

ANKARA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

UZAKTAN ALGILAMA GÖRÜNTÜLERİ KULLANAN HASSAS TARIM  
UYGULAMALARI İÇİN TRANSFORMER TABANLI ANLAMSAL  
BÖLÜTLEME MODELİ TASARIMI

Uğur ALTINDAĞ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA  
2025

Her hakkı saklıdır

## ÖZET

Yüksek Lisans Tezi

### UZAKTAN ALGILAMA GÖRÜNTÜLERİ KULLANAN HASSAS TARIM UYGULAMALARI İÇİN TRANSFORMER TABANLI ANLAMSAL BÖLÜTLEME MODELİ TASARIMI

Uğur ALTINDAĞ

Ankara Üniversitesi  
Fen Bilimleri Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalı

Danışman:Dr. Öğr. Üyesi İrem ÜLKÜ

Anlamsal bölütleme bilgisayarla görme alanında yapılan çalışmaların başında gelir. Piksel düzeyinde görüntülerin sınıflandırılmasını yani bir bakıma yorumlanmasını sağlayan bu yöntem günümüzde nesne tanıma, medikal uygulamalarda hastalık tespiti gibi çeşitli alanlarda kullanılmaktadır. Hassas tarım günümüzde giderek uygulaması artan bir alan olup bu alanda da anlamsal bölütlemeye yönelik çalışmalar yapılmakta ve gerekli görülmektedir. Transformer teknolojisine geldiğimizde ise doğal dil işleme teknolojilerinde öne çıkmış bugün kullanılmakta olan sohbet robotlarının arkaplanında tercih edilen bir model haline gelmiştir. Daha sonra bilgisayarla görme alanında da transformer mimarisi kullanılmaya başlanmış ve oldukça başarılı sonuçlar elde edilmiştir. Bu çalışmada görüntü işleme teknolojilerinde özellikle anlamsal bölütleme alanında başarılı sonuçlar vermesiyle öne çıkan UNet mimarisinden esinlenilerek farklı bir mimari ortaya koyulmaya çalışılmıştır. Özellikle de hassas tarım uygulamalarında kullanılan uzaktan algılama görüntüleri için anlamsal bölütleme yapmak adına Transformer tabanlı bir model geliştirilmeye çalışılmıştır. Tasarlanan model için farklı dikkat mekanizmaları da denenmiş ve kullanılan veri setleri için bunlar da test edilmiştir. Sonuç olarak, elde edilen veriler neticesinde önerilen tasarımın uzaktan algılama görüntülerinin anlamsal bölütlemesinde gayet başarılı sonuçlar verdiği ve bu alanda yapılacak çalışmalarda tercih edilebilecek bir model olduğu düşünülmektedir.

**Şubat 2025, 59 sayfa**

**Anahtar Kelimeler:** U-Net, Transformers, Vision Transformer, UnetFormer, ikili bölütleme, anlamsal bölütleme, dikkat mekanizmaları, uzaktan algılama görüntüleri

## ABSTRACT

Master Thesis

### DESIGN OF A TRANSFORMER BASED SEMANTIC SEGMENTATION MODEL FOR PRECISION AGRICULTURE APPLICATIONS USING REMOTE SENSING IMAGES

Uğur ALTINDAĞ

Ankara University  
Graduate School of Natural and Applied Sciences  
Department of Computer Engineering

Supervisor:Dr. Öğr. Üyesi İrem ÜLKÜ

Semantic segmentation is at the forefront of research in the field of computer vision. This method, which enables the classification of images at the pixel level, essentially allowing for their interpretation, is currently used in various fields such as object recognition and disease detection in medical applications. Precision agriculture is an increasingly applied field today, and in this area, studies aimed at semantic segmentation are being conducted and deemed necessary. When it comes to transformer technology, it has emerged in natural language processing technologies and has become a preferred model in the background of the chatbots currently in use. Later, the transformer architecture was also used in the field of computer vision, yielding quite successful results. In this study, a different architecture has been attempted by drawing inspiration from the UNet architecture, which stands out for its successful results, particularly in the field of semantic segmentation within image processing technologies. Especially for remote sensing images used in precision agriculture applications, an attempt has been made to develop a Transformer-based model for semantic segmentation. Different attention mechanisms were also tried for the designed model, and these were tested on the used datasets. As a result, based on the obtained data, it is considered that the proposed design has yielded quite successful results in the semantic segmentation of remote sensing images and could be a preferred model for future studies in this field.

**February 2025, 59 pages**

**Keywords:**U-Net, Transformers, Vision Transformers, UnetFormer, binary segmentation, semantic segmentation, attention mechanisms, remote sensing images

## TEŐEKKÜR

Bu tez alıőmasının her aőamasında verdiđi bilgi, öneri ve yönlendirmelerle desteđini esirgemeyen danıőmanım Dr. Öğretim Üyesi İrem Ülkü'ye ve bu süreçte desteklerini hiç esirgemeyen aileme ve dostlarıma en içten duygularıyla teşekkürlerimi sunarım.

Uđur ALTINDAĐ

Ankara, Őubat 2025

## İÇİNDEKİLER

<b>TEZ ONAYI</b>	
<b>ETİK.....</b>	<b>i</b>
<b>ÖZET.....</b>	<b>ii</b>
<b>ABSTRACT.....</b>	<b>iii</b>
<b>TEŞEKKÜR.....</b>	<b>iv</b>
<b>KISALTMALAR DİZİNİ .....</b>	<b>vii</b>
<b>ŞEKİLLER DİZİNİ .....</b>	<b>viii</b>
<b>ÇİZELGELER DİZİNİ .....</b>	<b>ix</b>
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>1.1 Motivasyon ve Problem Tanımı.....</b>	<b>3</b>
<b>2. KAYNAK TARAMASI .....</b>	<b>6</b>
<b>3. MATERYAL ve YÖNTEM.....</b>	<b>12</b>
<b>3.1 Veri Setleri .....</b>	<b>12</b>
<b>3.2 Sinir Ağları .....</b>	<b>12</b>
<b>3.3 Derin Öğrenme .....</b>	<b>13</b>
<b>3.4 Evrimsel Sinir Ağları .....</b>	<b>14</b>
<b>3.4.1 Evrim katmanı .....</b>	<b>14</b>
<b>3.4.1.1 Padding.....</b>	<b>15</b>
<b>3.4.1.2 Stride .....</b>	<b>15</b>
<b>3.4.2 Havuzlama katmanı .....</b>	<b>16</b>
<b>3.4.3 Aktivasyon katmanı .....</b>	<b>17</b>
<b>3.4.4 Tam bağlantılı katman .....</b>	<b>18</b>
<b>3.5 U-Net Mimarisi.....</b>	<b>18</b>
<b>3.6 ResNet.....</b>	<b>20</b>
<b>3.7 Transformers .....</b>	<b>22</b>
<b>3.8 Vision Transformer.....</b>	<b>24</b>
<b>3.9 U-Netformer Mimarisi.....</b>	<b>25</b>
<b>3.9.1 Küresel-yerel dikkat yapısı.....</b>	<b>26</b>
<b>3.9.2 Öznitelik iyileştirme başlığı.....</b>	<b>28</b>
<b>3.10 SENet Modülü .....</b>	<b>29</b>
<b>3.11 CBAM Modülü .....</b>	<b>30</b>
<b>3.12 CCNet Modülü .....</b>	<b>32</b>
<b>3.13 DANet Modülü .....</b>	<b>33</b>

<b>3.14 Atlama Bağlantısında Dikkat Modülü İçeren Model (AiSCUNetFormer).....</b>	<b>35</b>
<b>3.15 Hiperparametreler .....</b>	<b>36</b>
<b>3.15.1 Kayıp fonksiyonları.....</b>	<b>37</b>
<b>3.15.2 Optimizasyon algoritmaları .....</b>	<b>38</b>
<b>3.15.3 Mini-batch boyutu.....</b>	<b>38</b>
<b>3.15.4 Başlatıcı .....</b>	<b>39</b>
<b>3.15.5 Değerlendirme metrikleri.....</b>	<b>39</b>
<b>3.15.5.1 Doğruluk .....</b>	<b>40</b>
<b>3.15.5.2 Kesinlik.....</b>	<b>40</b>
<b>3.15.5.3 Duyarlılık .....</b>	<b>40</b>
<b>3.15.5.4 F1 skoru.....</b>	<b>41</b>
<b>3.15.5.5 Jaccard indeksi .....</b>	<b>41</b>
<b>4. BULGULAR ve TARTIŞMA.....</b>	<b>42</b>
<b>4.1 Model Değerlendirme Parametreleri .....</b>	<b>42</b>
<b>4.2 Rit-18 Veri Seti .....</b>	<b>43</b>
<b>4.3 Cxef Veri Seti.....</b>	<b>43</b>
<b>4.4 Rit-18 Veri Setiyle Elde Edilen Sonuçlar .....</b>	<b>44</b>
<b>4.5 Cxef Veri Setiyle Elde Edilen Sonuçlar.....</b>	<b>48</b>
<b>4.6 Tartışma .....</b>	<b>51</b>
<b>5. SONUÇLAR .....</b>	<b>53</b>
<b>5.1 Öneriler .....</b>	<b>53</b>
<b>KAYNAKLAR.....</b>	<b>54</b>
<b>ÖZGEÇMİŞ.....</b>	<b>59</b>

## KISALTMALAR DİZİNİ

CNN	Evrişimsel sinirsel ağ (Convolutional Neural Network)
FCN	Tamamen evrişimsel ağlar (Fully Convolutional Networks)
ReLU	Düzeltilmiş doğrusal ünite (Rectified Linear Unit)
RNN	Yinelenen sinirsel ağlar (Recurrent Neural Network)
MLP	Çok katmanlı algılayıcı (Multi Layer Perceptron)
NLP	Doğal dil işleme (Natural Language Processing)
ViT	Vision Transformer

## ŞEKİLLER DİZİNİ

Şekil 3.1 Derin sinir ağı .....	13
Şekil 3.2 Evrişimsel sinir ağları mimarisi .....	14
Şekil 3.3 Padding işlemi.....	15
Şekil 3.4 Stride işlemi .....	16
Şekil 3.5 Maksimum havuzlama ve ortalama havuzlama işlemi .....	16
Şekil 3.6 Aktivasyon fonksiyonları.....	18
Şekil 3.7 U-net mimarisi .....	20
Şekil 3.8 ResNet-18 mimarisi .....	21
Şekil 3.9 Transformers mimarisi.....	23
Şekil 3.10 Ölçeklendirilmiş nokta çarpımı dikkat ve çok-başlı dikkat yapıları.....	23
Şekil 3.11 Vision transformer’da kullanılan transformer encoder modeli.....	24
Şekil 3.12 Unetformer mimarisi.....	25
Şekil 3.13 Standart transformer bloğu ve küresel-yerel transformer bloğu.....	26
Şekil 3.14 Etkili küresel-yerel dikkat mekanizmasının yapısı .....	27
Şekil 3.15 Öznitelik Geliştirme Başlığı .....	29
Şekil 3.16 SENet Bloğu .....	30
Şekil 3.17 Konvolüsyonel blok dikkat modülünün yapısı .....	31
Şekil 3.18 Criss-cross dikkat modülü bloğu .....	33
Şekil 3.19 DANet modülünün yapısı .....	34
Şekil 3.20 Yapılan çalışmada ortaya koyulan mimari.....	36
Şekil 4.1 RIT-18 veri seti için farklı dikkat mekanizmalarıyla yapılan anlamsal bölütlemenin renklendirilmiş sonuçları .....	48
Şekil 4.2 Cxef veri seti için farklı dikkat mekanizmalarıyla yapılan anlamsal bölütlemenin renklendirilmiş sonuçları .....	51

## ÇİZELGELER DİZİNİ

Çizelge 3.1 Veri setlerinin özellikleri .....	12
Çizelge 3.2 Karmaşıklık matrisi.....	40
Çizelge 4.1 Rit-18 veri setinde iki katlı çapraz doğrulama için seçilen parametreler.....	43
Çizelge 4.2 Rit-18Veri seti 5 katlı çapraz doğrulama ile elde edilen sonuçlar .....	44
Çizelge 4.3 Rit-18 veri seti 5 katlı çapraz doğrulamadan elde edilen model değerlendirme parametrelerinin karşılaştırılması .....	45
Çizelge 4.4 Rit-18 veri setinde iki katlı çapraz doğrulama ile elde edilen Jaccard ve F1 skoru .....	46
Çizelge 4.5 Rit-18 veri seti, iki katlı çapraz doğrulama ve ResNet18 mimarisi kullanılarak elde edilen değerlendirme parametreleri .....	46
Çizelge 4.6 Rit-18 veri seti, iki katlı çapraz doğrulama ve ResNet34 mimarisi kullanılarak elde edilen değerlendirme parametreleri .....	47
Çizelge 4.7 Rit-18 veri seti, mini-batch boyutu 16, iki katlı çapraz doğrulama ve ResNet50 mimarisi kullanılarak elde edilen sonuç ve değerlendirme parametreleri değerleri .....	47
Çizelge 4.8 Cxef veri setinde ResNet18 ve ResNet34 kodlayıcı mimarileri ve dikkat mekanizmalarının kıyaslanması .....	49
Çizelge 4.9 Cxef veri seti değerlendirme parametreleri.....	49
Çizelge 4.10 Cxef veri seti, mini-batch boyutu 16, iki katlı çapraz doğrulama ve ResNet50 mimarisi kullanılarak elde edilen sonuç değerleri .....	50
Çizelge 4.11 Cxef veri seti, kodlayıcı ResNet50 mimarisinin değerlendirme parametreleri .....	50

## 1. GİRİŞ

Teknoloji, sürekli deęişmekte ve gelişmekte, geline noktada her gün yerini alan yeni teknolojiler her geçen gün insan yaşamını daha fazla kolaylaştırmaktadır. Teknolojik cihazlar, araçlar daha kolay erişilebilir bir hale gelmekte ve günlük yaşamın bir parçası haline bile gelmektedir. Bu teknolojik gelişmelerle birlikte farklı yöntem ve teknikler de ortaya koyulmakta ve birçok alanda çeşitli uygulamalarda kullanılmaya başlanmış durumdadır. Yapay zeka uygulamaları da bu teknolojilerin başında, giderek artan bir öneme sahip olmuş, karşılaşılan çeşitli sorunların çözümüne ulaşmak için faydalanılan bir araç haline gelmiştir.

Bilgisayar bilimlerinin babası olarak bilinen Alan Turing (1950) yayınlamış olduđu makalede “Makineler düşünebilir mi?” sorusunu sormuş ve daha sonrasında bir insan sorgulayıcının, bir bilgisayar ve insanın yazılı yanıtlarını ayırt etmeyi denediđi ünlü Turing Testini ortaya atmış ve bu yapay zeka tarihinin önemli bir parçası olarak tarihte yerini almıştır. 1956 yılında John McCarthy vermiş olduđu bir konferansta yapay zeka kavramını ortaya atmıştır. Rosenblatt (1958) Perceptron kavramını ortaya koymuş ve Mark I Perceptron adıyla, deneme yanılma ile öğrenen ilk bilgisayar tabanlı sinir ađları makinesini üretmiştir. İlk uygulaması bir yazılım iken devamında özel donanımlarla görsel algılama yapan bir makine projesi ortaya konmuştur. Daha sonra ilerleyen yıllarda geri yayılım yapan sinir ađlarının kullanımı yapay zeka uygulamalarında giderek artmıştır. Geline noktada yapay zeka teknolojileri, sohbet robotları, robotlar, otonom araçlar, konuşma tanıma, yazı tanıma başta olmak üzere hassas tarım, nesnelerin interneti gibi farklı alanlarda da kullanılmaya başlanmış ve bu alanlarda karşılaşılan problemlere çözümler üretilmesine olanak sağlamıştır.

Makine öğrenmesi, yapay zekanın bir alt dalıdır. Genel olarak bilgisayarların ve makinelerin insanların öğrenme sürecini taklit ederek, otonom bir şekilde çeşitli görevleri yerine getirmeye, daha fazla deneyim ve veriyle performansını ve doğruluđunu artırmaya odaklanmaktadır. Makine öğrenmesi algoritmaları, bir karar süreci, bir hata fonksiyonu ve optimizasyon süreci olmak üzere üç ana parçadan oluşmaktadır. Sinir ađları, makine öğrenmesinin alt dalıdır. Derin öğrenme de sinir

ağlarının alt dalı olarak kabul edilmektedir. Sonuçta hepsi yapay zekanın alt dallarıdır. Derin öğrenme, çok katmanlı sinir ağlarıyla insan beyninin kompleks karar verme gücünün taklit edilmesidir. Geleneksel makine öğrenmesi yöntemleri basit bir veya iki katmanlı sinir ağları kullanır. Derin öğrenme modelleri ise genellikle üçün üzerinde hatta binlerce katman ile modellerin eğitilmesini sağlarlar.

Evrişimli sinir ağlarından(CNN) önce zaman alıcı özellik çıkarım metotları görsellerdeki nesnelere tanımlamak için kullanılmaktaydı. Geline nokta CNN özellikle matris çarpımı ilkelerinden faydalanarak görüntü sınıflandırma, nesne tanıma işlemlerine daha ölçeklenebilir bir yaklaşım sunmaktadır. CNN yüksek doğruluk sunarak nesne tanıma, görüntü tanıma gibi en popüler alanlarda yer edinmiştir (Ajit vd. 2020).

Transformer mimarisi Vaswani vd. tarafından (2017) ilk olarak “Attention is All You Need” makalesi ile tanıtılmıştır. Kısa süre içinde büyük dil modelleri gibi uygulamalarda kullanılmaya başlanmıştır. Transformers mimarisi tekrarlayan sinir ağlarının(RNN) aksine verileri sıralı olarak değil, paralel olarak işleyerek eğitim sürecinin çok daha verimli ve kısa sürede olmasını sağlamaktadır. Transformer mimarileri, doğal dil işleme, ve anlık yazı ve konuşma çevirmelerinin yapılmasında kullanılırken, Vision Transformers(ViT) daha sonrasında görüntü işleme için ortaya atılmıştır.

Görsel anlamsal bölütleme, makine öğrenmesi ve bilgisayarla görme alanında araştırmacılar için giderek daha ilgi çekici hale gelmektedir. Otonom araçlar, iç mekan konumlandırma gibi çeşitli uygulamalarda etkili ve doğru bölütleme mekanizmaları gerekmektedir(Garcia vd., 2017). Görsel sınıflandırma bir görselde var olan nesnelere tanıma ve sınıflandırma da kullanılmaktadır. Nesne lokalizasyonu ise kabaca bu nesnelere yerini saptamakla alakalıdır. Anlamsal bölütleme bunların devamı niteliğinde ince ince işlenmiş bir sonucu elde etmekte, görsel üzerinde sınıflandırma yapmakta ve bunu piksel piksel çıkarımda bulunarak nesnelere uzaysal konumunu da elde etmektedir.

Bir görüntünün önemli kısımlarına dikkati yönlendirmek ve ilgisiz kısımları göz ardı etmek için kullanılan yöntemler dikkat mekanizmaları olarak adlandırılır. İnsanın görme sistemi de karmaşık sahneleri etkili ve verimli bir şekilde analiz edebilmek için böyle bir yol kullanır (Rensink 2000). Araştırmacılar bu tarz dikkat mekanizmalarını bilgisayarlı görme sistemlerinde kullanmaya ve bu sistemlerin performansını artırmaya çalışmaktadırlar. Dikkat mekanizması, girdinin önemine göre öznelikleri uygun bir şekilde ağırlıklarından dinamik bir seçim süreci olarak görülebilir. Geline nokta anlamsal bölütleme, nesne tanıma, görsel oluşturma ve daha birçok görsel çalışmada oldukça başarılı olmaktadır.

### **1.1 Motivasyon ve Problem Tanımı**

Sürekli artan insan nüfusu, tarıma ve doğal kaynaklara öngörülemez bir talebi beraberinde getirmiştir (Foley vd., 2011). Bunun yanı sıra yaşlanan nüfus yapısı, işlenebilir arazilerin azalması, gittikçe öngörülemez hale gelen iklimsel koşullar ve üretim maliyetleri ile tarım, global çapta büyük sorunlarla karşı karşıyadır (Su vd. 2023). Bu yüzden verimli, sürdürülebilir ve kârlı tarım sistemlerinin ortaya konması elzemdir. Tarıma yönelik bu talebe karşılık farklı stratejiler ve pratik çözümler aranmaktadır. Hassas tarım, sürdürülebilir tarım sistemleri için anahtar bir bileşen konumundadır. Geleneksel tarım işlenen tüm sahayı bir bütün olarak görürken hassas tarım bu alanı küçük parçalar olarak ele alır. Hassas tarım için tercih edilen yöntemlerden birisi uzaktan algılamadır. Uzaktan algılama en basit tabiriyle verilerin uzaktan elde edilmesidir. Veri sensörleri içeren bir hava aracı ya da uydu aracılığıyla veriler elde edilebilir. Uzaktan algılama kavramı henüz ortaya atılmadığı 1930'lu yıllarda ise hava fotoğrafçılığı gibi bugün konvansiyonel kalan yöntemlerde kullanılmıştır. Ancak bu yöntemler ayrıca analiz ve detaylı saha çalışması gerektirdiğinden çok pahalı ve zaman alıcıdır (Sishodia vd. 2020). Uzaktan algılama teknolojisi ile farklı dalga boylarındaki görsellerle mahsuller ve bitkiler kolayca gözlenebilmektedir (Tsouros vd., 2019).

