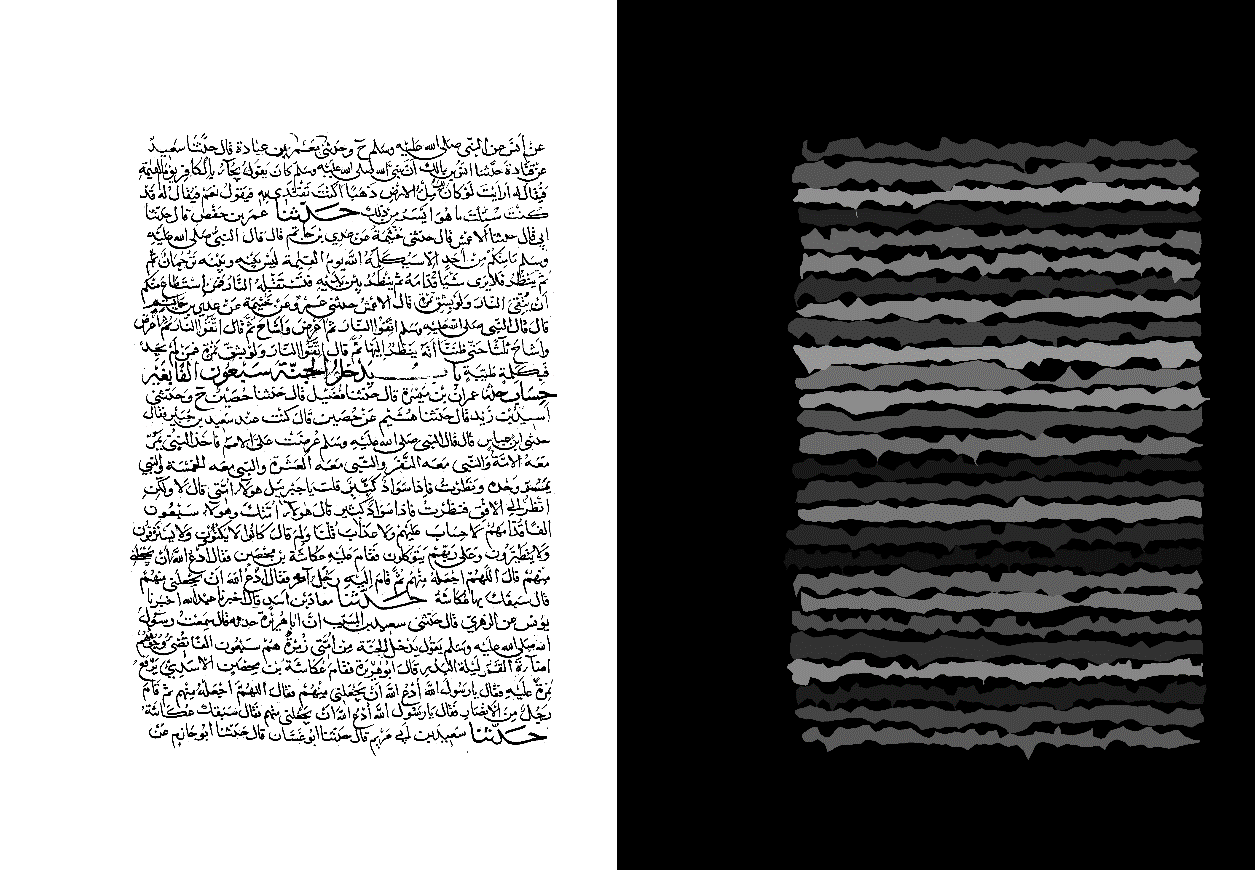
## VML-MOC: Çok Yönlü ve Eğimli El Yazması Veri Seti



Şekil 5.2. Eğimli veri seti ve etiketleri

Yüksek eğimli ve yüksek yönelimli bir el yazması Arapça veri setidir. 30 sayfadan oluşan yüksek çözünürlüklü imgelerdir. Veri setinde 20 tanesi eğitim 10 tanesi test için işaretlenmiştir. Veri seti içinde işaretlenmiş satırlar, XML tabanlı piksel kesin referansları ve satır çizgileri kesin referansları vardır.

## VML-AHTE: Arapça El Yazısı Veri Seti



Şekil 5.3. Veri setinden bir sayfa ve etiketleri

Dokunma ve örtüşme içeren bir Arapça veri setidir. Toplam 30 sayfa yüksek çözünürlüklü imge içermekledir. 20 tanesi eğitim 10 tanesi test için işaretlenmiştir. Kesin referanslar XML ve işaretlenmiş satır olarak bulunmaktadır.

