



FASTER R-CNN DERİN ÖĞRENME ALGORİTMASI KULLANARAK TAKIM
DÜZEYİNDE BÖCEK SINIFLANDIRILMASI

Musa Selman KUNDURACI

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Temmuz - 2020

FASTER R-CNN DERİN ÖĞRENME ALGORİTMASI KULLANARAK TAKIM
DÜZEYİNDE BÖCEK SINIFLANDIRILMASI

Musa Selman KUNDURACI

Kütahya Dumlupınar Üniversitesi

Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliği Uyarınca

Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalı

YÜKSEK LİSANS

Olarak Hazırlanmıştır.

Danışman : Dr. Öğr. Üyesi Durmuş ÖZDEMİR

Temmuz - 2020

KABUL VE ONAY SAYFASI

Musa Selman KUNDURACI tarafından hazırlanan tarafından hazırlanan "Faster R-Cnn Derin Öğrenme Algoritması Kullanarak Takım Düzeyinde Böcek Sınıflandırılması" adlı tez çalışması, aşağıda belirtilen jüri tarafından Dumlupınar Üniversitesi Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek OY BİRLİĞİ / OY ÇOKLUĞU ile Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir."

17/07/2020

Prof.Dr.Şahmurat ARIK

Enstitü Müdürü, Lisansüstü Eğitim Enstitüsü

Doç. Dr. Doğan AYDIN

Bölüm Başkanı, Bilgisayar Mühendisliği Bölümü

Dr. Öğr. Üyesi Durmuş ÖZDEMİR

Tez Danışmanı, Dumlupınar Üniversitesi

Sınav Komitesi Üyeleri

Doç. Dr. Doğan AYDIN

Dumlupınar Üniversitesi

Doç. Dr. Humar KAHRAMANLI ÖRNEK

Selçuk Üniversitesi, Bilgisayar Mühendisliği Bölümü

Dr. Öğr. Üyesi Durmuş ÖZDEMİR

Dumlupınar Üniversitesi

FASTER R-CNN DERİN ÖĞRENME ALGORİTMASI KULLANARAK TAKIM DÜZEYİNDE BÖCEK SINIFLANDIRILMASI

Musa Selman Kunduracı

Bilgisayar Mühendisliği, Yüksek Lisans Tezi, 2020

Tez Danışmanı: Dr.Öğr.Üyesi Durmuş ÖZDEMİR

ÖZET

Böcekler, eklem bacaklılar şubesinin sınıfı ve tür ve takson bakımından en kalabalık hayvan grubudur. Öyle ki böceklerde tanımlanan 750 bin, tahmin edilen tür sayısı ise 1,5 milyon civarındadır. Özellikle yapılan tahribatlar, orman yangınları türleri belirlenmeden ortadan kaybolmalarına neden olmaktadır. Bu nedenle biyolojik çeşitlilik ve akademik çalışmalar son derece önemlidir. Ayrıca böceklerin takım düzeyinde yararlarının ve zararlarının bilinmesi ve daha fazla bilgi sahibi olunması özellikle tarım alanında yapılan çalışmalara katkı sağlayacaktır. Öte yandan böceklerin tanınması sayesinde, doğayı teknolojik alanlarda taklit ederek çeşitli alanlarda robot teknolojileri, sensörler ve akıllı sistemleri geliştiren insanoğluna oldukça faydalı ve ilham kaynağı olmaktadır.

Son yıllarda yapay zeka, derin öğrenme, görüntü işleme tekniklerinde gelişen metot ve yöntemler sayesinde nesnelere tespiti, sınıflandırılması ve tanımlanması gerçekleştirilmektedir. Birçok alanda araştırmacılar bu kütüphaneleri kullanarak farklı özelliklerdeki nesnelere tespiti, sınıflandırılması ve tanımlanması için algoritma geliştirmektedirler. Bilimsel açıdan; Entomoloji ile uğraşan bilim insanları bu kriterleri baz alarak takım belirlenmesi yapmaktadır. Oldukça fazla kriterin olması zaman kaybına ve yanlış teşhislere yol açabilmektedir. Ekonomik açıdan; Özellikle tarımda mahsule zarar veren böcek türlerinin tespit edilmesi tarım verimliliği açısından son derece önemlidir.

Bu çalışmada Tensorflow kütüphanesi kullanılarak Faster R-CNN Derin Öğrenme metodu ile Böcekler üzerinde takım düzeyinde tespiti ve sınıflandırılması için bir model önerilmiştir. Çalışmada böcekler sınıfının 32 takımından 3246 böcek resmi derin öğrenmede kullanılmıştır. Oluşturulan yazılım literatüre böceklerin hızlı bir şekilde takım düzeyinde tespit edilmesi sağlayan bir sistem sunulmuştur.

Anahtar Kelimeler: Böcek sınıflandırılması, Derin Öğrenme, Faster R-CNN, Yapay Sinir Ağları

INSECT ORDER CLASSIFICATION USING FASTER R-CNN DEEP LEARNING ALGORITHM

Musa Selman KUNDURACI

Computer Engineering, MS.Thesis, 2020

Thesis Supervisor: Asist. Prof. Durmuş ÖZDEMİR

SUMMARY

Insects are the most crowded animal group in terms of class of the arthropod branch and in terms of species and taxa. In fact, 750 thousand identified in insects, the estimated number of species is around 1.5 million. In particular, destructions cause them to disappear without specifying forest fires. Therefore, biodiversity and academic studies are extremely important. Also, knowing the benefits and harms of insects at the team level and having more information will contribute to the studies especially in agriculture. On the other hand, thanks to the recognition of insects, by imitating nature in technological fields, it is very useful and inspiration for human beings who develop robot technologies, sensors and smart systems in various fields.

In recent years, thanks to the methods and methods developed in artificial intelligence, deep learning, image processing techniques, detection, classification and identification of objects have been carried out. In many areas, researchers are using these libraries to develop algorithms for the detection, classification and identification of objects of different properties. Scientifically; Scientists dealing with entomology determine teams based on these criteria. Having too many criteria can lead to time loss and misdiagnosis. In addition, there is no software to check the diagnoses made.

In this study, a model for the detection and classification of insects on Insects was proposed with the Faster R-CNN Deep Learning method using the Tensorflow library. The software created is presented to the literature with a system that enables fast detection of insects at the team level.

Keywords: Artificial neural networks, Deep Learning, Faster R-CNN , Insect classification,

TEŞEKKÜR

Bu çalışmamda bana yardımcı olan başta danışman hocam Dr.Öğr.Üyesi Durmuş ÖZDEMİR'e, desteğini her daim hissettiren Eşim Sema Kunduracı'ya, beni yetiştiren Annem Hatice ve Babam Mustafa Ziya Kunduracı'ya, anne karnından beri yanımda olan ikizim Mehmet Furkan Kunduracı'ya ve her sorunumda yardımcı olan abim Bekir Sami Kunduracı ve ablam Büşra Bardakçı'ya teşekkürü bir borç bilirim.



İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	v
SUMMARY	vi
ŞEKİLLER DİZİNİ.....	x
ÇİZELGELER DİZİNİ	xi
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ	1
1.1. Problem Durumu ve Gerekçesi	2
1.2. Çalışmanın Amacı	2
1.3. Çalışmanın Önemi.....	2
1.4. Varsayımlar	3
1.5. Sınırlılıklar	3
2. KURAMSAL ÇERÇEVE VE LİTERATÜR TARAMASI.....	4
2.1. Literatür Taraması.....	4
3. MATERYAL VE YÖNTEM	9
3.1. Materyal	9
3.1.1. Böcek takımları.....	9
3.2. Metot ve Yöntem	11
3.2.1. Tensorflow	11
3.2.2. OpenCV	11
3.2.3. Python	12
3.2.4. Nesne Tanıma Mimarileri	12
3.2.5. Çalışma ortamının hazırlanması	14
4. UYGULAMA	17
4.1. Çalışılacak Böcek Sınıf Seviyesinin Belirlenmesi	18
4.2. Resim Toplama	19

İÇİNDEKİLER (devam)

4.2. Veri eğitimi öncesi ayarlamalar	21
4.3. Verinin eğitilmesi.....	22
5. SONUÇ VE ÖNERİLER	24
5.1. Öneriler	33
KAYNAKLAR DİZİNİ	35
ÖZGEÇMİŞ	

ŞEKİLLER DİZİNİ

<u>Sekil</u>		<u>Sayfa</u>
2.1.	AlexNet Yapısı	4
2.2.	GoogleNet Mimarisi	5
3.1.	Böcek Takımları	10
3.2.	Fast R-CNN Yapısı	13
3.3.	Faster R-CNN Mimarisi	14
3.4.	Çalışma ortamı hazırlama aşamaları	15
4.1.	Uygulama Süreci	17
4.2.	Uygulama Aşamaları	18
4.3.	Eğitim Öncesi Ayarlamalar	21
4.4.	Eğitim	22
4.5.	Adım Sayısı ile Loss değeri ilişkisi	23
5.1.	Böceklerin Doğru Sınıflandırılma Oranları	25
5.2.	Örnek Sonuçlar	30
5.3.	Tespit Edilemeyen Ephemeroptera Böceği	31
5.4.	Doğru Tespit Edilen Ephemeroptera Böceği	31
5.5.	Yanlış Tespit Edilen Ephemeroptera Böceği	32
5.6.	Plecoptera Böceğinin Doğru Tespit Edilmiş Resmi	32

ÇİZELGELER DİZİNİ

<u>Cizelge</u>	<u>Sayfa</u>
4.1. Böcek Takımlarına göre eğitimde kullanılan resim sayısı	20
5.1. Böcek Takımlarının Sınıflandırma Sonuçlar	24
5.2. Konfüsyon Matris	27
5.3. Böcek Takımlarına göre Test Doğruluğu, Precision, Recall ve F1 Skor değerleri	29



SİMGELER VE KISALTMALAR DİZİNİ

<u>Kısaltmalar</u>	<u>Açıklama</u>
AI	Artificial İntelligence
API	Application Programming Interface
CNN	Convolution Neural Network
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
Faster RCNN	Faster Region Convolution Neural Network
FN	False Negative
FP	False Positive
GB	Gigabyte
Ghz	Gigaheartz
GNU	GNU's Not Unix
GPU	Graphical Processing Unit
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
JPEG	Joint Photographic Experts Group
KB	Kilobyte
MB	Megabyte
Ms	Milisaniye
OCR	Optical Character Recognition
OpenCV	Open Source Computer Vision
RAM	Random Access Memory (Rastgele Erişimli Bellek)
RFID	Radio frequency identification
RoI	Reigon of Interest (İlgi Alanı)
RPN	Region Proposal Network
SSD	Solid State Disk
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
TSU	Tensor Processing Unit XML
XML	Extensible Markup Language

1. GİRİŞ

Entomoloji literatürde “Böcek Bilimi” olarak geçmektedir. Entomoloji, temelde Zooloji'nin bir alt dalıdır. (Gullan ve Cranston, 2014) Böcek popülasyonunun fazlalığı nedeniyle, böcekler ile ilgili daha kapsamlı bir araştırma dalının olması uygun görülmüştür. Son yıllarda bu alandaki çalışmaların artışının en büyük nedeni, böceklerin bitkilere, hayvanlara ve insanlara olan zararlarının önüne geçilme isteği ile beraber doğanın taklit edilmesindeki arayıştır. Dünya mahsulünün üçte biri böcekler tarafından yağmalanmakta ve yok edilmektedir. Bu nedenle çok ürün kaybı nedeniyle ticari zararlar yaşanmaktadır. Böcek Bilimine olan katkı bu noktada büyük önem taşımının yanında ayrıca robotik, sensör teknolojisi ve akıllı sistemlere ilham kaynağı olmaktadır. (Şahin 2016) Dünya üzerinde 5 ile 100 milyon tür olduğu tahmin edilmektedir. Bilimsel olarak ise bu türlerin yalnızca 1.7 milyonu tür bazında teşhis edilerek isimlendirilmişlerdir. Her geçen gün de yeni bir tür keşfedilip isimlendirilmektedir. Bilinmeyen türlerin birçoğunu da böcek topluluğu oluşturmaktadır. Böceklerde tahmin edilen tür sayısı 1,5 milyon, isimlendirilip tanımlanan tür sayısı ise 750 bindir (Mayer, 2020)

Özellikle yapılan tahribatlar, orman yangınları türleri belirlenmeden ortadan kaybolmalarına neden olmaktadır. Bu nedenle biyolojik çeşitlilik ve akademik çalışmalar son derece önemlidir.

Böceklerin sınıflandırılması yapılırken en önemli kriterlerden birisi de ait olduğu takım düzeyini belirlemektir. Takım düzeyi belirlenmeden tür belirlenmesi yapılması mümkün olmamaktadır. Yapılan Bilimsel çalışmalar sonucunda Doğada 32 böcek takımı belirlenmiştir. En yenisi 2002 yılında bulunmuştur. Takım belirlenmesinde Kanat sayısından vücut şekline ayak sayısından baş şekline kadar yaklaşık 21 kriter vardır (Gullan ve Cranston, 2014).

Son yıllarda yazılım alanında yapay zeka algoritmalarının gelişmesiyle obje tespitinde önemli gelişmeler yaşanmıştır. Özellikle Google Brain Team tarafından geliştirilen Tensorflow kütüphanesi sayesinde yapay zeka algoritmaları kullanılarak çeşitli alanlarda farklı obje tanıma sistemleri geliştirilmiştir.

