

Derin Sinir Ağları ile Derinlik İmgelerinden İnsan Sezimi

Didem Özüpek Taş

**YÜKSEK LİSANS TEZİ**

Elektrik Elektronik Mühendisliği Anabilim Dalı

Eylül 2018

Human Detection from Depth Images with Deep Neural Networks

Didem Özüpek Taş

**MASTER OF SCIENCE THESIS**

Department of Electrical Electronics Engineering

Sept 2018

Derin Sinir Ağları ile Derinlik İmgelerinden İnsan Sezimi

Didem Özüpek Taş

Eskişehir Osmangazi Üniversitesi  
Fen Bilimleri Enstitüsü  
Lisansüstü Yönetmeliği Uyarınca  
Elektrik Elektronik Mühendisliği Anabilim Dalı  
Telekomünikasyon - Sinyal İşleme Bilim Dalında  
YÜKSEK LİSANS TEZİ  
Olarak Hazırlanmıştır

Danışman: Dr. Öğr. Üyesi Helin Dutagacı

Eylül 2018

## ONAY

Elektrik Elektronik Mühendisliđi Anabilim Dalı Yüksek Lisans öđrencisi Didem Özüpek Taş'ın YÜKSEK LİSANS tezi olarak hazırladıđı “Derin Sinir Ağları ile Derinlik İmgelerinden İnsan Sezimi” başlıklı bu çalışma, jürimizce lisansüstü yönetmeliđin ilgili maddeleri uyarınca deđerlendirilerek oybirliđi ile kabul edilmiřtir.

**Danışman** : Dr. Öğr. Üyesi Helin Dutađacı

**İkinci Danışman** : -

### **Yüksek Lisans Tez Savunma Jürisi:**

**Üye** : Dr. Öğr. Üyesi Helin Dutađacı

**Üye** : Prof. Dr. Hakan Çevikalp

**Üye** : Dr. Öğr. Üyesi Muammer Akçay

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ..... tarih ve  
..... sayılı kararıyla onaylanmıştır.

Prof. Dr. Hürriyet ERŞAHAN  
Enstitü Müdürü

## ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Dr. Öğr. Üyesi Helin Dutağacı danışmanlığında hazırlamış olduğum “Derin Sinir Ağları ile Derinlik İmgelerinden İnsan Sezimi” başlıklı YÜKSEK LİSANS tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim. 06/09/2018

Didem Özüpek Taş

İmza

## ÖZET

Bu tez çalışmasında, derin sinir ağları kullanarak, iç ortamda elde edilmiş derinlik imgelerinden, her türlü pozisyonda bulunabilen insanların baş-omuz bölgelerinin algılanması hedeflenmiştir. Aday baş-omuz bölgelerini doğrulayacak ya da eleyecek şekilde kurgulanmış iki sınıflı derin mimariler tasarlanmıştır. Bu mimarilerin eğitilmesi ve test edilmesi için Kinect Takip Hassasiyeti Veri Kümesi, RGB-D İnsan Veri Kümesi ve UR Düşme Algılama Veri Kümesi'ndeki derinlik imgelerinden baş-omuz bölgelerini içeren kutucuklar işaretlenip kesilmiştir. Aynı şekilde, baş-omuz bölgesine benzeyen negatif kutucuklar da bu derinlik imgelerinden elde edilmiştir. Pozitif ve negatif örnekleri doğru bir şekilde sınıflandırmak üzere çeşitli derin sinir ağı mimarileri tasarlanmıştır. Sinir ağlarının eğitiminde katman arttırımının ve sinir ağlarına özel parametrelerin değişiminin sınıflandırma başarımı üzerindeki etkileri incelenmiştir.

**Anahtar Kelimeler :** Yapay sinir ağları, Görüntü işleme, Derinlik imgeleri, İnsan sezimi, Baş omuz bölgesi, Derin öğrenme

## SUMMARY

The objective of this work is to detect humans in arbitrary poses from depth images acquired in indoor environment through recognition of head-shoulder regions with deep neural networks. In order to train and test deep neural networks, rectangular boxes containing head-shoulder regions were extracted from depth images in Kinect Tracking Precision Dataset, UR Fall Detection Dataset and RGB-D People Dataset. Rectangular regions containing negative examples resembling head-shoulder shape were also extracted. Several network architectures were designed to correctly classify positive and negative examples. Effects of increasing number of layers in neural networks and effects of changing special parameters on the classification performance were examined.

**Keywords:** Artificial neural networks, Image processing, Depth images, Human detection, Head shoulder region, Deep learning

## TEŞEKKÜR

Bu çalışmanın konusunun belirlenmesinde ve gerçekleştirilmesinde yardımlarını esirgemeyen, değerli bilgilerini benimle paylaşan, kendisine ne zaman danışmak istesem bana vaktini ayıran ve çözümler sunan danışman hocam Sayın Dr. Öğr. Üyesi Helin DUTAĞACI'na, birlikte yapılan toplantı anlarında bilgileriyle ve fikirleriyle bana yardımcı olan Sayın Arş. Gör. Kaya TURGUT'a, bu çalışmada bana merak etmeyi, araştırmayı ve gerçekleştirmeyi öğreten bütün üniversite hocalarıma teşekkür ederim.

Ayrıca bu süreçte bana destek olan ve sabır gösteren eşime, ablama, anneme, babama ve arkadaşlarıma teşekkür ederim.



## İÇİNDEKİLER

### Sayfa

|   |           |
|---|-----------|
| ÖZET .....  | vi        |
| SUMMARY .....                                       | vii       |
| TEŞEKKÜR .....                                      | viii      |
| İÇİNDEKİLER .....                                   | ix        |
| ŞEKİLLER DİZİNİ .....                               | xi        |
| ÇİZELGELER DİZİNİ .....                             | xii       |
| SİMGELER VE KISALTMALAR DİZİNİ .....                | xiii      |
| <b>1. GİRİŞ VE AMAÇ .....</b>                       | <b>1</b>  |
| <b>2. LİTERATÜR ARAŞTIRMASI .....</b>               | <b>3</b>  |
| <b>3. MATERYAL VE YÖNTEM .....</b>                  | <b>8</b>  |
| 3.1. Yapay Sinir Ağları.....                        | 8         |
| 3.2. Aktivasyon Fonksiyonları .....                 | 10        |
| 3.3. Kayıp Fonksiyonları .....                      | 12        |
| 3.4. Yapay Sinir Ağlarının Eğitilmesi .....         | 13        |
| 3.5. Tezin Gerçeklenmesi .....                      | 14        |
| 3.5.1. Veri kümelerinin hazırlanması .....          | 15        |
| 3.5.1.1. RGB-D insan veri kümesi .....              | 16        |
| 3.5.1.2. UR düşme algılama veri kümesi .....        | 17        |
| 3.5.1.3. Kinect takip hassasiyeti veri kümesi ..... | 18        |
| 3.5.1.4. Watch-n-patch veri kümesi .....            | 18        |
| 3.5.2. Yapay sinir ağının oluşturulması .....       | 19        |
| <b>4. BULGULAR VE TARTIŞMA .....</b>                | <b>22</b> |
| 4.1. İki Katmanlı Eğitim.....                       | 23        |

**İÇİNDEKİLER (devam)**

|  | <b><u>Sayfa</u></b> |
|--|---------------------|
| 4.2. Üç Katmanlı Eğitim.....                                       | 26                  |
| 4.3. Dört Katmanlı Eğitim.....                                     | 30                  |
| 4.4. Beş Katmanlı Eğitim .....                                     | 33                  |
| 4.5. Üç Veri Kümesinin Birleştirilerek Eğitimde Kullanılması ..... | 37                  |
| <b>5. SONUÇ VE ÖNERİLER .....</b>                                  | <b>45</b>           |
| <b>KAYNAKLAR DİZİNİ .....</b>                                      | <b>46</b>           |

## ŞEKİLLER DİZİNİ

| <u>Sekil</u>  | <u>Sayfa</u> |
|---|--------------|
| 3.1. Biyolojik ve bilgisayar ortamında sinir hücresi (Karpathy, 2015a) .....    | 8            |
| 3.2. Doğrusal skor fonksiyonu gösterimi (Karpathy, 2015b) .....                 | 9            |
| 3.3. Çok katmanlı sinir ağı .....   | 10           |
| 3.4. Sol:Standart sinir ağı. Sağ:Düğüm atma kullanılan sinir ağı.....           | 13           |
| 3.5. UR Veri Kümesi'nden bir imgenin uygulama ile etiketlenmesi .....           | 15           |
| 4.1. 2 katmanlı sinir ağı gösterimi .....                                       | 24           |
| 4.2. 1. duruma göre 3 katmanlı sinir ağı gösterimi.....                         | 27           |
| 4.3. 2. duruma göre 3 katmanlı sinir ağı gösterimi.....                         | 27           |
| 4.4. 3. duruma göre 3 katmanlı sinir ağı gösterimi.....                         | 28           |
| 4.5. 4 katmanlı sinir ağı gösterimi .....                                       | 31           |
| 4.6. 2. duruma göre 4 katmanlı sinir ağı gösterimi.....                         | 31           |
| 4.7. 5 katmanlı sinir ağı gösterimi .....                                       | 34           |
| 4.8. 5 katmanlı sinir ağı konolüsyonel katman özellik görünümleri .....         | 38           |
| 4.9. 5. duruma göre sinir ağı gösterimi.....                                    | 39           |
| 4.10. Üç veri kümesinin birleştirilerek eğitiminden çıkan sonuç görselleri..... | 40           |
| 4.11. 2 katmanlı sinir ağı mimarisi için duyarlılık doğruluk eğrisi .....       | 42           |
| 4.12. 3 katmanlı sinir ağı mimarisi için duyarlılık doğruluk eğrisi .....       | 42           |
| 4.13. 4 katmanlı sinir ağı mimarisi için duyarlılık doğruluk eğrisi .....       | 43           |
| 4.14. 5. katmanlı sinir ağı mimarisi için duyarlılık doğruluk eğrileri .....    | 43           |
| 4.15. Katmanlara ait duyarlılık doğruluk eğrileri .....                         | 44           |

## ÇİZELGELER DİZİNİ

| <u>Çizelge</u>   | <u>Sayfa</u> |
|--|--------------|
| 4.1. KTP Veri Kümesi ile eğitilmiş sinir ağı sonuçları .....                               | 24           |
| 4.2. UR Düşme Algılama Veri Kümesi ile eğitilmiş sinir ağı sonuçları .....                 | 25           |
| 4.3. RGB-D İnsan Veri Kümesi ile eğitilmiş sinir ağı sonuçları .....                       | 26           |
| 4.4. KTP Veri Kümesi ile eğitilmiş 3 katmanlı sinir ağı sonuçları .....                    | 28           |
| 4.5. UR Düşme Algılama Veri Kümesi ile eğitilmiş 3 katmanlı sinir ağı sonuçları .....      | 29           |
| 4.6. RGB-D İnsan Veri Kümesi ile eğitilmiş 3 katmanlı sinir ağı sonuçları .....            | 30           |
| 4.7. KTP Veri Kümesi ile eğitilmiş 4 katmanlı sinir ağı sonuçları .....                    | 32           |
| 4.8. UR Düşme Algılama Veri Kümesi ile eğitilmiş 4 katmanlı sinir ağı sonuçları .....      | 33           |
| 4.9. RGB-D İnsan Veri Kümesi ile eğitilmiş 4 katmanlı sinir ağı sonuçları .....            | 33           |
| 4.10. RGB-D İnsan Veri Kümesi ile eğitilmiş 5 katmanlı sinir ağı sonuçları .....           | 35           |
| 4.11. UR Düşme Algılama Veri Kümesi ile eğitilmiş 5 katmanlı sinir ağı sonuçları .....     | 36           |
| 4.12. RGB-D İnsan Veri Kümesi ile eğitilmiş 5 katmanlı sinir ağı sonuçları .....           | 37           |
| 4.13. Birleştirilmiş veri kümeleri 5 katmanlı sinir ağı sonuçları .....                    | 38           |
| 4.14. Birleştirilmiş veri kümeleri farklı katman sayılarına göre sinir ağı sonuçları ..... | 39           |
| 4.15. 4.duruma göre eğitilmiş sinir ağında Watch-n-Patch Veri Kümesi sonuçları .....       | 40           |
| 4.16. Hata matrisi .....   | 41           |

**SİMGELER VE KISALTMALAR DİZİNİ****Kısaltmalar**

TP

TF

KTP

UR

RGBD

**Açıklamalar**

Pozitif Doğruluk Oranı

Negatif Doğruluk Oranı

Kinect Takip Hassasiyeti Veri Kümesi (Kinect Tracking Precision)

UR Düşme Algılama Veri Kümesi

RGBD İnsan Veri Kümesi

## 1. GİRİŞ VE AMAÇ

Geçmişten günümüze insan sezimi, imge işleme alanındaki önemli problemlerden biri olmuştur. İnsan sezimi, karmaşık arka plandan insan ayırt etme ve yerini belirleme problemidir. Eğlence sektöründen otomotiv teknolojisine, arama kurtarma çalışmalarından sağlık hizmetlerine kadar birçok alanda insan sezimi yöntemlerine ihtiyaç duyulmaktadır.

İnsan seziminde, algılayıcıdan gelen veriye göre farklı yöntemler kullanılmaktadır. Bu veri, ışık kaynağına ihtiyaç duyan renkli imgeler olabildiği gibi karanlık ortamlarda da sahne hakkında geometrik bilgi veren derinlik imgeleri de olabilir.

İnsan sezimi yöntemlerinin büyük çoğunluğu, aydınlık ortamda dik pozisyonda duran insanları (örneğin trafikte yaya algılama, ya da güvenlik uygulamaları) bulmaya yöneliktir. Ancak iç ortamlarda yaşlı ya da hasta insanların takiplerinin yapılması (pozisyon anomalilikleri, düşme), doğal afetlerde kurtarma ekiplerinin ulaşamadığı ortamlarda afetzedelerin bulunması, servis robotlarının ev ve ofis ortamlarında her türlü pozda bulunabilecek insanlarla etkileşimde bulunabilmesi gibi görevler için sahnede her türlü insan pozunun bulunabileceği varsayılmalıdır. Ayrıca bu durumlarda ışık kaynağı her zaman bulunmayabilir. Gece kullanılan bir araçta yayanın aniden aracın önüne çıkması, ışıkların gece kapandığı bir hastane ya da bakım evinde hastaların kontrolü, doğal afetlerde ekip ışıklarının ulaşamadığı derinlikteki yerlerde renkli kameraların kullanımı yetersiz kalacaktır. Bu tür durumlar için, literatürde yaygın olarak kullanılan renkli imgeler yerine, sadece derinlik imgeleri kullanarak insan sezimi yapabilecek yöntemlere ihtiyaç duyulmaktadır.