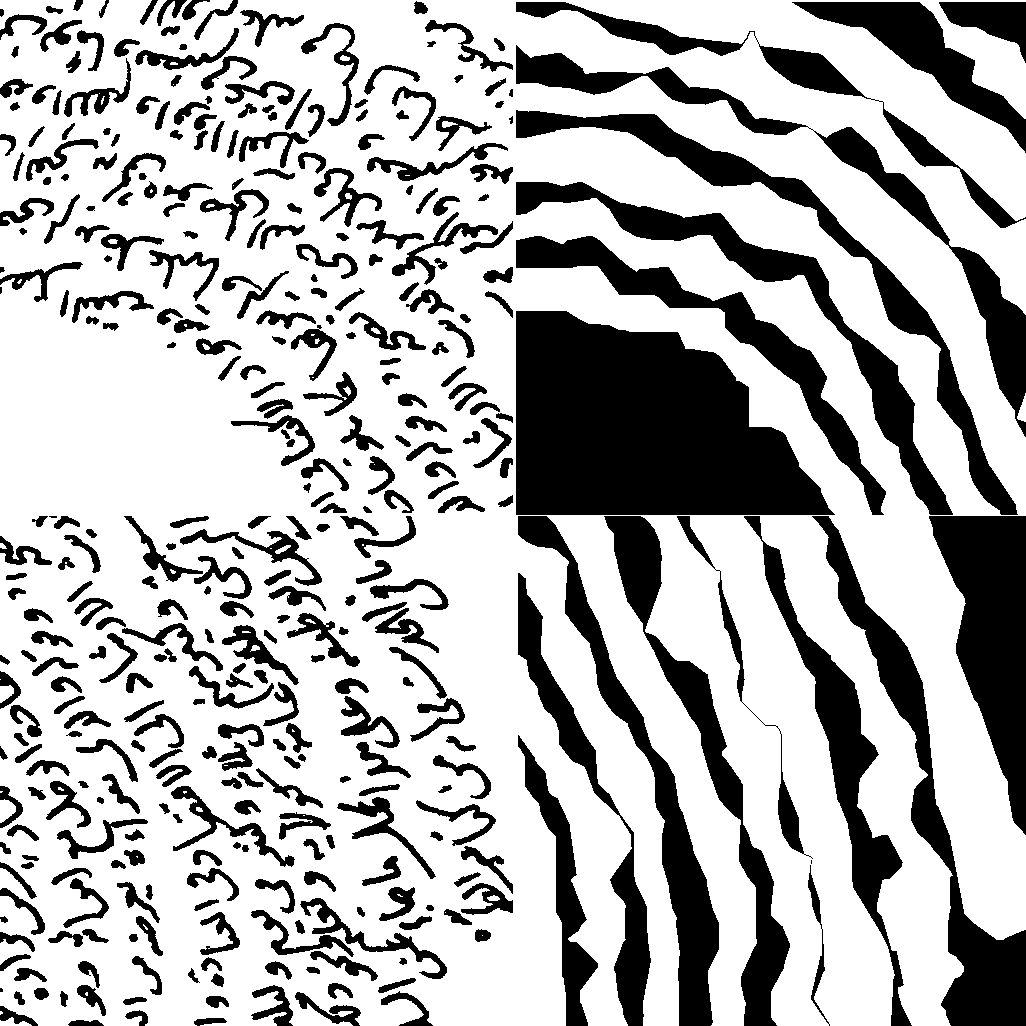
# GAN MODELİ İLE SATIR BÖLÜTLEME

Satır bölütleme işleminde GAN yapısı kullanılırken bir imgeden imgeye dönüşüm problem yapısı olarak ele alınmıştır.

## Eğitim ve Test Verilerinin Hazırlanması

Çalışmada kullanılan veri seti “VML-AHTE” ve “VML-MOC” seti birleştirerek elde edilmiştir. Her birinde 30 sayfa olmak üzere toplam 60 adet yüksek çözünürlüklü sayfa verisi elde dilmiştir. Bu sayfalar dan 8 tanesi test için ayrılırken veri seti içerindeki sorunlu veya çok az içerikli sayfalarda ayıklanarak 50 tanesi eğitim için kullanılmıştır.