Uzaktan algılama görüntülerinin kullanımı giderek artmakta tarım başta olmak üzere çevre yönetimi, kentlerin altyapı planlaması ve güvenlik amaçlı stratejik analizler gibi

alanlarda kullanılmaktadır. Hassas tarım uygulamaları da gittikçe yaygınlaşmakta ve uzaktan algılama görüntülerinin bu alandaki kullanımı da aynı şekilde artış göstermektedir. Hassas tarımın olabilecek en verimli şekilde gerçekleştirilmesi doğal olarak istenen bir durumdur. Ancak karşılaşılan bitkilerde su stresi gibi çeşitli problemler hassas tarımın verimsiz ya da istenenden ve beklenenden daha az verimli olacak şekilde gerçekleştirilmesine neden olabilmektedir. Uzaktan algılama görüntülerinin bölütlenmesi ile mahsullerdeki hastalıklar, mahsullerdeki haşereler, ve yabancı ot istilaları gibi görüntülerdeki anormalliklerin başarılı bir şekilde belirlenmesi ve bunlara alınacak önlemler hassas tarımın daha verimli bir şekilde gerçekleştirilmesini sağlayacaktır. Bunun yanı sıra hasat zamanı kararı gibi konularda da kolaylık sağlayabilmekte ve daha kolay ve doğru bir karar verilmesini sağlamaktadır. Anlamsal bölütleme görüntüdeki her bir pikselin belirli nesne sınıflarına ayrılmasını sağlamaktadır. Bu piksel tabanlı ayrılma görüntünün çok daha kapsamlı incelenmesi ve yorumlanmasına olanak sağlamaktadır. Günümüzde uzaktan algılama görüntülerinin doğru ve hızlı bir şekilde bölütlenmesi üzerine çalışmalar devam etmektedir.

Bu çalışmada iki farklı veri setinde ikili anlamsal bölütleme işlemi gerçekleştirilmiştir. Veri setlerinin birinde piksel piksel ağaç tespiti yapılmakta diğesinde ise yine piksel bazında bitkilerde sarı pas hastalığının tespiti yapılmaktadır. Bu tespitlerin iyileştirilmesi ve varolan sonuçlardan daha iyi sonuçlar elde edilmesi hedeflenmiştir. Bununla hassas tarım alanında yukarıda bahsedilen anormallikler başta olmak üzere varolan problemlerin tespitinin doğru bir şekilde yapılmasına olanak sağlayacak ve yine farklı alanlarda da kullanılabilir bir derin öğrenme modeli ortaya konulması planlanmaktadır. Bu sayede ortaya konulan model hassas tarım başta olmak üzere farklı alanlarda da kullanılabilir ve o alandaki mevcut uygulamaların daha başarılı olmasını sağlayacaktır.

Bu çalışmanın amacı, uzaktan algılama görüntüleri kullanılarak, kodlayıcı kısmında kalıntı sinir ağları olarak bilinen ResNet ve kod çözücü kısmında ise bir Transformer mimarisi ile anlamsal bölütleme gerçekleştirmektir. Ortaya konulan bu model atlama bağlantıları da kullanan U-Net benzeri bir mimari olup, mimarinin atlama bağlantılarına farklı dikkat modülleride eklenerek mevcut veri setleri için ve hassas tarım

uygulamaları için görüntü analizinin doğruluğunu ve etkinliğini artıracak, verimli bir modelin ortaya konması da amaçlanmaktadır. Bununla birlikte tarımsal özelliklerin hassas bir şekilde belirlenmesini sağlamak ve hassas tarım alanında daha iyi kararlar alınmasına olanak sağlamak hedeflenmektedir.

## 2. KAYNAK TARAMASI

Görsel bölütlemede kullanılması için Comaniciu ve Meer (1997) yapmış oldukları çalışmada ortalama kaydırma algoritması ve yoğunluk gradyanları analizine dayanan önemli görsel özelliklerin kurtarılmasını içeren bir teknik önermişlerdir. Yu ve Bajaj (2002) görüntü bölütlemek için aktif yılan modeli ve bölge birleştirme kullanmışlardır. Chen ve Murphey (2002) renkli görsel bölütlemesi için renk uzayında bulanık kümeleme işlemi ve görüntü uzayında bölge bölütlemesinden oluşan iki aşamalı bir bölütleme sistemi ortaya koymuşlardır.

Abadpour ve Kasaei (2008) renkli görüntü bölütlemesi için yenilikçi bir mesafe fonksiyonuyla genel kümeleme algoritmasını kullanan yeni bir yöntem önermişlerdir. Bu yöntem yerel minimum tuzaklarına yakalanma olasılığını da azaltmaktadır. Wu vd. (2018) çoklu lojistik regresyon kullanarak tek ağaç bölütlemesi yapmışlar ve ortaya koydukları model 220 ağaçtan 184'ünü tespit edebilmiştir. Immitzer vd. (2012), ağaç tepelerinin elle belirlenmiş olan ve güneş ışığı alan bölgelerinin spektrumlarını kullanarak rastgele orman algoritmasıyla sınıflandırma gerçekleştirmişlerdir. 10 ağaç türünde %82 genel doğruluk oranı elde etmişlerdir. Jin vd. (2014), el yapımı özelliklerle makine öğrenimi özelliklerin kombine edilmesiyle nesne tanıma çalışması gerçekleştirmişlerdir. Wu vd. (2021), K-ortalama metodunu kullanan yeni bir varyasyonel bölütleme modeli önermişlerdir. İlk önce renk seti insan eliyle veya K-ortalama yöntemiyle belirlenmektedir. Daha sonra değişken model bölütleme doğruluk değeriyle her piksel için en uygun rengin seçimini sağlanarak daha rafine bir sonuç elde edilmektedir. Plankman vd., (2020) ağaç tepelerinin yerel maksimum filtre ile tespit edildiği bir yapay kanopi yüksekliği modelinden ağaç tepelerini otomatik olarak belirlemişlerdir. Referans veriyi nesne tabanlı rastgele orman modelinde kullanmışlar ve sonuçları bağımsız bir veri setiyle doğrulanmıştır. Sonuç olarak ağaç türlerinin sınıflandırmalarında %84.5 doğruluk oranı elde etmişlerdir. Liu vd. (2024) zaman ölçekli veri setiyle dağlık ağaç türleri üzerinde çalışma yapmışlardır. Aylık veri setleri ilkbahar ve sonbaharda meydana gelen, hızla değişen fenolojik bilgileri yakalama avantajıyla mevsimsel ve yıllık veri setlerine göre daha yüksek doğruluk değerlerine ulaşmıştır. Rastgele orman algoritması, aynı veri setinde destek vektör makinesi

algoritmasına göre her zaman daha iyi sonuç vermiştir. Bilgisayarla görme teorisine dayanan yeni bir bireysel ağaç tepesi segmentasyon yöntemi Ting vd., (2021) tarafından önerilmiştir. Bu yöntem çift gaussian filtresi ve ağaç tepesi tarama stratejisini birleştirerek farklı ağaç boyutları için esnek bir filtreleme süreci ve yan dallar tarafından oluşturulan sahte ağaç tepelerinin hariç tutulmasını sağlar.

Culvenor (2002) yayınlamış olduğu çalışmada yüksek uzamsal çözünürlüklü uzaktan algılama görüntülerini kullanarak ağaç tepesi belirlemek için TIDA adını verdiği bir algoritma önermektedir. Algoritma, görüntü boyunca yerel maksimum ve yerel minimumun tespiti ve ağaç tepesi piksellerinin gruplanmasını içermektedir. Yapılan testler algoritmanın eş yaşlı orman örtülerinde en yüksek doğruluk oranını elde ettiğini göstermiştir.

Yukarıda bahsedilen çalışmalardan da görülebileceği üzere yakın bir geçmişe kadar uzaktan görüntüleme ağaç tespiti ve bölütleme alanında makine öğrenimi yöntemleri ve çeşitli algoritmalar hakim oldu. Bu alanda yapılan çalışmalar elle hazırlanmış özellik çıkarımlarının yanı sıra özellik kodlamasına ve özellik havuzlamasına da dayanıyordu. Ancak çıkarılan özelliklerin düşük seviyeli karakteristikleri nedeniyle bu yöntemler mekansal, zamansal ve spektral yüksel seviyeli özellikleri yakalamayı başaramıyordu. Uzaktan algılama görüntülerini piksel bazlı sınıflandırmada derin öğrenme şeması üç ana aşamadan oluşur: Veri girişi, hiyerarşik derin öğrenme eğitimi ve sınıflandırma (Zhang vd. 2016). Günümüzde derin öğrenme, uzaktan algılamada veri analizinin her noktasında yer almaktadır. Görüntü ön işleme, hedef tanıma, piksel bazlı sınıflandırmadan son zamanlardaki zorlayıcı çalışmalar olan yüksek düzeyde anlamsal öznitelik çıkarımı ve uzaktan algılama görüntülere kadar (Zhang vd., 2016). Evrimsel sinir ağları evrişim ve havuzlama katmanlarının serpiştirilerek ham görsellerden orta ve yüksek seviye soyut özelliklerin çıkarılabileceğini göstermiştir. Yeni çalışmalar CNN'lerin büyük ölçekli görsel tanıma, nesne tanıma ve anlamsal bölütlemeye oldukça etkili olduğunu göstermektedir (Zhu vd. 2017).

Long vd. (2015) anlamsal bölütleme için tam evrimsel ağları(FCN) kullanmışlardır. Yaptıkları çalışma pikselden piksele, uçtan uca eğitilmiş olan evrimsel ağların

anlamsal bölütlemeye çok başarılı sonuçlar verdiğini göstermiştir. Mostajabi vd. (2014), ileri beslemeli bir mimari ile evrişimli sinir ağlarını kullanmışlardır. Çalışmaları görsel sınıflandırma, tanıma gibi alanlarda kullanılan derin evrişimsel ağların bölütlemeye de kullanılabileceğini göstermiştir. Noh vd., (2015) derin dekonvolüsyon(ters evrişim) ağlarıyla bir anlamsal bölütleme algoritması önermişlerdir. Çalışmaları örnek bazlı tahmine dayalı algoritmalarının, sabit boyutlu algılama alanı sınırlamasını ortadan kaldırarak nesne ölçeği varyasyonları için avantajlı olduğunu göstermiştir. Lucas vd. (2024) yapmış oldukları çalışmada Mask R-CNN (maske alan tabanlı evrişimsel sinir ağları) kullanmışlardır. Yöntem, bireysel ağaçları tespit eder, tepelerini belirler ve onları canlı ya da ölü olarak sınıflandırır. Bu yaklaşım Harz Dağları'nda ormanların kurumasından etkilenen bir bölgede hava görüntüleri ve tepe yüksekliği modelleri kullanılarak değerlendirilmiş, 0.88-0.91 arası F1 skoru ile 0.80-0.88 arası hassasiyet skoru elde edilmiştir. Chen Liang-Chieh vd. (2016), DeepLab mimarisini ortaya koymuşlardır. Yoğun tahminleme için "Atrous convolution" adında yukarı örneklenmiş filtrelerle evrişim yöntemini öne çıkarmışlar. Bu yöntem parametre sayısını ve hesaplama miktarını artırmadan filtrelerin görüş alanını büyütürken daha geniş bir bağlamda etkili bir çalışmayı ortaya koyar. Ayrıca nesnelere birden fazla ölçekte bölütlemek için ASPP önermişlerdir. ASPP, evrişimli özellik katmanını farklı örnekleme oranlarında ve görüş alanlarında tarayarak nesnelere ve görüntü bağlamını çok ölçekli yakalamaktadır. Ortaya koydukları mimari bahsedilen yöntemleri kullanarak ve derin evrişimli sinir ağları ile tam bağlantılı koşullu rastgele alanlar yaklaşımını birleştirerek görüntü sınıflandırma için eğitilmiş ağları anlamsal bölütlemeye yeniden uyarlamıştır. Volpi ve Ferrari (2015) yüksek çözünürlüklü uydu ve havadan görüntülerden kentsel sınıfların anlamsal bölütlenmesi için ayrımcı adını verdikleri bir şekilde eğitilmiş koşullu rastgele alanlar modellerini önermişlerdir. Yerel sınıf tercihlerini dikkate almak için halka yapıları temel alan ikili potansiyel öğrenme stratejisi, standart bağlamlara kıyasla sonuçlarda büyük iyileştirmeler sağlamıştır. Sonuçta, ayrımcı eğitildikleri için bölütleme doğruluğu maksimize edilmiştir.

Krizhevsky vd. (2012) ortaya koydukları büyük, derin evrişimsel sinir ağlarıyla 2012 yılında gerçekleştirilmiş olan ImageNet büyük ölçekli görsel tanıma yarışmasını kazanmışlardır. CNN mimarisi ile ilk defa ilk beş testin hata oranı %15.3 olmuştur.

AlexNet olarak adlandırılan bu mimari şu katkıları sağlamıştır: Düzeltilmiş doğrusal birimin(ReLU) nonlinearer fonksiyon olarak kullanılarak eğitim zamanının geleneksel hiperbolik tanjant fonksiyonuna kıyasla birkaç kat azaldığı görülmüştür. Dropout katmanlarının overfitting problemi için uygulanabileceği ve veri artırma teknikleriyle eğitim setinin yapay olarak artırılabilmesi ortaya konulmuştur.

Ronneberger vd. (2015), biyomedikal görüntüler üzerinde görüntü bölütlemesi yapmak üzere U-Net mimarisini ortaya koymuşlardır. U-Net kodlayıcı olarak bağlamsal bilgiyi yakalamak ve mekansal çözünürlüğü azaltmak için bir contracting path(daralan yol) içermektedir. Kod çözücü kısmı kodlayıcıdan gelen verilerle öznelik haritalarını örnekleyerek artırırken aynı zamanda evrişimsel işlemlerde gerçekleştirirler. Atlama bağlantıları daralan yolda kaybolan mekansal bilginin korunmasına yardımcı olmaktadır. Bu da kod çözücü katmanların öznelikleri daha doğru bir şekilde bulmasını sağlamaktadır. Kod çözücü bu atlama bağlantılarından ve daralan yoldan gelen verilerle bölütleme haritasını oluşturmaktadır.

Wang vd. (2021), gerçek zamanlı kentsel alan bölütlemesi için transformer tabanlı bir kod çözücü ve CNN tabanlı bir kodlayıcı kullanarak Unet benzer bir mimari ile yeni bir model ortaya koymuşlardır. Kodlayıcı kısmında öne sürülen verimli bir küresel-yerel dikkat mekanizması ile yapılan çalışma birkaç veri seti üzerinde denenmiş ve hafif modeller arasında oldukça başarılı sonuçlar göstermiştir.

Dosovitskiy vd. (2021), yapmış oldukları çalışmada kodlayıcı kısmında bulunan saf transformer ile öne sürülen mimari, doğrudan görüntü parçaları dizilerinin görüntü sınıflandırmada oldukça iyi performans göstermiştir. CNN mimarilerinde yerellik, çeviri eşdeğerliliği ve iki boyutlu komşuluk yapısı tüm model boyunca her katmana entegre edilirken öne sürülen ViT mimarisinde yalnızca çok katmanlı perceptronlar yerel ve eşdeğerken kendine dikkat katmanları globaldir. Bu sayede CNN mimarilerinden çok daha az tümevarım yanlılığına sahip olmaktadır. Mimariye doğal dil işlemede kullanılan standart transformer kodlayıcısı kullanılmış ve kod çözücü kısmındaki bir MLP ile sınıflandırma işlemi gerçekleştirilmiştir.

Su vd. (2021) multispektral görüntüleme, insansız hava aracıyla algılama, bitki segmentasyonu ve derin öğrenme U-Net gibi en son teknik ve algoritmaları entegre bir çalışma yapmışlardır. Çalışmada öncelikle yanıtıcı renk RGB görüntüsünün oluşturulması ve sarı pas içeren alanların etiketlenmesinden sonra optimizasyonlu toprak ayarlı bitki örtüsü endeksini hesaplayarak bitki örtüsü bölütlemesi yapılır ve buğday içeren, içermeyen bölgeler elde edilmiştir. Daha sonra sarı pas hastalığı olan, buğday olan ve arka plan pikselleri elde edilmiştir. Çalışmada farklı bant aralıkları kullanılarak elde edilen sonuçlar karşılaştırılmış ve sonuçta RGB görünür görsellerle yakın kızılötesi ve kırmızı kenar bantların bir arada kullanılmasının bölütleme performansını artırdığı gözlenmiştir.

Kemker vd. (2017) çok yüksek çözünürlüklü multispektral görüntüler(MSI) içeren RIT-18 veri setini kullanmışlar ve multispektral görüntülerde bulunan etiket yetersizliğini aşmak için sentetik MSI kullanışlardır. Sentetik görüntüler aşırı öğrenmeye daha az yatkın oldukları için son teknoloji çalışmalar için temel sağlamaktadır. Yapılan çalışmada ilk defa derin evrişimli sinirsel ağlar, multispektral görüntülerde anlamsal bölütleme için kullanılmış ve bu ağların sentetik görüntülerle önceden eğitilmesiyle performansının artırılabilirdiği gösterilmiştir.

Ülkü vd. (2022) yaptıkları çalışmada iki farklı setini ele almışlar ve çok bantlı görsellerle derin anlamsal bölütleme gerçekleştirmişlerdir. Uydu görüntüsü özellik algılama görsel setinde(DSTL) farklı mekansal çözünürlüklerle hem 3 hem de 16 bantlı formatlar sağlanmıştır. Veri seti 10 etiketli sınıftan oluşmaktadır. 11 ve 14 bit renk derinliğine sahip olan bu veri seti bu sayede her bir pikselden derin öğrenme modellerinde kullanılabilecek daha fazla bilgi elde edilmesini sağlamaktadır. Diğer veri seti(RIT-18) için bir drone üzerine eklenmiş multispektral sensör kamera kullanılmıştır. Veri seti altı yüksek çözünürlüklü multispektral bant içermektedir bunlardan üçü görünür RGB bantlarken diğer üçü yakın kızılötesi bantlardır. Çalışmada U-Net, SegNet, DeepLabv3+ ve rastgele orman algoritması, anlamsal bölütleme mimarileri olarak kullanılmış ve karşılaştırılmıştır. Spektral bantların korelasyonunun incelendiği çalışmada farklı bant aralıkları gruplanarak eğitilmiş ve sonuçları kıyaslanmıştır. Elde edilen sonuçlar uzaktan algılama görüntülerinin mekansal çözünürlüğünün spektral

görüntüler ve kullanılan mimarilere kıyasla bölütleme performansında çok daha büyük etkiye sahip olduğunu yani güçlü bir mimari kullanılsa dahi yüksek çözünürlüklü görüntü seti olmadan dikkate değer bir doğruluğa ulaşamayacağını göstermiştir. Sonuçlar şunları da göstermiştir: Yakın kızılötesi bantları yeterince iyi değilse, ağaçların ayırt edilmesi için oldukça değerli olan yakın kızıl ötesi yansıma bilgisi yeterli doğrulukta ağaç bölütlemesini sağlayamamaktadır. Mekansal çözünürlüğün yetersiz kaldığı durumlarda bitki indekslerinin kullanılması avantajlı olabilmektedir.

Su vd. (2023) yaptıkları araştırmada hassas tarım için yapay zeka destekli insansız hava araçları algılama sistemlerini incelemişlerdir. İnsansız hava araçları algılama sistemleri 2010 yılından itibaren hassas tarım dahil birçok sivil uygulamada uygulanmaktadır. Araştırma İHA algılama sistemleriyle gerçekleştirilebilecek tipik hassas tarım unsurlarını içermektedir. İnsansız hava araçları küçük ölçekli ekim alanı tahmini için sensörleri ve esnek mekansal çözünürlüğü sayesinde kullanışlıdır. Ekim alanı tahmini hem derin öğrenme hem de makine öğrenmesi yöntemlerinin kullanılabilmesi denetimli sınıflandırma problemi olarak ele alınmaktadır. Toprak koşulları mahsullerin gelişimi için oldukça önemlidir ve düzenli olarak izlenmelidir. İHA tabanlı uzaktan algılama, toprak yüzeyi yansıması bu koşulları yansıtabildiği takdirde izleyebilir. Toprak koşullarının değerlendirilmesi genellikle bir regresyon problemi olarak ele alınır ve İHA'nın multispektral, hiperspektral ve termal görüntülerinden elde edilen özelliklerle birlikte yer ölçümleri kullanılır. Bunların dışında bitki gelişimi için tahminler, mahsul hastalıkları, mahsul haşereleri, mahsul su stresi, yabancı ot istilasısı, verim tahmini ve hasat kararı gibi konularda insansız hava aracı ve uzaktan algılamayla elde edilen görüntüler yapay zeka uygulamalarında kullanılabilir.