Sinir ağları insan beyninin çalışma yapısını taklit ederek geliştirilen yapay zeka tekniklerinden biridir. Yapay sinir ağlarında bilgi, ağdaki sinirlerin bağlantılarının ağırlıklarında tutulur. Bu ağırlıklar eğitimde kullanılan veri kümesine göre değişerek optimal değerler almalıdır. Ağın bu ağırlık değerlerine ulaşılabilmesi için yapılan işleme “ağın eğitilmesi” denir.

Bu çalışmadaki amaç, sadece derinlik imgeleri kullanılarak her pozisyonda bulunabilecek insanların algılanması için önemli problemlerden birine çözüm sunmaktır. Derinlik imgesindeki baş-omuz bölgesi adayları bulunduğu varsayılarak, bu adayların doğrulanması problemine çözüm olarak derin sinir ağlarının kullanımı önerilmektedir. Söz konusu doğrulamayı yapacak, diğer bir deyişle aday bölgeleri baş-omuz bölgesi içeren ve içermeyen olarak sınıflandıracak sinir ağlarının eğitimi için sınıf etiketlerinin bilindiği pozitif ve negatif örneklere ihtiyaç duyulmaktadır. Bu tez çalışmasında, Kinect Takip Hassasiyeti Veri Kümesi, RGB-D İnsan Veri Kümesi ve UR Düşme Algılama Veri Kümesi'nde bulunan derinlik imgelerinden baş-omuz bölgelerini içeren ve içermeyen kutucuklar elle işaretlenip kesilmiştir. Bu örnekler üzerinde veri arttırımı yapılmış; her bir veri kümesinin örnekleri kullanılarak eğitilen sinir ağı diğer veri kümesinin örnekleri üzerinde test edilmiştir. Katman ve filtre sayıları değiştirilerek farklı mimariler ile deneyler yapılmış; aynı zamanda ağın eğitiminde kullanılan parametrelerin sınıflandırma başarımı üzerine etkileri araştırılmıştır. Yapılan testlere göre belirlenen optimum sinir ağının, veri kümeleri birleştirildikten sonra da kullanılmasının önceki testlere paralel sonuçlar çıkardığı gözlemlenmiştir.

## 2. LİTERATÜR ARAŞTIRMASI

Günümüze kadar, elde edilen görüntülerden insan bulmak için çok sayıda çalışma yapılmıştır. Çalışmaların bir kısmında renk ve derinlik imgeleri kullanılırken bir kısmında yalnızca derinlik imgeleri kullanılmıştır. Aynı şekilde çalışmaların bir kısmında kenar bulucu gibi filtreler kullanılırken bir kısmı yapay sinir ağlarını kullanmıştır. Bu bölümde ilk olarak derinlik verilerinden insan sezimi için sinir ağlarından farklı yöntemler kullanan çalışmalardan daha sonra derin öğrenme ile insan sezimi üzerine yapılan çalışmalardan bahsedilmektedir.

Ikemura ve Fujiyoshi (2010), çalışmalarında ToF sensörlerinden sağlanan derinlik imgeleri kullanarak İlişkisel Derinlik Benzerliği Özellikleri çıkarmak üzerine çalışmışlardır. Yapılan çalışmada girdi olarak kullanılan imge satır satır tarama ile belirli sayıda dikkörtgen bölgelere ayrılır ve iki bölgenin derinlik histogramlarına bakılarak bölgeler arasındaki ilişkiler hesaplanır. Her bölgenin diğer bölgeler ile kombinasyonlarına bakılarak ilişkisel özellikler çıkarılır. Ayrılan bölgelere Gerçek AdaBoost Metodu uygulanarak bölgelerde insan ve ya insan-değil şeklinde sınıflandırmalar yapılır. Sınıflandırılan bölgeler ile ilişkisel özelliklere bakılarak kümeleme yapılır ve insan lokasyonları belirlenir.

Spinello vd (2010) tarafından yapılan çalışmada lazer sensörü kullanılarak ortamdan nokta bulutu verisi elde edilmiştir. Toplanan nokta bulutu eşit yükseklik aralıklarındaki katmanlara ayrılmıştır. Aynı yükseklikteki her nokta bulutu yatay yöndeki süreksizlikler bulunarak alt parçalara bölünmüştür. Alt parçalardan çıkarılan geometrik öznitelikler bir AdaBoost sınıflandırıcısına girdi olarak verilmiştir. Her bir yükseklik katmanı için ayrı bir AdaBoost sınıflandırıcısı eğitilmektedir. Böylece yüksekliğe göre alınan nokta kümeleri insan vücut merkezinden uzaklıklarına göre yerleştirilerek kümeler arasındaki geometrik uyumlar öğrenilmektedir. Farklı yükseklik katmanlarından gelen kararlar, daha sonra bir oylama yöntemi ile birleştirilerek alt parçaların bir insana ait olup olmadığına karar verilmiştir.



Xia vd. (2011) sadece derinlik imgeleri ve baş konturu kullanarak insan sezimi yapmayı amaçlamaktadır. İlk olarak Kinect Xbox 360 derinlik kamerasından alınan imgelere, 0 değerli gürültüleri gidermek için en yakın komşu aradeğerleme algoritması uygulanmaktadır. Daha sonra medyan filtre kullanılarak imge yumuşatılmaktadır. İmgelerde algılanacak baş objelerinin kameraya olan mesafeden bağımsız olması için resim piramidi oluşturulmaktadır. İmge, 2 boyutlu baş konturuna uygun hale gelene kadar 3/4 oranında küçültülmektedir. Yumuşatılmış imgede derinlik bilgilerine göre kenar-sınır bölgeleri arasında 2 boyutlu baş konturu kullanılarak aday bölgeler çıkartılmaktadır. Çıkarılan aday bölgeler 3 boyutlu hale getirilerek çalışmada tasarlanmış 3 boyutlu baş yüzeyi modeli ile karşılaştırılmaktadır. Bulunan baş bölgesinden hareketle büyüme fonksiyonu kullanılarak insanın vücudu algılanmaktadır.

Nghiem vd. (2012), çalışmalarında Kinect kamera ile sağlanan derinlik videolarını kullanarak baş algılamanın ve düşme tespitinin yapılmasını amaçlamaktadırlar. Baş algılama insan sezimi için kullanılmaktadır. Video sırasında hareket eden objeler tespit edilmekte ve tespit edilen objeler üzerinde baş algılama algoritması denenmektedir. Algoritmaya göre insan başı tam bir daire olarak kabul edilmiştir. Hareket eden objeler algılanıp bu objeler üzerinde verilen yöntemle göre baş daireleri aranıp, daireye göre aranılan bölgenin konturu çıkarılmıştır. Çıkarılan konturun orta noktası baş merkezi kabul edilmiş ve belirlenen merkezlere HOG özellik çıkarıcı ve SVM uygulanarak bölgenin baş olup olmadığına karar verilmiştir. Bu çalışmada insan sezimi için baş ve omuzlar birbirine dik olmak zorundadır. Yapılan çalışmaya göre insanlar kameraya dönük ve iki omuzu da görünecek şekilde durmalıdırlar.

Chen vd. (2016) çalışmaları sadece derinlik resimleri kullanılarak baş konumunun bulunması üzerinedir. Çalışmada Kinect2 kameradan alınan veriler kullanılmaktadır. Burada, diğer bazı çalışmaların aksine insan anatomisinde başa bağlı bulunan omuzların bulunmasına ve çok-ölçekli pencerelerde arama yapılmasına gerek duyulmamaktadır; bu nedenle önerilen yöntem daha esnektir. Derinlik imgelerindeki gürültüyü azaltmak için 3×3 medyan filtre uygulanmaktadır. Medyan filtre uygulandıktan sonra Doğrusal Ayırıcı Analiz (LDA) ile öğrenilen baş tanımalıcı ile imgedeki her piksel, baş merkezi ya da baş-merkezi-değil olarak sınıflandırılmaktadır. Daha sonra yanlış pozitif filtreden geçirerek en

olası baş-merkezleri saptanmaktadır. Derinlik verilerine göre baş-merkez kümelemesi kullanarak bulunan noktada baş-merkezi olup olmadığına karar verilmektedir. Bu karar verme mekanizması için kullanılan çember tamamen başın içinde kalan noktalardan oluşan küçük bir çemberdir. Çemberin merkezi hareket ettikçe alınan yeni noktaların merkez olma olasılığı azalmaktadır. Seçilen merkezler matematiksel bir denklemden geçirilerek baş-merkezi bulunmaktadır.

Xu ve Fujimura (2003) çalışmalarında derinlik ve gri ölçekli resimler kullanarak iç ortamda insan algılama üzerine çalışılmışlardır. Çalışmada ilk olarak derinlik ve gri ölçekli resimler kullanılarak dokusuz büyük nesnelere resimden çıkarılır. Bu nesnelere iç ortamlarda genel olarak duvarlara ve zeminlere denk geldiği varsayılmıştır. Duvarlar ve zemin sahnedan çıkarıldıktan sonra derinlik imgesi farklı derinlik katmanlarına bölütlenmiştir. Bir katmanda bulunan bir bağlı bileşenin gövde olup olmadığını sezmek için dikey elips uygulaması yapılmıştır. Elips, uygulanan bağlı bileşenin tamamen içine girene kadar küçültülmüştür. En küçük halde olan elips dikey değilse ya da çok kısa ise o bölgede gövde bulunmadığına karar verilmiştir. Bu yöntem, sadece ayakta duran ve arka plandan tamamen ayrılabilen insanları bulmak için kullanılabilir.

Khan vd. (2016), çalışmalarında derinlik imgeleri kullanarak insan sezimini amaçlamaktadır. Derinlik sensörlerinden kaynaklanan gürültünün giderilmesi için önce en yakın komşuluk aradeğerleme algoritması kullanılmakta, daha sonra imgelere medyan filtre uygulanmaktadır. Canny kenar bulucu ile imgelerdeki derinlik süreksizliklerine karşılık gelen pikseller belirlenmektedir. Kenarlar üzerinde 2 boyutlu bir baş kalıbı gezdirilmektedir. Baş kalıbı ile karşılaştırılan bölgelerde mesafe kareleri toplamına ve çapraz korelasyona dayalı eşleşme algoritmaları kullanılmaktadır. Algoritmaların çıktuları 3 boyutlu baş yüzeyi ile karşılaştırılarak baş sezimi tam olarak gerçekleştirilmektedir. Baş konumu bulunduğundan sonra belirli bir eşik değeri ile birlikte bölge büyüme algoritması kullanılarak insan vücudunun geri kalanı sezilmektedir. Bölge büyüme algoritması sırasında birbirine yakın derinlik bilgilerine sahip olan pikselleri ayırmak için gradyan filtre kullanılmaktadır.

Derinlik verileri ile insan sezimi problemi çeşitli yöntemler ile aşılmaya başlandıktan sonra yeni problem, insan seziminin kısa sürede yapılması ve gerçek zamanlı projelerde

kullanılması olmuştur. Bunun için imgelerin ve ya video sahnelerinin sürekli olarak insan sezimini sağlayan filtrelerden geçirilmesi yerine öğrenilmiş filtrelerden geçirilmesi üzerine çalışmalar yapılmaya başlanmıştır. Derin öğrenme yöntemlerinin derinlik imgelerine ilk uygulamaları, insan sezimini değil başka objelerin bulunması ya da sınıflandırılmasını hedeflemiştir.

Örneğin Eitel vd. (2015), çalışmalarında renk ve derinlik imgeleri kullanarak eğitilen konvolüsyonel sinir ağları ile sınıflandırma yapmayı amaçlamaktadırlar. Çalışmada ImageNet Veri Kümesi ve Washington RGB-D Obje Veri Kümesi kullanılmaktadır. RGB imgeler ve derinlik imgeleri iki farklı ağ katmanlarından geçerek son iki füzyon katmanında birleştirilmektedir. Sinir ağlarının eğitimi ImageNet Veri Kümesi ile yapıp testleri RGB-D Obje Veri Kümesi ve gerçek zamanlı iç ortam görüntüleri ile yapılmaktadır.

Daha sonra, derin öğrenme yöntemleri, insan sezimi için de kullanılmaya başlanmıştır.

Wang vd. (2015), yaptıkları çalışmada obje tanıma için RGB imgeleri ile derinlik imgelerini kullanmaktadırlar. Yapılan çalışmada dört farklı yöntem karşılaştırılmaktadır. İlk yöntem sinir ağının eğitilmesinde sadece RGB imgelerinin kullanıldığı yöntem, ikinci yöntem sinir ağının eğitimi için sadece derinlik imgelerinin kullanıldığı yöntem, üçüncü yöntem derinlik imgelerinin ve RGB imgelerin ağa girdi olarak birlikte verilme yöntemidir. Bu işlem sinir ağına 4 kanallı girdi olarak yapılmaktadır. Dördüncü yöntem, derinlik ve RGB imgeleri için ayrı ayrı konvolüsyonel katmanlar oluşturulup daha sonra çok modlu tam bağlantılı bir katman aracılığı ile sınıflandırma yapma yöntemidir. Tasarlanan bu katman sadece özellik öğrenmek için değil iki farklı formattaki imgelerin arasındaki bağlantıyı öğrenmek için de kullanılmaktadır. Yapılan testlerin incelenmesi ve literatürde yapılan diğer sonuçlar ile karşılaştırılması ile iki farklı formattaki imgenin sinir ağına ayrı girmesi yönteminin tatmin edici sonuçlar verdiği görülmektedir.