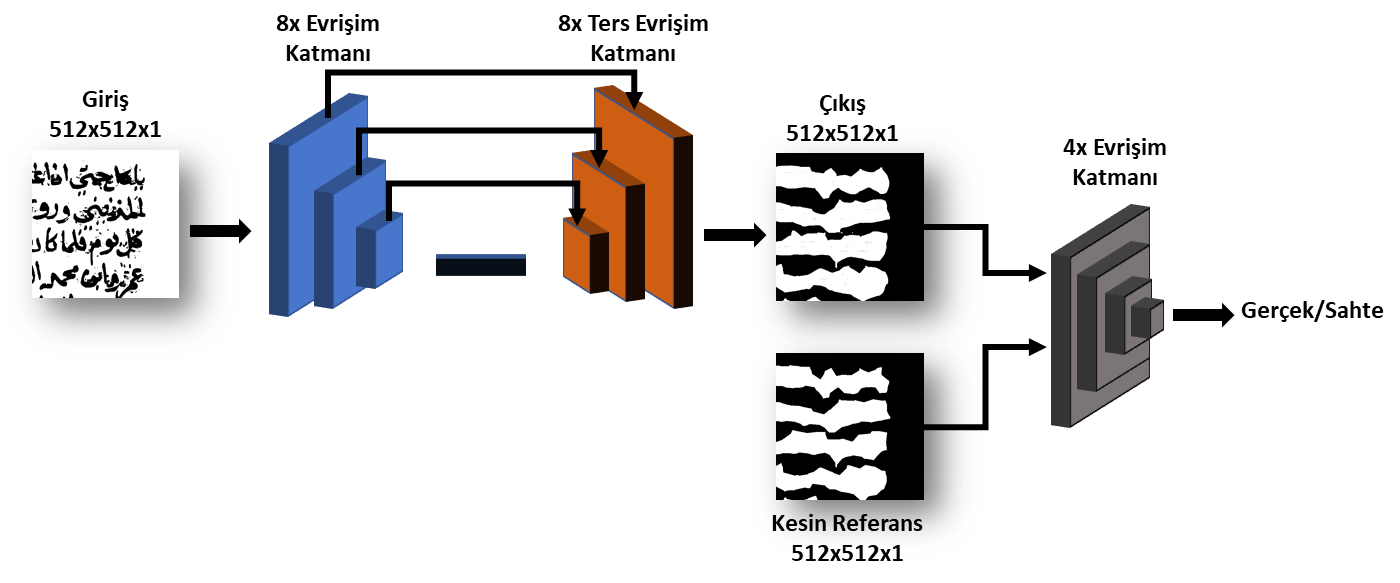
Yaklaşık olarak 4K çözünürlüğüne sahip her bir sayfadan sayfanın yazı içeriği yoğunluğu hesaplanarak bu yoğunluğa doğru orantılı şekilde rasgele noktalardan alınan parçanın içeriğinin belirli bir yazı yoğunluğunun üzerinde olmasına dikkat edilerek 512x512 boyutunda toplamda 50.000 adet eğitim oluşturulmuştur.



Şekil 6.1. Oluşturulan veri setinden örnekler

Test için ayırılan 8 sayfa eğitim verilerindeki gibi 512x512 parçalara ayrılmıştır. Tek farkı eğitimdeki gibi rasgele değil tekrardan birleştirilmek üzere düzenli bir şekilde hazırlanmıştır.

## Model ve Eğitim



Şekil 6.2. Bölütleme için kullanılan U-Net ve PatchGAN içeren GAN modeli

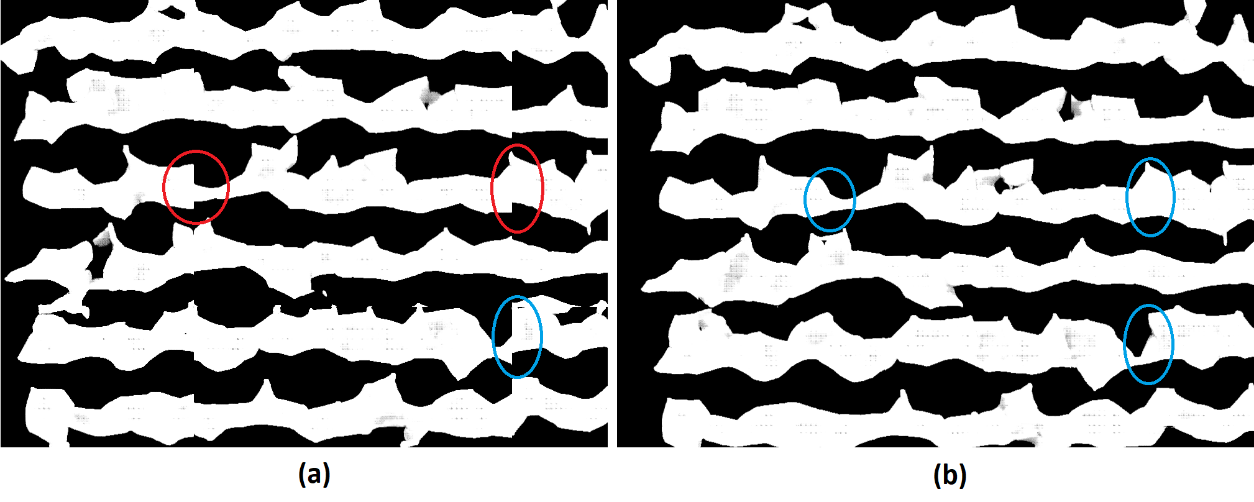
Şekil 6.2’de görüldüğü üzere GAN modelinde üretici katmanı 8 adet evrişim katmanı ile giriş verisi bir boğazdan geçirildikten sonra 8 adet ters evrişim ile tekrardan imge oluşturulmuştur. Bu aktarım sırasında her evrişim katmanında alınan kopya ters evrişim katmanında birleştirilmiştir. 4 adet evrişim katmanından oluşan bir modelimiz üretici çıkışı ve kesin referansı işleyerek verinin gerçek veya sahte durumlarına göre geri yayılım sağlamaktadır.