### 3. MATERYAL ve YÖNTEM

#### 3.1 Veri Setleri

Bu çalışmada ortaya konulan model iki farklı veri seti üzerinde denenmiştir. RIT-18 veri seti Rochester Teknoloji Enstitüsünün dijital görüntüleme ve uzaktan algılama laboratuvarı tarafından hazırlanmıştır. Kemker vd. (2017) bu veri seti için sentetik multispektral görüntüler kullanmışlar ve buna karşılık 18 etiketli nesne türü oluşturmuşlardır. Yapılan çalışmada veri setinin ağaç tespiti için Ülkü vd. (2022) tarafından hazırlanan ikili anlamsal bölütleme olan hali kullanılmıştır. 2. veri seti Northwest A&F Üniversitesi'nin Cao Xinzhuang Deney İstasyonu'nda yapılan çalışmadan elde edilmiştir (Su vd. 2021) . Bu veri setinde buğdaylarda bulunan sarı pas hastalığının tespiti yapılmak istenmektedir. Bu veri setinde de ikili anlamsal bölütleme yapılmıştır. Çizelge 3.1'de veri setleri için kullanılan ekipmanlar ve veri setlerinin özellikleri gösterilmiştir. Her iki veri seti için yapılan deneylerde yalnızca RGB görüntüler kullanılmış multispektral görüntüler kullanılmamıştır.

Çizelge 3.1 Veri setlerinin özellikleri

Özellikler	Veri Setleri	
	RIT-18	CXEF
Yer Örnekleme Mesafesi (cm/pixel)	4.7 cm/pixel	1.3 cm/pixel
224 x 224 Görsel Sayısı	1778	1299
Yıl	2017	2019
Sensörler	Tetracam MicroMCA6	RedEdge, Downwelling Işık Sensörü
Drone	DJI-S1000 octocopter	DJI M100

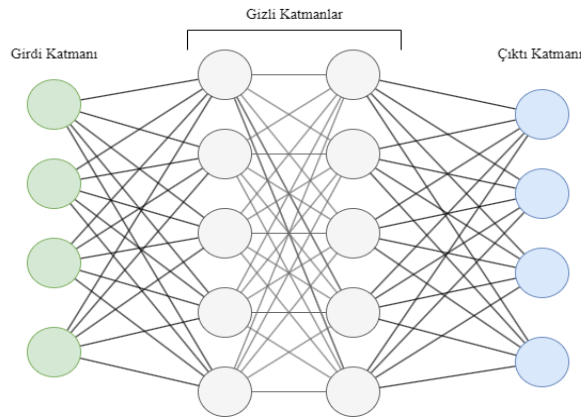
#### 3.2 Sinir Ağları

Düşünen bir makine fikrinin antik çağlardan beri varolduğunun izi sürülmektedir. McCulloch ve Pitts (1943) yayınladıkları çalışmada insan beyninin karmaşık desenleri nasıl ortaya koyabildiğini anlamayı amaçlamışlar ve nöronları ikili bir eşik mekanizmasıyla “doğru ya da yanlış” (boolean) mantığına benzetmişlerdir. İlerleyen

yıllarda perceptron kavramı ortaya atılmış onu takiben geri yayılım fikri ortaya çıkmıştır. Geri yayılımın sinir ağı mimarisine nasıl adapte edilebileceği üzerindeki çalışmalardan sonra sinir ağı bugünkü halini almıştır.

### 3.3 Derin Öğrenme

Makine öğrenmesinin ve çoğu zaman sinir ağlarının alt dalı olarak bilinen derin öğrenme, temelde insan beyninin kompleks karar verme mekanizmasının bir taklidi olarak görülebilir. Kısaca üç veya daha fazla sinir ağı katmanının, genellikle binlerce katmanın modelleri eğitmek için kullanıldığı modellerdir. Derin öğrenme modelleri, gözetimsiz öğrenme yapabilir bu sayede ham, yapısız verilerden doğru çıktılar elde etmek için gereken özellikleri ve bağları ortaya çıkarabilir. Birbirine bağlı birçok katman kendisinden önceki katmanı geliştirerek tahmini ya da sınıflandırmayı iyileştirir. Bu hesaplamaların bütün ağ boyunca ilerlemesi ileri yayılım olarak adlandırılır. Geri yayılım ise hataları hesaplamak için çeşitli algoritmalar kullanır. Ağırlıkları ve yanlılıkları hesaplamak için katmanlar boyunca geriye doğru hareket ederek modeli eğitir. İleri yayılım ve geri yayılım birlikte sinir ağının düzeltilmesini yaparak modeli daha doğru hale getirir. Şekil 3.1’de bir derin öğrenme sinir ağı gösterilmiştir. Girdi katmanı ve çıktı katmanı parametrelerinde değişime kolay bir şekilde gidilebilir ya da bu katmanlardaki parametrelerin değişimi rahatça gözlenebilirken gizli katmanlarda bunu yapmak bir hayli zor olmaktadır. Zaten bu genellikle tercih edilen bir durum değildir. Derin öğrenme adını bu gizli katmanların varlığından almaktadır.



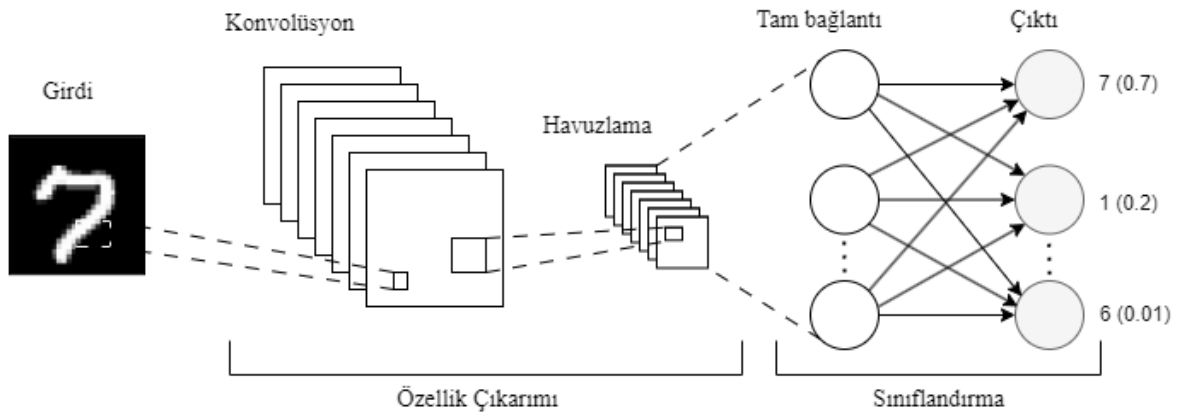
Şekil 3.1 Derin sinir ağı

### 3.4 Evrişimsel Sinir Ağları

Bu ağ yapısı ilk olarak Fukushima (1980) tarafından ortaya atılmıştır. Ancak ortaya atıldığı dönemde donanımsal sınırlar bu ağ yapısının az kullanılmasına neden olmuştur. LeCun vd. (1989) gradyan tabanlı bir öğrenim algoritmasını CNN'e uygulayarak el yazısı rakamların sınıflandırma problemi için başarılı sonuçlar elde etmişlerdir. Bundan sonra CNN mimarisi araştırmacılar tarafından giderek kullanılmaya başlamış ve birçok tanıma görevinde alanındaki en iyi sonuçları vermiştir. Konvolüsyon üç önemli fikri kullanmaktadır: Parametre paylaşımı, seyrek etkileşimler, eşdeğişken temsiller (GoodFellow vd. 2014). Sıklıkla tercih edilen CNN tipi, MLP'ye benzemektedir. Havuzlama katmanlarından önce birçok evrişim katmanı içerir ve son katmanı tam bağlı katmanlardır. CNN'ler bugün nesne tanıma, bilgisayarla görme, bölütleme, yüz tanıma vb. birçok yapay zeka yaklaşımında kullanılmaktadır. Şekil 3.2'de bir CNN mimarisinin genel yapısı görülmektedir.

#### 3.4.1 Evrişim katmanı

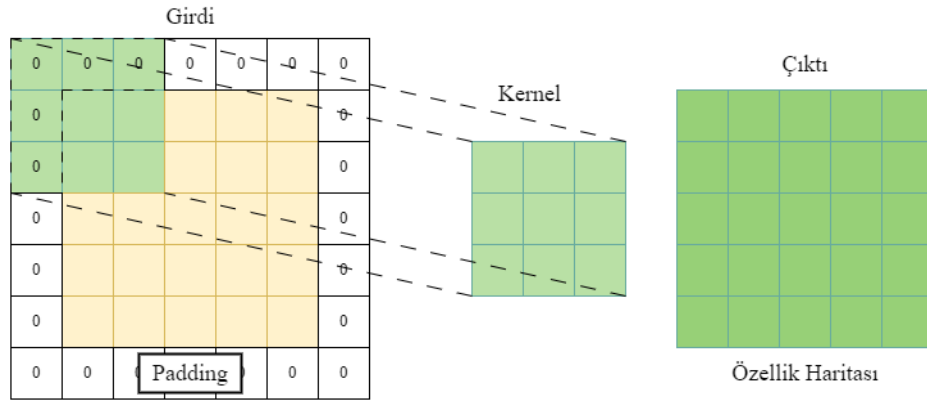
Evrişim katmanı CNN mimarisinin ana bileşenidir. Ana işlevi görüntüden öznitelik çıkarımı gerçekleştirmektir. Evrişim katmanı evrişimsel filtreler adı verilen bir dizi kernelden oluşmaktadır. Öznitelik çıkarımı pikselden matrislere dönüştürülmüş görüntüler üzerinde yapılan işlemler ile gerçekleştirilmektedir.



Evrişimsel katmanında kullanılan ve evrişim modeli eğitilmeden önce seçilmesi gereken hiperparametreler vardır.

### 3.4.1.1 Padding

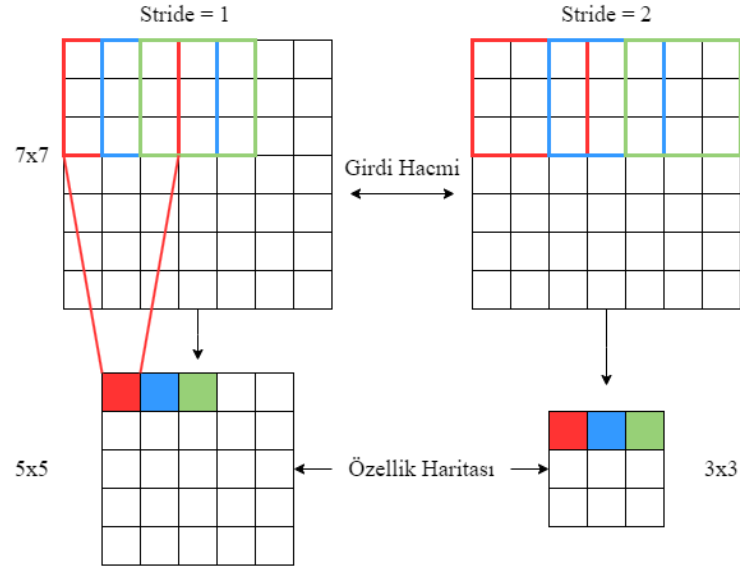
Evrişim işlemi öncesi girdinin kenarlarına sıfırlar ekleyerek çıktının boyutunun küçülmesini engelleyen işlemidir. Padding parametresi “valid padding” ise girdi özellik haritasına hiç padding uygulanmaz böylece çıktı katmanı, girdi katmanından daha küçük olur. Same padding’te çıkış ile giriş boyutunu aynı tutmak için gerektiği kadar sıfır eklenmektedir. Şekil 3.3’te bir Padding işlemi örneği görülmektedir.



Şekil 3.3 Padding işlemi

### 3.4.1.2 Stride

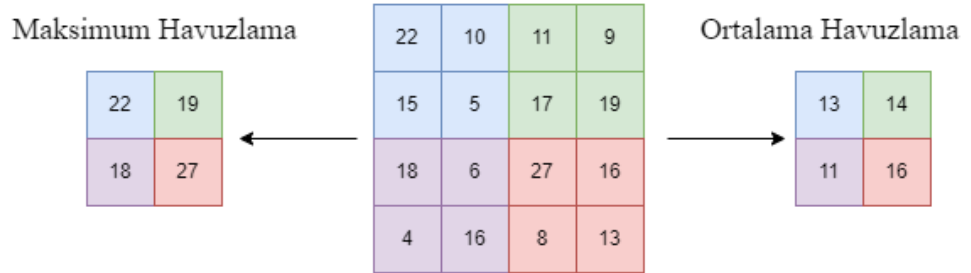
Kernel(çekirdek) matrisinin giriş matrisi üzerinde hareket ederken aldığı mesafeyi ifade eder. Daha büyük bir stride değeri daha küçük bir çıktı matrisi oluşmasına neden olur. Şekil 3.4’te Stride değerlerinin 1 ve 2 olduğu durumlarda 7x7 matristen elde edilecek çıktı özellik haritası gösterilmiştir.



Şekil 3.4 Stride işlemi

### 3.4.2 Havuzlama katmanı

Evrişimsel katmanların çıktı boyutunu küçültebileceği yukarıda anlatılan özelliklerle birlikte çok açıktır. Evrişimsel katmanların esas görevi öznitelik çıkarımıdır. Havuzlama katmanının ana amacı ise boyut azaltmadır. Bu yaklaşımla büyük boyutlu öznitelik haritalarından daha küçük özellik haritaları oluşturulmaktadır. Çeşitli havuzlama yöntemleri bulunmaktadır. Bu yöntemler baskın özniteliklerin devamını sağlar niteliktedir. Şekil 3.5'te havuzlamada kullanılan iki işlemden elde edilebilecek sonuçlar gösterilmiştir.



Şekil 3.5 Maksimum havuzlama ve ortalama havuzlama işlemi

Ortalama havuzlama yöntemiyle bir bölgedeki verilerin ortalaması alınıp veriler küçültülür. Bazı durumlarda bu yöntemin maksimum havuzlamadan daha iyi sonuç verdiği görülmektedir. Maksimum havuzlama bir bölgedeki en büyük değeri seçerek küçültme işlemini gerçekleştirir. Özellikle görüntülerde belirgin özelliklerin korunmasını sağlamaktadır.

### 3.4.3 Aktivasyon katmanı

Aktivasyon fonksiyonlarının ana amacı girdiyle çıktıyı eşleştirmektir. Evrişim sonucunda ortaya çıkan özelliklerin sayısal değerlerinin dağılımını değiştirmek değerlerin doğrusal olmayan bir şekilde temsilini sağlayarak değer dağılımını daha gerçekçi bir şekilde yansıtabilir. CNN mimarilerinde, tüm ağırlıklı katmanların (tam bağlı katmanlar ve evrişim katmanları) ardından doğrusal olmayan aktivasyon katmanları kullanılır. Bu katmanlar, CNN'in daha karmaşık şeyleri öğrenme yeteneğini artırır. CNN mimarisinde yaygın olarak kullanılan aktivasyon fonksiyonları ve formülleri şekil 3.6'da gösterilmiştir. Derin sinir ağlarında ve CNN'de kullanılan aktivasyon fonksiyonları şunlardır.

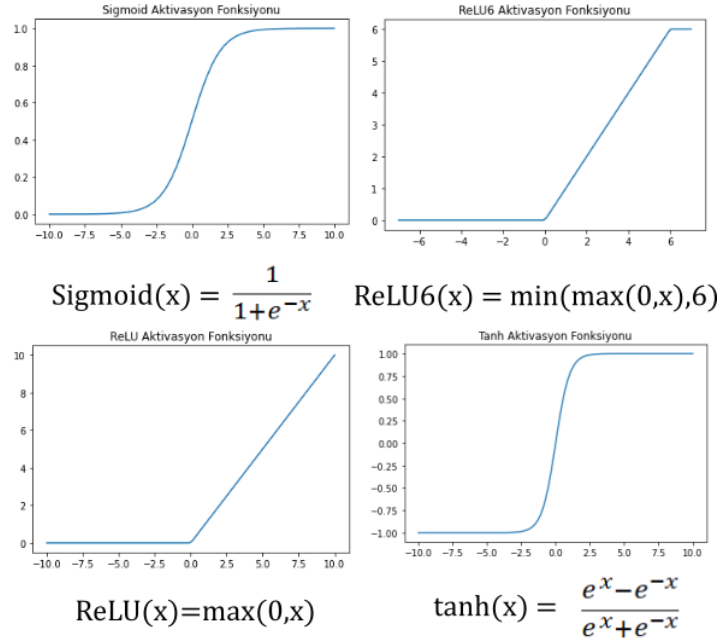
**Sigmoid:** Girdi değerleri reel sayılar olan bu fonksiyonun çıktı değerleri 0 ile 1 arasındadır. "S" şekilli bu fonksiyon en çok kullanılan fonksiyonlardan biridir.

**Tanh:** Sigmoid fonksiyona benzerlik gösterir ve gerçek sayıları kullanır. Çıktı değerleri ise -1 ile 1 arasındadır.

**ReLU:** Doğrutulmuş doğrusal birim, CNN'lerde en çok kullanılan aktivasyon fonksiyonudur. Düşük hesaplama yükü nedeniyle diğerlerinden üstün olduğu görülmüştür. Bütün girdi değerlerini pozitif sayılara çevirmektedir.

**ReLU6:** CNN mimarilerinde tercih edilen bu fonksiyon kimi zaman ReLU'da karşılaşılabilen sayıların sonsuza gitmesini durumunu önlemiştir. Elde edilen değerler [0,6] aralığında olmaktadır.

**Softmax:** Birden fazla sınıfın sınıflandırması için kullanılan bu fonksiyonda toplamı 1 olacak şekilde tahminleme değerleri sınıflar arasında paylaşılır.



Şekil 3.6 Aktivasyon fonksiyonları

### 3.4.4 Tam bağlantılı katman

Tam bağlantılı katman genellikle CNN mimarilerinin en sonunda yer almaktadır. Bu katmandaki bütün nöronlar bir önceki katmandaki her nöronla bağlantılıdır. CNN mimarisinin sınıflandırıcısı olarak kullanılırlar. Katmanın girişi havuzlama ya da evrişim katmanından gelmektedir. Bu giriş, genelde öznetelik haritalarının düzleştirilmesiyle oluşturulmuş vektörler şeklindedir. Uygun bir sınıflandırma yapmak için genellikle softmax aktivasyon fonksiyonunu kullanarak 0 ile 1 arası bir değer üretirler.

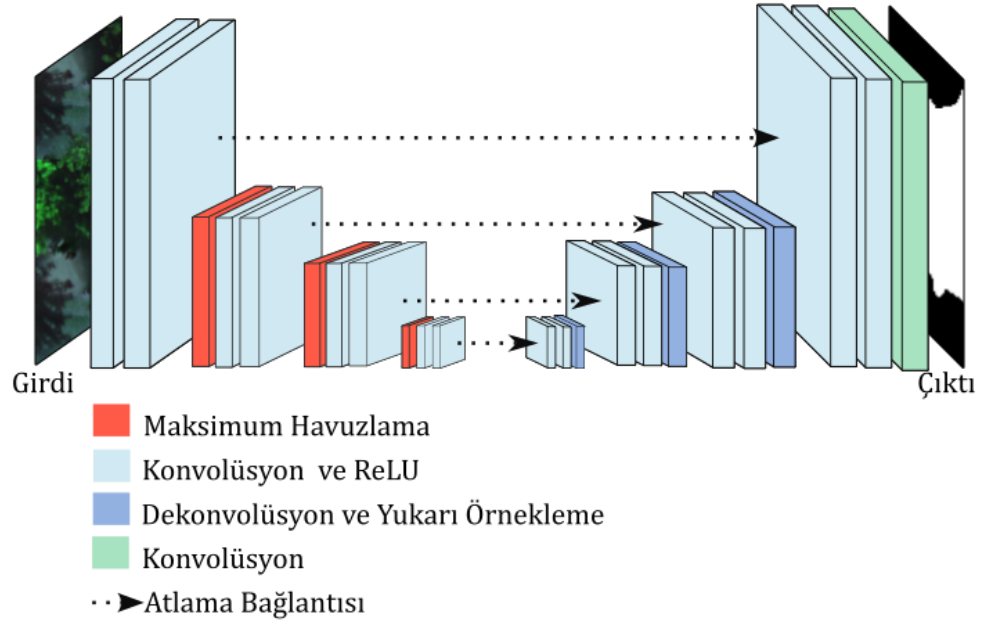
### 3.5 U-Net Mimarisi

Özellikle görüntü bölütlemeye kullanılan U-Net, anlamsal bölütleme alanında yüksek doğruluk oranıyla tercih edilen bir modeldir. CNN mimarisinin son katmanında kullanılan tam bağlantılı katmanlar mekansal bilginin yok olmasına dolayısıyla modelin performansında düşüşe neden oluyordu. Long vd. (2015), FCN önererek bu kısıtlamaya çözüm üretmişlerdir. FCN yapısı anlamsal temsili elde etmek için evrişim, aktivasyon

ve havuzlama içeren birçok evrişim bloğunu kodlayıcı kısmında kullanırken piksel seviyesinde tahminleme yapmak için kod çözücü kısmındaki üst örnekleme işleminde de konvolüsyon katmanını kullanmaktadır (Azad vd. 2022).