Işık bulunmayan ortamlarda da insan sezimi çalışmalarının yapılabilmesi ve derin sinir ağları ile sadece derinlik verilerinin kullanılması için Borghi vd. (2017) derinlik imgeleri ile baş konumlandırma ve poz tahmini için yeni bir derin öğrenme ağının

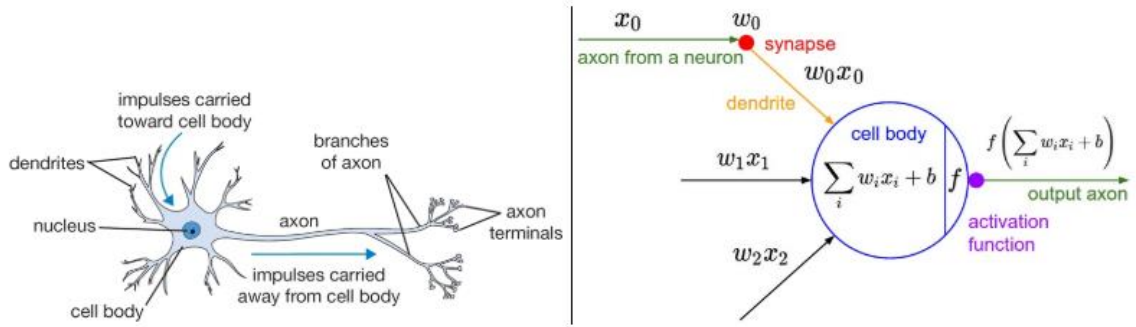
tasarlanması üzerine çalışmışlardır. Amaç, araç içinde bulunan düşük maliyetli sensörler ile sürücünün baş pozisyonunu ve yönelimini algılamaktır. İlk olarak sensörlerden alınan imge geriye doğru (regressive) konvolüsyonel sinir ağı ile işlenerek başın bulunduğu bölgede kırılmaktadır. Kırılan imge üç bağımsız konvolüsyonel ağı beslemektedir. Burada kullanılan ana ağ Face-from-Depth ağıdır. Diğer iki ağ poz tahmini hatalarını azaltmak için kullanılmaktadır. Genel ağın POSEidon olarak adlandırılmasına sebep olan füzyon katmanı değişik boyutlarda üç tam bağlı katmandan oluşmaktadır. Ağ çıktı olarak başın yönelimini 3 dönme açısı ile vermektedir. Aynı mimari kullanılarak omuz bölgesi de çıktı olarak alınmaktadır. Ağ 2 farklı veri kümesi üzerinde denenmektedir. Bunlar ICT3DHP Veri Kümesi ve Biwi Kinect Head Pose Veri Kümeleri'dir. Bu çalışmanın bir diğer amacı, Pandora adı verilen yeni bir veri kümesi oluşturmaktır. Oluşturulan veri kümesi Microsoft Kinect One cihazı ile alınan derinlik ve RGB imgelerini, eklem (birleşme yeri) bilgilerini ve IMU sensörü kullanılarak elde edilen omuz açılarını içermektedir.

Ballotta vd. (2018a), yaptıkları çalışmada derin sinir ağı kullanarak derinlik imgelerini baş içeren ve içermeyen imgeler olarak sınıflandırmayı hedeflemişlerdir. Watch and Patch Veri Kümesi ve Pandora Veri Kümesi kullanılmaktadır. Girdi boyutları 64×64 piksel olarak kabul edilmiş ve beş konvolüsyon katman kullanılmıştır. Sinir ağında hiperbolik tanh aktivasyon fonksiyonu kullanılmaktadır.

Ballotta vd. (2018b), çalışmalarında derinlik imgeleri ile "baş" sezimi için tam evrişimsel ağ metodu kullanmaktadır. VGG-19 ağının ilk 4 evrişimsel kısmı ile oluşturulan ağın çıktısı bir olasılıksal haritadır. Ağın girdisi 512×424 boyutunda bir imgedir. İmge ağ için 50cm ile 450cm arasında normalize edilmektedir. Ağın referans çıktısı olarak iki değişkenli Gauss fonksiyonu ile tanımlanan 64×53 piksellik imgedir Çalışmada eğitim için Watch-n-Patch veri kümesi kullanılmış, ayrıca veri çoğaltması uygulanmıştır. Ağın ürettiği olasılıksal haritada, değeri 0.3 eşliğinin üstünde bulunan pikseller baş konumlandırma için aday olarak belirlenmiştir. Kesin baş merkezini bulmak için yatay, dikey ve yüz genişliği için saptanan sabit aralıklar ile başın merkez noktası kullanılmaktadır. Çalışmada eğitilen ağ Pandora Veri Kümesi ile test edilmektedir. CNN metodu, SVM-HOG ve Derinlik Tabanlı Baş Tanımlayıcı-LDA metotları ile karşılaştırılmaktadır.

### 3. MATERYAL VE YÖNTEM

Derin öğrenme adı altında kullanılan yapay sinir ağları insan beyninin incelenmesi ile ortaya çıkmıştır. İnsan beyninde bulunan biyolojik sinir ağları birbirine bağlı nöronların elektrik sinyalleri ile haberleşmesine dayanır. Nöronun baş kısmında bulunan dentritler giriş sinyallerini alıp girdilere uygun bir çıkış sinyali oluşturur. Bu çıkış sinyali de belirli bir eşik geriliminin üzerinde ise başka bir nöronun dentritlerinden girerek o nöronda yorumlanıp diğer nöronu tetikler. Şekil 3.1.'de görüldüğü gibi yapay sinir ağları, bu biyolojik sistemdeki nöronların giriş sinyallerini yorumlayıp çıkış sinyallerine karar vermesini taklit ederek oluşturulmuştur (Karpathy, 2015a).



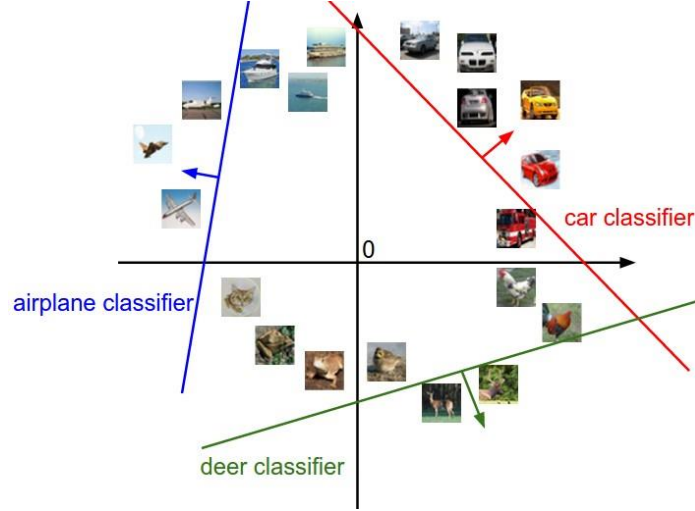
Şekil 3.1. Biyolojik ve bilgisayar ortamında sinir hücresi (Karpathy, 2015a)

#### 3.1. Yapay Sinir Ağları

Bilgisayardaki bir yapay sinir ağının en küçük birimi perseptrondur. Bu perseptron nöronlardaki gibi birden fazla girdiyi vektör olarak alabilir. Bir perseptron aşağıdaki doğrusal fonksiyondaki gibi girdileri ağırlık parametreleri ile çarpıp belirli bir yanlılık (bias) değeri ile toplar. Yapay sinir ağlarını eğitirken amaç, her bir perseptron için en optimum çıktı değerini veren parametreleri hesaplamaktır.

$$y = w * x + b \quad (3.1.)$$

Yapay sinir ağına bilgisayarla görü ile ilgili bir eğitimin uygulandığını varsayarsak yukarıdaki işlemde  $w$  öğrenilen ağırlık parametrelerini,  $x$  eğitimin uygulandığı imgenin girdi piksellerinden oluşan vektörü ve  $b$  yanlılık (bias) değişkenini belirtir (Karpathy, 2015b).



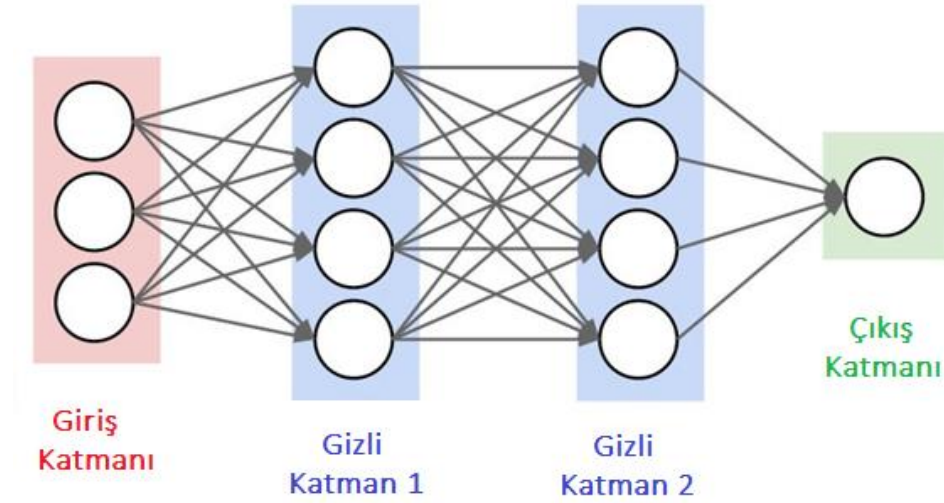
Şekil 3.2. Doğrusal skor fonksiyonu gösterimi (Karpathy, 2015b)

Ağırlık parametreleri görüntüye uygulanan filtreler gibi nitelendirilebilir. Ağırlık parametreleri ve piksel girdilerinin çarpımı bir doğrusal skor fonksiyonu oluşturur. Bu doğrusal skor fonksiyonun parametreleri, bir sınıflandırma probleminde, girdinin sınıfını belirlemek için kullanılan ayırıcı doğrunun denklem parametreleri olarak düşünülebilir. Yanlılık değeri ise yukarıda gösterilen Şekil 3.2.'deki gibi elde edilen doğrunun ötelenmesinde kullanılır.

Yapay sinir ağları çok katmandan oluşabilir. Aynı katmanda bulunan perseptronlar birbirlerinden bağımsızdır. Bir katmandaki perseptronlar bir önceki katmandaki perseptronlardan çeşitli aktivasyon değerleriyle etkilenir. Girdi ve çıktı katmanları arasında kalan katmanlara gizli katman denir.

Literatürdeki çalışmalara göre, bir katmandaki perseptronların çok olması ve ağından birden fazla katmandan oluşması, bulunması gereken parametre sayısını arttırdığı gibi öğrenme kabiliyetini de belirli bir seviyeye kadar artırır (Karpathy, 2015a). Eğitim sırasında perseptronların ağırlık parametrelerinin başlangıç değerlerini belirlemek için en sık kullanılan yöntem, bir rastgele dağılımdan değerler seçmektir. Eğitim sırasında,

parametreler, bir kayıp fonksiyonunun güdümünde + ve – yönlerde ilerleyerek optimum değerlere ulaşırlar. Sınıflandırma için kullanılan sinir ağlarında eğitim tamamladığında her bir sınıf için ağırlık matrisi belirlenmiş olur. Eğitimde kullanılmayan girdilerle test sırasında, sınıflandırma skorlarını elde etmek için bu ağırlık matrisleri kullanılır.



Şekil 3.3. Çok katmanlı sinir ağı

Şekil 3.3 de görülen her bir beyaz çember bir nöronu temsil etmektedir. Her bir nöron, bir alt katmandaki nöronlardan gelen girdileri, kendine ait ağırlık değerleri ile çarpıp bir yanlılık değeri ekler. Sonuç değeri bir aktivasyon fonksiyonundan geçirilerek nöronun çıktısı olarak belirlenir. Her bir nöronun çıktısı bir sonraki katmana girdi değeri olarak gönderilir.

### 3.2. Aktivasyon Fonksiyonları

Aktivasyon fonksiyonu, hangi nöronların öğrenmede daha etkili olacağını belirlemek için kullanılır. Sık kullanılan aktivasyon fonksiyonları sigmoid fonksiyonu, tanh fonksiyonu ve ReLU fonksiyonudur (Karpathy, 2015a).

Sigmoid fonksiyonu girdi değerlerini 0 ile 1 arasında dağıtır.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.2)$$

Yukarıdaki işlem ile sigmoid fonksiyonu negatif büyük sayıları 0'a, pozitif büyük sayıları 1'e öter. Böylece bir katmandaki büyük negatif değerler üreten nöronlar diğer katmanları etkilemezler. Sigmoid fonksiyonunun, geri yayılma sürecinde olumsuz bir etkisi vardır. Çoğunlukla değerler 0 ve 1'e çok yakınsandığı için 0 değeri alan nöronlar geri yayılım kullanıldığı zaman sönmülerek eğitim aşamasında bir sonraki ileri yayılım için de kullanılamaz hale gelirler. Bu fonksiyon geri yayılım sırasında ağı etkileyip eğitim esnasında iletişim halindeki nöronların hepsinin negatif ya da pozitif değerler almasını etkileyebilir.

Tanh fonksiyonu nöron çıktılarını -1 ile 1 arasında dağıtır. Bu dağılım Sigmoid fonksiyonunun aksine 0 merkezli bir dağılımdır.

$$\tanh(x) = 2\sigma(2x) - 1 \quad (3.3)$$

ReLU(Rectified Linear Unit) Düzeltilmiş Doğrusal Birim aktivasyon fonksiyonu

$$f(x) = \max(0, x) \quad (3.4)$$

negatif değerleri 0 ile sönmülerek pozitif değerleri doğrudan geçirir. Yapılan çalışmalarda ReLU aktivasyon fonksiyonunun ağı eğitim hızını sigmoid ve tanh aktivasyon fonksiyonlarına göre arttırdığı tespit edilmiştir. Nöronların değerlerine diğer fonksiyonlardan daha az müdahale eder, bu sebeplerle ReLU aktivasyon fonksiyonu diğer aktivasyon fonksiyonlarından daha çok tercih edilir.

ReLU fonksiyonunun bir dezavantajı Sigmoid Fonksiyonunda olduğu gibi negatif değerlerin "0" kabul edilmesi ile ölen bir nöronun bir daha eğitimde kullanılmama ihtimalidir. Bu dezavantajın giderilmesi için sızdıran (Leaky) ReLU aktivasyon fonksiyonu önerilmiştir.

Sızdıran (Leaky) ReLU fonksiyonunda

$$f(x) = 1(x < 0)(ax) + 1(x \geq 0)(x) \quad (3.5)$$



$\alpha$  küçük sabit bir değerdir. Bu işlem ile negatif değerler 0'a eşitlenmek yerine küçük bir  $\alpha$  ile çarpılarak 0'a yakın değerlere çekilirler.