Model üretilen 50.000 adet veri ile NVIDIA RTX 8000 üzerinde her turda 1 yığın kullanılarak 300 tur eğitilmiştir. Eğitim 11 saat sürmüştür.

## İşleme Sonrası ve Test Verilerinin Uygulanması

Test verileri uygulanırken belirlenen 8 sayfa 512x512 boyutunda parçalara bölünmüştür. Bölünen parçalardan piksel yoğunluğuna bakılarak yazı içeriği olanlar GAN modelinde işlenir ve parçanın maskesi elde edilir. Elde edilen parçalar doğrudan birleştirildiğinde parça kenarlarında kalan alanlarda farklı maske tahminleri çıkmasından dolayı problemler ortaya çıkmaktadır.

Bu problemin çözümü için test sayfaları GAN modelinden geçirilirken kayan 512x512 pencereler kullanılmıştır. İşlem sırasında her pencere geçişinde işlenen parçanın 300x300 büyüklüğündeki orta kısımları eklenerek bütün sayfa işlenir.



Şekil 6.3. Kayan pencere uygulanmadan(a) ve uyguladıktan sonrası(b)

# PERFORMANS METRİKLERİ

Performans metrikleri için keskinlik (P), hassaslık (R) ve f1 skorlar kullanılması hedeflenmiştir. Fakat bu metrikler piksel tabanlı yapıldığında maske karakteri bulabilecek durumda olsa bile metriklerin çok düşük olduğu görülmüştür. Çıktı ve kesin referans maskelerinden yapılan bir karşılaştırmanın çalışma için anlamsız olacağı görülmüştür.

Barakat, B. (2018) çalışmasında piksel tabanlı bir yapı yerine yazı hatlarının kapsadığı karakterler üzerinden metrikler elde edilmiştir. Yapılan çalışmada kullanılan veri setinde yazı hatları satırın ortasından geçen hatlar şeklinde temsil edildiği için satırlar satısı örtüşen veya dokunan karakterler metrik hesaplamalarında bir sorun olmamaktadır. Bu çalışmada yazının ortasında geçen güvenli bir hat bulmak yerine maskelerde satırın tamamını kapsayacak bir maske ele alınmıştır. Bu sebeple metrikler hesaplanırken sınıra çok yakın karakterde diğer satırların içerdiği karakterlerle örtüşme yaşanmakta ve metrikleri anlamsız hale getirmektedir. Bu sebeple metrikler üzerinde geliştirmeler yapılmıştır ve en büyük kesişim kapsadığı karakterler ve bu karakterlerin içerdiği pikseller olmak üzere iki farklı yöntem ile hesaplama yapılmıştır.



Şekil 7.1. Örnek bir kesin referans ve çıktı

Karakter tabanlı metriklerde maskelerin içerdiği karakterler bağlı bileşenler olarak ayrılır ve her birine bir etiket atanır. Kesin referansın içerdiği karakterler Gk = [g1, g2, g3, …, gn] ve hesaplanan maskenin içerdiği karakterler Ck = [c1, c2, c3, …, cn] olmak üzere:

Bu yöntemde sadece karakterlerin tamamı dikkate alınmıştır fakat Arapça ve Osmanlıca gibi dillerde hareke denilen ek ufak işaretler bulunmaktadır. Be sebeple metrikler karakter büyüklüklerini dikkate alacak şekilde ikinci bir metrik seti türetilmiştir. Satır maskesi tarafından tanımlanan her bir bileşenin içerdiği pikseller Bp = [b1, b2, b3, …, bn] olarak tanımlanırsa kesin referansın kapsadığı bileşenler Gp = [Bg1, Bg2, Bg3, …, Bgn], çıkış maskesinin kapsadığı bileşenler Cp = [Bc1, Bc2, Bc3, …, Bcn] tanımlanmaktadır. Yeni oluşturulan C ve G değerleri ile (4), (5), (6) eşitlikleri ile metrikler yeniden hesaplanır.