U-Net mimarisi, FCN ağlarını kullanarak kodlayıcı ve kod çözücü kısımlarıyla biyomedikal görüntü bölütlemesi için tasarlanmıştır. FCN ağların limitli sayıda veri setinde dahi görüntülerden özniteliklerin elde edilmesinde başarılı olduğu görülmektedir. U-Net mimarisinin en önemli özelliği atlama bağlantılarıdır. Bu bağlantılar her kodlayıcı bölümünde elde edilen veriyi ona karşılık gelen kod çözücü kısmına iletir. Bu sayede yüksek çözünürlüklü bağlamsal bilgi ağ boyunca yayılır. U-Net mimarisi onlarca mimari için ilham kaynağı olmuş ve üzerinde yapılan geliştirmelerle kullanılan bir mimari haline gelmiştir.

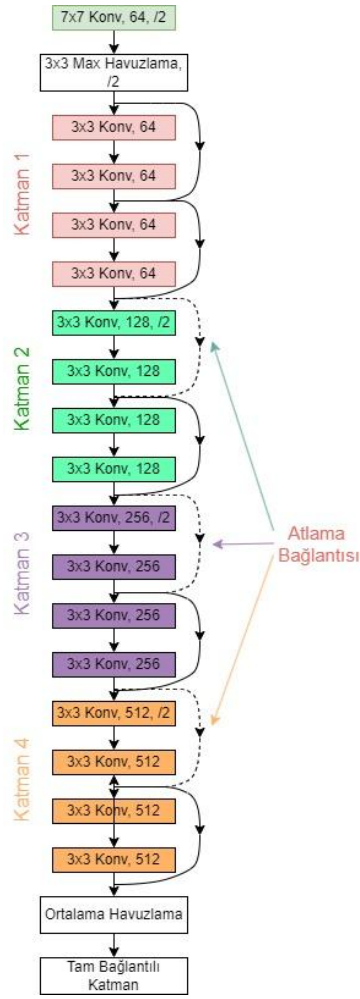
U-Net mimarisi Şekil 3.7’de görülmektedir. Tipik bir evrişimli ağ mimarisine benzemektedir. Kodlayıcı kısmı daralan yol olarak adlandırılmıştır. Bu kısmın her adımında evrişim işlemi ve onu takiben ReLU aktivasyon fonksiyonu uygulanmaktadır. Daha sonra maksimum havuzlama işlemi ile alt örnekleme(downsampling) yapılmış olur. Kod çözücü kısmı ise genişleyen yol olarak kabul edilmiştir. Öznitelik haritası yukarı örneklenir ve ters evrişim işlemi gerçekleştirilir. Atlama bağlantısıyla daralan yoldan gelen uygun bir şekilde kopyalanmış öznitelik haritası karşılık gelen aynı katman seviyesindeki öznitelik haritası ile birleştirilir. Kod çözücü kısmında da aynı şekilde evrişim ve ReLU aktivasyonu kullanılır. En son çıktı noktasında 1x1 evrişim uygulanarak piksel piksel sınıflandırma işlemi gerçekleştirilir.



Şekil 3.7 U-net mimarisi

### 3.6 ResNet

He vd. (2016) tarafından tasarlanan artık ağ mimarisi olarak adlandırabileceğimiz ResNet, ILSVRC 2015 yarışmasının kazanan mimari olmuştur. Kaybolan gradyan sorununu yaşamayan bir ağ olması amaçlanan bu mimaride artık bağlantılar kullanılarak derin ağların eğitimi kolaylaştırılmıştır. Katmanlar içinde kısa yol bağlantıları ile katmanlar arası bağlantı mümkün kılınmıştır. Şekil 3.8’de ResNet mimarisi görülebilir.

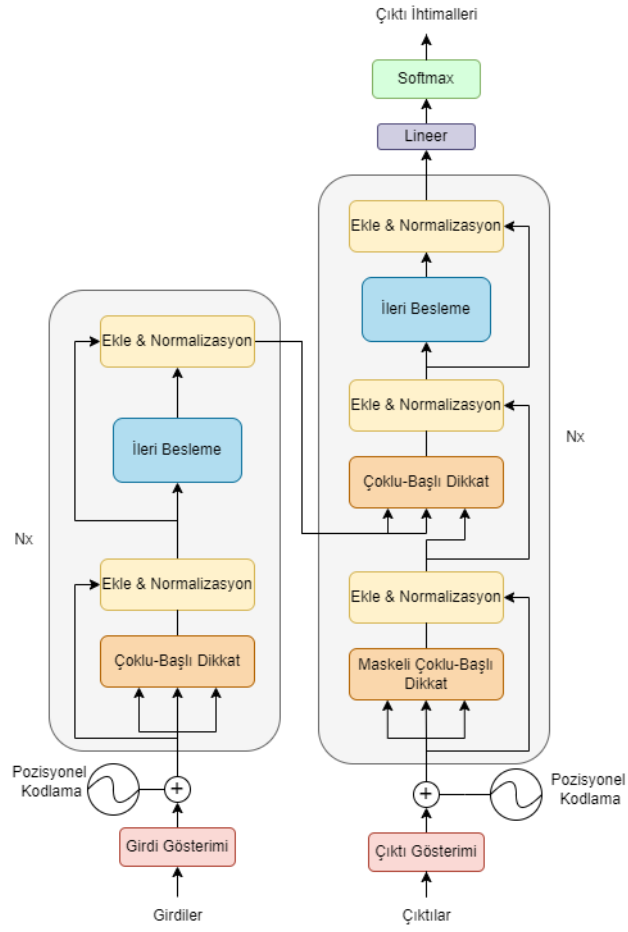


Şekil 3.8 ResNet-18 mimarisi

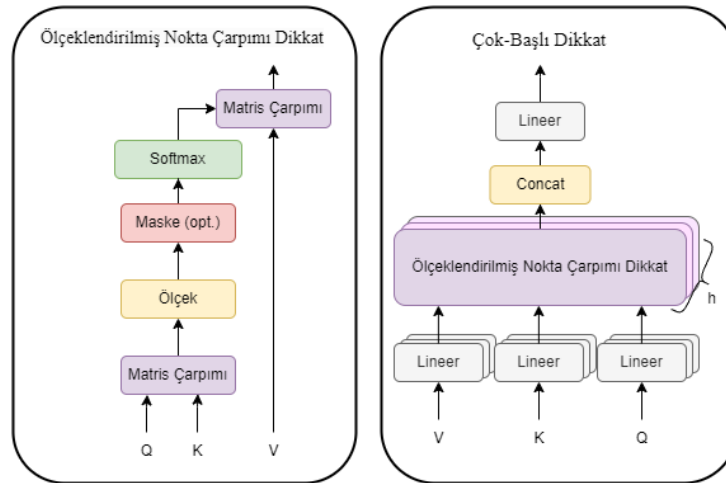
Kısa yol bağlantıları herhangi bir parametre ya da hesaplama karmaşıklığı ortaya koymamaktadır. Girdi ve çıktı boyutlarının aynı olduğu durumlarda kısa yol bağlantıları direkt olarak kullanılırken boyutların arttığı durumlarda iki seçenek düşünülmüştür. Padding ile kısa yol işleminin gerçekleştirilmesi ya da boyutları eşleştirmek için projeksiyon kısa yolunun kullanılmasıdır. İkinci işlem  $1 \times 1$  evrişim ile gerçekleştirilir ve öznitelik haritalarının boyutlarını eşitlemektedir. ResNet-18, ResNet-34, ResNet-50 ve ResNet-101 gibi farklı ResNet mimarileri bulunmaktadır. Yapılan çalışmada farklı ResNet mimarileri kullanılarak elde edilen sonuçlar karşılaştırılmıştır.

### 3.7 Transformers

Dikkat mekanizmaları, ardışık modelleme ve aktarım modellerinin vazgeçilmez bir parçası haline gelmiştir. Bu mekanizmalar, bağılıkları giriş ya da çıkış dizilerinin mesafelerine bakılmadan modelleyebilme imkanı sunmaktadır. Genellikle yinelenen ağlarla birlikte kullanılırlar. Tekrar kullanmadan tamamen dikkat mekanizmasına dayalı olan Transformers mimarisi Vaswani vd. (2017) tarafından ortaya atılmış ve kısa sürede oldukça popüler olmuştur. Öz-dikkat(Self-attention) mekanizması, bir dizi içerisinde farklı pozisyonlar arasında ilişki kuran dikkat mekanizmasıdır. Bir dizinin temsili oluşturulurken farklı konumların birbirleriyle etkileşimini sağlayan bu mekanizma okuma anlama, metinsel çıkarım ve bağımsız cümle temsilleri öğrenme gibi çeşitli konularda başarıyla kullanılmıştır (Cheng vd., 2016). Şekil 3.9'da görülebileceği gibi Transformers mimarisi iki kısımdan oluşmaktadır. En bilinen sinirsel seri aktarım modelleri genellikle kodlayıcı-kod çözücü yapıdadır (Cho vd. 2014). Kodlayıcı N sayısı kadar birebir aynı katmanların olduğu bir yığından oluşmaktadır. Her katmanda iki alt katman bulunmakta, ilk katman çok başlı kendine dikkat mekanizması, ikinci katman konum bazlı tam bağlı besleme ağıdır. Bu iki alt katman etrafında bir kalıntı bağlantısı kullanılır ve ardından katman normalizasyonu da uygulanmaktadır. Kod çözücü kısmı da aynı şekilde birebir aynı katmanların olduğu bir yığından meydana gelmektedir. Yine kalıntı bağlantısı ve ardından katman normalizasyonu da kullanılır. Her kodlayıcı katmanındaki iki alt katmana ek olarak, kod çözücüde kodlayıcı yığınının çıktısı üstünde çoklu başlı dikkat işlemi gerçekleştiren üçüncü bir alt katman eklenmiştir. Dikkat mekanizması, sorgu, anahtar ve değer vektörlerinin sorgu ile anahtar ve değer çiftine haritalanmasını sağlar. Transformers'ın temelini bu dikkat mekanizmasıdır. Şekil 3.10'da dikkat mekanizması gösterilmiştir. Dikkat mekanizması, her bir giriş pozisyonunun diğer pozisyonlara olan önemini belirler ve giriş pozisyonlarını ağırlıklandırarak çıkışa katkıda bulunmasını sağlar. Çok-başlı dikkat mekanizmasında, birden fazla başlık eş zamanlı olarak kullanılır. Öğrenilebilir ağırlıklarla her bir başlık, birbirlerinden bağımsız olarak öğrenirler. Transformers, giriş sırasının sıralı bilgisini içermediği için pozisyon bilgisini eklemek için pozisyonel kodlama kullanır. Pozisyonel kodlama, giriş pozisyonlarına benzersiz bir vektör ekleyerek sıralı bilgiyi temsil eder.



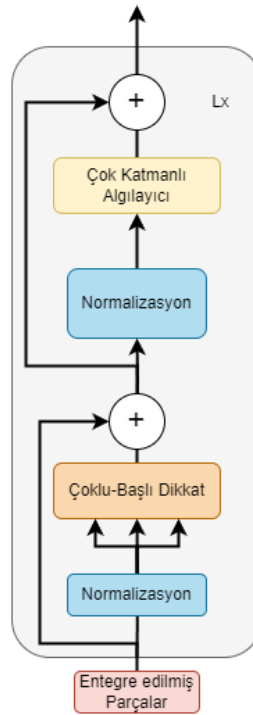
Şekil 3.9 Transformers mimarisi



Şekil 3.10 Ölçeklendirilmiş nokta çarpımı dikkat ve çok-başlı dikkat yapıları

### 3.8 Vision Transformer

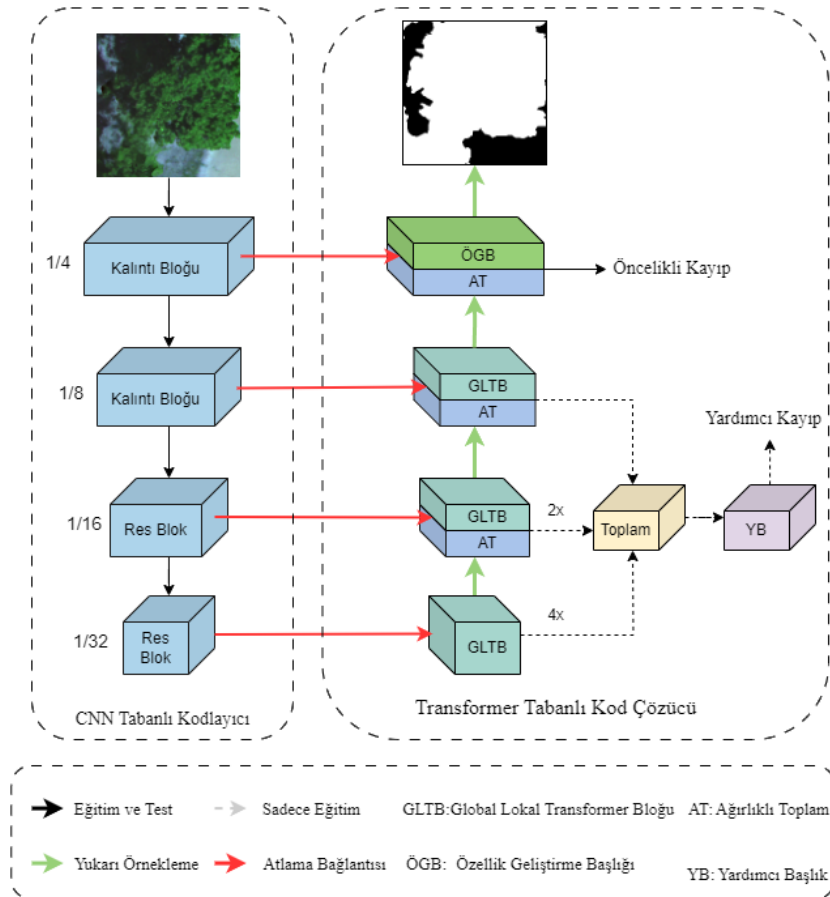
Transformers kısa sürede doğal dil işlemede tercih edilen bir mimari olmuştur. Dosovitskiy vd. (2020) tarafından bundan ilham alınarak geliştirilen Vision Transformerda ise yapılabilecek en az değişimle standart Transformer uygulaması görüntülere uygulanmıştır. Bunun için görüntüler parçalara bölünerek parçaların doğrusal gömülü dizisi Transformer'a girdi olarak verilir. Görüntü parçaları, bir NLP uygulamasında kullanılan tokenler ile aynı şekilde ele alınır. Mimari, görüntü sınıflandırması için denetimli bir şekilde eğitilir. Model için kullanılan encoder yapısı Şekil 3.11'da gösterilmiştir. ViT mimarisi birkaç adımdan oluşmaktadır. Giriş görüntüsü bir dizi görüntü parçalarına dönüştürülür ve parçaların hangi sırayla geldiği bilgisi olan konum numarası eklenir. Görüntü parçaları görüntünün eğitimle öğrenilebilir temsili olan entegre parçalara dönüştürülür. Bu işlemlerden elde edilen vektörler serisi standart transformer kodlayıcına girdi olarak verilir. Sınıflandırma işlemi için seriye ekstra sınıflandırma tokeni eklenmiştir. Mimaride en son kısımda çok katmanlı algılayıcı ile sınıflandırma işlemi gerçekleştirilmiş olur.



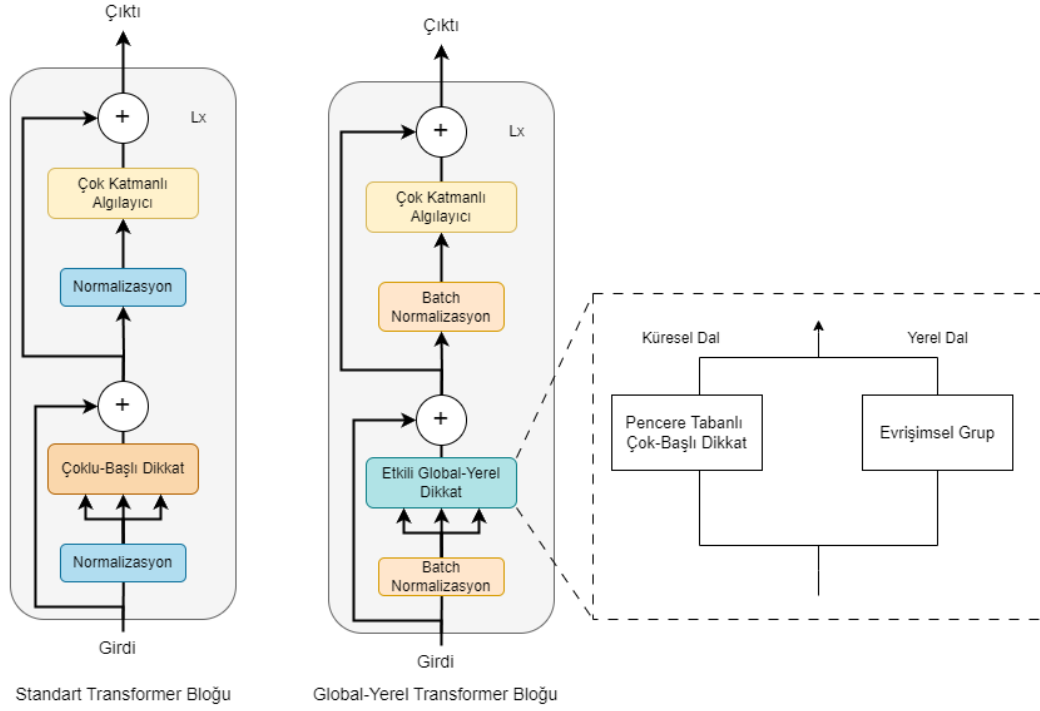
Şekil 3.11 Vision transformer'da kullanılan transformer encoder modeli

### 3.9 U-Netformer Mimarisi

Uzaktan görüntüleme görüntülerinin anlamsal bölütlenmesi, kentsel değişimin tespiti, arazi örtüsü haritalama gibi birçok pratik uygulama için gereklidir. Evrişimli sinir ağı uzun yıllar boyunca anlamsal bölütleme alanında öne çıkan yapı olmuştur. CNN mimarisindeki evrişim katmanının yerel yapısı, ağın küresel bağlamını yakalamasını kısıtladığı düşünülmektedir. Son yıllarda ortaya atılan Transformer ise küresel bilginin modellenmesinde büyük bir potansiyel ortaya koymuş ve nesne tanıma, anlamsal bölütleme gibi birçok alanda gelişmeler kaydedilmiştir. Şekil 3.12’de Unetformer mimarisinin genel yapısı görülmektedir. Bu çalışmada kullanılan Unetformer mimarisinde yardımcı kayıp kısmı işlem yükünü artırdığından tercih edilmemiştir.



Unetformer mimarisinde Transformer tabanlı bir kod çözücü önerilmiş ve bu U-Net benzeri bir mimaride sunulmuştur. Kodlayıcı olarak ResNet18 mimarisi seçilmiş ve kod çözücü kısmında hem küresel hem de yerel bilgiler için küresel-yerel dikkat mekanizması ortaya koyulmuştur. Şekil 3.13’de Standart bir transformer mimarisi ile bu mimaride kullanılan transformer yapısı görülmektedir.