### 3.3. Kayıp Fonksiyonları

Sinir ağlarının son katmanı hata oranını ölçen fonksiyonun tanımlandığı kayıp fonksiyonunun bulunduğu katmandır. Kayıp (Loss) fonksiyonu sinir ağının son katmandaki çıktı değerleri ile gerçek değer arasındaki hatayı tanımlar. Sinir ağındaki amaç kayıp fonksiyonunun değerini 0'a yaklaştırmaktır. Kayıp fonksiyonunda sigmoid, çok sınıflı SVM, softmax fonksiyonları kullanılabilir (Karpthy, 2015b). Sigmoid fonksiyonu 1 ve 0 arasında değerler verir ve 0.5 ile eşiklenir, bu sebeple iki sınıflı ağlar için kullanılır. Çok sınıflı SVM (Support Vector Machine) çok sınıflı sınıflandırma problemleri için kullanılır ve aşağıdaki formülle tanımlanır.

$$L_i = \sum_{j \neq y_i} \max(0, w_j^T x_i - w_{y_i}^T x_i + \Delta) \quad (3.6)$$

Yukarıdaki işlemde kullanılan  $w_j^T x_i$ , i'ninci imge örneği için j'ninci sınıfın skor değeridir. Skor hesaplamada kullanılan bias vektörü w matrisinin son satırına eklenmiştir.  $w_{y_i}^T x_i$  i'ninci örneğin gerçek sınıfının skoru ve  $L_i$  i'ninci örnek için elde edilen kayıp değeridir. Kayıp fonksiyonu elde edilirken doğru sınıfın skoru ile doğru olmayan sınıfların skoru arasındaki farka sabit bir marj ( $\Delta$ ) eklenir. Bu sayede, doğru sınıfın skorunun diğer sınıfların skorlarına göre bu marjdan daha büyük olması zorlanmış olur.

Bir diğer kayıp fonksiyonu ise Softmax Sınıflandırıcı'dır. Softmax SVM kayıp fonksiyonundan farklı olarak olasılık temelli kayıp değerleri üretir. Softmax'ın ürettiği skor değerleri en çok olabilirlik (maximum likelihood) fonksiyonu ile normalize edilerek olasılıksal hale getirilir.

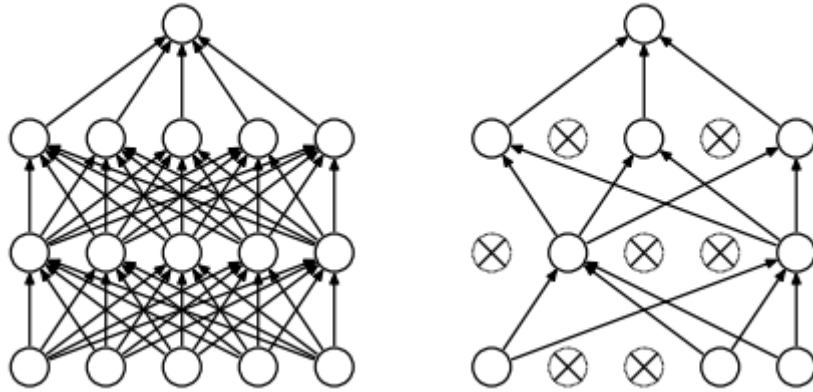
$$L_i = -\log(e^{f_{y_i}} / \sum_j e^{f_j}) \quad (3.7)$$

Yukarıdaki işlemde  $f_{y_i}$  i'ninci örneğin doğru sınıf için hesaplanmış skor değeri,  $f_j$  ise j'ninci sınıfın skor değeridir Softmax sınıflandırıcı, gerçek sınıfın dağılımı ve kestirilen sınıf olasılıkları arasındaki çapraz entropinin en aza indirilmesine dayanır.

### 3.4. Yapay Sinir Ağlarının Eğitilmesi

Skor fonksiyonunun elde edilmesi için yapılan işlemlere ileri yayılım denir. Eğitim sırasında, ağın başarısını arttırmak ve ağırlık değerlerinin en optimum halini bulmak için ileri yayılıma ek olarak geri yayılım kullanılmaktadır. Geri yayılım skor değeri elde edilirken gradyanları hesaplamaları Zincir Kuralı kullanılarak yapılmaktadır. Zincir Kuralı birden fazla fonksiyonun türevlerinin çoklu değişkenlere bağlı olarak alınmasını kolaylaştırır (Li, 2018a).

Aşırı öğrenme (overfitting) yapay sinir ağlarının eğitiminde oluşan bir sorundur. Veri kümesine de bağlı olarak tasarlanan ağın veri kümesini ezberlemesi ile oluşur. Eğitim sırasında kullanılan doğrulama kümesindeki performansın düşmesi ile ağın eğitiminin durdurulması gerekir ki bu durum ağın öğrenmediğini, ezberlediğini gösterir. Ağın ezberlemesini engellemek için ağın karmaşıklığını azaltmak ya da eğitimde kullanılan veri kümesini değiştirmek yerine düğüm atma kullanılır (Karpathy, 2015c).



Şekil 3.4. Sol:Standart sinir ağı. Sağ:Düğüm atma kullanılan sinir ağı

Şekil 3.4.'de görüldüğü gibi düğüm atma eğitim sırasında kullanılan nöronlardan bazılarının kimi eğitim verileri için düşmesi (kullanılmaması) demektir. Her nörona rastgele bir  $p$  olasılık değeri verilir ve belirli bir eşik değeri ile nöronların belli bir eğitim verisi için kullanılıp kullanılmayacağına bu şekilde karar verilir. Bu çeşit bir eğitimde, eğitim sırasında bütün nöronların bütün eğitim verisini görmemesi sağlanarak, ağın veriyi ezberlemesi bir

ölçüde engellenmiş olur. Eğitim aşaması bitip test aşamasına geçildiğinde ağı bütün nöronları kullanılır.

Yapay sinir ağlarının eğitiminde kullanılan 3 farklı öğrenme çeşidi vardır. Bunlar Denetimli (Supervised) Öğrenme, Denetimsiz (Unsupervised) Öğrenme ve Takviyeli (Reinforcement) Öğrenme'dir (Lison, 2012).

Denetimli Öğrenme çeşidinde yapay sinir ağının eğitiminde kullanılan girdi verileri ve bu verilere ait, yer gerçekliği (ground truth) olarak adlandırılan, gerçek çıktı değerleri ağa verilmektedir. Böylece ağ çıktı verilerine göre parametrelerini güncelleyerek öğrenimini gerçekleştirmektedir.

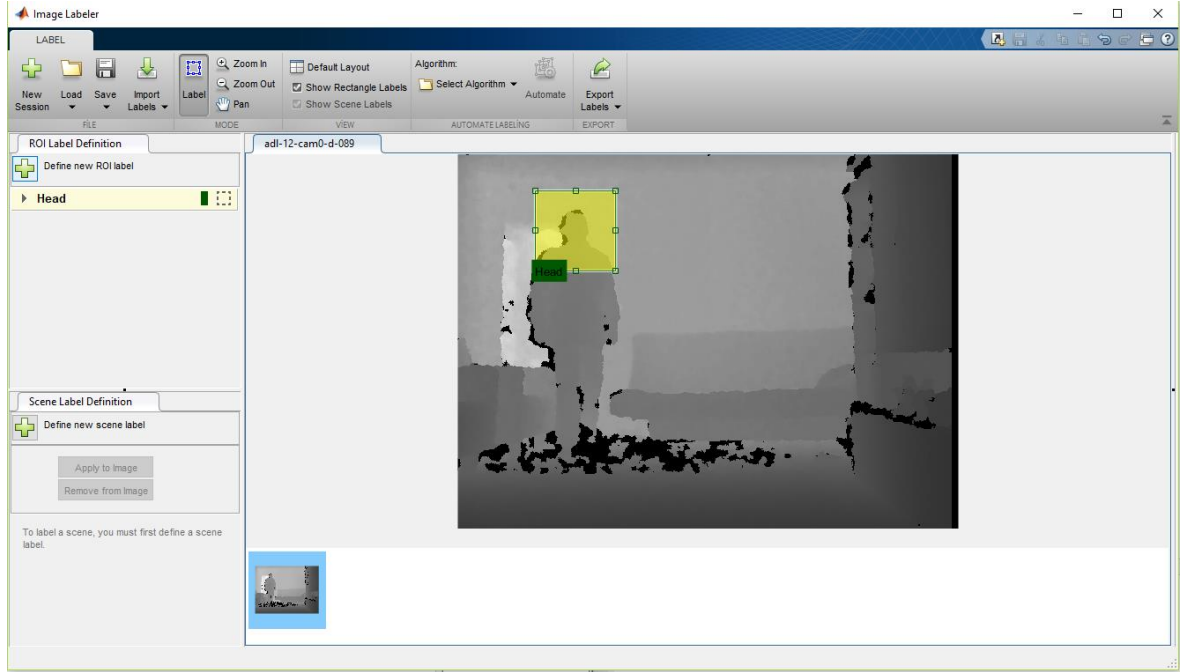
Denetimsiz Öğrenme çeşidinde yapay sinir ağına sadece girdi verileri verilmektedir. Eğitim devam ettikçe benzer girdi verilerinin kümelenmesi amaçlanmaktadır.

Pekiştirmeli Öğrenme çeşidi hareketlerin ya da alınan kararların öğrenilmesinde kullanılmaktadır. Burada, insanda gözlemlenen davranış psikolojisinde olduğu gibi çıktı verileri yerine bir ödül mekanizması kullanılmaktadır. Yapay sinir ağının girdisi, eylem ve çıktı ile alınan ödül arasında ilişki kurularak öğrenme gerçekleştirilmektedir. Bu öğrenme genellikle oyun ve kontrol alanlarında kullanılmaktadır.

### **3.5. Tezin Gerçeklenmesi**

Tezin gerçeklenmesinde MATLAB Programı'nın 2017-b versiyonu kullanılmıştır. Tez gerçeklenirken veri kümelerinin hazırlanması, yapay sinir ağı mimarisinin kurulması ve alınan çıktıları bakılarak ağı güncellenmesi adımları izlenmiştir.

### 3.5.1. Veri kümelerinin hazırlanması



Şekil 3.5. UR Veri Kümesi'nden bir imgenin uygulama ile etiketlenmesi

Veri kümeleri Matlab'ın "ImageLabeler" resim etiketleme aracı ile hazırlanmıştır. Varolan veri kümelerindeki derinlik imgelerindeki baş-omuz içeren dikdörtgenler ve baş-omuz görüntülerini andıran negatif dikdörtgenler bu etiketleme uygulaması ile elde edilmiştir. Şekil 3.6'da UR Düşme Veri Kümesi'ne ait bir imgeden baş-omuz bölgesini içeren bir dikdörtgen kutunun etiketlenmesi gösterilmiştir.

Eğitimin amacına göre veri kümesi hazırlanırken bütün resim etiketlenebilir. Yine eğitimin amacına göre uygulamaya eklenen resimler içerisinde etiketlenmesi istenen bölge ve ya bölgeler resim içerisinde seçilir ve etiket tanımlaması İlgi Bölgesi Etiket Tanımlaması (ROI Label Definition) kısmında tanımlanır. Bu uygulama ile aynı sahnede birden fazla alan birden fazla etiket ile etiketlenebilir.

Eğitimde kullanılması istenen bütün negatif ve pozitif dikdörtgenler seçilip etiketlendikten sonra oluşturulan resim dizisi Referans Objesi (Ground Truth Object) olarak

kaydedilir. Referans Objesi içerisinde etiketlenen resimlerin dizinlerini, etiket ve etiket açıklamalarını ve resim içerisinde seçilen alanların dikdörtgen ölçülerini sırası ile belirtir.

Yeni veri kümesi hazırlanırken diğer veri kümelerinden alınan derinlik imgeleri üzerinden baş ve omuz bölgeleri el ile seçilip Head etiketi ile etiketlenmiş her bir veri kümesinden elde edilen resimler ayrı .mat dosyalarına kaydedilmiştir.

Tezin gerçekleşmesinde dört farklı veri kümesi kullanılmıştır. Bunlar RGB-D İnsan Veri Kümesi, UR Düşme Algılama (UR Fall Detection) Veri Kümesi, KTP (Kinect Tracking Precision) Veri Kümesi ve Watch-n-Patch Veri Kümesi dir.

### **3.5.1.1. RGB-D insan veri kümesi**

RGB-D İnsan Veri Kümesi iç ortamda çoğunlukla dik pozisyonda duran ve ya yürüyen insan görüntülerini içerir. Ortamda yatay şekilde yerleştirilmiş üç ayrı Kinect Kamera bulunur. Veri kümesi bu kameralardan alınan 3399 tane derinlik imgesinden, 3399 tane renkli imgeden ve 31 tane insan takibi sırasında çıkartılan metin belgesinden oluşur (Spinello vd. 2011).

RGB-D People Veri Kümesi'nde MATLAB'in ImageLabeler arayüzü kullanılarak 501 adet baş-omuz bölgesi içeren kutucuk, 235 adet baş-omuz bölgesi ile karşılaştırılabilecek negatif kutucuk el ile işaretlenmiş ve kesilmiştir.

İmgelerin Sinir Ağı Araç Kutusu (Neural Network Toolbox) ile oluşturulan yapay sinir ağında girdi olarak kullanılabilmesi için rastgele boyutlarda seçilip etiketlenmiş alanların büyük kenarları seçilip etiketlenen alan kare hale getirilmiştir. Kare haline getirilip kesilen resimlerin sabit bir genişlik ölçüsü olmadığı için yapay sinir ağının girdi boyutu sabit seçilip girdi olarak kullanılacak resimler sabit bir boyuta küçültülmüştür.

Sabit boyuta küçültülen imgeler 1, 2 ve 3 piksel sağa, sola, yukarı, aşağı ve 4 köşeden çapraz olarak ötelenmiş ve saat yönünde ve saat yönünün tersi yönde 5, 10, 15, 20 derece döndürülerek veri arttırımı yapılmıştır. Yapılan veri arttırımı sonucu olarak arttırılmış veri

kümesi 17034 adet baş-omuz bölgesi içeren imgelerden ve 7990 adet baş-omuz bölgesi içermeyen imgelerden oluşur.