Klasik görüntü işleme tekniklerinde nesne tespiti yapılabilmesi için nesnenin sabit kamera , iyi bir ışık ile net bir fotoğrafının çekilmesi gerekmektedir. Derin Öğrenme metodunun klasik görüntü işleme tekniklerine göre sağladığı en önemli avantaj nesnenin farklı açılardan , farklı ışığa sahip fotoğraflarda bile nesne tespitini öğrenerek yapabilmesidir. Eğer bu çalışma sadece görüntü işleme teknikleri kullanılarak gerçekleştirilseydi her böcek takımı ve barındırdığı her sınıf için, her bir açı ve her bir ışık için ayrı bir algoritma oluşturulması gerekmekeydi.

1.1. Problem Durumu ve Gerekeçesi

Bilimsel açıdan; Entomoloji ile uğraşan bilim insanları bu kriterleri baz alarak takım belirlenmesi yapmaktadır. Oldukça fazla kriterin olması zaman kaybına ve yanlış teşhislere yol açabilmektedir. Ayrıca yapılan teşhislerin kontrol edilmesini sağlayan herhangi bir yazılım bulunmamaktadır. Derin Öğrenme algoritmalarının gelişmesiyle birlikte Böceklerin sınıflandırılması veya tanımlanması noktasında çeşitli çalışmalar gerçekleştirilmiştir. Bu çalışmada böcek sınıfının tüm takımlarına ait böceklerin sınıflandırılması için bir karar destek sistem yazılımı oluşturulmuştur. Böceklerin sınıflandırılması için bu karar destek sistem yazılımları büyük önem arz etmektedir.

1.2. Çalışmanın Amacı

Bu çalışmada da Biyoloji anabilim dalının Zooloji bilim dalının bir alt dolu olan Entomoloji biliminde Böcek sınıflandırılması için yardımcı derin öğrenmeye sahip bir yazılım sistemi geliştirilmiştir. Bu sistem sayesinde böcek sınıflandırılması daha kolay ve hızlı yapılacaktır. Ayrıca böcek bilimine meraklı fakat yeterince bilgisi olmayan insanlar da geliştirilen bu yazılım sayesinde doğadaki bir böceğin hangi sınıfa ait olduğunu öğrenebileceklerdir. Çalışmanın bir sonraki aşamasında bu böcekler ile bilgiler , yararları, alt familyaları gibi bilgiler yazılımı kullanan son kullanıcıya sunulacaktır. Geliştirilen yazılımın Zooloji alanındaki akademik çalışmalarda aktif bir şekilde kullanılması hedeflenmektedir. Bunun dışında zararlı böceklerin tespiti için tarımda da kullanılması amaçlanmaktadır. Çalışmanın, böceklerde tür-cins sınıflandırması için temel oluşturması amaçlanmaktadır.

1.3. Çalışmanın Önemi

Yapılan Literatür taramaları ve Entomoloji ile ilgili araştırmacılarla yapılan görüşmelerde Böcek sınıflandırılmasında Takım düzeyinde bu kadar kapsamlı bir çalışmanın olmadığı görülmüştür. Entomoloji alanındaki araştırmacıların ve bu alanda eğitim gören öğrencilerin Böcek Takım sınıflandırılmasında kullanılmak üzere bir Karar Destek Sistemi olarak kullanılabilir. Bu sayede eğitimde ve bilimde verim ve kalite artar.

Geliştirilen Yazılım Tarımda aktif bir şekilde kullanıldığı takdirde zararlı ve yararlı böceklerin tanınmasını sağlayacaktır. Bu da tarımın daha verimli olmasını sağlayacak, bu sayede ülke ekonomisine katkı sağlayacaktır.

Geliştirilen yazılımdan başarılı sonuçlar elde edilirse böcekler için kapsamlı bir tür-cins sınıflandırması için temel oluşturabilir.

1.4. Varsayımlar

Bu uygulamada sınıflandırma için kullanılacak böcek fotoğrafının görüntüsünün sabit ve net bir görüntü olduğu varsayılmıştır.

Böceğin karakteristik özelliklerini yansıtan bir resim olduğu varsayılmıştır.

Sistemin özellikle bu alanla ilgili araştırmacılara yönelik bir Karar Destek Sistemi olarak kullanılacağı varsayılmıştır.

1.5. Sınırlılıklar

Böcek familyasında yaklaşık 32 takıma mensup 750 bin tür olduğu yukarıda belirtilmişti. Tür sayısının bu kadar fazla olması her takımdaki tür, cins, sınıf sayısının fazla olmasına neden olmaktadır. Takımdan türe inildikçe farklılıklar azalmakta hatta bazı durumlarda iki böceğin herşeyi aynı olsa bile genetik yapılarına göre türler değişmektedir. Bu nedenle tür düzeyinde sınıflandırma neredeyse imkansızdır. Takım düzeyinde farklılıklar daha net olsa da aynı takımdaki türler arasında bile bir çok farklılık bulunmaktadır. Örneğin arı ve karınca dışarıdan bakıldığında oldukça farklı görünseler de aynı takıma aittirler. Bir başka takımdan örnek verirsek Coleoptera takımına ait türlerden 0.3 mm ‘den 15 cm uzunluğa sahip olanlar bulunmaktadır (Gullan ve Cranston, 2014). Her iki örnekte sınıflandırma işlemi zorlaştıran etmenlerden birisidir.

Bir diğer önemli sınırlılık ise sınıflandırma eğitimi için gerekli böcek sınıfının 32 takımına ait böcek türlerinin görsellerin elde edilmesidir. Bazı böcek takımları nadir görüldüklerinden gerekli görsellere erişim zorluğu yaşanmaktadır. Teoride Her bir takımın eğitimi için en az 100 görselin kullanılması hedeflenirken Uygulamada bazı takımlarda bu sayının altına düşmüştür. Ayrıca analiz aşaması için gerekli görseller eğitim için kullanılan görsellerden farklı olması gerekliliği bu görsele erişim zorluğunu daha da artırmaktadır.

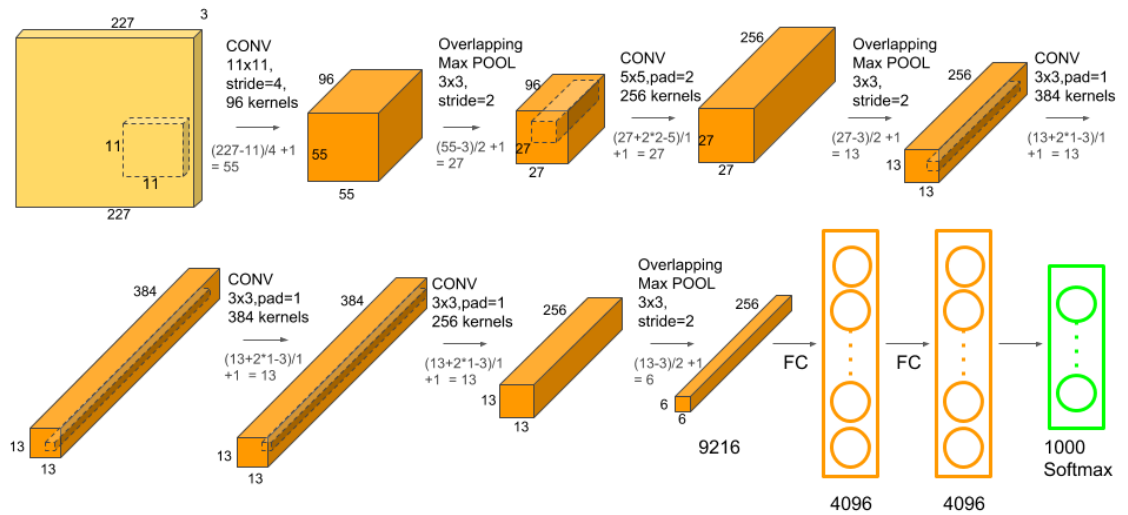
Bazı Böceklerin çok küçük olması ve hızlı hareket etmesi nedeniyle net fotoğraf elde edilmesinde zorluklar yaşanmaktadır. Bu durum da böceğin fotoğrafta karakteristik özelliğinin yansıtılamamasına neden olmaktadır.

2. KURAMSAL ÇERÇEVE VE LİTERATÜR TARAMASI

2.1. Literatür Taraması

Araştırmada literatür taraması, 2 farklı temel yapı üzerine odaklanarak gerçekleştirilmiştir. Bunlardan ilki fotoğraflardan nesne tanıma işlemi gerçekleştiren çalışmalar, ikincisi ise yazılım temelli olan ve böceklerin tanınması üzerine çalışmalardır.

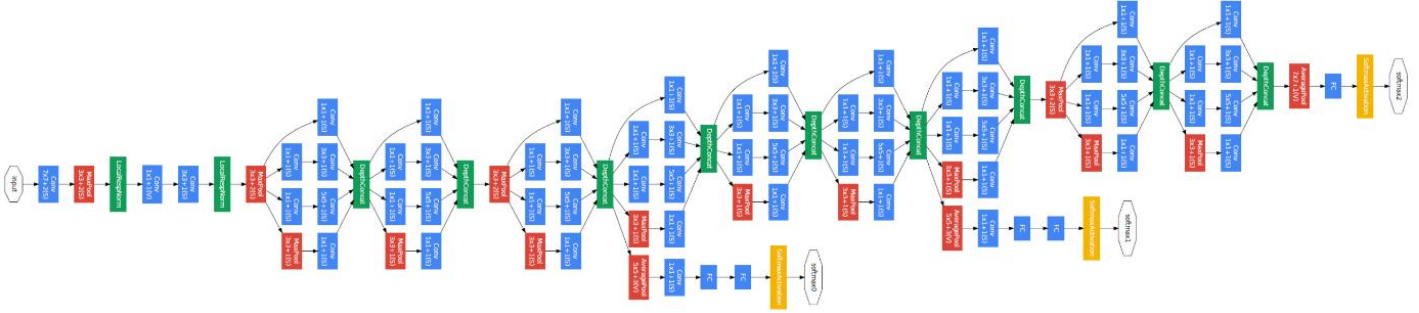
Fotoğraf üzerinden nesne tanıma ile ilgili çalışmalara bakıldığında ilk olarak AlexNet çalışması öne çıkmaktadır. AlexNet bir Image sınıflama yarışması olan ILSVRC 'yi 2012 yılında kazanmıştır. AlexNet'in mimarisi incelendiğinde bir önceki Computer Vision CNN'lerden çok daha büyük olduğunu görmekteyiz. 60 milyon parametreye ve 650.000 nörona sahip ve iki GTX 580 3GB GPU'da eğitilmesi beş ila altı gün sürdü. Günümüzde, çok büyük veri kümelerinde bile daha hızlı GPU'larda çok verimli bir şekilde çalışabilen çok daha karmaşık CNN'ler bulunmaktadır. Ama 2012 yılında bu büyük bir şeydi. AlexNet'in mimari yapısı aşağıdaki görselde görünmektedir. AlexNet 'in mimarisi 8 katmandan oluşmaktadır (Krizhevsky vd., 2012).



Şekil 2.1. AlexNet Yapısı.

AlexNet fotoğraf üzerinde bu denli başarılı olması diğer araştırmacıları da heyecanlandırmış, bu alanla ilgili araştırmalar, ar-ge çalışmaları artmıştır. Google AlexNet'in bu mimarisini geliştirerek 22 katmana sahip daha hızlı bir mimari yapı oluşturdu. Bu sayede

ImageNet üzerinde bulunan resimler üzerinden başarılı sonuçlar elde etti ve ILSVRC 2014'ü kazandı. Aşağıdaki görselde GoogleNet mimarisi görülmektedir (Szegedy vd., 2015).



Şekil 2.2. GoogleNet Mimarisi.

AdaBoost adı verilen bir algoritma Yapay Sinir Ağları yöntemlerini kullanarak tahıl ambarlarındaki depolanan böceklerin sınıflandırılmasını gerçekleştirmiştir. Geliştirilen algoritma standart sinir ağları yöntemleri ile karşılaştırılmış, deney sonuçlarına göre yeni yöntemin verimli olduğu ve sınıflama doğruluğunda önemli bir iyileşme elde edilmiştir (Zhang vd., 2008).

Ucuz sensörler kullanarak yapılan bir çalışmada Kanatlı Böcekler üzerinde sınıflandırma işlemi yapmışlardır. Bir kutunun bir tarafına yerleştirilen hareket sensörü ile kutuya konulan kanatlı böceklerin kanat hareketleri ile sınıflandırma işlemi gerçekleştirmektedir (Batista vd. 2011).

Buğdaylara zarar veren böcekler üzerinde yapılan çalışmada; Zararlı olan Pentatomidea ve Heteroptera türlerinin Gömülü bir Sistem ile Knn ve SVMa algoritmalarını kullanarak tespiti ve sınıflandırılmasını gerçekleştirmişlerdir (Yazgaç ve Mürvet, 2017).

Bir çalışmada, Kelebeklerin ve uğur böceklerin sınıflandırılmasını gerçekleştiren bir algoritma geliştirmişlerdir. Görüntü işleme tekniklerini kullanarak renk ,boyut ve pozisyon parametrelerine göre sınıflandırma işlemi gerçekleştirmişlerdir (J. Lim vd., 2006).

2009 yılında yapılan bir çalışmada Faster R-CNN mimarisini kullanarak Araçlar üzerinde Marka Model tespiti gerçekleştiren bir algoritma geliştirilmiştir. Geliştirdikleri algoritmanın daha önce yapılan çalışmalara ve daha farklı yöntemlere kıyasla daha başarılı olduğu görülmektedir (Kunduraci ve Örnek, 2019).

Bir diğer çalışmada Computer Vision teknolojilerini kullanarak bazı böceklerin tespit ettikleri 14 özelliğini makine öğrenme algoritmalarıyla eğitime sokarak bu böceklerin

tanımlanması ve sınıflandırılmasını gerçekleştirmişlerdir. Algoritmayı geliştirmek için bir Computer Vision kütüphanesi olan OpenCV kullanılmıştır. Tanımlama ve sınıflama işlemi 7 tür için gerçekleştirilmiştir (Yang vd., 2010).