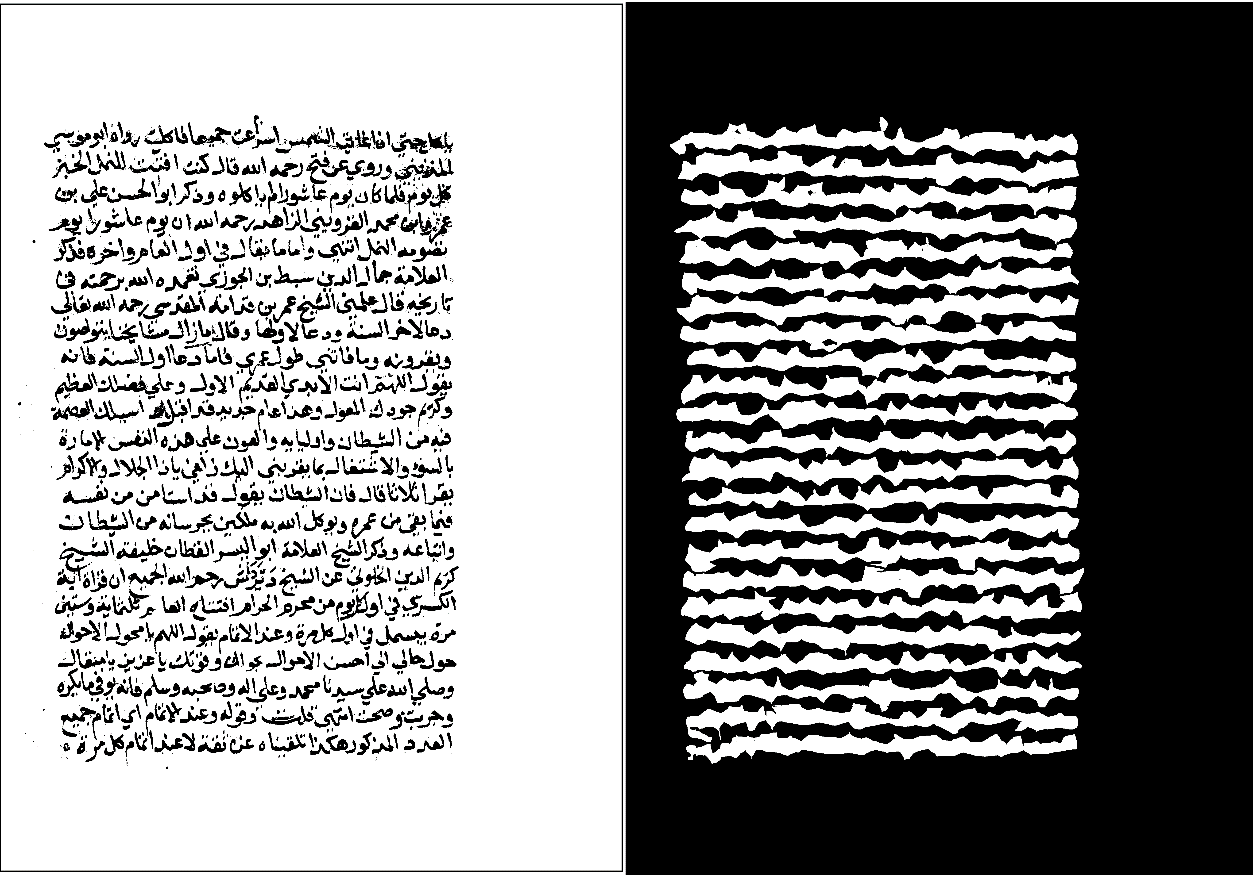


Şekil 7.2. Karakter tabanlı metrikler için metrik hesaplama yönteminin piksel tabanlı gösterilmesi

Şekil 7.2 incelendiğinden kelimenin büyük karakterleri başarılı bir şekilde yakalanırken harflere eklenen küçük noktalama işaretlerinde hatalar gözükmektedir. Sadece karakterler dikkate alındığında bu noktalama işaretlerinin ve harflerin temeli ve fazla piksel içeren karakterlerin metriklere etkisi aynı varsayılmaktadır. Yapılan değişiklik ile bu noktalamaların metriklere etkisi noktalamanın ve başarılı bulunan karakterlerin büyüklükleri oranında etkilenecektir. Ayrıca bu yöntem kullanılarak birleşiğin kesişimi (IoU) hesaplaması yapılabilmektedir.