### **3.5.1.2. UR düşme algılama veri kümesi**

UR Düşme Algılama (UR Fall Detection) Veri Kümesi iki farklı veri kümesinden oluşur. Bunlar düşme aksiyonun gerçekleştiği ve günlük aktiviteleri içeren veri kümeleridir. Günlük aktivite veri kümesi bir odaya kapıdan girip oturma, kalkma, eğilme, çömelme, yatma gibi hareketlerin hareket dizilerini içerir. Düşme veri kümesi yürürken ya da oturma anında gerçekleşen düşüş hareket dizilerini içermektedir. Bu veri kümesi toplanırken iki farklı Kinect Kamera kullanılmıştır. Birinci kamera hareket dizilerini karşıdan kaydederken ikinci kamera hareket dizilerini yukarıdan kaydeder. Böylece veri kümesi sadece dikey insan pozisyonlarını içermemektedir (Kwolek ve Kepski, 2014).

Düşme veri kümesi dikey ve yatay yerleştirilen Kinect kameralar ile kaydedilmiş 30 farklı düşme aksiyonu içermektedir. Veri kümesi içerisinde bu 30 farklı düşme aksiyonuna ait birinci ve ikinci kameraların derinlik imgeleri, renkli imgeleri, iki kameranın görüntülerinin senkronize edildiği .csv dosyaları, düşüşlere ait ivme dosyaları ve her iki kameraya ait 30 ayrı düşüşün videoları bulunmaktadır. Toplamda iki kameradan kullanılan 5990 düşme derinlik imgesi vardır.

Günlük aktivite veri kümesi yatay yerleştirilmiş Kinect Kamera ile kaydedilmiş 40 farklı aksiyon içerir. Veri kümesi içerisinde 40 farklı aktiviteye ait kaydedilmiş derinlik imgeleri, renkli imgeler, senkron ve ivme dosyaları ve videolar bulunur. Toplamda 8941 adet günlük aktivite bulunmaktadır.

Bu tez çalışması kapsamında, UR Düşme Algılama Veri Kümesi'nden MATLAB'ın ImageLabeler arayüzü kullanılarak 5958 adet baş-omuz bölgesi içeren kutucuk, 269 adet baş-omuz bölgesi ile karşılaştırılabilecek negatif kutucuk el ile işaretlenmiş ve kesilmiştir. RGB-D People Veri Kümesi'nde olduğu gibi sabit boyutlara getirilen imgelere veri arttırma uygulanmıştır. Veri arttırma uygulanırken RGB-D People Veri Kümesi'nde uygulanan yöntemler UR Düşme Algılama Veri Kümesi için de kullanılmıştır. Arttırılmış veri kümesi

202572 adet baş-omuz bölgesi içeren imgeden ve 9146 adet baş-omuz bölgesi içermeyen imgeden oluşur.

### **3.5.1.3. Kinect takip hassasiyeti veri kümesi**

Kinect Takip Hassasiyeti (Kinect Tracking Precision) Veri Kümesi bir mobil robot üzerine yerleştirilmiş Kinect Kamera kullanılarak kaydedilmiş dört videodan oluşur. Dört video robotun dört farklı hareketi için kaydedilmiştir. Her videoda, beş kişi ve beş farklı senaryo vardır. Veri kümesinde kullanılan toplam 8475 adet derinlik imgesi, 8475 adet renkli imge, robotun dört farklı hareketine özel kaydedilmiş konum belgeleri ve robotun dört farklı hareketine özel kaydedilmiş iki boyutlu ve üç boyutlu referans noktaları ve kameranın kalibrasyonu için kullanılan parametreleri içeren bir belge bulunur (Munaro vd 2012).

Yine bu tez çalışması kapsamında, Kinect Takip Hassasiyeti Veri Kümesi'nde MATLAB'in ImageLabeler arayüzü kullanılarak 3941 adet baş-omuz bölgesi içeren kutucuk, 1340 adet baş-omuz bölgesi ile karşılaştırılabilir negatif kutucuk el ile işaretlenmiş ve kesilmiştir. RGB-D People Veri Kümesi'nde olduğu gibi sabit boyutlara getirilen imgelere veri artırma uygulanmıştır. Veri artırma uygulanırken RGB-D People Veri Kümesi'nde uygulanan yöntemler KTP Veri Kümesi için de kullanılmıştır. Arttırılmış veri kümesi 133994 adet baş-omuz bölgesi içeren imgeden ve 45560 adet baş-omuz bölgesi içermeyen imgeden oluşur.

### **3.5.1.4. Watch-n-patch veri kümesi**

Watch-n-Patch Veri Kümesi Kinect v2 Kamera kullanılarak kaydedilmiş farklı nesnelere ile iletişime giren aksiyonları içeren renkli görüntülerden, derinlik görüntülerinden ve 25 eklemlilik iskelet takibinden oluşur. Veri kümesine ait 458 adet video içerisinde 7 farklı insan, 23 farklı nesne, 21 farklı aksiyon ve mutfaklar ve ofisler olmak üzere 13 farklı ortamın kombinasyonlarını içerir.

Bu çalışma için veri kümesinden Mutfak-1 adlı kısım kullanılmıştır. MATLAB'ın ImageLabeler arayüzü kullanılarak 266 adet baş-omuz bölgesi içeren kutucuk el ile işaretlenerek kesilmiştir.

### 3.5.2. Yapay sinir ağının oluşturulması

Yapay sinir ağı mimarisinin gerçekleştirilmesi için MATLAB kullanılmıştır. MATLAB, sinir ağı mimarisinin katman sayısını, her bir katmanda kullanılan filtre sayısını ve diğer parametreleri seçebilmemizi sağlamaktadır.

Yapay sinir ağı eğitilirken çözücü ismi (solverName), maksimum epok sayısı (MaxEpoch), veri doğrulama frekansı (ValidationFrequency), başlangıç öğrenim oranı (InitialLearnRate) ve L2 düzenleme (L2Regularization) gibi parametrelerin belirlenmesi gerekmektedir (MathWorks, 2016a).

'solverName' parametresi ağırlık parametrelerinin güncellenmesi için ayarlanır. Stokastik gradyan alçalma momentum (sgdm), RMSProp iyileştirici (rmsprop) ve Adam iyileştirici (adam) gibi seçenekleri vardır. Bu tez çalışmasında sgdm çözücü kullanılmıştır.

'MaxEpoch' parametresi sinir ağının eğitimi yapılırken kayıp değerinin belirli bir değerin altına ulaşmadığı durumlarda eğitimin durdurulması için kullanılan bir tam ileri-geri yayılım sayısıdır.

'InitialLearnRate' parametresi stokastik gradyan alçalma metodunda kullanılan bir katsayıdır. Eğitim veri kümesine göre küçük seçilirse eğitim çok uzun sürebilir. Büyük seçilirse sinir ağı veri kümesini öğrenmeden eğitim tamamlanabilir.

'L2Regularization' parametresi ezberlemeyi engellemek için kullanılan çözümlerden biridir. Düzenlenmiş kayıp fonksiyonlarının hesaplanması için katsayı olarak kullanılır.



Sinir ağı mimarisi tasarlanırken imge girdi katmanı (imageInputLayer), yığın normalleştirme katmanı (batchNormalizationLayer), konvolüsyonel katmanı (convolution2dLayer), ReLU aktivasyon katmanı (reluLayer), maksimum ortaklama katmanı (maxPoolingLayer), tam bağlantılı katman (fullyConnectedLayer), düğüm atma katmanı (dropoutLayer), Softmax Fonksiyon katmanı (softmaxLayer) ve sınıflandırma katmanı (classificationLayer) gibi katmanlar kullanılmıştır. Katmanların açıklamaları aşağıda verilmiştir.

‘imageInputLayer‘ katmanı girdi katmanı olarak kullanılan bir katmandır. Bu katman tanımlanırken girdi imgelerinin ölçüleri belirtilmelidir. Bu çalışmada girdi bilgilerinin ölçüleri  $28 \times 28 \times 1$  şeklindedir. Bu katman normalizasyon parametresi içerir ve parametrenin varsayılan değeri ‘zerocenter’ dir. Yani imgeler sinir ağına girerken 0 merkezli normalize edilir (MathWorks, 2016b).

‘batchNormalizationLayer‘ katmanı ilk olarak girdi için kullanılan her kanalın aktivasyonlarını normalize eder. Bu katmanın konvolüsyonel katman ile aktivasyon katmanları arasında kullanılması tavsiye edilmiştir. Katman içerisinde standart sapma parametreleri varsayılan olarak ayarlanmıştır (MathWorks, 2017).

‘convolution2dLayer‘ sinir ağı içerisinde girdilere kayan filtreler uygulanmasını sağlar. Konvolüsyonel katmanı içerisinde filtre boyutunu, kullanılacak filtre sayısını, filtre uygulanırken imgenin etrafının nasıl doldurulacağını, filtre uygulanırken ne kadar pikselin atlanacağını belirten parametreler bulunur (MathWorks, 2016c).

‘reluLayer‘ katmanı girdiye eşik değeri uygulanmasını sağlayan aktivasyon fonksiyonunun katmanıdır. Aktivasyon fonksiyonunda olduğu gibi eşik değeri 0’dır ve değiştirilemez (MathWorks, 2016d).

‘maxPoolingLayer‘ katmanı bir önceki katmandan gelen girdiyi bölgelere ayırarak her bir bölgenin maksimumunu hesaplar. Maksimumlardan oluşan katman çıktısı yeni katmanın girdisi olarak ayarlanır. Katman uygulanırken katmana ait bazı parametrelerin tanımlanması gerekir. PoolSize parametresi girdiye uygun ortaklamının uygulanacağı

pencerenin boyutunun belirlenmesini, Stride parametresi pencere kaydırılırken atlanması gereken piksel sayısının belirlenmesini ve Padding parametresi ortaklama uygulanırken girdinin sınırlarının nasıl doldurulacağını belirlemesini sağlar. Eğer girdiye ait maksimumlar belirlendikten sonra girdiye ait boyutun değişmesi istenmiyorsa PaddingMode parametresi ve Padding parametresi 'same' olarak ayarlanmalıdır (MathWorks, 2016e).

'fullyConnectedLayer' katmanı girdinin bir ağırlık matrisi ile çarpılıp bias vektörü ile toplandığı katmandır. Bu katmanın girdi boyutu bir önce kullanılan katmana göre değişebileceği için 'auto' olarak seçilebilir. Katmanın çıktı boyutu kullanılan etiket sayısı ile aynı olmalıdır. İsteğe göre ağırlık ve bias değerleri katmanın Weights ve Bias bileşenleri kullanılarak el ile belirlenebilir (MathWorks, 2016f).

'dropoutLayer' katmanı yapay sinir ağında kullanılan nöronların belirli bir olasılık değeri ile eğitimde kullanılmamasını sağlar (MathWorks, 2016g).

'softmaxLayer' katmanı yapay sinir ağına Softmax Kayıp Funksiyonu'nun uygulandığı katmandır (MathWorks, 2016h).

'classificationLayer' katmanı bir sınıflandırma katmanıdır. Bir önceki katmanın çıktı boyutları bu katmanın çıktı boyutunu belirler (MathWorks, 2016ı).

#### 4. BULGULAR VE TARTIŞMA

Bu tez çalışmasında, ilk olarak, bir veri kümesi ile eğitilen yapay sinir ağları, diğer veri kümeleri üzerinde test edilmiştir. Veri kümeleri MATLAB ile oluşturulmuş özdeş yapay sinir ağlarında ayrı ayrı denenmektedirler. Sinir ağları eğitildikten sonra, eğitilen ağlar üzerinde diğer eğitimde kullanılmayan veri kümeleri test edilmiştir. Sonuçlar, sinir ağlarında kullanılan katmanlarda değişiklik yapıp veri kümelerine en uygun sinir ağının ve ilgili parametrelerin belirlenmesi için kullanılmıştır. Bu deneysel süreç sonunda belirlenen optimum ağ, üç veri kümesinden seçilen eğitim kümesi ile tekrar eğitilmiş ve yine aynı üç veri kümesinden elde edilen test kümesi üzerinde denenmiştir.

Testler sırasında kullanılan veri kümeleri ile ilgili bilgiler şu şekildedir: KTP Veri Kümesi'nin kullanıldığı sinir ağı eğitiminde 107200 pozitif ve 36448 negatif, RGB-D İnsan Veri Kümesi'nin kullanıldığı sinir ağı eğitiminde 13627 pozitif ve 6392 negatif, UR Düşme Algılama Veri Kümesi'nin kullanıldığı sinir ağının eğitiminde 162060 pozitif ve 7317 negatif örnek, derinlik imgelerinden baş bölgesi içeren ve içermeyen bölgelerin derinlik imgelerinden elle kesilmesi ve veri artırımı yoluyla elde edilmiştir. Son olarak bu veri kümelerinin dışında örnekler sağlaması için Watch-n-Patch Veri Kümesi'ne ait 266 adet pozitif imge kullanılmıştır

Testler sırasında kullanılan yapay sinir ağları genel olarak girdi katmanı, normalizasyon katmanı, gizli katmanlar (konvolüsyonel katman, normalizasyon katmanı, aktivasyon katmanı, ortaklama katmanı), düğüm atma katmanı, tam bağlantılı katman, kayıp fonksiyonu katmanı ve sınıflandırma katmanlarından oluşmuştur. İki çeşit test yapılmıştır. Birincisi katman sayısını ve yapısını sabit tutulup Başlangıç Öğrenim Oranı (InitialLearnRate) ve L2 Düzenleme (L2Regularization) parametrelerinin değiştirilmesi, ikincisi ise yukarıda bahsedilen parametrelerin sabit tutulup katman yapılarının değiştirilmesidir.

$$\theta_{\ell+1} = \theta_{\ell} - \alpha \nabla E(\theta_{\ell}) \quad (4.1)$$

4.1 denkleminde gösterilen  $\alpha$  parametresi `initialLearnRate`'i temsil etmektedir.  $\theta$  ağırlık ve bias vektörlerini,  $\nabla E(\theta)$  ise kayıp fonksiyonundaki gradyanı temsil etmektedir.

$$E_R(\theta) = E(\theta) + \lambda \Omega(w) \quad (4.2)$$

4.2 denkleminde gösterilen  $\lambda$  parametresi L2Regularization'ı,  $\Omega(w)$  fonksiyonu ağırlık vektörlerinin birbiri ile çarpılmış halini,  $E(\theta)$  ise kayıp fonksiyonunu temsil etmektedir.