Bu makalenin konusuna yakın olan bir çalışmada Orman böceklerinin sınıflandırılması ile ilgili çalışmada CNN mimarisi kullanarak bir Uygulama geliştirilmiştir. Derin Öğrenme, Görüntü İşleme kütüphanelerinin kullanıldığı mobil uygulamada; fotoğrafı çekilen orman böceği web tabanlı sunucudaki sınıflandırıcıda sınıflandırıldıktan sonra sonucu uygulamaya gönderilir. Uygulama 30 tür için gerçekleştirilmiş olup, sınıflandırılmada kullanılan isimlendirme bilimsel isimlendirmeler değil yaygın bilinen isimlendirmelerdir. (Örneğin Bal Arısı gibi) Eğitim için toplamda 22877 , Onaylama için toplamda 3861, test için de toplamda 2984 resim kullanılmıştır. Derin öğrenme için kullanılan resimlerin fazla olması eşleştirme sayısını artırmıştır. Çalışma sonucunda yapılan deneyde %95 başarılı elde edilmiştir. Ayrıca bilinen böcek türlerinin seçilmesi de eğitim için gerekli resim sayısının elde edilmesini artırmıştır. Geliştirilen Uygulama Android cihazlarda çalışmaktadır. (S. Lim et al. 2018) Bu çalışmadan farklı olarak bu makalenin konusu olan çalışmada böcek sınıflandırılması bilimsel adları ile yapılmaktadır. Ayrıca çalışmada bilinen böcek türleri ile çalışma gerçekleştirilmiş olup takım düzeyinde bir ayırım söz konusu değildir. Gerçekleştirdiğimiz çalışmada ise böcekler takım düzeyinde sınıflandırılmaktadır. Bu nedenle bazı ender bulunan böcek takımları üzerinde de sınıflandırılma yapılabilmektedir (Lim vd., 2018).

Yapılan başka bir çalışmada Squeeze-and-Excitation Networks modülünü kullanarak bir böcek sınıflama uygulaması geliştirmişlerdir. SERAN adı verilen uygulamada 123 sınıf için 34525 resim eğitimde kullanılmıştır. Geliştirilen algoritma benzer sınıflama algoritmaları ile karşılaştırıldığında daha başarılı olduğu görülmüştür (Park vd., 2019).

Bir bilim projesi ve dünyadaki biyoçeşitlilik gözlemlerinin haritalanması ve paylaşılması kavramına dayanan doğalcıların, vatandaş bilim adamlarının ve biyologların çevrimiçi sosyal ağ olan iNaturalist web sitesindeki türlerin sınıflandırılması ve tespit veri seti oluşturulmuştur. Çalışma 5000 farklı bitki ve hayvan türünden sınıflandırılması ve tespit veri setini 859 000 resim ile sağlamaktadır. Resimler iNaturalist ile Global Biyoçeşitlilik Bilgi Tesisi(GBIF) yardımıyla toplanmıştır. iNat2017 karşılaşmasında geliştirilen Algoritmalarından IncResNetV2 SE diğer algoritmalara göre daha başarılı sonuçlar vermiştir (Horn vd., 2017).

CNN Kullanarak Böcek Sınıflandırması yapılan bu çalışmada Performans Etki Analizi gerçekleştirilmiştir. Geliştirilen algoritmada 27 sınıf ve her bir sınıfta 1300 resmin eğitim için 256x256 boyutlarına kırılmış, filtrelenmiştir. Eğitim için hazırlanan resimler CNN mimarisinde kullanılmıştır. Daha sonra eğitilmiş olan model ile böcek sınıflaması yapılarak performansı

ölçülmüştür. Sonuç olarak sınıflama performansında %20 ila %25 arasında artış görülmüştür (Lim vd., 2017).

Sinir Ağı ve Derin Öğrenme yaklaşımına dayalı böcek sınıflandırma için CDNN modeli önerilmektedir. İlk olarak, böcek görüntüleri, Yoğun Ölçekli Değişmeyen Özellik Dönüşümü temel alınarak toplanmış ve özellikler çıkarılmıştır. Ardından, Özellikler kesesi, özellik vektörleri olarak görüntü gösterimi için kullanılır. Son olarak, bu özellik vektörleri Derin Sinir Ağına dayalı CDNN modeli kullanılarak eğitilmiş ve sınıflandırılmıştır. Bu yaklaşım pirinç üretiminin yapıldığı Mekong Deltasındaki Pirinçlere zarar veren Kahverengi Yaprak Piresi ve Uğur Böceğinin tespit edilmesi amacıyla geliştirilmiştir (Huynh vd., 2019).

Bir başka böcek sınıflandırma çalışmasında, 24 farklı Böcek Türünün Derin Öğrenme metotlarından CNN üzerinden geliştirilen VGG19 sınıflama mimarisi kullanılarak sınıflandırılması gerçekleştirilmiştir. Çalışmada 24 farklı tür için internet üzerinden elde edilen 540 fotoğraf derin öğrenmede kullanılmıştır. Veri artırımı aşamasında ise bu sayı 4800'e çıkarılmıştır. Seçilen türler incelendiğinde aynı takıma sahip böcekler olduğu görülmektedir. Çalışmada tarım alanlarındaki en yaygın 24 tür seçilmiştir (Xia vd., 2018).

Insectify adı verilen Android uygulamasında bir CNN kullanılarak böcek tür tanımlaması gerçekleştirilmiştir. Çalışmada CNN olarak Inception v3 kullanılmıştır. Çalışmada en bilinen böcek çeşitlerinden 10 sınıf seçilmiştir. Bee, Bettle, Butterfly, Cicada, Cockroach, Dragonfly, Fly, Grasshopper, Mantis, Wasp seçilen türlerdir. Çalışmamızda 32 takım türünün sınıflandırılması gerçekleştirilmiştir. Çalışmamızın en büyük farkı da budur. Sınıf sayısının az olması yakın takımların sınıflandırılmasındaki hata oranını azaltmaktadır. Ayrıca takım sayısının az olması çalışmanın hazırlık , uygulama ve eğitim aşamalarında daha az vakit harcanmasını sağlamaktadır (Guiam ve Bawagan, 2017) .

Kelebekler ile ilgili 22 türün sınıflandırılmasının gerçekleştirildiği bu çalışmada Derin CNN ile SVM (Karar destek makinesi) metotlarının kombine ederek hibrit bir metot kullanılmıştır. 22 türün sınıflandırılması için gerekli CNN eğitimi için 1301 fotoğraf kullanılmıştır (Zhu vd., 2017).

Bir başka çalışmada Böceklerin tanımlanmasına yönelik bir Karar Destek Sistemi gerçekleştirilmiştir. Çalışmada iki zorlu görev test edilmiştir. İlki daha alt gruplara ait böcek görselleri daha üst gruplara tanımlanmış, ikinci görevde ise uzmanlar bile zor olan benzer türlerin ayırt edilmesine olanak sağlayan bir SVM oluşturulmuştur. İlk görev için Diptera böcek takımına ait 11 familyadan 884 yüz görüntüsünde %96 doğruluk ve Coleoptera böcek takımına ait 14

famlyadan 2936 grntden %90 dođruluk sađlanmıřtır. İkinci grevde; Bir veri kmesinde Coleoptera cinsi Oxythyrea'nın ç trnn 339 grntsnde %96 dođruluk ve diđer bir veri kmesinde Plecoptera larvalarının trlerinde %98.6 dođruluk sađlanan bir karar destek sistemi oluřturulmuřtur. Geliřtirilen yaklařımın byk lde uygulanabilir olduđu sonucuna varılmıřtır (Valan vd., 2019).

Bcek zararlıları, tarımsal rn retimini etkileyen ana faktrlerden biridir. Bilgisayar algoritmalarının ve yapay zekanın geliřmesiyle, bcek zararlılarının erken ařamalarda dođru ve hızlı bir řekilde tanınması, kısa ve uzun vadede ekonomik kayıpların nlenmesine yardımcı olabilir. alıřmada bcek zararlılarının Tarıma verdiđi zararı en aza indirmek iin Derin đrenme metotları ile bcek tanıma sistemi geliřtirilmiřtir. Veri bytme tekniđi ile 102 sınıflandırma tr iin 27500 fotođrafi eđitimde kullanılmıřlardır. alıřmada tr dzeyinde, tarımda grlen bcek trleri zerinde bir sınıflandırma gerekleřtirilmiřtir. alıřmamamızda ise 32 sınıf vardır; Bu sınıflar bceklerin takım dzeylerindeki eřitlerine gre oluřturulmuřtur ve sadece belirli bir blge deđil tm bcek takımlarını sınıf olarak alıřmaya dahil edilmiřtir. Bu alıřmada veri bytme tekniđinin nemine vurgu yapılmıřtır (Khalifa vd., 2020).

3. MATERYAL VE YÖNTEM

3.1. Materyal

Bu bölümde veri eğitiminde kullanılan materyaller açıklanmıştır. Bu çalışmada Böcek sınıfının takımları üzerinde çalışmalar gerçekleştirilmiştir.

3.1.1. Böcek takımları

Böcekler, eklem bacaklılar (Arthropoda) şubesinin bir sınıfıdır ve tür bakımından en kalabalık hayvan topluluğudur. Böcek sınıfına bağlı da 32 takım vardır (Mayer, 2020).

Böceklerin sınıflandırılması yapılırken en önemli kriterlerden birisi de ait olduğu takım düzeyini belirlemektir. Takım düzeyi belirlenmeden tür belirlenmesi yapılması mümkün olmamaktadır. Yapılan Bilimsel çalışmalar sonucunda Doğada 32 böcek takımı belirlenmiştir. En yenisi 2002 yılında bulunmuştur. Takım belirlenmesinde Kanat sayısından vücut şekline ayak sayısından baş şekline kadar yaklaşık 21 kriter vardır. Anahtar olarak adlandırılan bu kriterler sayesinde böcek bilimciler sınıflandırma yapmaktadırlar. Bu kriterler aşağı gibi sıralanabilir (Gullan ve Cranston 2014).

✓ Kanat Sayısı	✓ Tarsomer	✓ Vücut
✓ Kanat Şekli	numarası	bölgeleri
✓ Kanat Deseni	✓ Pretarsus	✓ Vücut Şekli
✓ Kanat gelişimi	✓ Kafa Şekli	✓ Pronotum şekli
✓ Kanat Damarı	✓ Bileşik gözler	✓ Karın apeksi
✓ Kanat Tabanı	✓ Ağız parçaları	✓ Karın tabanı
✓ Ön Ayaklar	✓ Dokungaçlar	✓ Cerci
✓ Arka bacaklar		

Bu kriterlerin aldıkları değerlere göre böceğin ait olduğu takım belirlenebiliyor. Her bir kriterde de bir çok değer bulunmaktadır. Geliştirilen bazı yazılımlar sayesinde bu kriterleri seçip böceğin ait olduğu takım belirlenmektedir.

Entomologlar bu kriterlere göre böceği 32 böcek takımından birine ait olduğunu söylemektedirler. Eğer bu kriterlerin alabileceği değerlerden farklı bir özellikte ise yeni bir böcek

takımı olabilir. Zira en son keşfedilen Mantophasmotodea böcek takımı 2002 yılında bu şekilde bulunmuştur. Tabi ki bu durum oldukça nadir olan bir durumdur(Gullan ve Cranston, 2014).

Geliştirilen bu yazılımda tüm takım türlerine ait sınıflandırma işlemi yapılmıştır. Yazılımda tanımlanan böcek takımları fotoğrafları ile birlikte aşağıda sıralanmıştır



Şekil 3.1. Böcek Takımları.

- Archaeognatha - Microcoryphia - Bristletails
- Blattodea – Cockroaches
- Coleoptera - Beetles
- Collembola - Springtails
- Dermaptera - Earwigs
- Diplura
- Diptera – True Flies
- Embioptera - webspinners
- Ephemeroptera - Mayflies
- Grylloblatodea - Rockcrawlers
- Hemiptera – Aphids, Cicadas, Leafhoppers, Scale insect
- Hymenoptera – Ants, Wasps, Bees
- Isoptera – Termites
- Lepidoptera – Moths, Butterflies
- Mantodea
- Mantophasmotodea - Gladiators
- Mecoptera - Scorpionflies
- Megaloptera
- Neuroptera - Lacewings
- Odonata – Dragonflies, Damselflies
- Orthoptera–Grasshoppers, Crickets
- Phasmatodea -Phasmida - Walkingsticks
- Phthiraptera - Lice

- Plecoptera – Stoneflies
- Protura - Coneheads
- Psocoptera – Booklice, Barklice
- Siphonaptera - Fleas
- Strepsiptera–Twisted-wing
Parasites
- Thysanoptera - Thrips
- Thysanura - Silverfish
- Trichoptera - Caddisflies
- Zorapter

Bu çalışmada geliştirilen yazılımda ise bu kriterlerin öğrenilmesi derin öğrenme ve yapay zeka ile sağlanmaktadır.

3.2. Metot ve Yöntem

Bu çalışmada Python programlama dili, Tensorflow Derin Öğrenme kütüphanesi ve Faster R-CNN Derin Öğrenme Ağ Modeli kullanılmıştır.