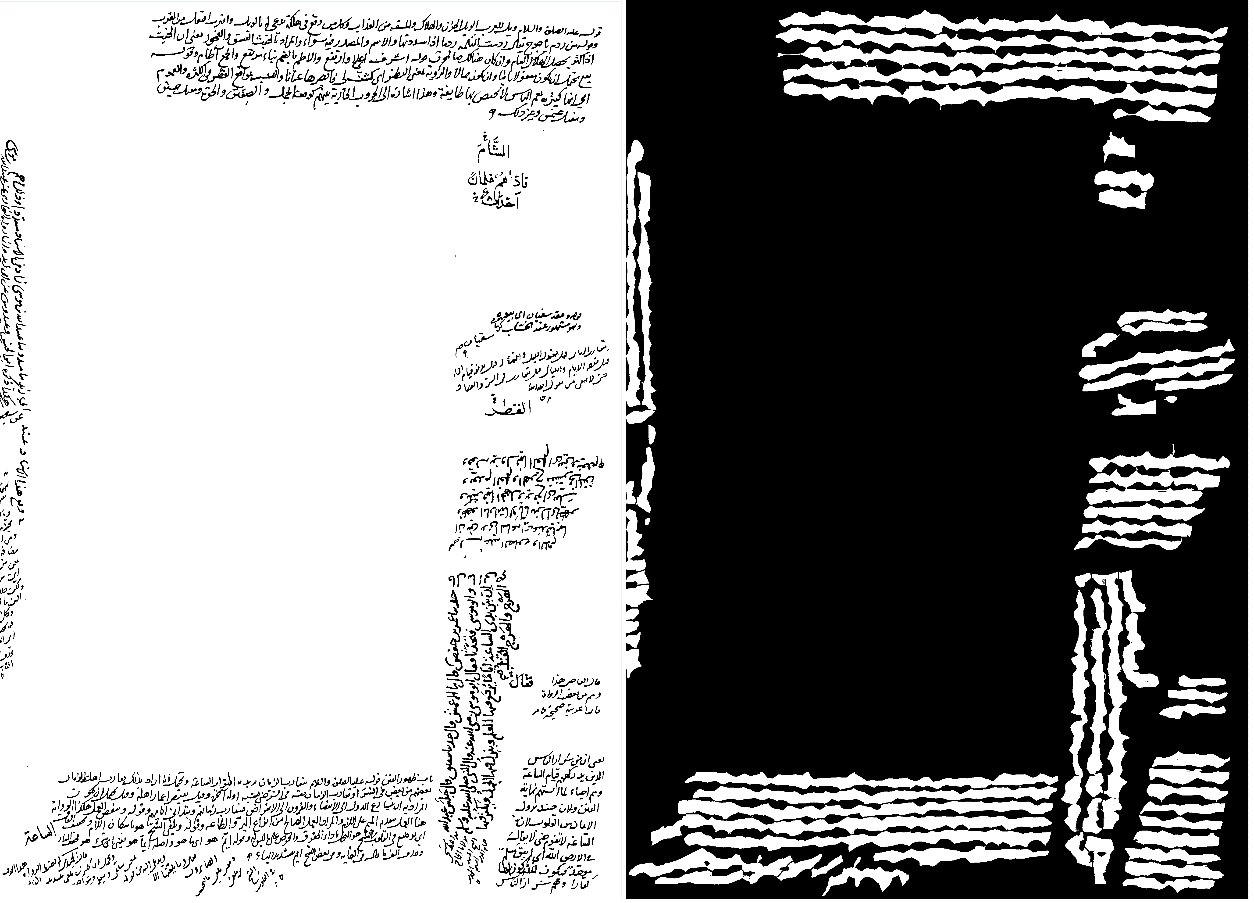
# SONUÇLAR

İlk aşama sonuçları sayfaların GAN çıktılarının 512x512’lik kayan pencere ile elde edildikten sonra ağdan kaynaklanan gürültüyü temizlemek için çıktılar bağlı bileşen analizi ile belirli bir piksel boyutundan aşağıda kalan parçalar temizlenmiştir. Saf GAN performansının hesaplanabilmesi için çıktıların bağlantı noktalara el ile ayrılmıştır.