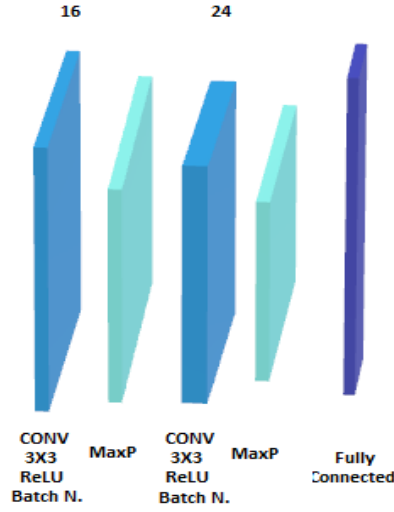
Testlere ait çizelgeler, ağırlık eğitiminin sonuna gelindiğinde test edilen veri kümelerine ait pozitif ve negatif imgelerin doğruluk oranlarını içerir. Pozitif doğruluk oranı (TP) ve negatif doğruluk oranı (TF)'nin hesaplanması aşağıdaki gibidir. Pozitif doğruluk oranı hesabında kullanılan etiket 'Head', negatif doğruluk oranı hesabında kullanılan etiket 'NotHead' etiketidir.

$$\text{Pozitif Doğruluk Oranı} = \frac{\text{'Head' etiketine sahip doğru sınıflandırılmış örnek sayısı}}{\text{'Head' etiketine sahip toplam örnek sayısı}} \quad (4.3)$$

$$\text{Negatif Doğruluk Oranı} = \frac{\text{'NotHead' etiketine sahip doğru sınıflandırılmış örnek sayısı}}{\text{'NotHead' etiketine sahip toplam örnek sayısı}} \quad (4.4)$$

#### 4.1. İki Katmanlı Eğitim

Bu alt bölümde sonuçları verilen sinir ağı, iki gizli katmandan, bir düğüm atma katmanından ve bir tam bağlantılı katmandan oluşur. Birinci gizli katman 16 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına ve bir ortaklama katmanına sahiptir. İkinci gizli katman 24 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına ve bir ortaklama katmanına sahiptir. Şekil 4.1'de bu 2 katmanlı sinir ağının gösterimi verilmiştir.



Şekil 4.1. 2 katmanlı sinir ağı gösterimi

Çizelge 4.1’de KTP Veri Kümesi ile eğitilmiş 2 katmanlı sinir ağının parametre değişimlerine göre doğruluk oranları gösterilmiştir. Arttırılmış KTP Veri Kümesi toplamda 143648 imgeye sahiptir. İki test veri kümesi için de, eğitim sırasında ‘InitialLearnRate’ parametresi 0.001 olarak seçildiğinde, 0.01 değerine oranla ağın veri kümesini tam olarak öğrenmeden eğitimini tamamladığı görülmektedir. 0.001 parametresi kullanılan veri kümesi için küçük bir öğrenme hızı sağlamıştır. ‘L2Regularization’ parametresinin etkisi incelendiğinde, bu parametrenin 0.01 değerinden 0.0001 değerine düşürülmesinin veri kümesinin öğrenilmesinde daha etkili olduğu görülmüştür.

Çizelge 4.1. KTP Veri Kümesi ile eğitilmiş sinir ağı sonuçları

|   | RGB-D İnsan |        | UR     |        |
|---|-------------|--------|--------|--------|
|   | TP          | TN     | TP     | TN     |
| InitialLearnRate=0.01<br>L2Regularization=0.01    | 0,9122      | 0,6723 | 0,4515 | 0,9145 |
| InitialLearnRate=0.01<br>L2Regularization=0.0001  | 0,9361      | 0,6000 | 0,5074 | 0,8922 |
| InitialLearnRate=0.001<br>L2Regularization=0.0001 | 0,7844      | 0,8894 | 0,4354 | 0,8894 |

Çizelge 4.2’de UR Düşme Veri Kümesi ile eğitilmiş 2 katmanlı sinir ağının parametre değişimlerine göre doğruluk oranları gösterilmiştir. Arttırılmış UR Veri Kümesi

toplamda 169377 imgeye sahiptir. Eğitim sırasında ‘InitialLearnRate’ parametresi için 0.01 ve 0.001 değerleri karşılaştırıldığında değerın küçülmesinin RGB-D İnsan Veri Kümesi’nde doğruluk oranlarında düşmeye sebep olurken, KTP Veri Kümesi’nde pozitif doğruluk oranında artışa sebep olmuştur. ‘L2Regularization’ parametresi incelendiğinde parametrenin 0.01 seçilmesinin ağır veri kümesini öğrenmesinde daha etkili olduğu görülmüştür. Bu durumda eğitimde kullanılan imge sayısı arttıkça ağırlık parametrelerinin güncellenmesi için L2Regularization parametresinin artırılması olumlu bir etki yaratmıştır..

Çizelge 4.2. UR Düşme Algılama Veri Kümesi ile eğitilmiş sinir ağı sonuçları

|   | RGB-D İnsan |        | KTP    |        |
|---|-------------|--------|--------|--------|
|   | TP          | TN     | TP     | TN     |
| InitialLearnRate=0.01<br>L2Regularization=0.01    | 0,8882      | 0,2170 | 0,9657 | 0,7052 |
| InitialLearnRate=0.01<br>L2Regularization=0.0001  | 0,8443      | 0,4426 | 0,8960 | 0,8336 |
| InitialLearnRate=0.001<br>L2Regularization=0.0001 | 0,7924      | 0,3532 | 0,9246 | 0,8179 |

Çizelge 4.3’de RGB-D Veri Kümesi ile eğitilmiş 2 katmanlı sinir ağının parametre değişimlerine göre doğruluk oranları gösterilmiştir. Arttırılmış RGB-D Veri Kümesi toplamda 20019 imgeye sahiptir. Eğitim sırasında ‘InitialLearnRate’ parametresi için 0.01 ve 0.001 değerleri karşılaştırıldığında UR Düşme Veri Kümesi için ciddi bir düşüş olduğu ve KTP Veri Kümesi için normal sayılabilecek oranda bir düşüş olduğu gözlemlenmiştir. RGB-D ve UR veri kümelerinin pozitif doğruluk oranlarının düşük olmasının sebebinin, veri kümelerinin imge sayısındaki ve içerdiği insan tiplerindeki farklılıklar olduğu düşünülmüştür. ‘L2Regularization’ parametresi incelendiğinde parametrenin 0.0001 seçilmesinin veri kümesinin öğrenilmesinde daha etkili olduğu görülmüştür. Bu durumda eğitimde kullanılan imge sayısı azaldıkça ağırlık parametrelerinin daha doğru güncellenmesi için L2Regularization parametresinin küçük tutulması etkili olmaktadır.

Çizelge 4.3. RGB-D İnsan Veri Kümesi ile eğitilmiş sinir ağı sonuçları

|   | UR     |        | KTP    |        |
|---|--------|--------|--------|--------|
|   | TP     | TN     | TP     | TN     |
| InitialLearnRate=0.01<br>L2Regularization=0.01    | 0,0651 | 0,8885 | 0,6752 | 0,9873 |
| InitialLearnRate=0.01<br>L2Regularization=0.0001  | 0,3317 | 0,7955 | 0,7973 | 0,9201 |
| InitialLearnRate=0.001<br>L2Regularization=0.0001 | 0,0984 | 0,8550 | 0,7384 | 0,9806 |

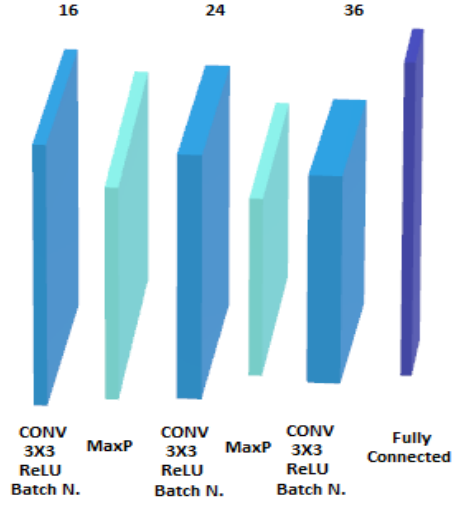
Yapılan testlere göre InitialLearnRate parametresinin 0.01 olduğu ve L2Regularization parametresinin 0.0001 olduğu sonuçlar KTP ve RGB-D İnsan Veri Kümesi için diğer parametrelere göre daha iyi sonuçlar vermiştir. UR Veri Kümesi için yukarıda bahsedilen parametreler ciddi bir eğitim problemine yol açmadığı için ağ yapısının değişimi bu parametreler üzerinden yapılacaktır.

#### 4.2. Üç Katmanlı Eğitim

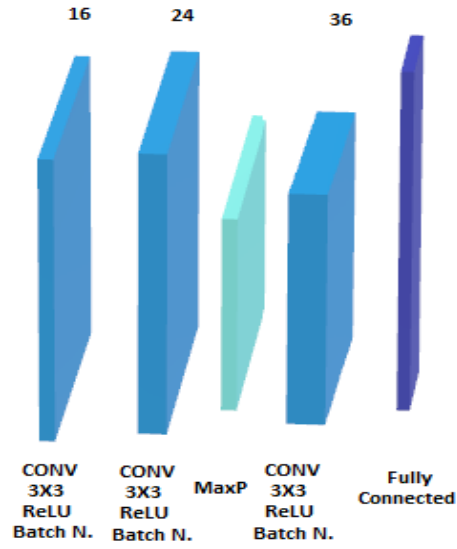
Bu bölümde sonuçları verilen sinir ağı üç gizli katmandan, bir düğüm atma katmanından ve bir tam bağlantılı katmandan oluşur. İlk gizli katman 16 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına ve bir ortaklama katmanına sahiptir. İkinci gizli katman 24 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına ve bir ortaklama katmanına sahiptir. Üçüncü gizli katman 36 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına sahiptir.

İki katmanlı eğitimden çıkan sonuçlara göre InitialLearnRate parametresi 0.01 ve L2Regularization parametresi 0.0001 olarak seçilip sinir ağı üç katmana çıkarılmıştır. Üç katmana çıkarılan sinir ağında katmanlara ait yapılar değiştirilip sonuçları incelenmiştir. Yeni sinir ağ yapıları Şekil 4.2, Şekil 4.3 ve Şekil 4.4'de gösterilmiştir.

1. durum: 3 katmanlı sinir ağı
2. durum: İlk ve ikinci konvolüsyonel katman arasındaki ortaklamanın kaldırılması
3. durum: Birinci ve ikinci konvolüsyonel katmandan sonra kullanılan normalizasyon katmanlarının kaldırılması

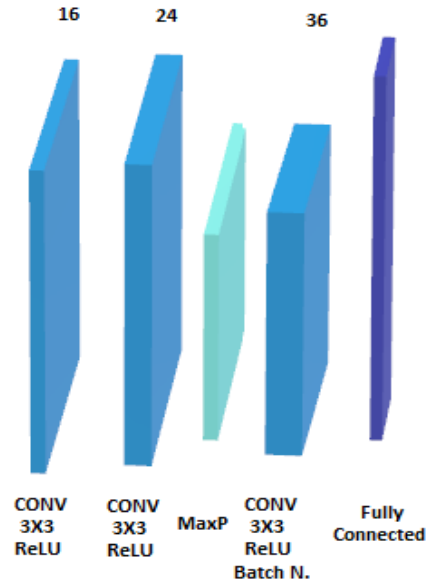


Şekil 4.2. 1. duruma göre 3 katmanlı sinir ağı gösterimi



Şekil 4.3. 2. duruma göre 3 katmanlı sinir ağı gösterimi





Şekil 4.4. 3. duruma göre 3 katmanlı sinir ağı gösterimi

Çizelge 4.4’de KTP Veri Kümesi ile eğitilmiş üç katmanlı sinir ağının yukarıda açıklanan 3 durum için doğruluk oranları gösterilmiştir. 1. durum ile 2 katmanlı ağ testi karşılaştırıldığında katman artırımının pozitif doğruluk oranlarında düşüşe, negatif doğruluk oranlarında artışa neden olduğu gözlemlenmiştir. 2. durum ile 1. Durum karşılaştırıldığında negatif doğruluk oranlarının arttığı ve KTP-RGBD veri kümeleri arasındaki uyumun azalıp, KTP-UR veri kümeleri arasındaki uyumun arttığı gözlemlenmiştir. 3 durum ile 2. Durum karşılaştırıldığında normalizasyon katmanlarının her katmanda uygulanmamasının olumlu etkileri olacağı görülürken, negatif doğruluk oranlarında düşüş gözlemlenmiştir.

Çizelge 4.4. KTP Veri Kümesi ile eğitilmiş 3 katmanlı sinir ağı sonuçları

|         | RGB-D İnsan |        | UR     |        |
|---------|-------------|--------|--------|--------|
|         | TP          | TN     | TP     | TN     |
| 1.durum | 0,8343      | 0,7532 | 0,4706 | 0,9257 |
| 2.durum | 0,7026      | 0,9191 | 0,5586 | 0,9740 |
| 3.durum | 0,8743      | 0,6596 | 0,6915 | 0,8699 |

Çizelge 4.5’de UR Veri Kümesi ile eğitilmiş üç katmanlı sinir ağının 3 farklı duruma göre doğruluk oranları gösterilmiştir. KTP Veri Kümesi’nde olduğu gibi 1. durum ile 2 katmanlı ağ testi karşılaştırıldığında katman arttırımının pozitif doğruluk oranlarında düşüşe, yalnızca KTP Veri Kümesi’ne ait negatif doğruluk oranında artışa neden olduğu gözlemlenmiştir. 2. durum ile 1. durum karşılaştırıldığında negatif doğruluk oranlarının arttığı ve UR-RGBD veri kümeleri arasındaki uyumun azalıp, UR-KTP veri kümeleri arasındaki uyumun arttığı gözlemlenmiştir. 3 durum ile 2. durum karşılaştırıldığında normalizasyon katmanlarının her katmanda uygulanmamasının olumlu etkileri olacağı görülürken, negatif doğruluk oranlarında ciddi düşüş gözlemlenmiştir.