3.2.1. Tensorflow

Tensorflow, Google Brain Team tarafından geliştirilen, yapay sinir ağları, derin öğrenme ve genetik algoritmalar gibi alanlarda kullanılabilen, ücretsiz ve açık kaynak kodlu bir yazılım kütüphanesidir. Google'da hem araştırma hem de üretim için kullanılır. Temelde Python programlama dili ile geliştirilen bu kütüphane C++, Java, Javascript ve R gibi programlama dilini desteklemektedir. CPU ve GPU olarak iki şekilde entegre edilebilir. CPU modeli verilerin işlenmesinde işlemci kullanırken, GPU modelinde ekran kartı işlemcisi kullanılmaktadır (Tensorflow, 2019).

Tensorflow'un GPU modelinin kurulabilmesi için CUDA destekli bir ekran kartına ihtiyaç vardır. Bu çalışmada Tensorflow GPU modelinin 1.14 sürümü kullanılmıştır.

3.2.2. OpenCV

OpenCV (Açık Kaynak Bilgisayarlı Görme Kitaplığı) açık kaynak kodlu gerçek zamanlı bir bilgisayarlı görme ve makine öğrenimi yazılım kütüphanesidir. Başlangıçta Intel firması tarafından geliştirilen ardından Willow Garage daha sonra itseez tarafından desteklenen OpenCV bilgisayarlı görme yazılımlarına bir altyapı sağladı ve ticari uygulamalarda sıkça kullanıldı. Açık Kaynak kodlu olduğu için geliştiriciler tarafından da geliştirildi de değiştirildi.

OpenCV'nin en önemli avantajlarından bir tanesi de platform bağımsız olmasıdır. C++, Python ve Java programlama dillerinde kullanıldığından Linux, MacOS, Windows, iOS ve Android işletim sistemlerinde kullanılmaktadır.

OpenCV kütüphanesi içinde klasik görüntü işleme , bilgisayarlı görü ve makine öğrenimini içeren 2500'ü aşkın optimize edilmiş algoritma vardır. Bu algoritmalar sayesinde yüz tanıma, nesne tanıma, sınıflandırma, nesnelerin takibi, nesnelerin 3 boyutlu modellerini ayıklamak gibi daha bir çok işlem gerçek zamanlı ya da fotoğraf üzerinden yapılabilir. OpenCV 47 binden fazla topluluğa ve 18 milyondan fazla indirmeye sahiptir. Kütüphane şirketler, kamu kurumları ve araştırma grupları tarafından kullanılmaktadır (OpenCV, 2019).

3.2.3. Python

Python, dinamik şemasıyla nesne yönelimli, yorumlamalı , yüksek seviyeli bir programlama dilidir. Akılda kalıcı girintilere dayalı basit söz dizimi yapısıyla vakit kazandıran , kolay öğrenilen bir dildir. Modüler yapısı ile her türlü veri alan girişini ve sınıf yapısını destekler. Platform bağımsız olması nedeniyle Unix , Linux, Mac, Windows, Amiga, Symbian işletim sistemlerinde kullanılabilir. Bu avantajının yanında diğer programlama dillerinden ayrılan en önemli avantajı ve bu kadar yaygın kullanılmasının nedeni; Python ile web uygulamaları, kullanıcı arabirim uygulamaları, mobil uygulamalar, sistem uygulamaları, veritabanı yazılımı gibi birçok alanda yazılım geliştiriliyor olmasıdır.

Python aslında bir programlama dili olan ABC'ye alternatif olarak 80'lerin sonunda tasarlandı. İlk versiyonundan sonra 2000 yılında Python 2 , 2008 yılında da Python 3 sürümü yayınlandı. Bu iki versiyon birbirinden farklı olduğundan birinde geliştirilen uygulama diğerinde çalışmamaktadır.

Bu çalışmada geliştirilen yazılım Python 3 sürümünde yazılmıştır. Bu nedenle sadece Python 3 sürümlerinde stabil olarak çalışabilir (Python, 2020).

3.2.4. Nesne Tanıma Mimarileri

Literatüre bakıldığında; Nesne tanımda klasik görüntü işleme metotlarından farklı olarak Derin Öğrenme Ağ Modellerinin kullanıldığını görüyoruz. Bu Derin Öğrenme Ağ Modelleri sayesinde bir nesne farklı ışıkta, farklı açılarda olsa bile tanıma işlemi başarılı olabilmektedir. Literatürde bazı derin öğrenme ağ modelleri öne çıkmaktadır. Bunlardan bazıları CNN, R-CNN, Fast R-CNN, Faster R-CNN ve SSD Ağ Modelleri.

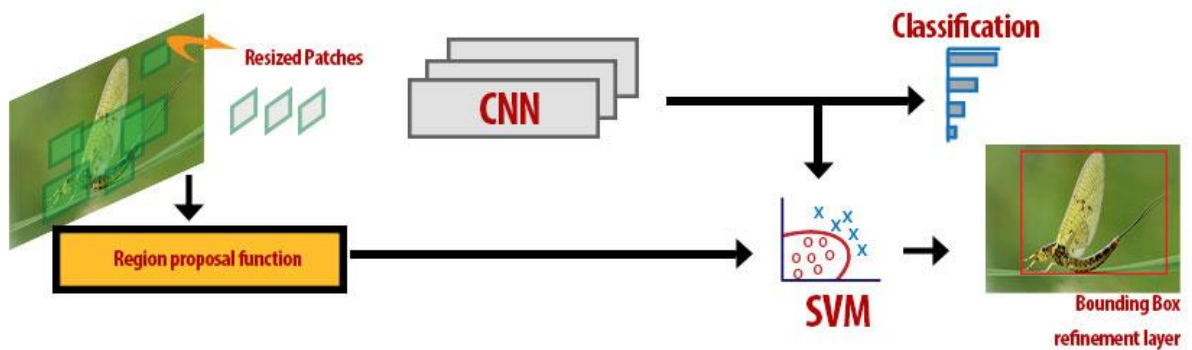
Bu çalışmada Derin Öğrenme Ağ Modellerinden Faster R-CNN modeli tercih edilmiştir.

CNN 2010 yılında geliştirilen resim ve video işleme kullanılan ve zamanla gelişme gösteren oldukça popüler bir derin öğrenme ağ modelidir. Bu model temelde 4 katmandan oluşmaktadır. Bu Katmalar; Convolution Katmanı, Relu Katmanı, Pooling Layer katmanı ve Flattening katmanıdır. Her bir katmanda görüntü üzerinde farklı işlemler gerçekleştirilerek sonuca ulaşılır (Shea, 2015).

R-CNN, CNN mimarisini baz alarak geliştirilen bir derin öğrenme ağ modelidir. Bu modelde ise resim 2000 tane bölge önerilerine bölünerek, her bölgeye sırasıyla CNN uygulanır ve çıkan “features map” ler için bir sınıflandırma algoritması olan Support Vector Machine (SVM) uygulanır.

Objenin varlık ve yokluğu belirlenerek linear regression modelleme yapılarak bölgelerin boyutu belirlenir ve doğru belirlenen bölge yapay sinir ağına sokulur. En büyük dezavantajı yavaş çalışmasıdır(Girshick vd., 2016) .

SSD gerçek zamanlı olarak obje tespiti yapmak için tasarlanmış bir modeldir. SSD, bölge teklif ağının ihtiyacını ortadan kaldırarak süreci hızlandırır. Faster R-CNN'deki Düşüşün doğruluğunu düzeltmek için SSD, çoklu ölçek özellikleri ve varsayılan kutular dahil olmak üzere birkaç geliştirme uygular. Bu gelişmeler, SSD'nin daha düşük çözünürlüklü görüntüler kullanarak daha hızlı R-CNN'in doğruluğunu eşleştirmesini sağlar ve bu da hızı daha da yükseltir. SSD'nin düşük çözünürlüklü görüntüler kullanması ve hızlı olması nedeniyle mobil uygulamalarda kullanılabilmesine olanak sağlamıştır(Liu, 2015).

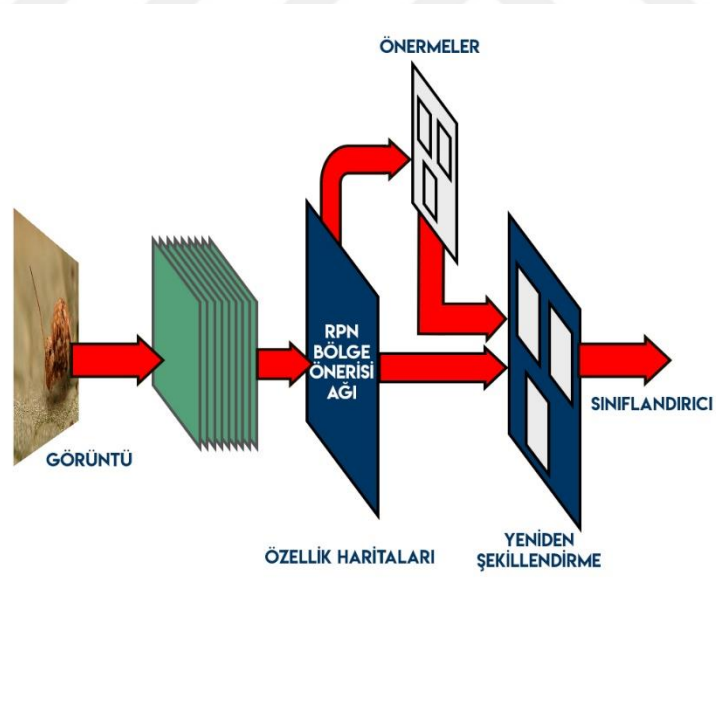


Şekil 3. 2. Fast R-CNN Yapısı.

R-CNN'den farklı olarak resme önce CNN uygulayıp oluşan COVN5 özellik haritası üzerinde bölge önerilerine ayırmasıdır. Sınıflama metodu olarak ise SVM algoritması yerine Softmax Classification kullanılmaktadır. Bu yöntemler sayesinde eğitim süresi yaklaşık %90 oranında azalmıştır(Girshick, 2015) .

Faster R-CNN

Fast R-CNN ile birçok benzer yanı olmakla birlikte Bölge önerileri kısmında Selective search yerine ayrı bir bölge önerisi ağı oluşturularak alınmaktadır. Fast R-CNN ile RPN (Bölge Önerisi Algoritması) modellerinin birleştirilerek oluşturulmuş bir yapay sinir ağı algoritmasıdır. Bu algoritmada giriş resmi konveksiyonel sinir ağlarından geçirilir ve özellik haritası çıkarılır. Bu aşamadan sonra RPN oluşturulur ve bölge önerileri bu ağ üzerinden yapılır. Belirlenen Ağ bölgeleri tekrar şekillendirildikten sonra tam bağlı katmanlarından geçirilir ve sınıflandırma işlemi gerçekleştirilir. Bu sayede daha hızlı bir tahmin süresi elde edilmektedir. Aşağıdaki görselde bu süreç şekillendirilmiştir (Ren vd., 2016) .



Şekil 3.3. Faster R-CNN Mimarisi.

3.2.5. Çalışma ortamının hazırlanması

Derin öğrenme gerçekleştirebilmek için öncelikle uygun bir çalışma ortamı oluşturmak gerekmektedir. Windows İşletim sisteminde Anaconda programı sayesinde bu ortam

oluşturulabilmekte ve gerekli kütüphaneler kurulabilmektedir (Anaconda 2019). Eğer ekran kartı ile çalışılacak ise Ekran kartı sürücüsü ve CUDA Geliştirme Kiti kurulur. Anaconda ve CUDA kurulumları gerçekleştirildikten sonra veri eğitimi ve sınıflandırma için gerekli kütüphanelerin eklenmesine olanak sağlayan bir ortamın oluşturulması gerekmektedir. Bu ortam herhangi bir isimle oluşturulabilir. Ardından bu ortamın içerisine veri eğitimi için gerekli kütüphanelerin kurulumları gerçekleştirilir. (Tensorflow-gpu, Pillow, Jupyter, OpenCV, Panda, Matplotlib). Tensorflow Object Detection klasörü içerisinde nesne tanıma için gerekli Python dosyaları oluşturulur. Tüm ayarlamalar yapıldıktan sonra verilen eğitilmesi gerçekleştirilir.



Şekil 3.4. Çalışma ortamı hazırlama aşamaları.

Çalışma ortamı hazırlanırken bir çok sorunla karşılaşılmıştır. Bu sorunlardan en önemlisi ekran kartının CUDA sürümünden dolayı Tensorflow GPU'nun kurulamaması ve veri eğitimine başlanılmaması; Tensorflow GPU sürümü donanımsal olarak CUDA özelliğine sahip bir ekran kartı gereksinimi duymaktadır. Her Tensorflow sürümünün de minimum CUDA sürümü ve Cudnn kütüphane gereksinimleri vardır. Bununla beraber her CUDA sürümünün de gerekli minimum ekran kartı Compute Capacity(CC) değeri vardır. Yani sadece CUDA desteği olan düşük CC değerine sahip bir ekran kartına CUDA'nın son sürümü ve Tensorflow GPU'nun son

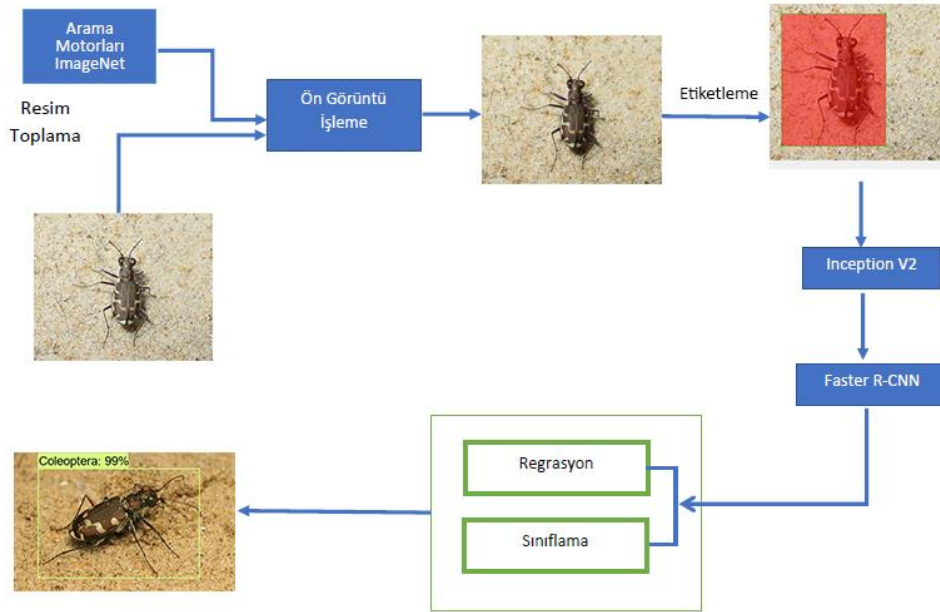
sürümü kurulduğunda hata oluşmakta, eğitim başlatılamamaktadır. Bu da ciddi bir zaman ve emek kaybına neden olmaktadır.