Şekil 8.1. Eğimsiz giriş sayfası ve işlemler sonrası elde edilen çıktı

Şekil 8.1 gözle incelendiğinde yatay satırlarda GAN performansının yüksek olduğu görülmektedir.



Şekil 8.2. Çok yönlü metin içeren bir sayfa ve sonuçları

Şekil 8.2 incelendiğinde satırların belirli olduğu bölgelerde farklı yönlerde satırlar içinde başarılı maskeleme gerçeklemiştir. Fakat tek satır içindeki farklı yönelimler ya da satır sonlarına eklenen not ve benzeri yapılar satırın başarısını bozabilmektedir.

Test için ayrılan 8 sayfa işlenerek 7. Bölümde bahsedilen metrikler elde dilmiştir.

Çizelge 8.1. Test verilerinin performans metrikleri

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sayfa** | **Pk** | **Rk** | **F1k** | **Pp** | **Rp** | **F1p** | **IOUp** |
| 1 | 0.9800 | 0.9879 | 0.9838 | 0.9897 | 0.9936 | 0.9916 | 0.9836 |
| 2 | 0.9786 | 0.9870 | 0.9826 | 0.9778 | 0.9953 | 0.9857 | 0.9732 |
| 3 | 0.7526 | 0.9605 | 0.7954 | 0.7530 | 0.9652 | 0.7962 | 0.7305 |
| 4 | 0.9712 | 0.8932 | 0.9195 | 0.9839 | 0.9100 | 0.9371 | 0.8977 |
| 5 | 0.9479 | 0.9556 | 0.9511 | 0.9490 | 0.9586 | 0.9516 | 0.9093 |
| 6 | 0.9477 | 0.9755 | 0.9610 | 0.9465 | 0.9879 | 0.9657 | 0.9351 |
| 7 | 0.8178 | 0.8950 | 0.8052 | 0.8256 | 0.9107 | 0.8176 | 0.7490 |
| 8 | 0.8521 | 0.9066 | 0.8626 | 0.8563 | 0.9173 | 0.8719 | 0.8197 |
| **Ortalama** | 0.9060 | 0.9452 | 0.9077 | 0.9102 | 0.9548 | 0.9147 | 0.8747 |

Çizelge 8.1 tablosunda Pk tekil karakter etiketini dikkate alarak hesaplanan keskinlik, Rk tekil karakter etiketini dikkate alarak hesaplanan duyarlılık ve F1k bu metriklerden elde edilen F1 skorudur. Aynı metriklerin karakterin içerdiği piksel sayısına bakılarak hesaplamaları Pp, Rp, F1p olarak hesaplanmıştır. Ayrıca piksel tabanlı yapılan çalışma IOU hesabı yapılmasını sağlamıştır.



Şekil 8.3. Yatay bir satırın kesin referans ve çıktı örtüşmesi



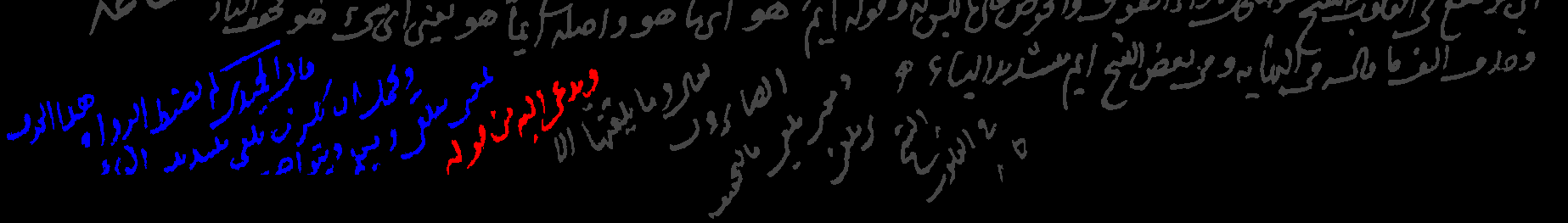
Şekil 8.4. Açılı bir satırın kesin referans ve çıktı kesişimi

Örnek satır örtüşmeleri incelendiğinde yatay satırlarda başarılı skorlar elde edilirken açılı satırlarda performans düşmektedir.