Çizelge 4.5. UR Düşme Algılama Veri Kümesi ile eğitilmiş 3 katmanlı sinir ağı sonuçları

|         | <b>RGB-D İnsan</b> |           | <b>KTP</b> |           |
|---------|--------------------|-----------|------------|-----------|
|         | <b>TP</b>          | <b>TN</b> | <b>TP</b>  | <b>TN</b> |
| 1.durum | 0,6128             | 0,4255    | 0,8457     | 0,8739    |
| 2.durum | 0,4391             | 0,7106    | 0,9361     | 0,8440    |
| 3.durum | 0,9381             | 0,1957    | 0,9168     | 0,5604    |

Çizelge 4.6’da RGB-D İnsan Veri Kümesi ile eğitilmiş üç katmanlı sinir ağının 3 farklı duruma göre doğruluk oranları gösterilmiştir. 1. durum ile 2 katmanlı ağ testi karşılaştırıldığında katman arttırımının diğer veri kümelerinde olduğu gibi pozitif doğruluk oranlarında düşüşe, negatif doğruluk oranlarında artışa neden olduğu gözlemlenmiştir. 2. durum ile 1. durum karşılaştırıldığında negatif doğruluk oranlarının düştüğü ve iki veri kümesi ile de uyumun arttığı gözlemlenmiştir. 3 durum ile 2. durum karşılaştırıldığında normalizasyon katmanlarının her katmanda uygulanmamasının olumlu etkileri olacağı görülürken, diğer veri kümelerinin aksine negatif doğruluk oranlarında artış UR pozitif doğruluk oranında düşüş gözlemlenmiştir. UR ve RGBD veri kümeleri için çizelgeler karşılaştırıldığında iki ağın uyumunun yakın olması beklenirken çok farklı çıkmasının sebebinin ağ eğitiminin yapıldığı veri kümesindeki imge sayısına bağlı olduğu düşünülmüştür.

Çizelge 4.6. RGB-D İnsan Veri Kümesi ile eğitilmiş 3 katmanlı sinir ağı sonuçları

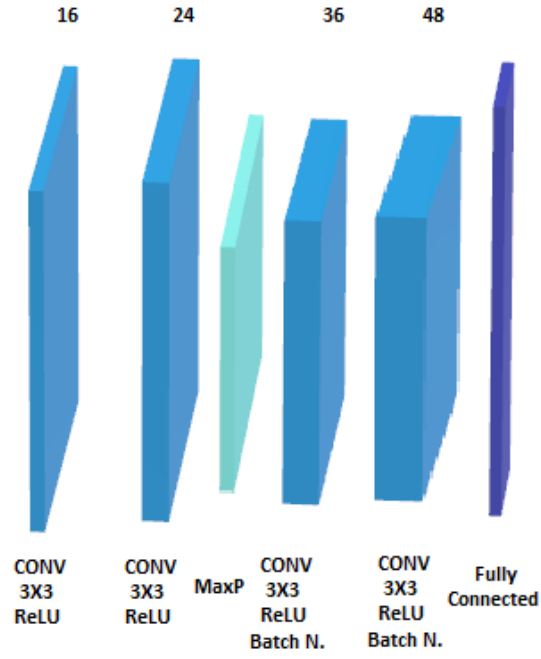
|         | UR     |        | KTP    |        |
|---------|--------|--------|--------|--------|
|         | TP     | TN     | TP     | TN     |
| 1.durum | 0,1396 | 0,8476 | 0,7293 | 0,9470 |
| 2.durum | 0,4434 | 0,8067 | 0,8648 | 0,9097 |
| 3.durum | 0,1177 | 0,8959 | 0,8292 | 0,9485 |

İki katmanlı eğitimlerin ikinci sonuçları ve üç katmanlı eğitimlerin birinci sonuçları karşılaştırıldığında genel olarak eğitilen veri kümesi ve test veri kümeleri arasında pozitif doğruluk oranlarında düşme, negatif doğruluk oranlarında artma görülmüştür.

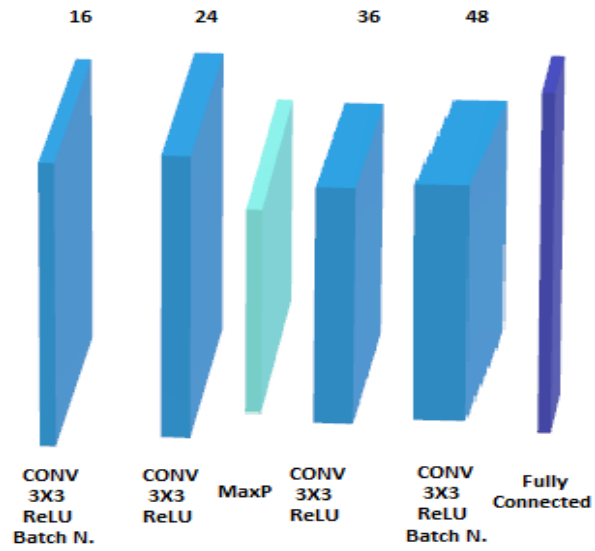
#### 4.3. Dört Katmanlı Eğitim

Üç katmanlı testlerde birinci ve ikinci katmanlar arasındaki ortaklama katmanının kaldırılmasının pozitif doğruluk oranlarında etkili olduğu gözlemlenip dört katmanlı yapılarda da bu şekilde kullanılmasına karar verilmiştir. Bu bölümde sonuçları verilen sinir ağı dört gizli katmandan, bir düğüm atma katmanından ve bir tam bağlantılı katmandan oluşmaktadır. Birinci gizli katman 16 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına ve bir aktivasyon katmanına sahiptir. İkinci gizli katman 24 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına ve bir ortaklama katmanına sahiptir. Üçüncü gizli katman 36 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına sahiptir. Dördüncü gizli katman 48 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına ve bir aktivasyon katmanına sahiptir. Yeni sinir ağı yapıları Şekil 4.5 ve Şekil 4.6'da gösterilmiştir.

1. durum: Birinci ve ikinci katmanlarda normalizasyon katmanlarının kaldırılması
2. durum: İkinci ve üçüncü katmanlarda normalizasyon katmanlarının kaldırılması
3. durum: Düğüm atma katmanının ağ içerisinde kaldırılması



Şekil 4.5. 4 katmanlı sinir ağı gösterimi



Şekil 4.6. 2. duruma göre 4 katmanlı sinir ağı gösterimi

Çizelge 4.7’de KTP Veri Kümesi ile eğitilmiş dört katmanlı sinir ağının belirlenen durumlara göre doğruluk oranları gösterilmiştir. 1. durum ile 3 katmanlı ağ testinin 3. durumu karşılaştırıldığında katman arttırımının bir önceki katman arttırımında olduğu gibi pozitif doğruluk oranlarında düşüşe, negatif doğruluk oranlarında artışa neden olduğu gözlemlenmiştir. 2. durum ile 1. Durum karşılaştırıldığında negatif doğruluk oranlarının düştüğü ve normalizasyon katmanının ilk katmanda kullanılması ile pozitif doğruluk oranlarının arttığı gözlemlenmiştir. 3. durum ile 2. durum düğüm atma katmanının kaldırılmasının doğruluk oranlarında artışa sebep olacağı beklense de RGBD Veri Kümesi için pozitif doğruluk oranında düşüş görülmüştür.

Çizelge 4.7. KTP Veri Kümesi ile eğitilmiş 4 katmanlı sinir ağı sonuçları

|         | RGB-D İnsan |        | UR     |        |
|---------|-------------|--------|--------|--------|
|         | TP          | TN     | TP     | TN     |
| 1.durum | 0,7265      | 0,8936 | 0,6620 | 0,9219 |
| 2.durum | 0,8224      | 0,8340 | 0,7226 | 0,8736 |
| 3.durum | 0,6447      | 0,8851 | 0,7499 | 0,9331 |

Çizelge 4.8’de UR Düşme Veri Kümesi ile eğitilmiş dört katmanlı sinir ağının belirlenen durumlara göre doğruluk oranları gösterilmiştir. 1. durum ile 3 katmanlı ağ testinin 3. durumu karşılaştırıldığında katman arttırımının KTP Veri Kümesi’nde olduğu gibi pozitif doğruluk oranlarında düşüşe, negatif doğruluk oranlarında artışa neden olduğu gözlemlenmiştir. 2. durum ile 1. durum karşılaştırıldığında KTP Veri Kümesine ters olarak negatif doğruluk oranlarının arttığı ve normalizasyon katmanının ilk katmanda kullanılması ile pozitif doğruluk oranlarının azaldığı gözlemlenmiştir. 3. durum ile 2. durum düğüm atma katmanının kaldırılmasının doğruluk oranlarında artışa sebep olacağı beklense de bütün doğruluk oranlarında azalma gözlemlenmiştir. Bu durumda UR Veri Kümesi için eğitimde normalizasyon katmanının kullanılması ve düğüm atma katmanının kullanılması öğrenimi arttırmıştır.

Çizelge 4.8. UR Düşme Algılama Veri Kümesi ile eğitilmiş 4 katmanlı sinir ağı sonuçları

|         | RGB-D İnsan |        | KTP    |        |
|---------|-------------|--------|--------|--------|
|         | TP          | TN     | TP     | TN     |
| 1.durum | 0,8204      | 0,2170 | 0,8962 | 0,7634 |
| 2.durum | 0,6667      | 0,4298 | 0,8594 | 0,8194 |
| 3.durum | 0,6387      | 0,3489 | 0,7836 | 0,7560 |

Çizelge 4.9’da RGB-D Veri Kümesi ile eğitilmiş dört katmanlı sinir ağının doğruluk oranları gösterilmiştir. 1. durum ile 3 katmanlı ağ testinin 3. durumu karşılaştırıldığında katman arttırımının yalnızca UR pozitif doğruluk oranında artışa neden olduğu gözlemlenmiştir. 2. durum ile 1. durum karşılaştırıldığında UR Veri Kümesi’nde olduğu gibi negatif doğruluk oranlarının arttığı ve normalizasyon katmanının ilk katmanda kullanılması ile pozitif doğruluk oranlarının azaldığı gözlemlenmiştir. 3. durum ile 2. durum düğüm atma katmanının kaldırılmasının doğruluk oranlarında artışa sebep olacağı beklense de bütün doğruluk oranlarında azalma gözlemlenmiştir.

Çizelge 4.9. RGB-D İnsan Veri Kümesi ile eğitilmiş 4 katmanlı sinir ağı sonuçları

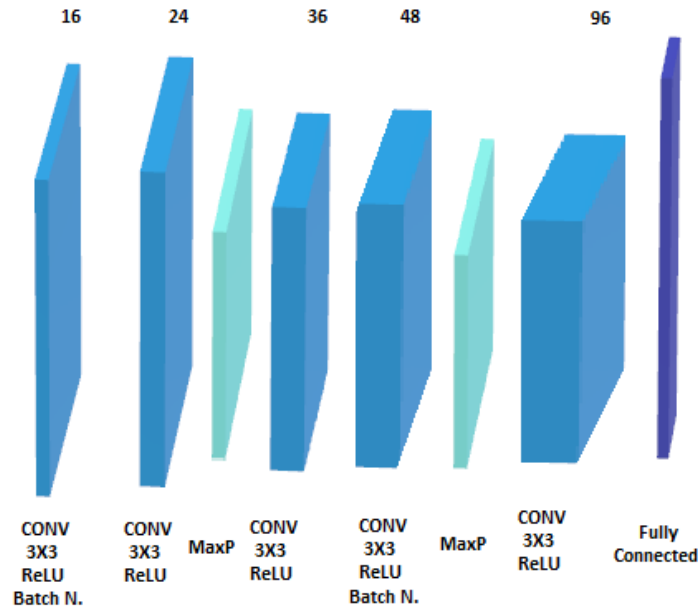
|         | UR     |        | KTP    |        |
|---------|--------|--------|--------|--------|
|         | TP     | TN     | TP     | TN     |
| 1.durum | 0,3741 | 0,8327 | 0,7947 | 0,9448 |
| 2.durum | 0,3048 | 0,8625 | 0,7711 | 0,9657 |
| 3.durum | 0,3949 | 0,8587 | 0,8036 | 0,9478 |

#### 4.4. Beş Katmanlı Eğitim

Dört katmanlı test sonuçları incelenip pozitif ve negatif doğruluk oranları karşılaştırıldığında yüksek değerlerin çoğunlukla 2. durumda olduğu görülmüş ve beş katmanlı testlerde 2. durum baz alınıp katman arttırımı yapılmasına karar verilmiştir.

Yapılan testler de katman artımı ile parametre değişimleri arasında ilişki olup olmadığına bakılmıştır.

Bu bölümde sonuçları verilen sinir ağı beş gizli katmandan, bir düğüm atma katmanından ve bir tam bağlantılı katmandan oluşur. Birinci gizli katman 16 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına sahiptir. İkinci gizli katman 24 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına ve bir ortaklama katmanına sahiptir. Üçüncü gizli katman 36 filtre genişliğine sahip bir konvolüsyon katmanına ve bir aktivasyon katmanına sahiptir. Dördüncü gizli katman 48 filtre genişliğine sahip bir konvolüsyon katmanına, bir normalizasyon katmanına, bir aktivasyon katmanına ve bir ortaklama katmanına sahiptir. Beşinci gizli katman katman 96 filtre genişliğine sahip bir konvolüsyon katmanına ve bir aktivasyon katmanına sahiptir. Yeni sinir ağı yapısı Şekil 4.7’de gösterilmiştir.



Şekil 4.7. 5 katmanlı sinir ağı gösterimi

Çizelge 4.10’da KTP Veri Kümesi ile eğitilmiş beş katmanlı sinir ağının doğruluk oranları gösterilmiştir. InitialLearnRate parametresi 0.01, 0.05, 0.005 değerleri arasında değiştirilirken L2Regularization parametresi 0.0001 0.001 değerleri arasında değiştirilmiştir. L2Regularization parametresi 0.0001 değerinde iken, InitialLearnRate parametresi 0.01

değerinden 0.005 değerine değiştirildiğinde pozitif doğruluk oranlarında ciddi bir düşüş olduğu görülmüştür. L2Regularization parametresi 0.0001 değerinden 0.001 değerine değiştirildiğinde RGBD Veri Kümesine ait pozitif doğruluk oranlarında ciddi bir düşüş gözlemlenmektedir. UR Veri Kümesi'ne ait pozitif doğruluk oranlarında da bir düşüş gözlemlenmiştir. L2Regularization parametresi 0.001 değerinde iken, InitialLearnRate parametresi 0.01'den 0.05'e değiştirildiğinde RGBD Veri Kümesi içi artış, UR veri kümesi için düşüş olduğu gözlemlenmiştir. Bu durum L2Regularization ve InitialLearnRate parametrelerin artırılmasının bazı filtrelerin öğrenilmesinde sorun yarattığını göstermiştir. Pozitif doğruluk oranları için en iyi sonuçlar tablonun ilk satırındaki parametre değerleri için elde edilmiştir.