Bir diğer sorun ise Tensorflow GPU sürümlerinin farklılarından kaynaklı olarak Tensorflow Object Detection kütüphanesindeki dosyalardaki kodlardan kaynaklı hatalardır. Bu hatalar da tek tek düzeltilerek veri eğitimine başarıyla geçilmiştir.



4. UYGULAMA

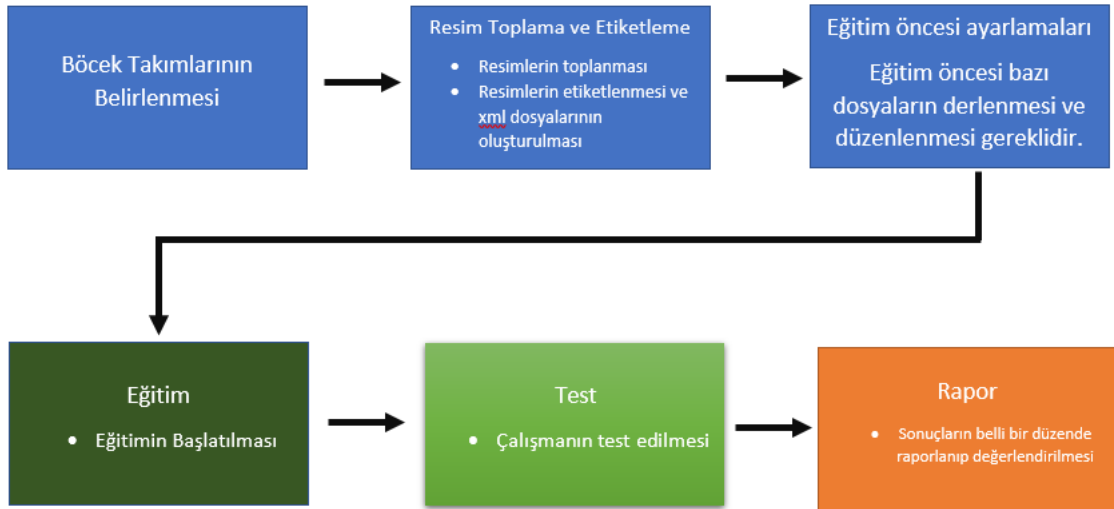
Uygulama sürecinde öncelikle materyal olarak belirlenen böcek takımları ile ilgili araştırma çalışmaları yapılmıştır. Böcek Bilimi ile ilgili çeşitli kaynaklardan bilgi edinilmiştir. Belirlenen araştırma sorularına göre süreç tasarlanmıştır. Uygulama süreci Şekil 4.1 de gösterilmektedir. Arama motorları ve ImageNet gibi veritabanlarından toplanan resimler Işık Dengesi ayarlama, Görüntü Boyutu düzenleme gibi Ön Görüntü İşleme tabi tutulmuştur. Eğitim için uygun hale getirilen resimler yardımcı bir yazılım ile etiketlendikten sonra Faster R-CNN Inception V2 derin öğrenme algoritmasıyla eğitime sokulmuştur. Yaklaşık 50000 adımdan sonra Yazılım sınıflandırma yapmaya uygun hale gelmiştir. Bu aşama aynı zamanda yapılan sınıflama hatalarının görüldüğü kısım olduğundan yapılan hatalı sınıflamalara göre tekrar en baştan itibaren tüm adımlar izlenir ve sonuca göre bu işlemler tekrarlanabilir. Eğer sınıflama sonuçları istenilen düzeye ulaşırsa tekrar eğitime sokulmaz ve yazılım sınıflama yapmaya hazır hale gelir.



Şekil 4.1. Uygulama Süreci.

Uygulama aşamasında gerçekleştirilen tüm adımlar, alt işlemleri ve sıraları aşağıdaki infografikte belirtilmiştir. Bu şemada gösterildiği üzere ilk aşamada çalışılacak Böcek Takımları belirlenir. Bu aşamadan sonra her böcek takımının resmi toplanır ve bilimsel adlarıyla isimlendirilir. Bu etiketlemeler sonucu XML dosyaları ve JPEG dosyaları eğitim için hazırlanmış olur. Eğitim Öncesinde gerekli ayarlamalar yapıldıktan sonra eğitim başlatılır. Eğitim istenilen

düzeğe geldiğinde sonlandırılır ve çalışma test edilir. Test sonuçları bir tabloya aktarıldıktan sonra eğer istenilen sonuçlar elde edilmediyse sorunlar tespit edilir ve gerekli görülürse tekrar edilir. İstenilen sonuçlar elde edildiğinde sonuçlar rapor haline getirilir.



Şekil 4.2. Uygulama Aşamaları.

4.1. Çalışılacak Böcek Sınıf Seviyesinin Belirlenmesi

Böceklerin Hayvan Familialarında en fazla türe sahip olan familyadır. Yaklaşık 750 Bin tanımlanmış tür mevcuttur. Böceklerde Tür belirlenmesinde çok ufak farklılıklar hatta bazı durumlarda genetiğine bakılarak tespit yapılmaktadır. Tür sayısının fazla olması ve tür tespitindeki zorluklar nedeniyle Böceklerde sadece böcek resminde Tür düzeyinde bir tanımlama yapılamamaktadır. Gerek farklılıkların türe göre daha fazla olması gerek çeşitlilikteki azlık nedeniyle, gerekse entomoloji alanındaki uzman görüşlerine dayanarak sadece görüntü üzerinden takım düzeyinde bir tespit yapılması daha mümkündür. Ayrıca bilimsel çalışmalarda bu alanla ilgili araştırma yapan bilim insanlarının böceğin tanımlanmasındaki ilk adımları böceğin ait olduğu takımının belirlenmesi aşamasıdır.

Çalışılacak sınıf düzeyi belirlendikten sonra her takıma ait özellikler bu alanla ilgili kaynaklardan araştırılmıştır.

4.2. Resim Toplama

Geliştirilen sınıflandırma sisteminin en önemli aşaması Belirlenen böcek takımlarının resimlerinin bulanması aşamasıdır. Çünkü elde edilen resimler böcek sınıflandırma algoritmasının eğitiminde kullanılacak, algoritma her bir resim ile defalarca eğitilecektir. Yanlış eğitilen bir takım hem o böcek takımının sınıflandırılmasında hem de diğer böcek takımlarının sınıflandırılmasında hata olmasına neden olacaktır. Bu yüzden böcek takımları için toplanan resimlerin o böcek takımına ait olması, resmin böcek takımının karakteristik özelliğini yansıtması ve mümkün olduğu kadar fazla açıdan resim olması büyük önem arz etmektedir.

Böcek takımının resimleri taranırken arama motorlarında böceğin bilimsel adı ve günlük yaşamda bilinen isimleri ile arama gerçekleştirilmiştir. Hymenoptera, Blattodea, Diptera, Mantodea gibi böcek takımları bilimsel adlarıyla çok bilinmeseler de günlük hayatta sık karşılaştığımız türlerdir.(Sırasıyla birer örnek vermek gerekirse; Arı, Hamam Böceği,Sinek, Peygamber Devesi) Bu takımların resimlerinin bulunmasında ve doğru tür olduğunu teyit etmede herhangi bir zorluk yaşanmadı. Fakat Zoraptera, Mantophasmotodea, Gylloblatodea gibi böcek takımlarının resimlerinin elde edilmesinde zorluk yaşanmıştır. Çünkü bu türler nadir bulunan ve fotoğraflanan türlerdir. Bu yüzden gerek arama motorlarında gerekse böcek bilimiyle ilgilenen internet sitelerinde bu türlerin çok fotoğrafları bulunmamaktadır. Bu olumsuzluklara rağmen mümkün olduğu kadar fazla resim eğitimde kullanılması için toplanmıştır.

Böcek takımlarının karakteristik özelliklerini bilmek ve buna göre araştırma yapmak resim toplama aşamasının en önemli adımıdır. Zira bir böcek takımına başka bir böcek takımının resmi konup tanımlanırsa; bu hem böcek takımının sınıflandırılmasında hem de diğer böcek takımlarının sınıflandırılmasında hata yapılmasına neden olacaktır. Bunun önüne geçmek için toplanan resimler birkaç defa tekrar gözden geçirilip yanlış resimler çıkarılmıştır. Buna rağmen eğitim gerçekleştirildikten sonra bile hatalı sonuçlar olduğunda bu aşamaya tekrar dönülüp hatalı resimler ayıklanmış ve tekrar eğitime sokulmuştur.

Toplanan resimlerin böcek takımının karakteristik özelliklerini yansıtması gerekmektedir. Eğer resim bulanıksa, böcek belirgin değilse, böceğin karakteristik özellikleri yansıtmıyorsa o resmi eğitimde kullanmamızın hiçbir yararı olmayacaktır.

Resimlerin boyutlarının çok düşük ve çok yüksek çözünürlükte olmaması gerekmektedir. Çok düşük çözünürlüklü resimlerde böceğin ayırt edici özellikleri kaybolur, çözünürlük düşük olduğundan hata oranı yüksek olabilir. Resimler Çok yüksek çözünürlükte olduğunda ise eğitim süresinin uzamasına neden olmaktadır. Resimlerin 740x740 piksel çözünürlükten fazla, 250x250

piksel çözünürlükten az olmasına özen gösterilmiştir. Yüksek çözünürlükteki resimler yardımcı yazılımlar vasıtasıyla istenilen çözünürlüğe indirgenmiş, düşük çözünürlükteki resimler ise elenmiştir.

Çizelge 4.1. Böcek Takımlarına göre eğitimde kullanılan resim sayısı.

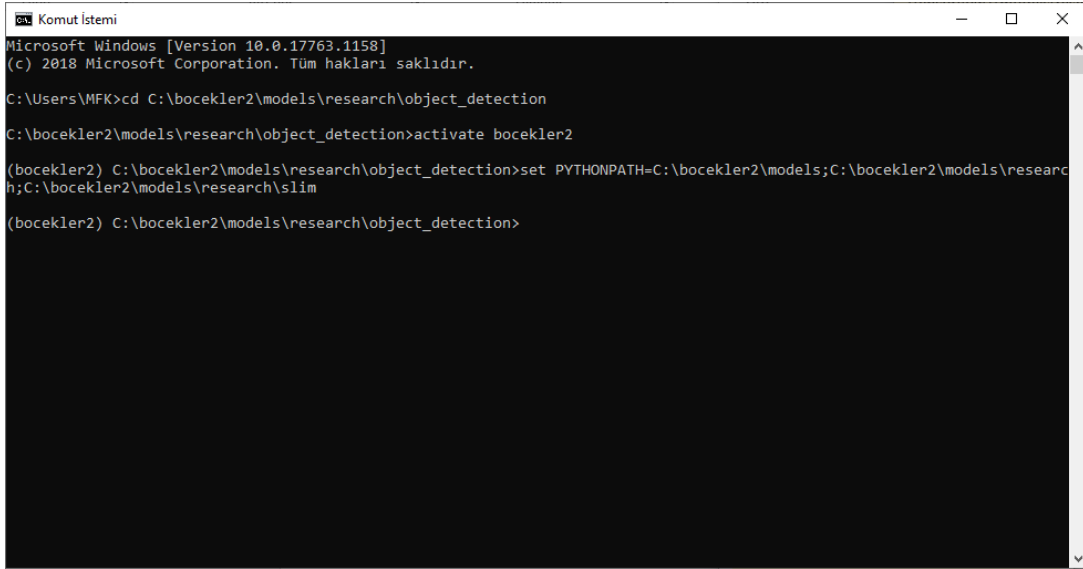
No	Takım Adı	Resim Sayısı	No	Takım Adı	Resim Sayısı
1	Archaeognatha	74	17	Mecoptera	102
2	Blattodea	99	18	Megaloptera	99
3	Coleoptera	124	19	Neuroptera	101
4	Collembola	96	20	Odonata	105
5	Dermaptera	100	21	Orthoptera	104
6	Diplura	100	22	Phasmatodea	95
7	Diptera	100	23	Phthiraptera	90
8	Embioptera	60	24	Plecoptera	95
9	Ephemeroptera	100	25	Protura	101
10	Grylloblatodea	96	26	Psocoptera	95
11	Hemiptera	100	27	Siphonaptera	68
12	Hymenoptera	319	28	Strepsiptera	97
13	Isoptera	125	29	Thysanoptera	85
14	Lepidoptera	137	30	Thysanura	81
15	Mantodea	102	31	Trichoptera	92
16	Mantophasmatodea	60	32	Zoreptera	21
				Toplam	3246

Yukarıdaki tabloda Böcek Takımlarının eğitim sırasında kullanılan görüntü sayısı verilmiştir. 32 Böcek Takımı için toplamda 3246 böcek fotoğrafı kullanılmıştır. Alt sınıfları fazla olan Böcek Takımlarında daha fazla görüntü kullanılırken, bazı nadir takımlarda ise az görüntü kullanılmak zorunda kalmıştır. Zoreptera gibi bazı böcek takımları sadece belirli bir bölgede yaşadığından fotoğraflarının elde edilmesi zordur.