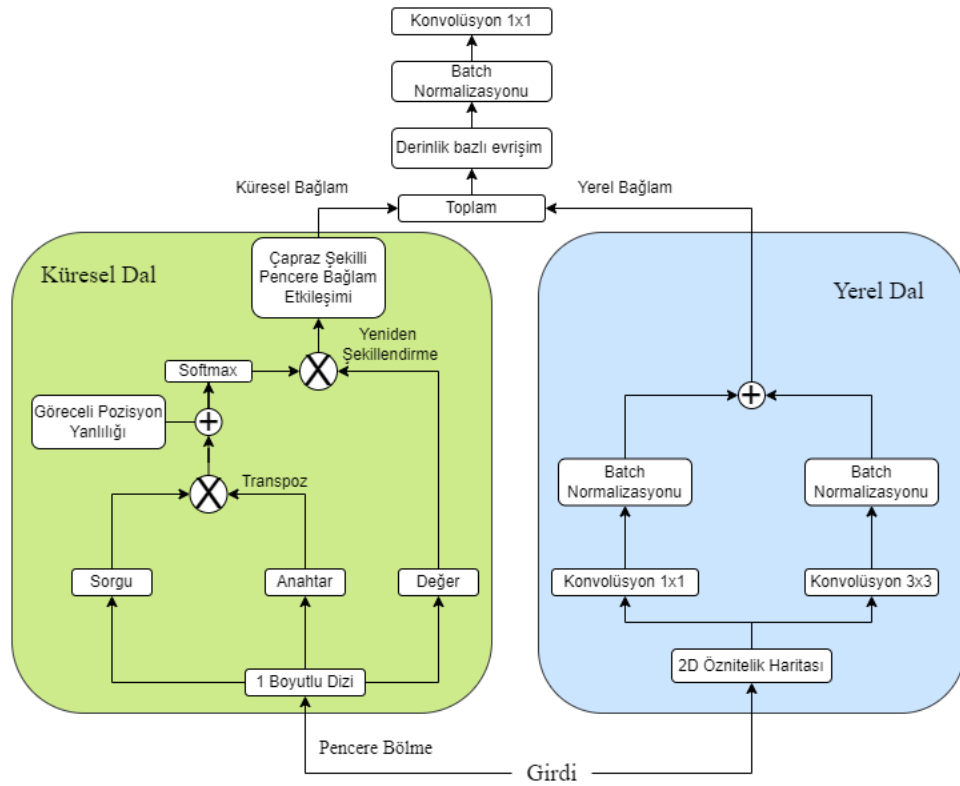
Şekil 3.13 Standart transformer bloğu ve küresel-yerel transformer bloğu

Standart transformer’da bulunan öz-dikkat bloğunun yerine GLTB kullanılmıştır. Görüntü algılamada küresel bağlamı yakalamak için dikkat temelli bir küresel dal ve yerel bağlamlar için evrişim tabanlı bir yerel dal içeren dikkat mekanizması önerilmiştir.

### 3.9.1 Küresel-yerel dikkat yapısı

Karmaşık kentsel görüntülerin anlamsal bölütlemesi için küresel bağlam oldukça önemlidir. Aynı şekilde yerel bilgi de uzamsal detayların korunması için önemli ve gereklidir. Bu durumda önerilen global-lokal dikkat mekanizması, global ve lokal bağlantıları çıkarmak için iki paralel daldan meydana gelmektedir. Şekil 3.14’te bu etkili küresel-yerel dikkat yapısı incelenebilir. Lokal dal farklı çekirdek boyutları olan (3x3 ve 1x1) iki paralel evrişim katmanını kullanmaktadır. Bunlar yerel verinin elde

edilmesini sağlar. Evrişim katmanları sonrasında toplu normalizasyon işlemleri gerçekleştirilir ve elde edilen değerler toplanır. Küresel dalda ise pencere tabanlı çok başlı öz dikkat mekanizması kullanılmaktadır. Standart 1x1 konvolüsyon, girdi 2D öznitelik haritasının kanal boyutunu üç katına çıkartmak için kullanılır. Daha sonra 1D seriler oluşturmak için öznitelik haritası bölünür ve bu seriler sorgu, anahtar ve değer vektörlerine dönüştürülürler. Bu mimaride pencere boyutu ve başlık sayısı 8 olarak seçilmiştir. Kısaca Liu vd. (2021) tarafından önerilen Swin Transformer mimarisinde kullanılan pencere tabanlı çok başlı öz dikkat yapısının kullanıldığı görülmektedir.



Şekil 3.14 Etkili küresel-yerel dikkat mekanizmasının yapısı

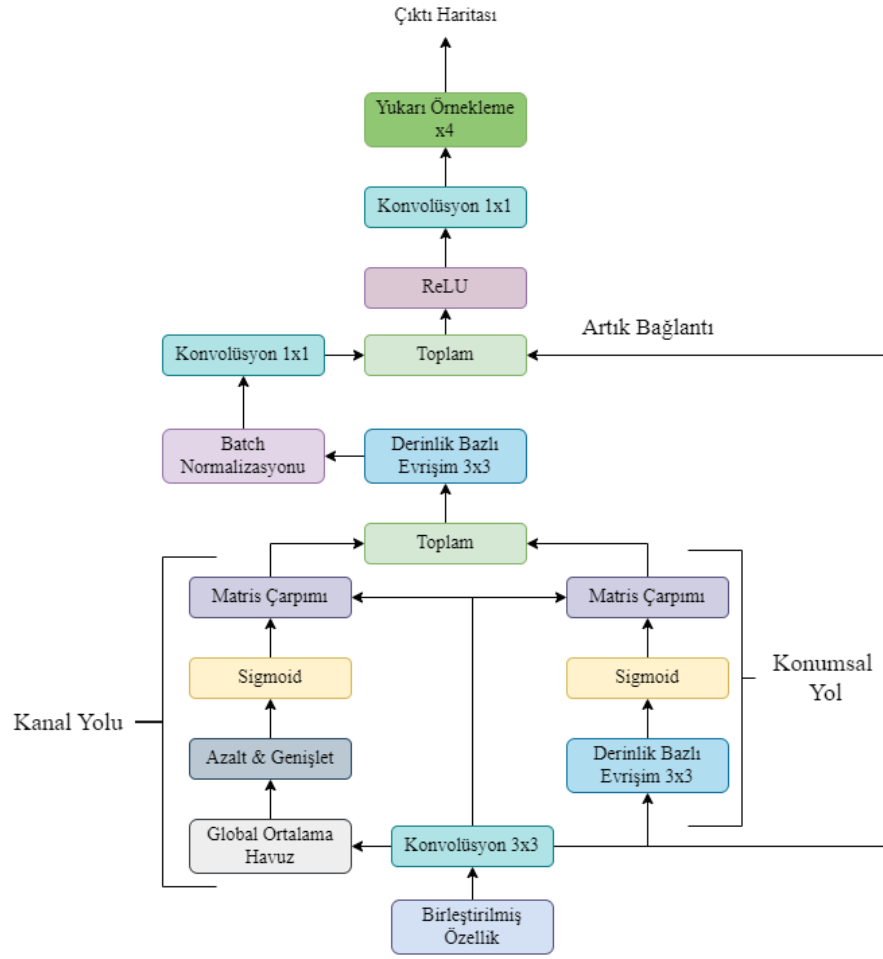
Pencereler arasında etkileşimin yokluğundan ötürü, yerel pencerelerde öz-dikkat işleminin uzamsal tutarlılığı bozabileceği öngörülmektedir. Buna çözüm olarak Swin Transformer yerel pencereler arasındaki ilişkileri ortaya koymak için kaydırılmış bir Transformer bloğu eklemektedir. Unetformer mimarisinde çapraz şekilli pencere bağlam etkileşim modülü önerilmiştir. Bu modülde yatay ortalama havuzu ve dikey

ortalama havuzu ile öznitelik değer haritası üretildikten sonra bu iki harita birleştirilerek bir küresel bağlam elde edilmiş olur.

### 3.9.2 Öznitelik iyileştirme başlığı

Görüntü bölütleme çalışmasında ilk resnet artık bloğu tarafından elde edilen öznitelikler zengin konumsal detayları korurken anlamsal içerik bakımından zayıftır. global-local özellikler ise oldukça kesin anlamsal bilgiler sunarken kaba bir konumsal çözünürlüğe sahiptirler. Bu iki özelliğin doğrudan toplamı hızlı olsa da genellikle anlamsal bölütleme doğruluğunu azaltmaktadır (Poudel vd. 2018). Şekil 3.15'te gösterilen öznitelik geliştirme başlığı bu probleme bir çözüm olarak önerilmiştir.

Anlamsal bilginin ve konumsal detayların birleştirilmesi için iki öznitelik haritası üzerinde ağırlıklı bir toplama işlemi gerçekleştirilir ve elde edilen öznitelikler öznitelik geliştirme başlığının girdisi olarak kullanılırlar. ÖGB'de kanal bazlı ve konumsal öznitelik temsillerini güçlendirmek için iki ayrı yol oluşturulur. Kanal yolunda global ortalama havuzlama ile kanal bazında dikkat oluşturulur sonrasında 1x1 konvolüsyon katmanı içeren azalt ve genişlet işlemi gerçekleştirilir. İlk önce kanal boyutu 4 kat azaltılır sonrasındaki katmanda ise boyutu tekrar eski haline getirilir. Uzamsal yolda ise öznitelik haritasının uzaysal çözünürlüğü üzerinden derinlik bazlı evrişim ile bir dikkat haritası oluşturulur. İki yoldan gelen dikkat özellikleri toplanır ve 1x1 evrişim işlemi ve sonrasında yukarı örnekleme ile bölütleme haritası üretilmiş olur. Burada kullanılan artık bağlantı da ağın performansını korumak ve bozulmasını önlemek amacını taşımaktadır.

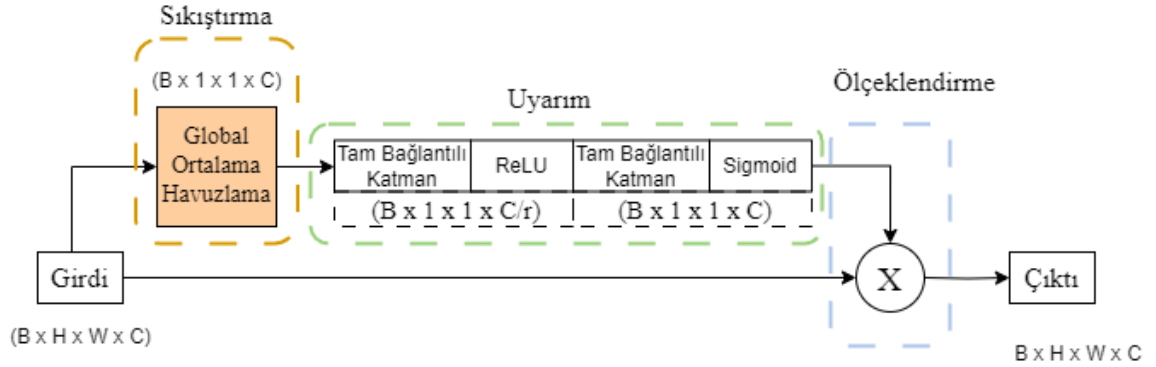


Şekil 3.15 Öznitelik Geliştirme Başlığı

### 3.10 SENet Modülü

Derin sinir ağlarında farklı öznitelik haritalarındaki farklı kanallar farklı nesnelere temsil etmektedir (Chen Long vd. 2016). Kanal dikkati her bir kanalın ağırlığını uyarlanabilir bir şekilde yeniden ölçeklendirir. Bu dikkat mekanizması hangi nesnelere dikkat edileceğini belirler. Hu vd. (2017) kanal dikkatini ilk kez önererek SENet adlı mimariyi sunmuşlardır. Mimarinin temelinde kanal bazında bağlantıları yakalamak ve temsiliyeti artırmak için kullanılan bir sıkıştırma ve uyarım (SE) bloğu vardır. Sıkıştırma modülü küresel ortalama havuzlama ile uzamsal bilgiyi toplar. Uyarım modülü kanal bazında bağlantıları yakalar ve tam bağlı katmanlar ile doğrusal olmayan katmanlar kullanarak bir dikkat vektörü oluştururlar. Giriş özniteliklerinin her kanalı dikkat vektöründe karşılık gelen eleman ile çarpılarak ölçeklendirilir. SE bloklarının bazı limitleri bulunmaktadır.

Küresel ortalama havuzlamanın karmaşık bilgiyi yakalamak için oldukça basit olduğu düşünülmektedir. Ayrıca tamamen bağlı katmanlar modelin karmaşıklığını artırıp kullanılan mimarinin performansını etkileyebilmektedir. Şekil 3.16’te SENet bloğu görülmektedir.

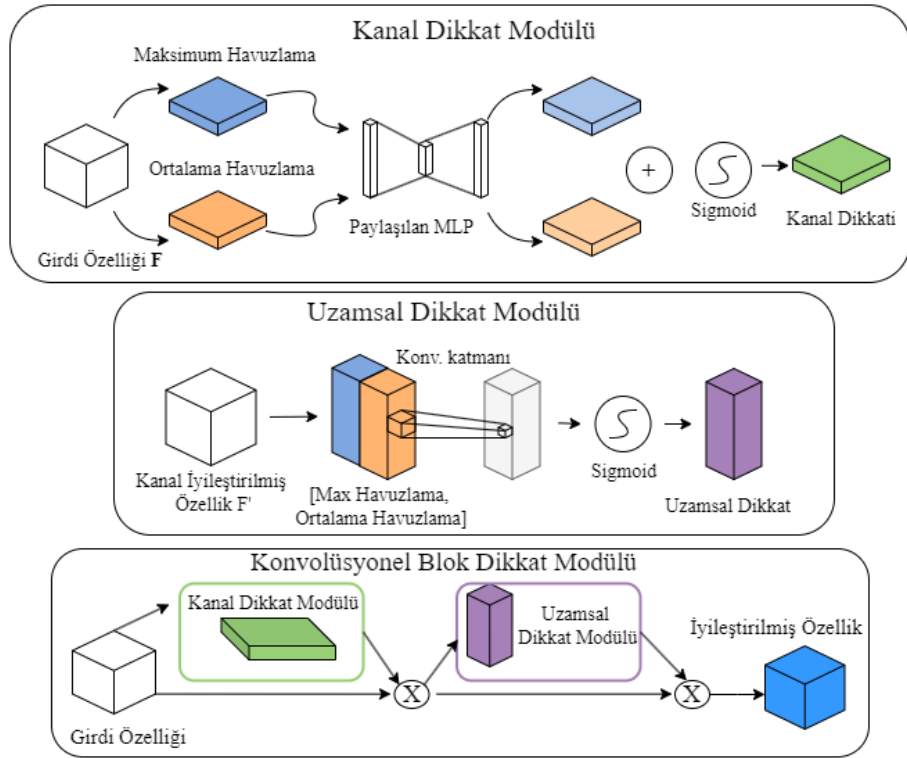


Şekil 3.16 SENet Bloğu

### 3.11 CBAM Modülü

Konvolüsyonel blok dikkat modülü olarak adlandırabileceğimiz bu modül Woo vd. (2018) tarafından önerilmiştir. Evrişim işlemleri çapraz-kanal ve mekansal bilgileri karıştırarak anlamlı özellikler elde etmektedir. Bu modül bu iki temel boyut boyunca anlamlı öznitelikleri vurgulamak için ortaya koyulmuştur. Şekil 3.17’de görüldüğü üzere kanal dikkat modülü ve uzamsal dikkat modülü sırasıyla uygulanmaktadır. Kanal dikkat modülünde özniteliklerin kanallar arası bağlantılarını kullanarak bir kanal dikkat haritası oluşturulur. Kanal dikkati verilen bir giriş görüntüsünde “ne” anlamlı onu vurgular. Mekansal bilgiyi elde etmek için ortalama havuzlama yöntemi sıkça kullanılmaktadır. Yukarıda bahsedilen SENet mimarisinde de mekansal istatistikleri hesaplamak için dikkat modülünde ortalama havuzlama kullanılmıştır. Bu çalışmada ise maksimum havuzlamanın nesne özniteliklerinden önemli noktaları çıkarmak için gerekli olduğu öne sürülmüştür. Hem ortalama havuzlama hem de maksimum havuzlama eşzamanlı olarak kullanılmıştır. Her iki kümelenmiş kanal özellikleri aynı anlamsal entegre uzayında yer aldığından parametrelerden tasarruf etmek amacıyla ortak bir MLP

kullanılmıştır. Parametre yükünü azaltmak amacıyla gizli aktivasyon boyutu tanımlanır ve azaltma oranı  $r$  ile işlem gerçekleştirilir. Bu ortak ağ uygulandıktan sonra çıktı öznitelik vektörleri eleman bazlı toplama işlemiyle toplanırlar. Uzamsal dikkat modülüne gelindiğinde ise özniteliklerin mekansal bağlantıları kullanılarak oluşturulur. Kanal dikkatinden farklı olarak uzamsal dikkat, bilgili kısımların “nerede” olduğu üzerine yoğunlaşır ve bu kanal dikkatini tamamlayıcı bir durumdur. Uzamsal dikkat için öncelikle kanal eksenini boyunca ortalama havuzlama ve maksimum havuzlama işlemleri uygulanmaktadır. Bu işlemlerden elde edilen öznitelik tanımlayıcıları birleştirilerek etkili bir öznitelik tanımlayıcısı oluşturulmuş olur. Kanal eksenini boyunca yapılan havuzlama işlemlerinin bilgi açısından zengin alanları öne çıkarmada etkili olduğu görülmüştür. Elde edilen etkili öznitelik tanımlayıcısına evrişim katmanı uygulanarak uzamsal dikkat haritası oluşturulur. Bu harita, hangi bölgelerin bastırılması gerektiğini ya da vurgulanması gerektiğini öne çıkarır.



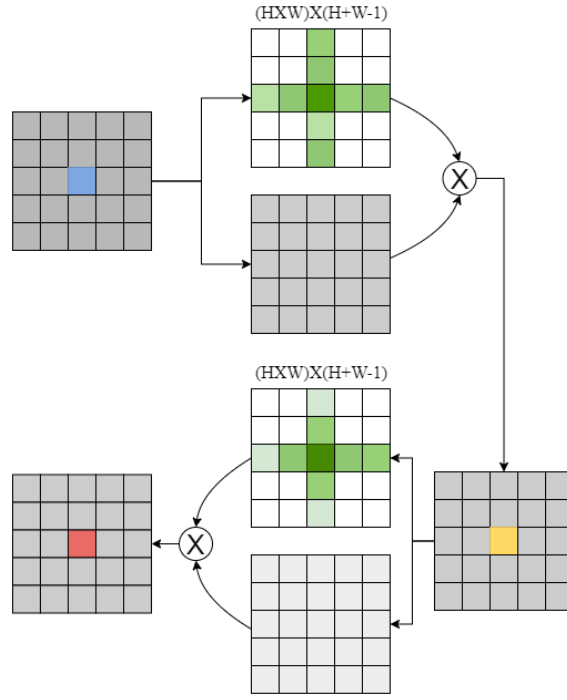
Şekil 3.17 Konvolüsyonel blok dikkat modülünün yapısı

Verilen bir giriş görseli için kanal ve uzamsal olmak üzere iki dikkat modülü arka arkaya birbirini tamamlayıcı şekilde dikkat hesaplar. Sırasıyla “ne” ve “nerede”

sorularının yanıtı olacak bilgilere odaklanılır. Bu modüllerin paralel düzenlenebileceği de önerilmiş ancak ardışık yapının daha iyi sonuç verdiği gözlenmiştir. Ayrıca kanal dikkatinin önce kullanılmasının da daha iyi performans gösterdiği tespit edilmiştir.

### 3.12 CCNet Modülü

Çapraz dikkat modülü olarak bilinen bu mimari Huang vd. (2018) tarafından anlamsal bölütleme için sunulmuştur. Öz dikkat işlemini bir grafik dönüştürme olarak ele alan yapıda yoğun bağlantılı grafik, çok daha seyrek bağlantılı bir grafiğe dönüştürülür. Bunun için mimari, satır ve sütun dikkatini tekrar tekrar inceleyerek küresel bilgiyi elde eden çapraz bir dikkat mekanizmasından oluşmaktadır. Her bir pikselin evrimsel öznitelik haritasında bir düğüm olarak kabul edildiği ve bu düğümler arasındaki içeriğin daha iyi düğüm öznitelikleri oluşturmak için kullanıldığı bir sinir ağıdır. Şekil 3.18’de çapraz dikkat modülü bloğu görülmektedir. Yerel özellikler bir çapraz dikkat modülünden geçirilir ve yatay ve dikey yönlerde bağlamsal bilgi toplanır. Bu modülden elde edilen öznitelik haritası bir başka çapraz dikkat modülüne gönderilebilir. Böylece elde edilecek ek bağlam bilgisinin tüm konumlar için tam görüntü bağıntılarını sağlayacağı düşünülmektedir. Criss-cross modülünde sorgu, anahtar ve değer anlamında üç 1x1 evrişim katmanı tanımlanmıştır. Bu katmanlar sorgu, anahtar ve değer matrislerini oluşturarak dikkat modülü için gerekli değerlerin elde edilmesini sağlar. Daha sonra dikkat katsayılarını temsil eden enerji matrisleri hesaplanmaktadır. Softmax fonksiyonu ile enerji matrisleri normalize edildikten sonra bu katsayılar ilgili projeksiyon matrisleri ile birleştirilerek çıkış değerleri hesaplanır. Bu yapı, Non-local Network ile benzer şekilde yoğun içeriksel bilgi sağlamaktadır ve bunu çok daha verimli bir şekilde yapmaktadır. Dikkat ağırlıklarını daha etkili bir şekilde güncellemekte ve benzer özniteliklere sahip pikselleri daha yüksek ağırlıklarla işleme almaktadır. Non-local Network’e benzerliği ise benzer olmayan özelliklere düşük dikkat ağırlıkları verirken benzer olanlara yüksek ağırlık vermesidir.