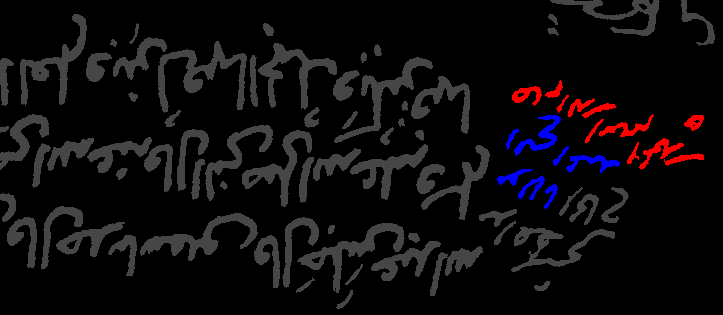
# TARTIŞMA

Bu çalışmada kullanılan imgeden imgeye transfer sağlayan GAN modeli başarısı detaylı bir şekilde incelenmiştir. Modelin yatay satırlarda karmaşık yapılar için iyi performanslar gösterdiği gözlemlenmiştir. Modelin performansının etkilendiği durumlar şunlardır.

* Bir satırdaki birden fazla yönelme,
* Aşırı Örtüşme,
* Yazı sonlarına ve diplerine eklenen notlar.



Şekil 9.1. Bir satır yapısındaki farklı yönelmelere örnek



Şekil 9.2. Satır sonuna eklenen notlar

Şekil 9.1 ve 9.2 görsellerinde kırmızı kesin referans ile örtüşme, mavi ise çıktı tarafından bulunan karakterleri temsil etmektedir.

# KAYNAKLAR

Arivazhagan, M., Srinivasan, H., & Srihari, S. (2007, January). A statistical approach to line segmentation in handwritten documents. In Document recognition and retrieval XIV (Vol. 6500, p. 65000T). International Society for Optics and Photonics.

Barakat, B., Droby, A., Kassis, M., & El-Sana, J. (2018, August). Text line segmentation for challenging handwritten document images using fully convolutional network. In 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR) (pp. 374-379). IEEE.

Cohen, R., Dinstein, I., El-Sana, J., & Kedem, K. (2014, October). Using scale-space anisotropic smoothing for text line extraction in historical documents. In International Conference Image Analysis and Recognition (pp. 349-358). Springer, Cham.

Demir, A. A., & Özkaya, U. (2019). Süperpiksel tabanlı satır bölütleme. Mühendislik Bilimleri ve Tasarım Dergisi, 7(4), 854- 868.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde- Farley, D., Ozair, S., Courville, A.C., & Bengio, Y., “Generative Adversarial Networks”, ArXiv, abs/1406.2661, 2014.

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to- image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

Koo, H. I., & Cho, N. I. (2011). Text-line extraction in handwritten chinese documents based on an energy minimization framework. IEEE Transactions on Image Processing, 21(3), 1169-1175.

Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

Kundu, S., Paul, S., Bera, S. K., Abraham, A., & Sarkar, R. (2020). Text-line extraction from handwritten document images using GAN. Expert Systems with Applications, 140, 112916.

Louloudis, G., Gatos, B., Pratikakis, I., & Halatsis, K. (2006, October). A block-based Hough transform mapping for text line detection in handwritten documents. In Tenth International Workshop on Frontiers in Handwriting Recognition. Suvisoft.

Louloudis, G., Gatos, B., Pratikakis, I., & Halatsis, C. (2008). Text line detection in handwritten documents. Pattern Recognition, 41(12), 3758-3772.

Mirza, Mehdi & Osindero, Simon. (2014). Conditional Generative Adversarial Nets.

Radford, Alec & Metz, Luke & Chintala, Soumith. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Ryu, J., Koo, H. I., & Cho, N. I. (2014). Language-independent text-line extraction algorithm for handwritten documents. IEEE Signal processing letters, 21(9), 1115-1119.

Vo, Q. N., & Lee, G. (2016, September). Dense prediction for text line segmentation in handwritten document images. In 2016 IEEE

# ÖZGEÇMİŞ

Adı Soyadı : İbrahim ÖZŞEKER

Taranmış

Fotoğraf

(3.5cm x 3cm)

Doğum Yeri ve Yılı : Seyhan 1994

Medeni Hali : Bekâr

Yabancı Dili : İngilizce

E-posta : ozsekeribrahim@gmail.com

Eğitim Durumu

Lise : Gazi Anadolu Lisesi,2012

Lisans : SDÜ, Mühendislik Fakültesi, Elektronik ve Haberleşme Mühendisliği, 2018

Yayınları

A. A. Demır, İ. Özseker and U. Özkaya, "Text Line Segmentation in Handwritten Documents with Generative Adversarial Networks," 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 2021, pp. 1-5, doi: 10.1109/INISTA52262.2021.9548523.