Resimlerin etiketlenmesi adımıyla toplanan resimlerdeki böcekler bir yardımcı yazılım vasıtasıyla bilimsel isimleriyle etiketlenir. Resimdeki böceğin boydan kuyruktan antene kadar , enden ise varsa kanatların sonuna kadar dikdörtgen şeklinde etiketlenir. Eğer böceğin herhangi bir kısmı bu alana dahil edilmezse eğitim hatalı gerçekleşebilir.

4.2. Veri eğitimi öncesi ayarlamalar

Eğitim gerçekleştirilmeden önce bunun için gerekli ortamın oluşturulması, gerekli araçların kurulması ve gerekli ayarlamaların yapılması gerekmektedir. Tensorflow yapay zeka kütüphanesinin çalışabilmesi için İşletim sisteminde Python'un kurulu olması gerekmektedir. Tensorflow kütüphanesini için Windows işletim sisteminde Python için Anaconda yazılımı tercih edilmiştir. Tensorflow GPU 1.13.2 sürümü çalışma ortamına kurularak , opencv, pillow, matlob gibi ek kütüphaneler de kurulmuştur. Gerekli tüm kütüphaneler kurulduktan sonra gerekli derleme ve yüklemeler yapılmıştır. Bu ayarlamalardan sonra eğitim başlatılmıştır.



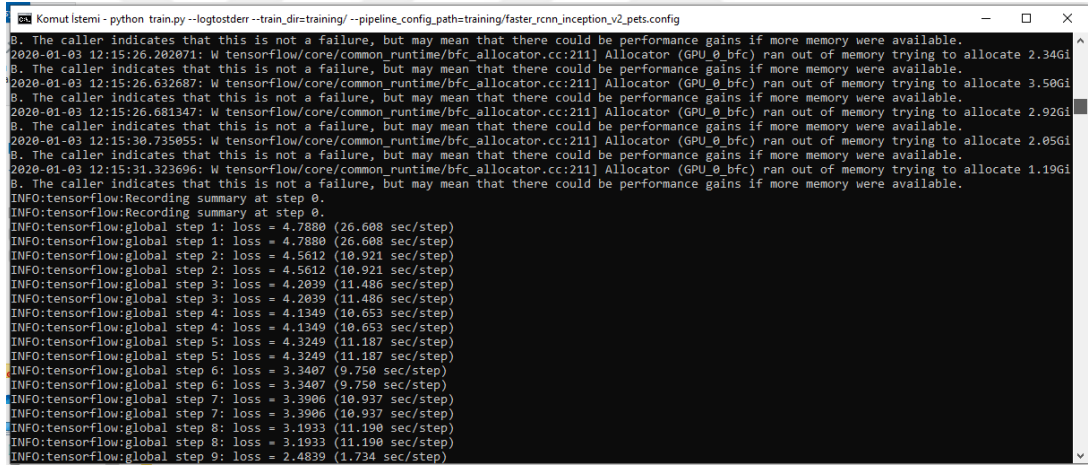
```
Komut İstemi
Microsoft Windows [Version 10.0.17763.1158]
(c) 2018 Microsoft Corporation. Tüm hakları saklıdır.

C:\Users\MFK>cd C:\bocekler2\models\research\object_detection
C:\bocekler2\models\research\object_detection>activate bocekler2
(bocekler2) C:\bocekler2\models\research\object_detection>set PYTHONPATH=C:\bocekler2\models;C:\bocekler2\models\research\h;C:\bocekler2\models\research\slim
(bocekler2) C:\bocekler2\models\research\object_detection>
```

Şekil 4.3. Eğitim Öncesi Ayarlamalar.

4.3. Verinin eğitilmesi

Eğitim öncesi ayarlamalar düzgün bir şekilde yapıldığı takdirde eğitim sorunsuz bir şekilde başlayacaktır. Eğitim süresi, kullanılan bilgisayarın hızı, sınıf sayısı ve resim sayısına göre değişkenlik göstermektedir. Sınıf ve resim sayısı sabit tutulduğunda bilgisayarın ekran kartı kalitesi, işlemci hızı ve RAM hızı artarsa eğitim süresi kısalmaktadır. Sınıflandırmanın düzgün gerçekleşmesi için eğitimdeki adım sayısının optimum düzeyde olmasıyla doğru orantılıdır. Eğitimdeki her adımdan sonra bir loss değeri oluşur. Bu loss değeri tahmin değerinin gerçek değerden ne kadar uzak olduğunu hesaplar ve başlangıçta yüksek iken zamanla düşer. Eğer bu loss değeri hiç düşmüyor hatta giderek yükseliyorsa eğitim sonlandırılır. Eğitim adımındaki optimum değer işte bu noktadır. Eğitim sırasındaki ekran görüntüsünde görüldüğü üzere başlangıçta 4.7 olan olan loss değeri 9. Adımda 2.4'e düşmüştür. (Resim 3.3). Bu değer adım sayısı arttıkça daha da düşmüştür.



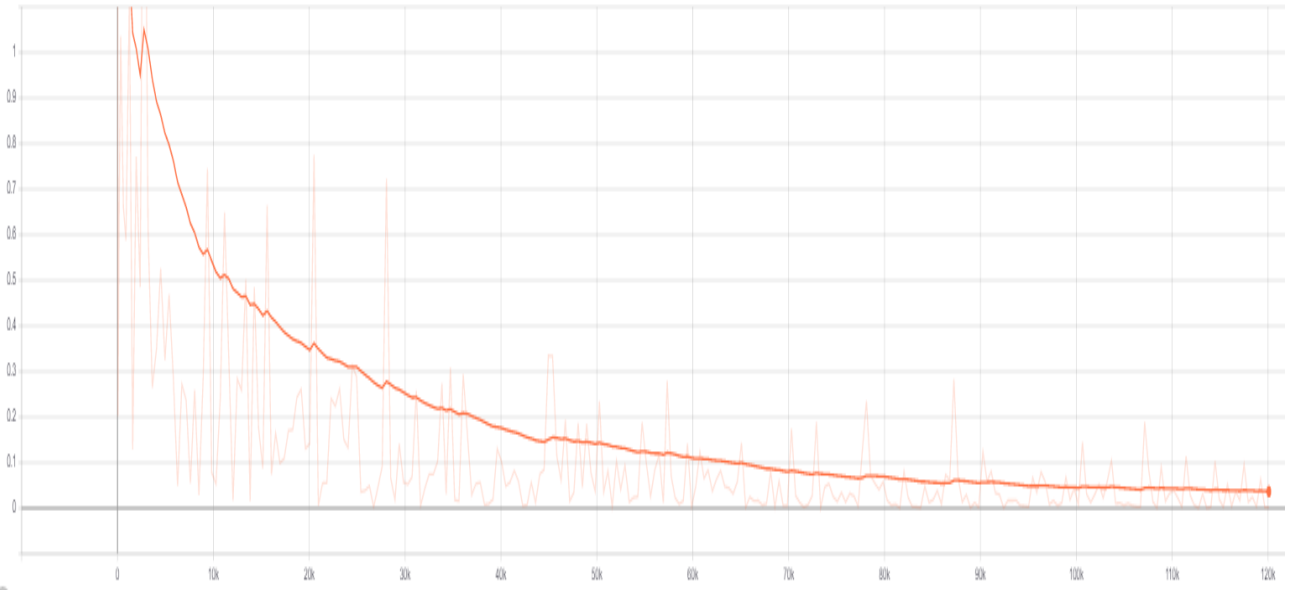
```

Komut İstemi - python train.py --logtostderr --train_dir=training/ --pipeline_config_path=training/faster_rcnn_inception_v2_pets.config
B. The caller indicates that this is not a failure, but may mean that there could be performance gains if more memory were available.
2020-01-03 12:15:26.282071: W tensorflow/core/common_runtime/bfc_allocator.cc:211] Allocator (GPU_0_bfc) ran out of memory trying to allocate 2.34GiB. The caller indicates that this is not a failure, but may mean that there could be performance gains if more memory were available.
2020-01-03 12:15:26.632687: W tensorflow/core/common_runtime/bfc_allocator.cc:211] Allocator (GPU_0_bfc) ran out of memory trying to allocate 3.50GiB. The caller indicates that this is not a failure, but may mean that there could be performance gains if more memory were available.
2020-01-03 12:15:26.681347: W tensorflow/core/common_runtime/bfc_allocator.cc:211] Allocator (GPU_0_bfc) ran out of memory trying to allocate 2.92GiB. The caller indicates that this is not a failure, but may mean that there could be performance gains if more memory were available.
2020-01-03 12:15:30.735055: W tensorflow/core/common_runtime/bfc_allocator.cc:211] Allocator (GPU_0_bfc) ran out of memory trying to allocate 2.05GiB. The caller indicates that this is not a failure, but may mean that there could be performance gains if more memory were available.
2020-01-03 12:15:31.323696: W tensorflow/core/common_runtime/bfc_allocator.cc:211] Allocator (GPU_0_bfc) ran out of memory trying to allocate 1.19GiB. The caller indicates that this is not a failure, but may mean that there could be performance gains if more memory were available.
INFO:tensorflow:Recording summary at step 0.
INFO:tensorflow:Recording summary at step 0.
INFO:tensorflow:global step 1: loss = 4.7880 (26.608 sec/step)
INFO:tensorflow:global step 1: loss = 4.7880 (26.608 sec/step)
INFO:tensorflow:global step 2: loss = 4.5612 (10.921 sec/step)
INFO:tensorflow:global step 2: loss = 4.5612 (10.921 sec/step)
INFO:tensorflow:global step 3: loss = 4.2839 (11.486 sec/step)
INFO:tensorflow:global step 3: loss = 4.2839 (11.486 sec/step)
INFO:tensorflow:global step 4: loss = 4.1349 (10.653 sec/step)
INFO:tensorflow:global step 4: loss = 4.1349 (10.653 sec/step)
INFO:tensorflow:global step 5: loss = 4.3249 (11.187 sec/step)
INFO:tensorflow:global step 5: loss = 4.3249 (11.187 sec/step)
INFO:tensorflow:global step 6: loss = 3.3407 (9.750 sec/step)
INFO:tensorflow:global step 6: loss = 3.3407 (9.750 sec/step)
INFO:tensorflow:global step 7: loss = 3.3906 (10.937 sec/step)
INFO:tensorflow:global step 7: loss = 3.3906 (10.937 sec/step)
INFO:tensorflow:global step 8: loss = 3.1933 (11.190 sec/step)
INFO:tensorflow:global step 8: loss = 3.1933 (11.190 sec/step)
INFO:tensorflow:global step 9: loss = 2.4839 (1.734 sec/step)

```

Şekil 4.4. Eğitim.

Eğitimin adım sayısı ve loss değerinin ilişkisini gösteren grafik aşağıda gösterilmiştir. Grafikte görüldüğü üzere başlangıçta yüksek olan loss değerleri adım sayısı 120 bine geldiğinde 0,5 değerinin altına inmiş, 120000 adımda ise daha fazla inmediğinden eğitim sonlandırılmıştır.



Şekil 4.5. Adım Sayısı ile Loss değeri ilişkisi.

Eğitim sistem özellikleri aşağıda verilen bilgisayarda gerçekleştirildiğinde yaklaşık 12 saat sürmüştür.

- Intel Core i5 3230M 2,60Ghz Quad-Core İşlemci
- NVIDIA GeForce GTX 1050ti 4 GB Ekran Kartı
- 20 GB DDR4 Ram
- SSD Hard Disk

5. SONUÇ VE ÖNERİLER

Gerçekleştirilen test çalışmaları neticesinde geliştirilen yazılımın birçok böcek türünde Başarılı sonuçlar verdiği söylenebilir. Yapılan testlerde %60 ve üstünde çıkan en yüksek değer sınıflama olarak kabul edilmiş yapılan karşılaştırmalara göre doğru sınıflama ise “Başarılı sınıflama” olarak eklenmiştir.