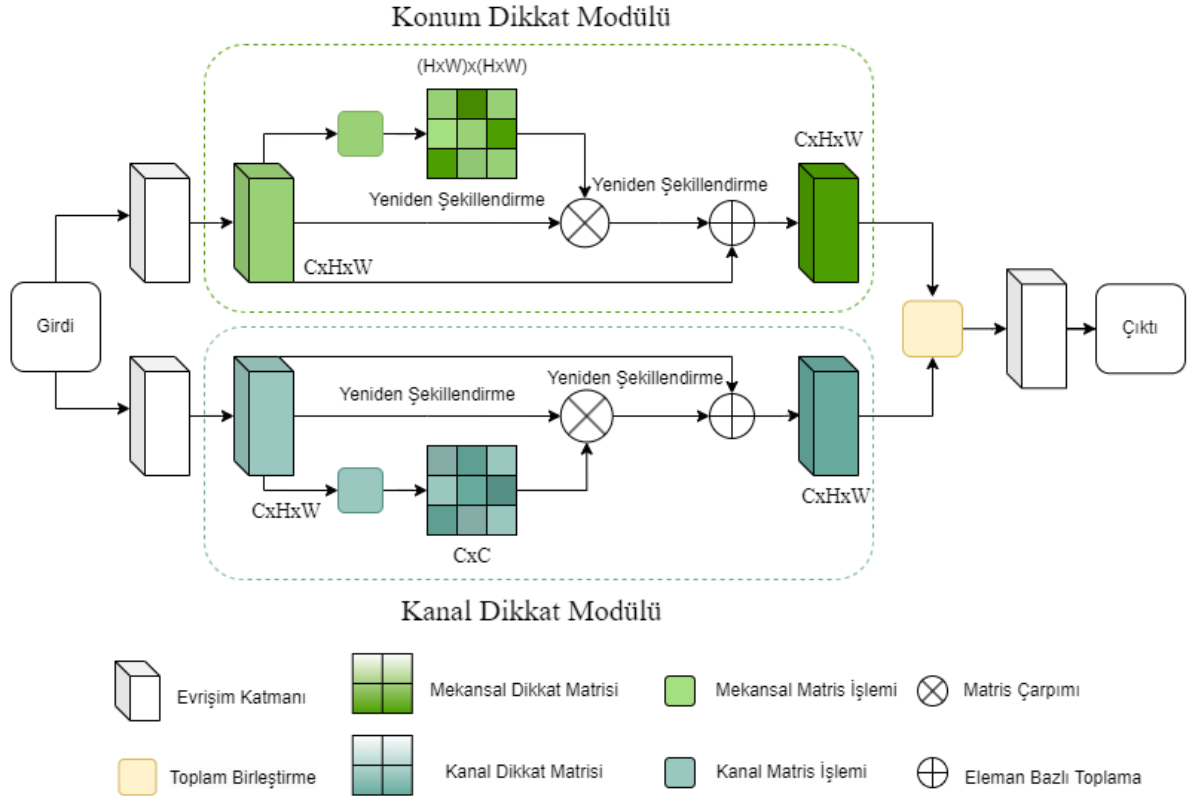


Şekil 3.18 Criss-cross dikkat modülü bloğu

### 3.13 DANet Modülü

Çift dikkat ağı olarak çevirebileceğimiz bu yapı Fu vd. (2018) tarafından önerilmiştir. Bu yöntem doğal alan görüntüsü bölütlemesi için geleneksel yöntemlerde varolan sınırları aşmayı hedeflemektedir. Bu problemler şunlardır: Kodlayıcı, kod çözücü yapılarında alandaki nesnelere arasındaki küresel bağlantılar iyi bir şekilde öğrenilememektedir. RNN tabanlı yapılar ise uzun vadeli hafıza çıktısına fazlasıyla bağımlıdır (Guo vd. 2022). Şekil 3.19’de DANet mimarisi görülmektedir. Bu modül, konum dikkati ve kanal dikkati modüllerini paralel olarak kullanarak hem mekansal hem de kanal alanlarında özellik bağılıkları modelleyebilmektedir. Bunu öz-dikkat mekanizması kullanarak gerçekleştirmektedir. Yeni öznelik haritalarının elde edilmesi için girdi öznelik haritalarına evrişim katmanları uygulanır. Konum dikkat modülü her bir pozisyondaki özneliklerin ağırlıklı toplamını kullanarak seçici olarak toplamaktadır. Bu ağırlıklar ilgili pozisyon çiftleri arasındaki öznelik benzerliği ile belirlenmiştir. Kanal dikkat modülü çapraz kanal ilişkilerini modellemek için boyut azaltma dışında benzer bir forma sahiptir. Bu iki modülün çıktıları birleştirilerek nihai temsiller elde edilir. Konum dikkat modülü, mimarinin geniş bir alandaki içeriksel

bigileri elde etmesini ve küresel bir bakış açısıyla herhangi bir ölçekte benzer özelliklerin uyumlu bir şekilde entegrasyonunu sağlar. Kanal dikkat modülü ise yararlı görülen kanalları güçlendirmekte ve gürültüyü bastırmaktadır.

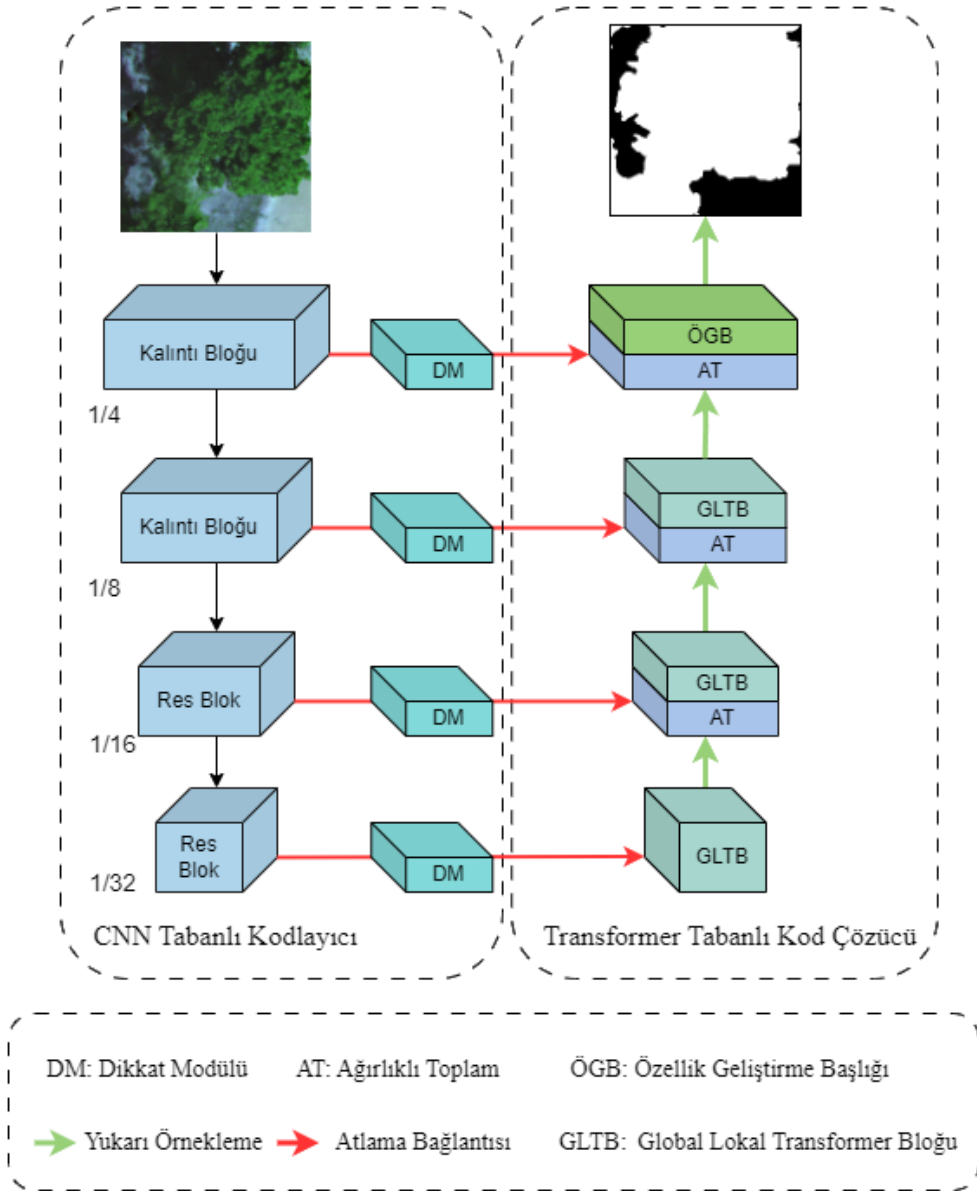


Şekil 3.19 DANet modülünün yapısı

Özniteliklerin herhangi iki pikseli arasındaki uzamsal bağlantıyı modelleyen bir mekansal dikkat matrisi oluşturulur. Ardından orjinal öznitelikler ile dikkat matrisi arasında bir matris çarpımı gerçekleştirilir. Elde edilen matris ile orjinal öznitelikler arasında bir eleman bazlı toplama işlemi gerçekleştirilir ve böylece geniş alandaki içerikleri yansıtan temsiller elde edilmiş olur. Aynı zamanda kanal boyutundaki geniş alandaki bağlantısal bilgiler kanal dikkat modülü tarafından elde edilmektedir. Kanal dikkat matrisi kanal boyutunca hesaplanır ve bu süreç konum dikkat modülüyle epey benzerdir. İki dikkat modülünden elde edilen çıktılar birleştirilir.

### 3.14 Atlama Bağlantısında Dikkat Modülü İçeren Model (AiSCUNetFormer)

Bu çalışmada ortaya koyulan esas mimaridir. Şekil 3.20’de mimari görülebilmektedir. Yukarıda bahsedilen SENet, CBAM, CCNet ve DANet dikkat mekanizmalarının mimariye eklenen dikkat modülü kısmında nasıl performans gösterecekleri incelenmiş ve birbirleriyle kıyaslanmıştır. Mimari genel olarak Unet mimarisine benzemekle birlikte verilerin kodlayıcı kısmından kod çözücü kısma geçişi yalnızca atlama bağlantıları ile olmaktadır. Unetformer mimarisinin üzerinde yapılan değişiklikler ile ortaya koyulmuş olan model aynı şekilde yalın unetformer modeliyle de kıyaslanmıştır. Şekil 3.8’de görülebildiği üzere Resnet mimarisi 4 katman olarak ele alınabilmektedir. Bu katmanlardan elde edilen öznitelikler atlama bağlantılarıyla kod çözücü kısmına gitmektedir. Kod çözücü kısmında bir alt katmandan gelen öznitelikler ile atlama bağlantısından gelen öznitelikler birleştirilir ve global-lokal transformer bloğu ile işlenirler. İşlenen veri bir üst katmana geçer. En üst katmanda özellik geliştirme başlığında son kez işlenen öznitelikler bu adımla birlikte anlamsal bölütleme işlemini gerçekleştirirler. Dikkat mekanizmalarının kod çözücü kısma iletecekleri özniteliklerin kod çözücünün daha başarılı bir şekilde öznitelik çıkarımında bulunmasını sağlayacağı öngörülmektedir.



Şekil 3.20 Yapılan çalışmada ortaya koyulan mimari

### 3.15 Hiperparametreler

Derin öğrenme modeli geliştirmek için öğrenme oranı gibi çeşitli hiperparametrelerin optimal değerlerine karar vermek gereklidir. Model parametreleri olarak adlandırılan bu parametreler model tarafından öğrenilmezler, kullanıcı tarafından verilirler. Hiperparametreler, modelin ağırlıklar ve yanlılıklar gibi diğer parametrelerini kontrol ederler (Shazida ve Zaheer, 2021). Hiperparametreler düzgünce ayarlanarak parametre

değerleri etkili şekilde öğrenilebilirler. Bu kısımda gene olarak hiperparametreler ve ortaya koyulan çalışmada kullanılan hiperparametreler kısaca anlatılacaktır.

Öğrenme oranı, ağırlıkların güncellenmesi sırasındaki adım boyutunu kontrol etmektedir. Öğrenme oranı çok yüksek seçildiğinde eğitim kararsız hale gelir ve kayıp yakınsaması gerçekleşmez. Çok düşük olduğu durumda ise eğitim süreci çok uzun sürer ve model yetersiz bir sonuca takılabilir. Epoch hiperparametresi ise veri setinin kullanılan modelden geçirilme sayısıdır. Epoch sayısı çok az olursa model yeterince öğrenmemektedir ve buna underfitting denir. Epoch sayısı çok fazla olursa model eğitim verisini ezberlemektedir ve bu durum aşırı öğrenme(overfitting) olarak adlandırılır. Adım boyutu hiperparametresi ise öğrenme oranının ayarlandığı epoch veya iterasyon sayısını ifade etmektedir. Genellikle StepLR gibi öğrenme oranı düzenleyicileriyle yaygın olarak kullanılmaktadır. Gamma ise bu öğrenme oranını düzenleyicilerde azaltmak için kullanılan çarpandır. Öğrenme oranı düzenleyicisinde seçilen adım boyutu kadar epoch gerçekleştikten sonra öğrenme oranı gamma sayısı ile çarpılarak yeni öğrenme oranı elde edilmiş olmaktadır.

### **3.15.1 Kayıp fonksiyonları**

Kayıp fonksiyonları modelin performansını ölçmek ve model ağırlıklarını optimize etmek için kullanılır. Kayıp fonksiyonları model tahminlerinin gerçek değerlerden uzaklığını ölçmektedir. BCEWithLogitsLoss ikili sınıflandırma problemlerinde kullanılırlar ve tahmin edilen her değer 0 ile 1 arasında bir olasılık olduğunu varsaymaktadır. Bu fonksiyondaki “logits” tahmin edilen verilerin aktivasyon fonksiyonuna uygulanmadan hemen önceki ham değerleridir. Bu fonksiyon sigmoid aktivasyon fonksiyonunu kayıp hesaplaması esnasında otomatik olarak uygular böylece sayısal kararlılık sağlar hem de sigmoid adımını da ortadan kaldırmaktadır. CrossEntropyLoss fonksiyonu ise çok sınıflı sınıflandırma problemlerinde kullanılırlar. Tahminler üzerinde softmax işlemini otomatik olarak uygulamaktadır. Ortalama hata karesi(MSE), fonksiyonu modelin tahminlerini gerçek değerlerle karşılaştırır ve her örneğin hata karelerinin ortalamasını hesaplar. Ortalama mutlak hata(MAE) fonksiyonu ise tahmin edilen değer ile gerçek değer arasındaki mutlak farkı ölçer. Bu fonksiyon

hata karesine göre uç değerlerden daha az etkilenir. İki fonksiyonda regresyon problemlerinde tercih edilmektedir. Burada bahsedilenlerden çok daha fazla kayıp fonksiyonu bulunmakta ve kullanım yeri ve amacına göre tercih sebebi olabilmektedirler.

### **3.15.2 Optimizasyon algoritmaları**

Modelin hata ya da kayıp fonksiyonunu minimize ederek performansını artırmak amacını taşımaktadırlar. Optimizasyon algoritmaları, modelin ağırlıklarını gradyanlara göre nasıl güncelleyeceğini belirlemektedir. Stokastik gradyan inişi(SGD) algoritmasında her yinelemede tüm veri kümesi yerine veriden rastgele seçilmiş küçük gruplar işlenmektedir. Bu da daha verimli ve hesaplama açısından daha uygulanabilir bir optimizasyon sağlamaktadır. Adam algoritması gradyanların birinci ve ikinci momentlerinin tahminlerinden, farklı parametreler için tekil uyarlanabilir öğrenme oranları hesaplamaktadır. İsmi uyarlanabilir moment tahmini ifadesinin kısaltmasından almaktadır. SGD'nin aksine Adam algoritması eğitim boyunca sabit bir öğrenme oranı kullanmaz ve geçmiş gradyanlar ile bunların ikinci momentlerine dayalı olarak her parametre için dinamik bir tekil öğrenme oranı hesaplamaktadır. RMSProp algoritması yakındaki gradyanlara göre öğrenme oranını ayarlamaktadır. Bu algoritmayla küçük değişikliklerin önemli olduğu parametreler için daha hassas öğrenme sağlanırken, daha büyük gradyanların olduğu parametreler için öğrenme hızı azaltılmaktadır. Burada bahsedilen optimizasyon algoritmaları günümüzde en çok tercih edilen algoritmalarıdır. Yine de kullanım yerine göre farklı algoritmaların da tercih edildiği bilinmektedir.

### **3.15.3 Mini-batch boyutu**

Hiperparametrelerden biri de mini-batch boyutudur. Mini batch boyutu, eğitim sırasında modelin aldığı veri miktarını ifade etmektedir. Model her bir eğitim adımında belirli bir veri kümesi parçasını kullanarak güncellemeleri gerçekleştirir. Veri setini tam bir şekilde eğitmek zaman alabilir, model küçük parçalara ayrılarak eğitim süreci hızlandırılmış olur. Mini-batch sayesinde optimizasyon süreci daha istikrarlı olmakta ve

model küçük veri parçalarını eğittiği için daha genel bir eğilim yakalamakta ve genel veri dağılımını daha iyi öğrenmektedir. Mini-batch boyutu çok küçük seçilirse model daha az genel öğrenme yapabilir hatta fazlaca eğitilmemiş gürültüde edinebilir. Çok yüksek mini-batch boyutu ise eğitim sürecini yavaşlatmakta ve yüksek bellek tüketimi ortaya çıkmaktadır. Mini batch boyutu için genellikle 16, 32, 64, 128 gibi değerler seçilmektedir.

#### **3.15.4 Başlatıcı**

Modeldeki ağırlıkların başlangıç değerini belirlemektedir. Derin öğrenme için önemli bir adımdır çünkü doğru bir başlangıç ile modelin daha hızlı öğrenmesi ve daha iyi performans göstermesi sağlanabilir. Sıfır başlatıcı, tüm ağırlıkların sıfırdan başlatılmasıdır ve genelde zor bir yaklaşımdır, öğrenme süreci zaman alabilmektedir. Xavier Initialization, katman giriş ve çıkış boyutlarına bağlı olarak ağırlıkların başlangıç değerlerini hesaplamaktadır. Derin sinir ağırları uygulamalarında yaygın olarak kullanılırlar. Kaiming Initialization ise özellikle ReLU ve benzeri doğrusal olmayan aktivasyon fonksiyonları için geliştirilmiştir ve giriş boyutlarına göre ağırlıkların dağılımını hesaplar.

#### **3.15.5 Değerlendirme metrikleri**

Modelin performansını ölçmek için kullanılan çeşitli metriklerdir. Her biri farklı sınıflandırma ya da regresyon çalışması için farklı değerlendirme özelliklerini sunmaktadır. Yapılan çalışmada Jaccard indeksi ve F1-skoru değerlendirme metrikleri kullanılmıştır. Jaccard genellikle nesne tanıma ve bölütleme için yapılan çalışmaları değerlendirmek için kullanılırken F1-skoru ikili veya çoklu sınıflandırma problemleri ile büyük dil modellerinin değerlendirilmesinde kullanılmaktadır. Çizelge 3.2'de karmaşıklık matrisi görülmektedir. Değerlendirme metriklerinde kullanılan kısaltmaların ve bu kısaltmalar kullanılarak elde edilen formüllerin neyi ifade ettiğini anlamak adına yardımcı olan bir gösterimdir.