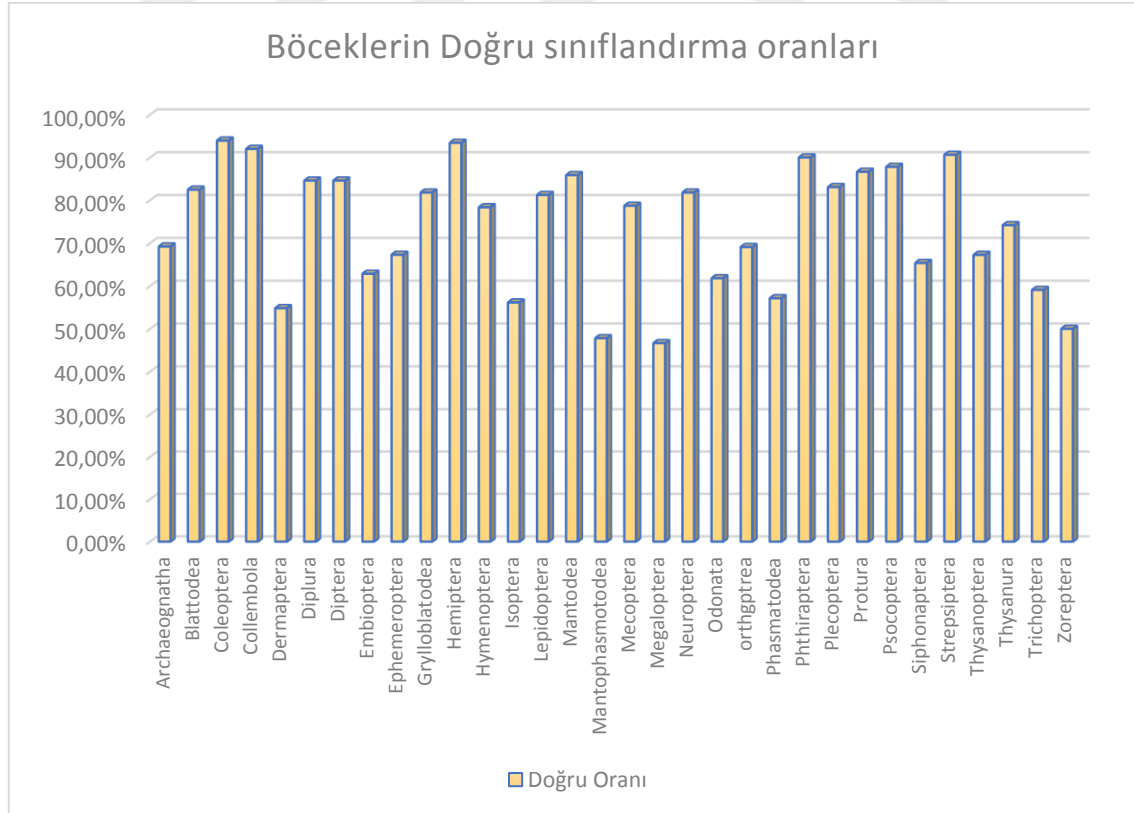
Çizelge 5.1. Böcek Takımlarının Sınıflandırma Sonuçları

Marka	S	DTS	YTS	DO	HO
Archaeognatha	52	36	16	69,2%	30,8%
Blattodea	40	33	7	82,5%	17,5%
Coleoptera	33	31	2	93,9%	6,1%
Collembola	25	23	2	92,0%	8,0%
Dermaptera	62	34	28	54,8%	45,2%
Diplura	39	33	6	84,6%	15,4%
Diptera	52	44	8	84,6%	15,4%
Embioptera	35	22	13	62,9%	37,1%
Ephemeroptera	107	72	35	67,3%	32,7%
Grylloblatodea	22	18	4	81,8%	18,2%
Hemiptera	76	71	5	93,4%	6,6%
Hymenoptera	37	29	8	78,4%	21,6%
Isoptera	73	41	32	56,2%	43,8%
Lepidoptera	64	52	12	81,3%	18,8%
Mantodea	85	73	12	85,9%	14,1%
Mantophasmotodea	23	11	12	47,8%	52,2%
Mecoptera	47	37	10	78,7%	21,3%
Megaloptera	30	14	16	46,7%	53,3%
Neuroptera	33	27	6	81,8%	18,2%
Odonata	55	34	21	61,8%	38,2%
orthgptrea	55	38	17	69,1%	30,9%
Phasmatodea	63	36	27	57,1%	42,9%
Phthiraptera	20	18	2	90,0%	10,0%
Plecoptera	71	59	12	83,1%	16,9%
Protura	30	26	4	86,7%	13,3%
Psocoptera	41	36	5	87,8%	12,2%
Siphonaptera	26	17	9	65,4%	34,6%
Strepsiptera	32	29	3	90,6%	9,4%
Thysanoptera	55	37	18	67,3%	32,7%
Thysanura	31	23	8	74,2%	25,8%
Trichoptera	44	26	18	59,1%	40,9%
Zoreptera	42	21	21	50,0%	50,0%
GENEL	1500	1101	399	73,4%	26,6%

Yukarıdaki tabloda her böcek takımı için sınanan resim sayısı(S), doğru tahmin sayısı (DTS), Yanlış Tahmin Sayısı (YTS), Doğru Oranı(DO), Hata Oranı(HO) gösterilmiştir. Bu tablo incelendiğinde doğru oranı en yüksek böcek takımı %93.9 ile Coleoptera'dır. Bu takımdan sonra %93.4 ile Hemiptera ve %92 ile Collembola takımları gelmektedir. Bu böcek takımlarının başarı oranlarının yüksek olması ayırt edici özelliklerinin diğer türlere göre daha fazla olmasından kaynaklanabilir. Günlük hayatta daha sık karşılaşılan böcek takımlarının doğru oranları da %60'ın üzerindedir.

Hata Oranının en fazla olduğu böcek takımı %47.8'lik Doğru Oranı ile Mantophasmotodea böcek takımıdır. Bu böcek takımının orthoptera böcek takımı ile karakteristik özellikleri bir çok açıdan benzerdir. Sınanan resimler incelendiğinde bu benzerlikten dolayı sistemin bu iki takımı ayırt etmede güçlük yaşadığı görülmektedir. Bununla beraber Mantophasmotodea böcek takımı nadir bir takım olmasından dolayı eğitimde kullanılabilir yeterli resim bulunmamaktadır. Bu da modelin Mantophasmotodea böcek takımı üzerinde yeterince eğitilememesine neden olmuş olabilir.

Aşağıdaki grafikte Böcek Takımlarının Doğru sınıflandırılma Oranları gösterilmiştir.



Şekil 5.1. Böceklerin Doğru Sınıflandırılma Oranları.

Test işlemleri Intel i5 3230M 2.60 GHZ işlemci, 8 GB Ram, 2 GB Dahili Ekran kartına sahip Bilgisayar üzerinde gerçekleştirilmiştir. Sınıflama süreleri yaklaşık 1 ila 3 sn arasında sürmektedir.

Geliştirilen yazılım bazı böcek takımlarını benzer özelliklere sahip yakın böcek takımları ile karıştırabilmektedir. Örneğin arı ve karıncanın ait olduğu hymenoptera takıma ait bir arı fotoğrafı yazılıma yüklediğinde Sineğin bağlı olduğu takım olan Diptera ile karıştırmaktadır.



Yukarıdaki çizelgede yazılımın hangi böcek takımını tespit ettiğine ait konfüsyon matrisi verilmiştir. Konfüsyon matrisi makine öğreniminde kullanılan ölçümlerden biridir. Böcek takımlarının tespitinde hangi böcek takımında daha başarılı olduğu bu çizelgede daha net görülmektedir. Yeşil renkte ifade edilen sayılar başarılı tespiti, kırmızı ile ifade edilenler ise başarısız tespitlerdir. Yanlış tespitler incelendiğinde genellikle fiziksel özellikleri benzer takımlar olduğu görülmektedir.

Derin Öğrenme alanındaki en yaygın performans ölçümleri testin doğruluğu, hassasiyet(precision), hatırlama(recall) ve F1 skorudur. (Goutte ve Gaussier, 2005) . Bu değerleri elde etmek için TP, TN, FP ve FN değerlerini konfüsyon matrisinden çıkarmak gerekmektedir. Bir sınıfın; TP değeri Doğru Tahmin sayısını, FN değeri yanlış tahmin sayısı, FP diğer sınıflarda çıkan o sınıfın sayısını, TN değeri ise o sınıfın ilgili olmadığı toplam sayıyı ifade etmektedir. Bu değerler elde edildikten sonra her sınıf ¹(yani böcek takımı) için Test Doğruluğu(Testing Accuracy) , Hassasiyet(Precision), Hatırlama(Recall) ve F1 skorları hesaplanır. Hesaplama için kullanılan formüller aşağıda verilmiştir. En sonunda elde edilen F1 Skoru (4) çalışmanın başarısını ölçmektedir.

$$\text{Test Doğruluğu(Testing Accuracy)} = \frac{(TN + TP)}{(TN + TP + FN + FP)} \quad (1)$$

$$\text{Precision (Hassasiyet)} = \frac{TP}{(TP + FP)} \quad (2)$$

$$\text{Recall (Hatırlama)} = \frac{TP}{(TP + FN)} \quad (3)$$

$$\text{F1 Skor} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Aşağıdaki tabloda her sınıf (böcek takımı) için hesaplanan TP, FN, FP, TN, Test Doğruluğu, Precision, Recall ve F1 Skor değerleri görülmektedir. Test Doğruluğu, Precision, Recall ve F1 skorlarında en yüksek yüzde değerlerinin tabloda altları çizilmiştir. Test Doğruluğunda en yüksek sınıf %99,76 değeri ile Blattodea iken, Precision da Blattodea, Dermaptera ve Zoreptera %100 ile en yüksek değere sahip sınıflardır. Recall değerinde ise %93,94 ile Coleoptera sınıfı en yüksek değere sahip iken, F1 skorunda ise %90,41 değerine sahip

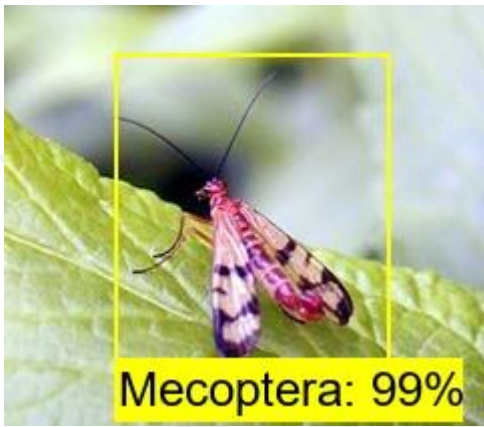
¹ Buradaki sınıf, Sınıflandırma algoritmalarında kullanılan her bir etikettir. Böcek taksonomisindeki sınıf değildir. Dolayısıyla bu çalışmadaki sınıflandırma sınıfları böcek takımlarıdır. Diğer çalışmalarda bu sınıflandırma sınıfı, böcek türü, araba markası, hayvan ismi vb olabilmektedir.

Hemiptera sınıfı en yüksek değere sahiptir. Tablo incelendiğinde tüm sınıflarda Test doğruluğu %97'den az değil ve ortalama olarak da %98,92 dir. Precision değerinde ortalama yüzdesi %78,15 iken Recall değerinde ortalama %73,90 olmuştur. F1 skor ortalamasına bakıldığında %74,46 olduğu görülmektedir.

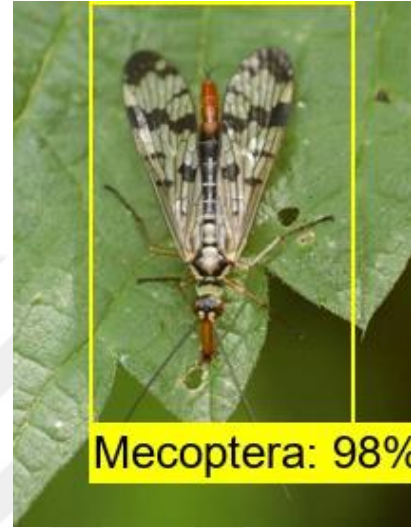
Çizelge 5.3. Böcek Takımlarına göre TP, TN, FP, FN, Test Doğruluğu, Precision, Recall ve F1 Skor değerler.

	TP	FN	FP	TN	Test Doğruluğu (%)	Precision (%)	Recall (%)	F1 Skor (%)
Archaeognatha	36	16	6	1442	98,53	85,71	69,23	76,60
Blattodea	33	7	0	2908	99,76	100,00	82,50	90,41
Coleoptera	31	2	6	2869	99,72	83,78	93,94	88,57
Collembola	23	2	12	2838	99,51	65,71	92,00	76,67
Dermaptera	34	28	0	2788	99,02	100,00	54,84	70,83
Diplura	33	6	21	2728	99,03	61,11	84,62	70,97
Diptera	44	8	12	2685	99,27	78,57	84,62	81,48
Embioptera	22	13	9	2653	99,18	70,97	62,86	66,67
Ephemeroptera	72	35	11	2544	98,27	86,75	67,29	75,79
Grylloblatodea	18	4	5	2528	99,65	78,26	81,82	80,00
Hemiptera	71	5	10	2447	99,41	87,65	93,42	90,45
Hymenoptera	29	8	48	2372	97,72	37,66	78,38	50,88
Isoptera	40	33	2	2345	98,55	95,24	54,79	69,57
Lepidoptera	52	12	4	2279	99,32	92,86	81,25	86,67
Mantodea	73	12	24	2174	98,42	75,26	85,88	80,22
Mantophasmotodea	11	12	5	2170	99,23	68,75	47,83	56,41
Mecoptera	37	10	11	2117	99,03	77,08	78,72	77,89
Megaloptera	14	16	8	2090	98,87	63,64	46,67	53,85
Neuroptera	27	6	9	2056	99,29	75,00	81,82	78,26
Odonata	34	21	15	1995	98,26	69,39	61,82	65,38
Orthoptera	38	17	2	1953	99,05	95,00	69,09	80,00
Phasmatodea	36	27	29	1863	97,14	55,38	57,14	56,25
Phthiraptera	18	2	9	1863	99,42	66,67	90,00	76,60
Plecoptera	59	12	28	1773	97,86	67,82	83,10	74,68
Protura	26	4	5	1766	99,50	83,87	86,67	85,25
Psocoptera	36	5	12	1718	99,04	75,00	87,80	80,90
Siphonaptera	17	9	2	1702	99,36	89,47	65,38	75,56
Strepsiptera	29	3	7	1665	99,41	80,56	90,63	85,29
Thysanoptera	37	18	8	1609	98,44	82,22	67,27	74,00
Thysanura	23	8	11	1575	98,82	67,65	74,19	70,77
Trichoptera	26	18	5	1537	98,55	83,87	59,09	69,33
Zoreptera	21	21	0	1500	98,64	100,00	50,00	66,67
ORTALAMA					98,92	78,15	73,90	74,46

Aşağıdaki resimlerde doğru tespit edilen böcek takımları dikdörtgen içine alınarak sınıflandırma sonucu ve tahmin oranı yazılmıştır. Sınamada kullanılan fotoğraflar ile eğitimde kullanılan fotoğraflar farklıdır. Burada önemli olan sınama fotoğrafları ile eğitimde kullanılan fotoğrafların benzer boyutlarda olmasıdır. Bu olduğunda doğru sonuçlar elde edilmektedir.