Metriklerde kullanılan bazı kısaltmalar şunlardır:

TP: Gerçek pozitif(True Positive)

TN: Gerçek negatif(True Negative)

FP: Yanlış pozitif(False Positive)

FN: Yanlış negatif(False Negative)

Çizelge 3.2 Karmaşıklık Matrisi

	Aslında Pozitif(1)	Aslında Negatif(0)
Pozitif Tahmin (1)	Gerçek Pozitif (TP)	Yanlış Pozitif (FP)
Negatif Tahmin (0)	Yanlış Negatif (FN)	Gerçek Negatif (TN)

### 3.15.5.1 Doğruluk

Doğru olarak tahmin edilen değerlerin tüm değerlere oranıdır.

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+FN+TN} \quad (3.1)$$

### 3.15.5.2 Kesinlik

Gerçekten pozitif olan değerlerin tüm pozitif tahminlere oranıdır. Yani pozitif tahminlerin ne kadarının pozitif olduğunu ölçmektedir.

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (3.2)$$

### 3.15.5.3 Duyarlılık

Gerçek pozitif değerlerin ne kadarının pozitif tahmin edildiğini göstermektedir.

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (3.3)$$

#### 3.15.5.4 F1 skoru

F1 skoru, duyarlılık ve kesinlik değerlerinin harmonik ortalaması olarak bilinmektedir.

$$F1 \text{ Skoru} = \frac{2TP}{2TP+FP+FN} \quad (3.4)$$

#### 3.15.5.5 Jaccard indeksi

IoU olarak da bilinen bu indeks, kaçırılan değerleri ve yanlışları aynı anda dikkate alır ve toplam yanlış etiketlenmiş piksel sayısını sayarak değerlendirmektedir.

$$IoU = \frac{TP}{TP+FP+FN} \quad (3.5)$$

## 4. BULGULAR ve TARTIŞMA

Yapılan çalışmalar iki farklı veri setiyle ve farklı ResNet mimarileriyle tekrar edilmiştir. RIT-18 veri seti için 5 katlı çapraz doğrulama ve 2 katlı çapraz doğrulama kullanılmıştır. CXEF veri seti içinse 2 katlı çapraz doğrulama tercih edilmiştir. Bütün çalışmalar kaggle üzerinden yapılmış olup, çalışmalarda kullanılan GPU olarak P100 seçilmiştir.

### 4.1 Model Değerlendirme Parametreleri

Modellerin nasıl performans sergilediğini değerlendirmek için çeşitli parametreler kullanılmaktadır. Bu çalışmada 3 farklı parametre her bir model için ayrı ayrı elde edilmiştir. Bu parametrelerden ilki olan GLOPs milyar düzeyinde FLOPs'u temsil etmektedir. Modellerdeki karmaşıklığı hesaplamak için kullanılan GLOPs bir saniyede model tarafından hesaplanan milyar tane nokta işlemini belirtmektedir. GLOPs değerinin yüksek olması modelin karmaşık olduğunu gösterirken düşük olması genellikle daha hızlı işlem yapabildiğini gösterir ancak bu doğruluk oranının daha iyi olacağı anlamına gelmemektedir. FPS parametresi ise bir modelin saniyede işleyebildiği kare sayısını ifade etmektedir. Genel olarak gerçek zamanlı işleme uygulamalarında FPS değerinin yüksek olması tercih edilmektedir. İşlenen modelin karmaşıklığı, girdinin boyutları ve batch boyutu gibi çeşitli faktörler FPS değerini etkileyebilmektedir. Parametre sayısı bir modeldeki öğrenilebilir parametrelerin sayısını göstermektedir. Yani kısaca modelin karmaşıklığının en bariz göstergesidir. Parametre sayısının yüksek olması modelin karmaşık olduğunu gösterir ve bu durumda öğrenme süreci daha uzun sürecektir.

## 4.2 Rit-18 Veri Seti

Veri seti 1778 adet (224x224x3) boyutlarında görselden oluşmaktadır. Yapılan çalışmada bu görsellerden piksel piksel ağaç tespitiyle ikili anlamsal bölütleme yapılmaktadır. Çizelge 4.1’de konfigürasyon için seçilen hiperparametreler ve diğer değerler görülmektedir. İki katlı doğrulamada eğitim seti görsel sayısı 801 iken doğrulama görsel sayısı 88 olmaktadır. Beş katlı doğrulamada ise 1281 görsel eğitim setinde 142 görsel doğrulama setindedir.

Çizelge 4.1 Rit-18 veri setinde iki katlı çapraz doğrulama için seçilen parametreler

Config	Value
Toplam Input	1778
Fold No	1
Fold Size	2
Validation Rate	0.1
Mini-Batch Size	32
Number of Epochs	70
Learning Rate	0.0001
Optimizer	Adam
Training Loss	BCEWithLogitsLoss
Validation Loss	BCEWithLogitsLoss
Accuracy	Jaccard
Initialization	xavier_uniform_
Step Size	5
Gamma	0.9
Input Size	224
Channel Index	RGBs (224x224x3)
Transfer type	notr

## 4.3 Cxef Veri Seti

Veri seti 1299 adet RGB (224x224x3) boyutlarında görselden oluşmaktadır. Yapılan çalışmada bu görsellerden sarı pas hastalığı olan buğdayların piksel piksel tespiti yani ikili anlamsal bölütleme yapılmaktadır. Rit-18 veri setiyle aynı hiperparametreler

kullanılmış olup sadece ResNet50 mimarisinin backbone olarak kullanıldığı durumda mini-batch boyutu 16 olarak seçilmiştir. Multispektral görüntüler bu çalışmada kullanılmamıştır. İkili katlı çapraz doğrulama yapılan bu veri setinde 585 görsel eğitim setinde 65 görsel ise doğrulama setindedir.

#### 4.4 Rit-18 Veri Setiyle Elde Edilen Sonuçlar

Çizelge 4.2’de Jaccard indeksiyle değerlendirilen sonuçlar görülmektedir. ResNet18 mimarisinin kullanıldığı durumda sonuçların görece birbirine yakın olduğu görülmekle birlikte DANet dikkat modülü ile elde edilen sonuçların daha iyi olduğu açıktır. CBAM ise 2. en iyi sonucu vermiştir. ResNet34 mimarisi kullanılarak elde edilen sonuçların incelendiği durumda ise CBAM haricinde bütün dikkat modüllerinin ResNet18’de elde edilen sonuçlara kıyasla daha iyi sonuçlar verdiği görülmektedir. ResNet50 mimarisinin kodlayıcı kısmında olduğu diğer sonuçlara baktığımızda da bütün değerlerin daha iyi sonuçlar verdiği görülmekle birlikte yine DANet en iyi sonucu vermiştir.

Çizelge 4.2 Rit-18 Veri seti 5 katlı çapraz doğrulama ile elde edilen sonuçlar

Kullanılan Mimari	ResNet Mimarisi	Jaccard(%)
Unetformer	ResNet18	0.8808 ± 0.23889(-)
Unetformer +CBAM	ResNet18	0.8881 ± 0.23028(%0.82)
Unetformer +CCNet	ResNet18	0.8838 ± 0.23539(%0.34)
Unetformer +SENet	ResNet18	0.8814 ± 0.23609(%0.06)
Unetformer +DANet	ResNet18	<b>0.8905 ± 0.22796(%1.1)</b>
Unetformer	ResNet34	0.8848 ± 0.23278(-)
Unetformer +CBAM	ResNet34	0.8854 ± 0.23125(%0.06)
Unetformer +CCNet	ResNet34	0.8886 ± 0.23252(%0.43)
Unetformer +SENet	ResNet34	0.8838 ± 0.23505(-%0.11)
Unetformer +DANet	ResNet34	<b>0.8939 ± 0.22372(%1.03)</b>
Unetformer	ResNet50	0.8919 ± 0.22580(-)
Unetformer +CBAM	ResNet50	0.8925 ± 0.21824(%0.06)
Unetformer +CCNet	ResNet50	0.8930 ± 0.22640(%0.12)
Unetformer +SENet	ResNet50	0.8870 ± 0.23479(-%0.55)
Unetformer +DANet	ResNet50	<b>0.9043 ± 0.20686(%1.39)</b>

Çizelge 4.3’de bu sonuçların elde edildiği modellerin değerlendirme parametreleri karşılaştırılmaktadır. DANet dikkat modülü diğerlerinden daha başarılı sonuçlar vermesine rağmen parametre sayısının çok fazla olduğu ve hesaplama süresinin de diğerlerine göre yüksek olduğu görülmektedir.

Çizelge 4.3 Rit-18 veri seti 5 katlı çapraz doğrulamadan elde edilen model değerlendirme parametrelerinin karşılaştırılması

Kullanılan Mimari	ResNet Mimarisi	GLOPs	FPS	Parametre Sayısı (M)
Unetformer	ResNet18	72.01	34.12	11.72 M
Unetformer	ResNet34	131.36	25.37	21.80 M
Unetformer	ResNet50	148.18	16.28	24.20 M
Unetformer +CBAM	ResNet18	72.03	30.11	11.81 M
Unetformer +CBAM	ResNet34	131.38	23.1	21.88 M
Unetformer +CBAM	ResNet50	148.28	13.75	25.60 M
Unetformer +CCNet	ResNet18	74.11	26.47	12.16 M
Unetformer +CCNet	ResNet34	133.46	20.81	22.23 M
Unetformer +CCNet	ResNet50	181.1	11.04	31.17 M
Unetformer +SENet	ResNet18	72.03	33.6	11.81 M
Unetformer +SENet	ResNet34	131.38	25.06	21.88 M
Unetformer +SENet	ResNet50	148.28	15.69	25.60 M
Unetformer +DANet	ResNet18	136.13	15.2	25.71 M
Unetformer +DANet	ResNet34	195.48	13.2	35.82 M
Unetformer +DANet	ResNet50	1.16	3.78	248.47 M

Çizelge 4.4’de gösterildiği üzere iki katlı çapraz doğrulama ile de farklı dikkat modülleri ile kullanılan mimari denenmiştir. Sonuçlar yine DANet dikkat modülünün başarılı sonuçlar verdiğini göstermektedir. Ancak genel olarak beş katlı çapraz doğrulamada elde edilen sonuçlardan daha düşük sonuçlar elde edilmiştir. ResNet18 mimarisi ile elde edilen sonuçlarda CCNet en yakın 2. sonucu verirken diğer dikkat modüllerinden alınan sonuçların görece başarısız oldukları da görülmektedir. ResNet34 ile alınan sonuçlarda ise SENet ve dikkat modülü kullanılmayan yalın halin ResNet18 ile alınan sonuçlara kıyasla çok az daha iyi sonuçlar verdiği gözlenmiş ancak genel olarak bakıldığında Jaccard indeksi değerinde bir artış görülemediği görülmüştür. Bu denemelerde DANet’in Resnet18 için  $0.8701 \pm 0.2523$  Jaccard indeksi ve  $0.90000$  F1 Skoru elde ettiği görülmektedir. Resnet34 içinse  $0.8704 \pm 0.2545$  Jaccard ve  $0.89980$  F1 Skoru

değerleri alınmıştır. Sonuçlarda farklı ResNet mimarileri kullanılmasına rağmen kayda değer bir gelişme sağlanmadığı açıkça gözlenmektedir.

Çizelge 4.4 Rit-18 veri setinde iki katlı çapraz doğrulama ile elde edilen Jaccard ve F1 skoru

Kullanılan Mimari	Kodlayıcı Mimarisi	Jaccard(%)	F1 Skoru
Unetformer	ResNet18	0.8546 ± 0.2665(-)	0.88693
Unetformer +CBAM	ResNet18	0.8547 ± 0.2593(-)	0.88924
Unetformer +CCNet	ResNet18	0.8664 ± 0.2542(%1.38)	0.89745
Unetformer +SENet	ResNet18	0.8595 ± 0.2723(%0.57)	0.88817
Unetformer +DANet	ResNet18	<b>0.8701 ± 0.2523(%1.81)</b>	<b>0.90000</b>
Unetformer	ResNet34	0.8614 ± 0.2535(-)	0.89507
Unetformer +CBAM	ResNet34	0.8485 ± 0.2663(-%1.52)	0.88323
Unetformer +CCNet	ResNet34	0.8586 ± 0.2606(-%0.32)	0.89103
Unetformer +SENet	ResNet34	0.8612 ± 0.2614(-%0.02)	0.89208
Unetformer +DANet	ResNet34	<b>0.8704 ± 0.2545(%1.04)</b>	<b>0.89980</b>

Çizelge 4.5 ve Çizelge 4.6’da sırasıyla ResNet18 ve ResNet34 mimarilerinden elde edilen değerlendirme parametreleri gösterilmiştir. Bu çizelgeler kıyaslandığında beklendiği üzere mimarilerin karmaşıklığı ve parametre sayısı artmıştır. Çizelge 4.5’e baktığımızda mimari DANet ile kullanıldığında parametre sayısında 2.2 kat artış görülmüştür ve FPS yalın mimaride elde edilenin %44’ü kadardır. Çizelge 4.6’ya baktığımızda DANet ile parametre sayısında 1.64 kat artış görülmektedir. CBAM ve SENet ile elde edilen değerlerin benzerlik gösterdiği görülmekte ancak SENet’in FPS değerinin daha yüksek olduğu da açıkça görülmüştür.

Çizelge 4.5 Rit-18 veri seti, iki katlı çapraz doğrulama ve ResNet18 mimarisi kullanılarak elde edilen değerlendirme parametreleri

Kullanılan Mimari	GLOPs	Fps	Parametre (M)
Unetformer	72.01	34.11	11.69
Unetformer +CBAM	72.03	30.14	11.78
Unetformer +CCNet	74.11	26.45	12.12
Unetformer +SENet	72.03	33.63	11.78
Unetformer +DANet	136.13	15.18	25.71

Çizelge 4.6 Rit-18 veri seti, iki katlı çapraz doğrulama ve ResNet34 mimarisi kullanılarak elde edilen değerlendirme parametreleri

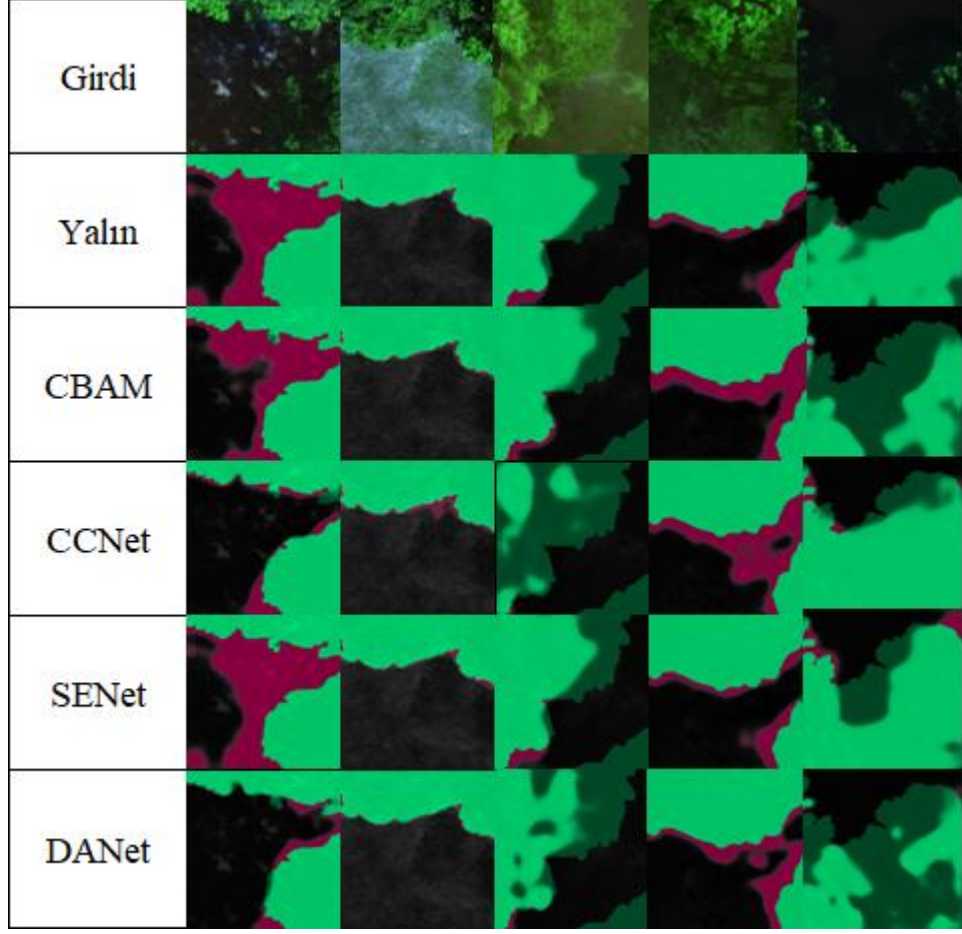
Kullanılan Mimari	GLOPs	Fps	Parametreler (M)
Unetformer	131.36	25.45	21.8
Unetformer +CBAM	131.38	23.16	21.88
Unetformer +CCNet	133.46	20.43	22.23
Unetformer +SENet	131.38	25.11	21.88
Unetformer +DANet	195.48	13.17	35.82

Çizelge 4.7’de gösterilen sonuçlar için mini-batch boyutu 16 olarak alınmış ve ResNet50 mimarisi kullanılmıştır. Sonuçlar beklenmedik bir şekilde mimarinin yalın halinin diğerlerine kıyasla çok az daha iyi sonuç verdiği göstermektedir. Mini-batch boyutu 32 olan ve ResNet18 ve ResNet34 ile elde edilen sonuçlarla(Çizelge 4.4) kıyaslandığında hepsinin daha iyi sonuç verdiği görülmektedir. Unetformer ile kıyaslandığında %0.32 ile %1.16 arasında daha kötü sonuçlar elde edildiği görülmektedir. DANet mimarisiyle parametre sayısının 10.2 kat arttığını ve Fps değerinin 4.26 kat düştüğünü görmekteyiz.

Çizelge 4.7 Rit-18 veri seti, mini-batch boyutu 16, iki katlı çapraz doğrulama ve ResNet50 mimarisi kullanılarak elde edilen sonuç ve değerlendirme parametreleri değerleri

Kullanılan Mimari	Jaccard(%)	F1 Skoru	GLOPs	FPS	Parametre (M)
Unetformer	<b>0.8755 ± 0.2565(-)</b>	<b>0.90252</b>	148.18	16.08	24.2
+CBAM	0.8669 ± 0.2721(-%0.99)	0.89305	148.28	13.76	25.6
+CCNet	0.8727 ± 0.2609(-%0.32)	0.89915	181.1	11.05	31.17
+SENet	0.8654 ± 0.2726(-%1.16)	0.89184	148.28	15.7	25.6
+DANet	0.8717 ± 0.2602(-%0.43)	0.89893	1.16	3.77	248.47

Şekil 4.1’de dikkat mekanizmalarının anlamsal bölütleme sonuçlarını görmekteyiz. İlk satır girdi olarak verilen farklı görselleri göstermektedir. Geri kalanı ise anlamsal bölütleme sonuçlarıdır. Sonuçlarda açık yeşil alanlar doğru tahminleri gösterirken kırmızıya yakın pembe renkli alanlar ise yanlış tahmin edilen noktalarıdır. Ayrıca koyu yeşil alanlar ise ıskalanan yani tahmin edilmeyen piksellerdir.



Şekil 4.1 RIT-18 veri seti için farklı dikkat mekanizmalarıyla yapılan anlamsal bölütlemenin renklendirilmiş sonuçları

#### 4.5 Cxef Veri Setiyle Elde Edilen Sonuçlar

Çizelge 4.8’de Cxef veri setiyle elde edilen sonuçlar görülmektedir. Kullanılan mimari 32 mini batch boyutuyla çalıştırılmıştır. ResNet18 backbone olarak kullanıldığında elde edilen veriler DANet’in açık ara diğerlerinden iyi performans gösterdiğini görmemizi sağlarken, ResNet34 ile elde edilen sonuçlara baktığımızda SeNet ve DANet’in daha kötü performans sergilediği diğerlerininse başarılı bir sonuç elde ettiği görülmektedir. CBAM dikkat modülü Resnet34’te diğerlerinden daha başarılı sonuçlar elde etmesine ve **0.81145** F1 Skoru almasına rağmen DANet’in ResNet18’de aldığı **0.81939** F1 Skoru daha iyi bir sonuçtur. Tabii DANet’in ResNet34’te aldığı 0.79676 F1 Skorunu da dikkate almak ve kötü sonuç elde edildiğini gözardı etmemek gerekmektedir.

Çizelge 4.8 Cxef veri setinde ResNet18 ve ResNet34 kodlayıcı mimarileri ve dikkat mekanizmalarının kıyaslanması

Kullanılan Mimari	ResNet Mimarisi	Jaccard(%)	F1 Skoru
Unetformer	ResNet18	0.7272 ± 0.3072(-)	0.79100
Unetformer +CBAM	ResNet18	0.7284 ± 0.3075(%0.16)	0.79211
Unetformer +CCNet	ResNet18	0.7376 ± 0.3014(%1.43)	0.80060
Unetformer +SENet	ResNet18	0.7294 ± 0.3126(%0.3)	0.79059
Unetformer +DANet	ResNet18	<b>0.7607 ± 0.2928(%4.6)</b>	<b>0.81939</b>
Unetformer	ResNet34	0.7377 ± 0.3031(-)	0.79997
Unetformer +CBAM	ResNet34	<b>0.7523 ± 0.3001(%1.97)</b>	<b>0.81145</b>
Unetformer +CCNet	ResNet34	0.7446 ± 0.3059(%0.93)	0.80376
Unetformer +SENet	ResNet34	0.7263 ± 0.3123(-%1.57)	0.78872
Unetformer +DANet	ResNet34	0.7319 ± 0.3011(-%0.79)	0.79676

Çizelge 4.9’da Cxef veri setinin değerlendirme parametreleri değerleri görülmektedir. İlk veri setinde de gördüğümüz gibi bu veri setinde de benzer değerler gözlenmektedir. DANet mimarisinin parametre sayısı ve karmaşıklığı diğerlerinden fazlayken diğer sonuçların hemen hemen birbirine yakın olduğu bariz bir şekilde ortadadır.

Çizelge 4.9 Cxef veri seti değerlendirme parametreleri

Kullanılan Mimari	ResNet Mimarisi	GLOPs	FPs	Parametre (M)
Unetformer	ResNet18	72.01	34.2	11.69
Unetformer +CBAM	ResNet18	72.03	30.2	11.78
Unetformer +CCNet	ResNet18	74.11	26.43	12.12
Unetformer +SENet	ResNet18	72.03	34.48	11.78
Unetformer +DANet	ResNet18	136.13	15.24	25.71
Unetformer	ResNet34	131.36	25.46	21.8
Unetformer +CBAM	ResNet34	131.38	23.19	21.88
Unetformer +CCNet	ResNet34	133.46	20.82	22.23
Unetformer +SENet	ResNet34	131.38	25.11	21.88
Unetformer +DANet	ResNet34	195.48	13.22	35.82

Çizelge 4.10’da Cxef veri setinin 16 mini batch boyutuyla ve ResNet50 mimarisinin kodlayıcı olarak kullanılmasıyla elde edilen değerler görülmektedir. DANet burada

**0.8125 ± 0.2653** Jaccard sonucuyla %5.31 artış göstermiş ve **0.8619** F1 Skoruyla yine yukarıda incelenen bir çok sonuçta olduğu gibi en iyi performansı sergilemiştir. Bir sonraki en iyi değerleri ise CCNet  $0.7762 \pm 0.2773$  Jaccard ve 0.8352 F1 Skoru ile göstermiştir. Önceki sonuçlardan farklı olarak diğer sonuçların DANet'in elde ettiği sonuçlara yakın olmadığı açık bir şekilde düşük oldukları görülebilmektedir. Çizelge 4.11'de Cxef veri setinin kodlayıcı kısmında ResNet50 kullanıldığında alınan değerlendirme parametreleri değerleri görülmektedir. Önceki değerlendirme parametreleri sonuçlarıyla büyük oranda benzer değerler gözlenmektedir. DANet mimarisinin parametre sayısı ve karmaşıklığı diğerlerinden büyük oranda fazlayken diğer sonuçların hemen hemen birbirine yakın parametre değerlerine sahip olduğu görülmektedir. DANet ile parameter sayısının 10.26 kat arttığını görmekteyiz. 4.7'de elde edilen sonuçlarla kıyaslanırsa değerlendirme parametrelerinde elde edilen sonuçların hemen hemen aynı olduğu da görülebilmektedir.

Çizelge 4.10 Cxef veri seti, mini-batch boyutu 16, iki katlı çapraz doğrulama ve ResNet50 mimarisi kullanılarak elde edilen sonuç değerleri

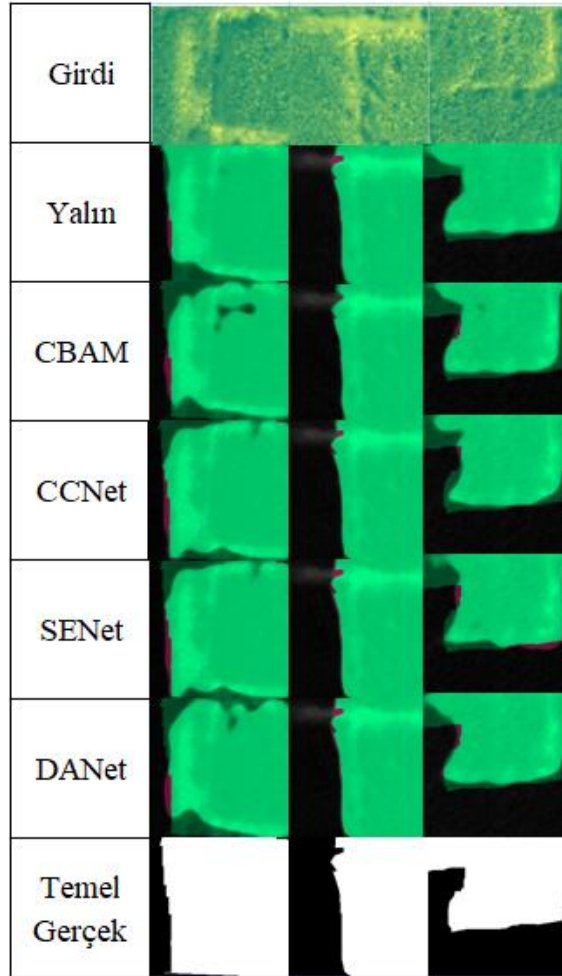
Kullanılan Mimari	Jaccard(%)	F1 Skoru
Unetformer	$0.7715 \pm 0.2856(-)$	0.8294
Unetformer +CBAM	$0.7645 \pm 0.2911(-\%0.91)$	0.8227
Unetformer +CCNet	$0.7762 \pm 0.2773(-\%0.61)$	0.8352
Unetformer +SENet	$0.7578 \pm 0.2918(-\%1.8)$	0.8177
Unetformer +DANet	<b><math>0.8125 \pm 0.2653(\%5.31)</math></b>	<b>0.8619</b>

Çizelge 4.11 Cxef veri seti, kodlayıcı ResNet50 mimarisinin değerlendirme parametreleri

Kullanılan Mimari	GLOPs	FPS	Parametreler(M)
Unetformer	148.18	16.08	24.2
Unetformer+CBAM	148.28	13.76	25.6
Unetformer+CCNet	181.1	11.07	31.17
Unetformer+SENet	148.28	15.71	25.6
Unetformer+DANet	1.16	3.76	248.47

Şekil 4.2'de Cxef veri seti için dikkat mekanizmalarının anlamsal bölütleme sonuçlarını görmekteyiz. İlk satır girdi olarak verilen farklı görselleri göstermektedir. Diğer veri setinden farklı olarak son satırda temel gerçek yani piksel piksel olması gereken esas

sonuç görülebilmektedir. Geri kalanı ise anlamsal bölütleme sonuçlarıdır. Sonuçlarda açık yeşil alanlar doğru tahminleri gösterirken pembe renkli alanlar ise yanlış tahmin edilen noktalar. Koyu yeşil alanlar ise ıskalanan yani hiç tahmin edilmeyen piksellerdir. Girdi kısmından da görülebileceği üzere veri setinin çıplak gözle yorumlaması bir hayli zordur. Ancak temel gerçekten de görülebileceği üzere veri setinin piksel piksel ayrımı biraz kaba kalmıştır.



Şekil 4.2 Cxef veri seti için farklı dikkat mekanizmalarıyla yapılan anlamsal bölütlemenin renklendirilmiş sonuçları

#### 4.6 Tartışma

4.4 ve 4.5 bölümlerinde ortaya koyulan sonuçlar hakkında her bir veri seti için ayrı ayrı yorumlaması yine bahsi geçen bölümlerde yapılmıştır. Genel olarak elde edilen sonuçlar

DANet'in diğçerlerine kıyasla daha başarılı sonuçlar gösterdiğđ olmuştur. Ancak değçerlendirme parametreleri göz önüne alındığında diğçerlerine kıyasla çok daha karmaşık ve çok daha hantal bir yapıda olduđu görülebilmektedir. Cxef veri seti etiketlerinin istenildiđi kadar düzgün gözükmediđi göz önüne alınırsa beklenenden iyi performans sergilediđi de açıktır.

Çalışmada bazı noktalarda benzerlikler ve bazı noktalarda farklılıklar iççeren 4 dikkat modülüyle ortaya koyulan mimaride öz-dikkat mekanizması kullanan DANet modelinin daha iyi sonuçlar verdiđi görülmektedir. SENet ve CBAM dikkat modülleri parametre sayısı olarak aynı çıkmışlar ve genel olarak birbirine yakın sonuçlar vermiş durumdadırlar. Grafik dikkat mekanizması iççeren CCNet modülü genellikle 2. en iyi sonucu veren modül olmuştur. DANet mimarisinin diğçerlerinden farkı evrişim tabanlı bir öz-dikkat mekanizması kullanması ve kanal dikkat ile konum dikkat modüllerinin paralel olması olduđu görülmektedir. Özellikle ResNet50 kodlayıcısıyla kullanıldığında parametre sayısında 10.2 kat kadar artış görülmektedir. Mimari bu haliyle donanımsal bir yük getirmekle birlikte doğruluk oranını artırdıđı da net bir sonuç olarak karşımızdadır.

## 5. SONUÇLAR

Yapay zeka teknolojilerinin kullanımı günümüzde gün geçtikçe artmakta ve insan hayatını giderek kolaylaştırmaktadır. Uzaktan görüntüleme teknolojileri de bir süredir hassas tarım başta olmak üzere birçok alanda bir durum incelemesi yapmaya ve gerekli adımların atılmasına olanak sağlamaktadır. Hassas tarımın giderek önem arz ettiği dünyamızda bu alanda kullanılacak bir derin öğrenme mimarisi ortaya konulmaya çalışılmıştır.

Yapılan çalışmada iki farklı veri seti üzerinde ortaya konulan mimari denenmiş ve farklı dikkat modülleri de aynı parametrelerle test edilerek hangisinin mimari için daha uygun olacağı ortaya konulmaya çalışılmıştır. Elde edilen sonuçlar ikili anlamsal bölütlemeye bu dikkat mekanizmalı ortaya konulan mimarinin uzaktan algılama görüntülerinde mimarinin yalın haline kıyasla daha iyi performans gösterebileceğini açık bir şekilde ortaya koymaktadır. Dikkat mekanizmaları genel olarak “nerede” ve “neye” dikkat edilmesi gerektiğine işaret etmektedir. Bu haliyle amaçlarına uygun sonuçlar ortaya koyulabilmiştir. DANet ile kullanıldığında mimarinin çok daha iyi ve CCNet ile genel olarak daha iyi sonuçlar verdiği görülmüştür. Ancak DANet mimarisinde görülen parametre sayısındaki artış ve FPs oranındaki düşüş göz ardı edilmemelidir.

### 5.1 Öneriler

Ortaya konulan mimari için daha farklı dikkat mekanizmaları da denenebileceği düşünülmektedir. Özellikle DANet modülünden elde edilen sonuçlar öz-dikkat mekanizması içeren veya kanal dikkat ve konum dikkat modüllerinin paralel olduğu farklı dikkat mekanizmalarının incelenmesi düşünülebilir. Ayrıca yapılan çalışma iki veri üzerinde denenmiş olsa da sadece ikili anlamsal bölütleme için kullanılmıştır. Farklı veri setleriyle de denenmesi gerekli görülebilir. Son olarak, kodlayıcı kısmında Resnet mimarisi yerine başka mimarilerle çalışmanın ilerletilebileceği düşünülmektedir.

## KAYNAKLAR

- Ajit, A., Acharya, K., Samanta, A. 2020. A Review of Convolutional Neural Networks. 1-5. 10.1109/ic-ETITE47903.2020.049.
- Abadpour, A., Kasaei, Shohreh. 2008. Principal Color and its Application to Color Image Segmentation. *Scientia Iranica*. 15. 238-245.
- Azad, R., Aghdam, E.K., Rauland, A., Jia, Y., Avval, A.H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J.P., Adeli, E., & Merhof, D. 2022. Medical Image Segmentation Review: The Success of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 10076-10095.
- Chen, L., Papandreou, G., Kokkinos, I., Murphy, K.P., Yuille, A.L. 2016. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40, 834-848.
- Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., & Chua, T. 2016. SCA-CNN: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6298-6306.
- Chen, T. Murphey, Y. 2002. Color image segmentation - An innovative approach. *Pattern Recognition*. 35. 395-405. 10.1016/S0031-3203(01)00050-4.
- Cheng, J., Dong, L., Lapata, M. 2016. Long Short-Term Memory-Networks for Machine Reading. *ArXiv, abs/1601.06733*.
- Cho, K., Merriënboer, B.V., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. *Conference on Empirical Methods in Natural Language Processing*.
- Comaniciu, D., Meer, P. 1997. "Robust analysis of feature spaces: color image segmentation," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, PR, USA, 1997, pp. 750-755, doi: 10.1109/CVPR.1997.609410.
- Csurka, G., Larlus, D. 2013. What is a good evaluation measure for semantic segmentation?. *IEEE Trans. Pattern Anal. Mach. Intell.*. 26. 10.5244/C.27.32.
- Culvenor D. S. . 2002. TIDA: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Comput. Geosci*. 28, 1 (February 2002), 33–44. [https://doi.org/10.1016/S0098-3004\(00\)00110-2](https://doi.org/10.1016/S0098-3004(00)00110-2)

- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv*, *abs/2010.11929*.
- Foley, J., Ramankutty, N., Brauman, K., Cassidy, E., Gerber, J., Johnston, M., Mueller, N., O'Connell, C., Ray, D., West, P., Balzer, C., Bennett, E., Carpenter, S., & Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Zaks, D. 2011. Solutions for a Cultivated Planet. *Nature*. 478. 337-342. [10.1038/nature10452](https://doi.org/10.1038/nature10452).
- Fu, J., Liu, J., Tian, H., Fang, Z., Lu, H. 2018. Dual Attention Network for Scene Segmentation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3141-3149.
- Fukushima, K. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics* 36, 193–202 (1980). <https://doi.org/10.1007/BF00344251>
- Garcia-Garcia, A., Orts, S., Oprea, S., Villena Martinez, V., Rodríguez, J. 2017. A Review on Deep Learning Techniques Applied to Semantic Segmentation. [10.48550/arXiv.1704.06857](https://arxiv.org/abs/1704.06857).
- Goodfellow, I., Bengio, I., Courville, A. 2016. *Deep Learning (Adaptive Computation and Machine Learning series)*. The MIT Press, Cambridge.
- Guo, MH., Xu, T., Liu, JJ., Liu, ZN., Jiang, PT., Mu, TJ., Zhang, SH., Martin, R., Hu, SM. 2022. Attention mechanisms in computer vision: A survey. *Computational Visual Media*. 8. [10.1007/s41095-022-0271-y](https://doi.org/10.1007/s41095-022-0271-y).
- He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 27-30 June 2016, 770-778*. <https://doi.org/10.1109/CVPR.2016.90>
- Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. 2017. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 2011-2023.
- Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Shi, H., & Liu, W. 2018. CCNet: Criss-Cross Attention for Semantic Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 603-612.
- Immitzer, M., Atzberger, C., & Koukal, T. 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, 4(9), 2661-2693. <https://doi.org/10.3390/rs4092661>

- Jin L, Gao S, Li Z, Tang, J. 2015. Hand-crafted features or machine learnt features? together they improve RGB-D object recognition. In: proceedings of the IEEE ISM-2014, pp. 311–319. Taichung, Taiwan
- Kemker, R., Salvaggio, C., & Kanan, C. (2017). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Krizhevsky, A., Sutskever, I., Hinton, G.E. 2012. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60, 84 - 90.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), 541–551. doi:10.1162/neco.1989.1.4.541
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9992-10002.
- Liu, P., Ren, C., Wang, Z., Jia, M., Yu, W., Ren, H., Xia, C. 2024. Evaluating the Potential of Sentinel-2 Time Series Imagery and Machine Learning for Tree Species Classification in a Mountainous Forest. *Remote. Sens.*, 16, 293.
- Long, J., Shelhamer, E., Darrell, T. 2015. Fully convolutional networks for semantic segmentation. 3431-3440. 10.1109/CVPR.2015.7298965.
- Lucas, M., Pukrop, M., Beckschäfer, P., Waske, B. 2024. Individual tree detection and crown delineation in the Harz National Park from 2009 to 2022 using mask R–CNN and aerial imagery. *ISPRS Open J. Photogramm. Remote Sens.* 2024, 13, 100071.
- McCulloch, W.S., Pitts, W. A 1943. logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133 (1943). <https://doi.org/10.1007/BF02478259>
- Mostajabi, M., Yadollahpour, P., & Shakhnarovich, G. 2014. Feedforward semantic segmentation with zoom-out features. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3376-3385.
- Noh, H., Hong, S., Han, B. 2015. Learning deconvolution network for semantic segmentation. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1520–1528. IEEE, Santiago, Chile (2015). <https://doi.org/10.1109/ICCV.2015.178>
- Poudel, R.P., Bonde, U.D., Liwicki, S., & Zach, C. 2018. ContextNet: Exploring Context and Detail for Semantic Segmentation in Real-time. *British Machine Vision Conference*.

- Rensink, R. A. 2000. The Dynamic Representation of Scenes. *Visual Cognition*, 7(1–3), 17–42. <https://doi.org/10.1080/135062800394667>
- Ronneberger, O., Fischer, P., Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- Shaziya, H., Zaheer, R. 2021. Impact of Hyperparameters on Model Development in Deep Learning. In: Chaki, N., Pejas, J., Devarakonda, N., Rao Kovvur, R.M. (eds) *Proceedings of International Conference on Computational Intelligence and Data Engineering*. Lecture Notes on Data Engineering and Communications Technologies, vol 56. Springer, Singapore. [https://doi.org/10.1007/978-981-15-8767-2\\_5](https://doi.org/10.1007/978-981-15-8767-2_5)
- Sishodia, R.P., Ray, R.L., & Singh, S. 2020. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote. Sens.*, 12, 3136.
- Su, J, Yi, D, Su, B, Mi, Z, Liu, C, Hu, X, Xu, X, Guo, L & Chen, W-H 2021. 'Aerial Visual Perception in Smart Farming: Field Study of Wheat Yellow Rust Monitoring', *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2242-2249. <https://doi.org/10.1109/TII.2020.2979237>
- Su, J., Zhu, X., Li, S., & Chen, W. 2023. AI meets UAVs: A survey on AI empowered UAV perception systems for precision agriculture. *Neurocomputing*, 518, 242–270.
- Tsouros, D., Bibi, S., Sarigiannidis, P. 2019. A Review on UAV-Based Applications for Precision Agriculture. *Information*. 10. 349. 10.3390/info10110349.
- Ülkü, İ., Akagündüz, E., Ghamisi, P., 2022. Deep Semantic Segmentation of Trees Using Multispectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. PP. 1-17. 10.1109/JSTARS.2022.3203145.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I., Kaiser, L. 2017 Attention Is All You Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, 5998–6008.
- Volpi, M., & Ferrari, V. 2015. Semantic segmentation of urban scenes by learning local class interactions. *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1–9.

- Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., Atkinson, P.M. 2021. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Woo, S., Park, J., Lee, JY., Kweon, I.S. 2018. CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Springer, Cham. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
- Wu, J., Yao, W., Polewski, P. 2018. Mapping Individual Tree Species and Vitality along Urban Road Corridors with LiDAR and Imaging Sensors: Point Density versus View Perspective. *Remote Sensing*, 10(9), 1403. <https://doi.org/10.3390/rs10091403>
- Wu T, Gu X, Shao J, Zhou R, Li Z. 2021. Colour image segmentation based on a convex K-means approach. *IET Image Process.* 2021; 15: 1596–1606. <https://doi.org/10.1049/ipr2.12128>
- Yu, Z., Bajaj, C. 2002. Normalized Gradient Vector Diffusion and Image Segmentation. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds) Computer Vision — ECCV 2002. ECCV 2002. Lecture Notes in Computer Science, vol 2352. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/3-540-47977-5\\_34](https://doi.org/10.1007/3-540-47977-5_34)
- Zhang, L., Zhang, L., B., Du. 2016. "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22-40, June 2016, doi: 10.1109/MGRS.2016.2540798
- Zhu, X., Tuia, D., Mou, L., Xia, GS., Zhang, L., Xu, F., Fraundorfer, F. 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*. 5. 8-36. 10.1109/MGRS.2017.2762307.