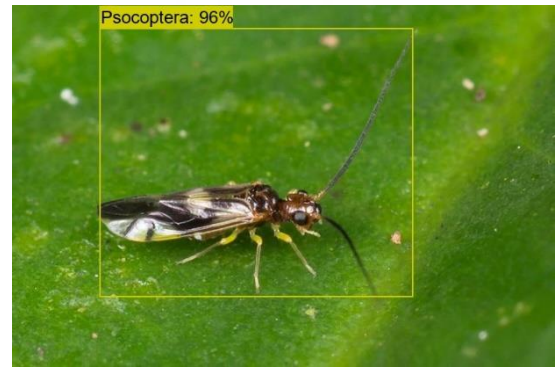
(a)



(b)



(c)



(d)

Şekil 5. 2. Örnek Sonuçlar



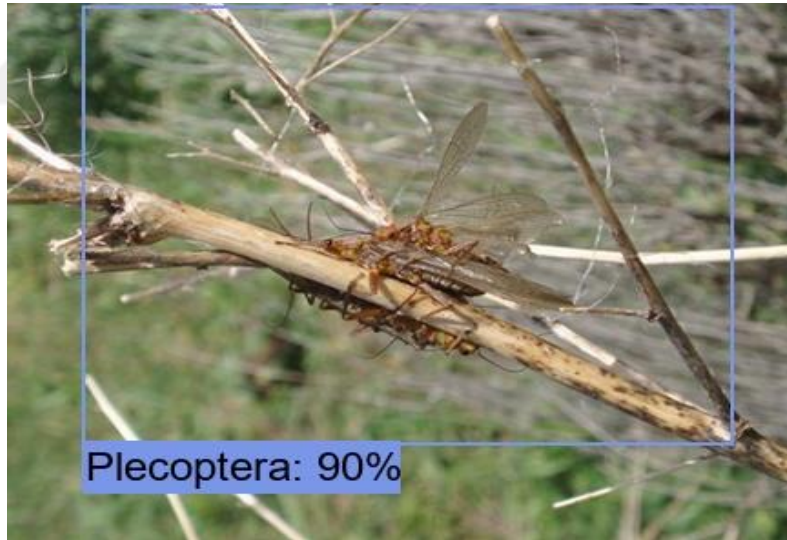
Şekil 5.3. Doğru Tespit Edilen Ephemeroptera Böceği.



Şekil 5.4. Tespit Edilemeyen Ephemeroptera Böceği.



Şekil 5.5. Yanlış Tespit Edilen Ephemeroptera Böceği.



Şekil 5.6. Plecoptera Böceğinin Doğru Tespit Edilmiş Resmi.

Yukarıdaki şekillerde Ephemeroptera ve Plecoptera böcek takımlarına ait resimlerdeki sınıflandırma sonuçları görülmektedir. Bu sonuçlarda Şekil 5.3’de yüzde 99 ile doğru tespit edilen bir Ephemeroptera böcek takımı görülmektedir. Bu resimde Ephemeroptera böcek takımının karakteristik özellikleri tam olarak yansımıştır. Bununla beraber resimde böcek net bir şekilde görünmektedir. Bu sebeple %99 gibi yüksek bir oranla doğru sınıflandırma gerçekleştirilmiştir. Şekil 5.4’de aynı böcek takımına ait bir başka sınıflandırma resmi görülmektedir. Burada böcek diğer resimde olduğu gibi net görünmemektedir. Aslında yazılım çok düşük yüzdelerde de olsa

sınıflandırma yapmaktadır. Ama düşük oranlar sınıflandırma sonucu olarak gösterilmemektedir. Şekil 5.5.'de ise yanlış tespit edilen bir Ephemeroptera böcek takımı görülmektedir. Yazılım Plecoptera böcek takımı ile Ephemeroptera böcek takımını karıştırmıştır. Karıştırdığı türe ait doğru sınıflandırma gerçekleştirilmiş sonuç Resim 5.6'da görülmektedir. Şekil 5.5'de böcek takımının karakteristik özellikleri Plecoptera böcek takımına benzediğinden yazılım yanlış sınıflandırma yapmıştır.

Bu sonuç resimleri, yazılımın hatalı noktalarını, geliştirilmesi gereken yerleri, sınırlılıkları göstermektedir. Bu sonuçlardan da anlaşılacağı üzere Yazılımın sınıflandırma yapabilmesi için böcek fotoğrafının net olması ve böceğin özelliklerini yansıtması gerekmektedir. Yazılımın eksik ve hatalı noktaları tekrar tekrar eğitime sokularak telafi edilebilir ama sınanan fotoğraftaki böcek net bir şekilde görünmüyorsa yazılım başarısız olur.

Sonuç resimleri ve istatistikler geliştirilen yazılımın genel olarak başarılı olduğunu göstermektedir. Yazılım, resimde böcek net bir şekilde görünüyorsa doğru sınıflandırma yapmaktadır. Ayrıca sınıflandırma yaptığı resimdeki oranlar da yüzde 80'in üzerindedir.

Sonuç olarak geliştirilen Faster R-CNN kullanılarak geliştirilen böcek sınıflandırması sistemi verilere de bakıldığında başarılı olarak çalışmaktadır. Bu yapı Entomoloji ile ilgilenen bilim insanları ve bu alanda çalışacak öğrenci ile araştırmacılara yardımcı olacak karar destek yazılımı olarak kullanılabilir. Özellikle nadir bulunan türlerin tespitinde karar destek yazılımı olarak kullanılarak araştırmacıya yardımcı olabilir.

Ayrıca yazılımın kolay kullanımı nedeniyle bilim ile ilgisi olmayan ama tarım gibi alanlarda çalışan ve böcek konusunda bir bilgisi olmayan kişiler bu çalışmada geliştirilen yazılımı kullanarak tarım ürünlerine zarar verebilecek böceklerin tespitini gerçekleştirebilirler. Bu sayede tarımsal verimlilik artar ve ülke ekonomisine doğrudan katkı sağlanır.

5.1. Öneriler

CNN yapılarını kullanarak böcek tespitinde yapmak isteyen araştırmacılar ise böcek takım düzeyini daha aza indirgeyip daha fazla böcek resmi eğitime sokarak daha başarılı sonuçlar elde edebilirler. Bu tezde geliştirilen yazılım özellikle nadir bulunan böcek takımları için daha fazla kaynaktan elde edilecek böcek resmi ile çok daha güçlü bilgisayarlar ile eğitime sokulup geliştirilebilir.

Geliştirilen çalışma bir sunucu bilgisayara kurulup bir web uygulaması ile sınıflandırma işlemi gerçekleştirilebilir. Web Uygulaması ile Akıllı telefonlara yönelik bir uygulama geliştirilmesi kolaylaşabilir.

Tezde sınıflandırma veri eğitimi için kullanılan veri setleri bu alan ile ilgili çalışma yapacak olan araştırmacılara yardımcı olması için Kaggle platformuna veri seti eklenmiştir. Araştırmacılar dilerse bu veri setini kullanabilirler.

Bu çalışma ile böceklerin takım düzeyinde sınıflandırılması gerçekleştirilmiştir. Bu çalışmadaki her takımın alt sınıfları için de yapıldığı takdirde familya düzeyinde sınıflandırma işlemi de gerçekleştirilir. Bu şekilde kademe kademe tür olmasa bile cins tasnifi gerçekleştirilebilir.

KAYNAKLAR DİZİNİ

Anaconda. (2019). Anaconda Open Source Community. *Anaconda*. <https://www.anaconda.com/anaconda-community/> (March 8, 2020).

Batista, G. E., Hao, Y., Keogh, E., ve Mafra-Neto, A. (2011, December). Towards automatic classification on flying insects using inexpensive sensors. In 2011 10th International Conference on Machine Learning and Applications and Workshops (Vol. 1, pp. 364-369). IEEE.

Girshick, R., Donahue, J., Darrell, T., ve Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1), 142-158.

Girshick, R. (2015). Fast R-CNN.

Goutte, Cyril, ve Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In *Lecture Notes in Computer Science*, , 345–59.

Guiam, Angelo, C., ve Bawagan, J. (2017). Insectify : An Android Application For Digital Insect Identification Using A Convolutional Neural Network. : 1–10.

Gullan, P.J., ve Cranston, P.S. (2014). *The Insects : An Outline of Entomology*. 5. Oxford: Wiley Blackwell.

Horn, G., Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Hartwig, A., Perona, P., Belongie, S. (2017). The INaturalist Species Classification and Detection Dataset. 8769–78.

Huynh, H., Lam, D., Van, H., Le, D., Le, L. (2019). CDNN Model for Insect Classification Based on Deep Neural Network Approach. In *Context-Aware Systems and Applications, and Nature of Computation and Communication*, eds. Phan Cong Vinh and Abdur Rakib. Cham: Springer International Publishing, 127–42.

Khalifa, N. E. M., Loey, M., Ve Taha, M. H. N. (2020). Insect pests recognition based on deep transfer learning models. *Journal of Theoretical and Applied Information Technology*, 98(01).

Krizhevsky, A., Sutskever, I., ve Geoffrey, E. H. (2012). ImageNet Classification with Deep Convolutional Neural Networks.

Kunduraci, M.F., ve Örnek, H.K. (2019). Vehicle Brand Detection Using Deep Learning Algorithms. *International Journal of Applied Mathematics Electronics and Computers*.

Lim, J., Jongman, C., Taewoo, N., ve Soohong, K. (2006). Development of a Classification Algorithm for Butterflies and Ladybugs. *TENCON 2006 - 2006 IEEE Region 10 Conference*: 1–3.

Lim, S., Seunghyun, K., Doyeon, K. (2017). Performance Effect Analysis for Insect Classification Using Convolutional Neural Network. (November): 24–26.

KAYNAKLAR DİZİNİ (devam)

Lim, S., Seunghyun, K., Sungwook, P., ve Doyeon, K. (2018). Development of Application for Forest Insect Classification Using CNN. *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*: 1128–31.

Liu, W. (2015). SSD : Single Shot MultiBox Detector SSD : Single Shot MultiBox Detector. (December).

Mayer, J. (2020). Index to the Compendium of Hexapod Classes and Orders. *ENT 425 | General Entomology / Resource Library / Compendium*. <https://projects.ncsu.edu/cals/course/ent425/library/compendium/index.html>.

OpenCV. (2019). OpenCV About <https://opencv.org/about/> (April 2, 2020).

Park, Y. J., Gervase, T., ve Jun, Z. (2019). Insect Classification Using Squeeze-And-Excitation And Attention Modules - A Benchmark Study Yoon Jin Park , Gervase Tuxworth , Jun Zhou. *2019 IEEE International Conference on Image Processing (ICIP)*: 3437–41.

Python. (2020). What Is Python? Executive Summary. <https://www.python.org/doc/essays/blurb/> (April 2, 2020).

Ren, S., He, K., Girshick, R., ve Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

Shea, K. T. (2015). *An Introduction to Convolutional Neural Networks An Introduction to Convolutional Neural Networks*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Tensorflow. (2019). Why TensorFlow. <https://www.tensorflow.org> (December 1, 2019).

Valan, M., Makonyi, K., Maki, A., Vondráček, D., ve Ronquist, F. (2019). Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Systematic biology*, 68(6), 876-895.

Xia, D., Chen, P., Wang, B., Zhang, J., ve Xie, C. (2018). Insect detection and classification based on an improved convolutional neural network. *Sensors*, 18(12), 4169..

Yang, H., Liu, W., Xing, K., Qiao, J., Wang, X., Gao, L., & Shen, Z. (2010, August). Research on insect identification based on pattern recognition technology. In *2010 Sixth International Conference on Natural Computation* (Vol. 2, pp. 545-548). IEEE.

Yazgaç, B. G., & Kırıcı, M. (2017, August). Embedded system application for sunn pest detection. In *2017 6th International Conference on Agro-Geoinformatics* (pp. 1-6). IEEE.

KAYNAKLAR DİZİNİ (devam)

Zhang, H., Quangong, H., ve Wei, D. (2008). The Application of AdaBoost-Neural Network in Storedproduct Insect Classification. : 973–76.

Zhu, L. Q., Ma, M. Y., Zhang, Z., Zhang, P. Y., Wu, W., Wang, D. D., Wang, H. Y. (2017). Hybrid deep learning for automated lepidopteran insect image classification. *Oriental Insects*, 51(2), 79-91.



ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Musa Selman KUNDURACI
Uyruğu : TC
Doğum Yeri ve Tarihi : Meram / 16.08.1990
Telefon : 0546 656 08 16
E-posta : mskunduraci@gmail.com

Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek Lisans	Dumlupınar Üniversitesi / Bilgisayar Mühendisliği	
Lisans	Selçuk Üniversitesi / Bilgisayar Mühendisliği	2015
Lisans	Selçuk Üniversitesi / Bilgisayar Sistemleri Öğretmenliği	2013
Lise	Fatih Teknik Lisesi / Bilgisayar Bölümü	2007

İş Deneyimleri

Yıl	Kurum	Görevi
2015 – Halen	İnegöl Hacı Sevim Yıldız – 1 Mesleki ve Teknik Anadolu Lisesi	Teknik Öğretmen