Çizelge 4.10. RGB-D İnsan Veri Kümesi ile eğitilmiş 5 katmanlı sinir ağı sonuçları

|   | RGB-D İnsan |        | UR     |        |
|---|-------------|--------|--------|--------|
|   | TP          | TN     | TP     | TN     |
| InitialLearnRate=0.01<br>L2Regularization=0.0001  | 0,9361      | 0,5957 | 0,7516 | 0,8810 |
| InitialLearnRate=0.005<br>L2Regularization=0.0001 | 0,6467      | 0,4891 | 0,4891 | 0,9257 |
| InitialLearnRate=0.01<br>L2Regularization=0.001   | 0,5888      | 0,9149 | 0,7101 | 0,9182 |
| InitialLearnRate=0.005<br>L2Regularization=0.001  | 0,6427      | 0,9191 | 0,6207 | 0,8885 |
| InitialLearnRate=0.05<br>L2Regularization=0.001   | 0,7405      | 0,8553 | 0,5267 | 0,9331 |

Çizelge 4.11'de UR Düşme Veri Kümesi ile eğitilmiş beş katmanlı sinir ağının doğruluk oranları gösterilmiştir. L2Regularization parametresi 0.0001 değerinde iken, InitialLearnRate parametresi 0.01 değerinden 0.005 değerine değiştirildiğinde, genel olarak doğruluk oranlarında düşüş olduğu görülmüştür. L2Regularization parametresi 0.0001 değerinden 0.001 değerine değiştirildiğinde yalnızca RGBD Veri Kümesi'ne ait pozitif doğruluk oranında düşüş, olduğu görülmüştür. L2Regularization parametresi 0.001 değerinde iken, InitialLearnRate parametresi 0.01 değerinden 0.05 değerine değiştirildiğinde doğruluk oranlarında artış olduğu görülmüştür. Doğruluk oranları karşılaştırıldığında en iyi sonuçlar tablonun beşinci satırında verilen parametre değerleri için elde edilmiştir



Çizelge 4.11. UR Düşme Algılama Veri Kümesi ile eğitilmiş 5 katmanlı sinir ağı sonuçları

|   | RGB-D İnsan |        | KTP    |        |
|---|-------------|--------|--------|--------|
|   | TP          | TN     | TP     | TN     |
| InitialLearnRate=0.01<br>L2Regularization=0.0001  | 0,7625      | 0,3489 | 0,7853 | 0,7030 |
| InitialLearnRate=0.005<br>L2Regularization=0.0001 | 0,5549      | 0,5191 | 0,7719 | 0,6687 |
| InitialLearnRate=0.01<br>L2Regularization=0.001   | 0,6527      | 0,4894 | 0,8358 | 0,7090 |
| InitialLearnRate=0.005<br>L2Regularization=0.001  | 0,6507      | 0,3362 | 0,8861 | 0,5709 |
| InitialLearnRate=0.05<br>L2Regularization=0.001   | 0,6627      | 0,6383 | 0,8376 | 0,8254 |

Çizelge 4.12' de RGB-D Veri Kümesi ile eğitilmiş beş katmanlı sinir ağının doğruluk oranları gösterilmiştir. L2Regularization parametresi 0.0001 değerinde iken, InitialLearnRate parametresi 0.01 değerinden 0.005 değerine değiştirildiğinde pozitif doğruluk oranlarında artış olduğu görülmüştür. L2Regularization parametresi 0.0001 değerinden 0.001 değerine değiştirildiğinde pozitif doğruluk oranlarında artış olduğu görülmüştür. L2Regularization parametresi 0.001 değerinde iken, InitialLearnRate parametresi 0.01 değerinden 0.05 değerine değiştirildiğinde pozitif doğruluk oranlarında artış olduğu görülmüştür. Eğitim de kullanılan imge sayısında azalma olmasına rağmen parametrelerin artışına göre doğruluk oranlarında artış görülmüşse de doğruluk oranları karşılaştırıldığında en iyi sonuçlar tablonun ilk satırında verilen parametre değerleri için elde edilmiştir.

Çizelge 4.12. RGB-D İnsan Veri Kümesi ile eğitilmiş 5 katmanlı sinir ağı sonuçları

|   | UR     |        | KTP    |        |
|---|--------|--------|--------|--------|
|   | TP     | TN     | TP     | TN     |
| InitialLearnRate=0.01<br>L2Regularization=0.0001  | 0,2954 | 0,8885 | 0,7183 | 0,9619 |
| InitialLearnRate=0.005<br>L2Regularization=0.0001 | 0,3218 | 0,7918 | 0,8480 | 0,9209 |
| InitialLearnRate=0.01<br>L2Regularization=0.001   | 0,3904 | 0,8253 | 0,8056 | 0,9321 |
| InitialLearnRate=0.005<br>L2Regularization=0.001  | 0,4193 | 0,7844 | 0,8224 | 0,9209 |
| InitialLearnRate=0.05<br>L2Regularization=0.001   | 0,4171 | 0,8216 | 0,8310 | 0,9164 |

#### 4.5. Üç Veri Kümesinin Birleştirilerek Eğitimde Kullanılması

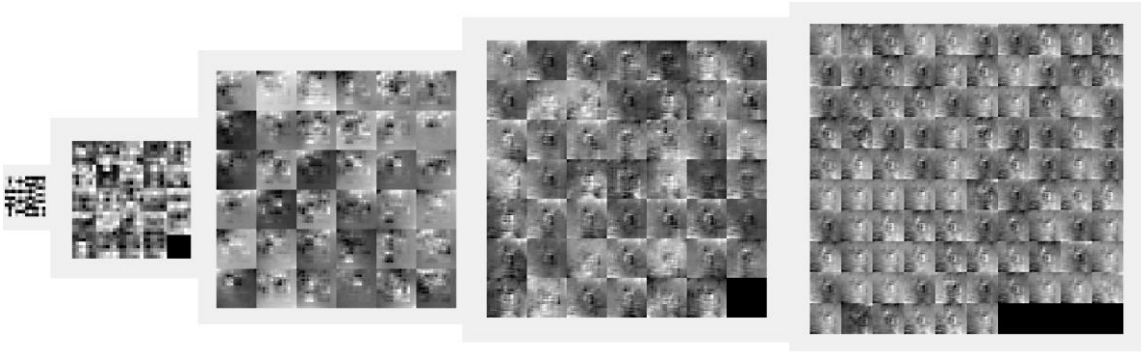
Beş katmanlı sinir ağı eğitimleri her bir veri kümesi için ayrı ayrı yapıldıktan sonra veri kümelerinin birleştirilip eğitimin tekrarlanmasına karar verilmiştir. Birleştirilen veri kümesinde pozitif ve negatif örneklerin %70'i eğitim ve %30'u test için kullanılmıştır. Toplamda pozitif ve negatif karışık olarak 291407 imge eğitimde, pozitif ve negatif karışık olarak 124889 imge doğrulamada kullanılmıştır.

Yeni testin Şekil 4.7'de görülen ağ yapısı ile yapılmasına karar verilmiştir. İlk olarak ağ yapısı sabit tutulup parametreler beş katmanlı testlerde olduğu gibi değiştirilmiştir.

Çizelge 4.13'de birleştirilmiş veri kümesinin parametrelere göre doğruluk oranları gösterilmiştir. Yukarıdaki testlerde olduğu gibi KTP ve RGBD veri kümelerinde daha iyi sonuç veren parametreler veri kümeleri birleştirildikten sonra da diğer parametrelere göre iyi sonuçlar vermiştir.

Çizelge 4.13. Birleştirilmiş veri kümeleri 5 katmanlı sinir ağı sonuçları

|   |        |
|---|--------|
|   |        |
| InitialLearnRate=0.01<br>L2Regularization=0.0001  | 0.9819 |
| InitialLearnRate=0.005<br>L2Regularization=0.0001 | 0.9816 |
| InitialLearnRate=0.005<br>L2Regularization=0.001  | 0.9807 |
| InitialLearnRate=0.05<br>L2Regularization=0.001   | 0.9702 |

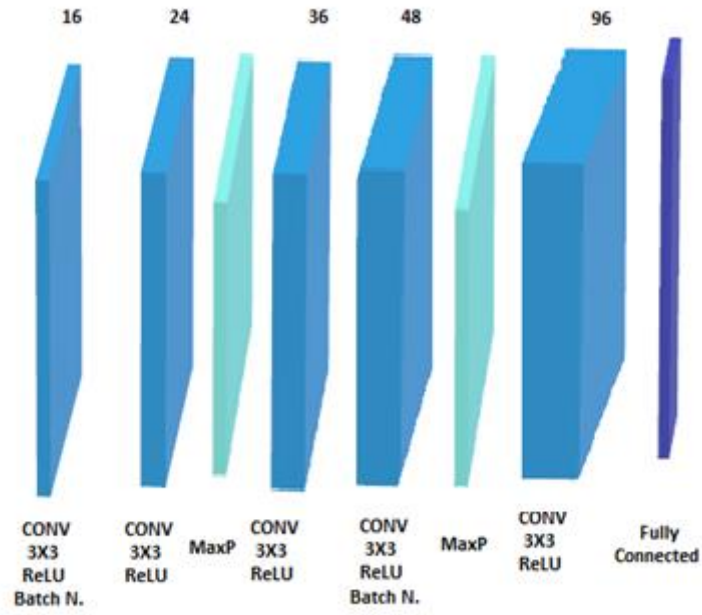


Şekil 4.8. 5 katmanlı sinir ağı konvolüsyonel katman özellik görüntüleri

Şekil 4.8. Sırasıyla 1., 2., 3., 4. ve 5. konvolüsyon katmanlarının özellik görüntüleri gösterilmiştir.

İkinci test InitialLearnRate parametresi 0.01 değerine ve L2Regularization parametresi 0.0001 değerinde sabit tutulup, sinir ağı katman sayısının önceki testlerde olduğu gibi değiştirilerek yapılmıştır.

1. durum 2 katmanlı mimari yapısının denenmesi
2. durum 3 katmanlı mimari yapısının denenmesi
3. durum 4 katmanlı mimari yapısının denenmesi
4. durum 5 katmanlı mimari yapısının denenmesi
5. durum 5 katmanlı mimari yapısı ile maksimum ortaklama boyutu sonrası boyutun aynı kalmasının denenmesi (Şekil 4.9)



Şekil 4.9. 5. duruma göre sinir ağı gösterimi

Çizelge 4.14. Birleştirilmiş veri kümeleri farklı katman sayılarına göre sinir ağı sonuçları

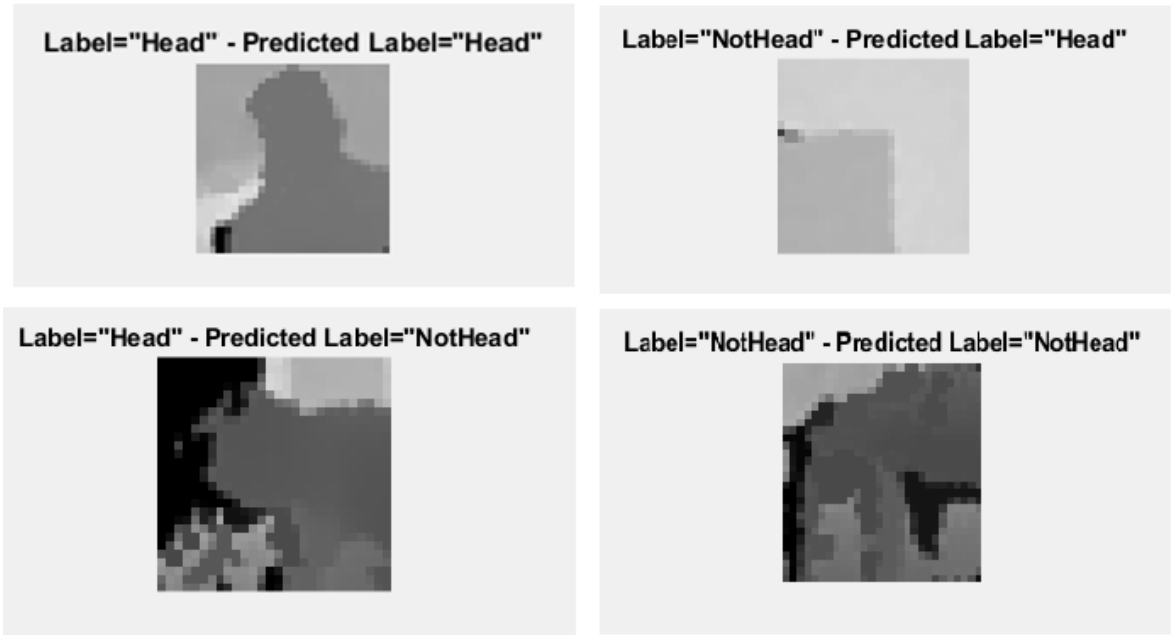
|          |        |
|----------|--------|
|          |        |
| 1. durum | 0,9740 |
| 2. durum | 0,9609 |
| 3. durum | 0,9805 |
| 4. durum | 0,9816 |
| 5. durum | 0,9856 |

Sonuçlar karşılaştırıldığında birleştirilmiş veri kümesi için sinir ağı katman sayısının artırılmasının ağı eğitilmesinde olumlu rol oynadığı görülmektedir. 5. durumun 4. duruma göre daha iyi sonuç verdiği görülmüştür. Bu duruma göre imge boyutlarının artırılması ile sonuçların iyileşebileceği düşünülmektedir.

Test veri kümesinin, eğitim veri kümesi ile aynı kaynakları kullanmasının imgelerin arka planlarının ya da imgelerde bulunan insanların ezberlenmiş olabileceği düşünülmüştür. Bu sebeple 4. duruma göre eğitilmiş ağın başka bir veri kümesi ile daha test edilmesine karar verilmiştir. 4. durum ile “Veri kümelerinin hazırlanması” başlığı altında bahsedilen Watch-n-patch Veri Kümesi test edilmiş ve sonuç Çizelge 4.15’de gösterilmiştir. Çıkan sonuca göre yapay sinir ağının veri kümelerini bir şekilde ezberlemiş olabileceği düşünülmüştür.

Çizelge 4.15. 4.duruma göre eğitilmiş sinir ağında Watch-n-Patch Veri Kümesi sonuçları

| Watch-n-Patch Veri Kümesi |
|---------------------------|
| 0.6053                    |



Şekil 4.10. Üç veri kümesinin birleştirilerek eğitiminden çıkan sonuç görselleri

4. duruma göre elde edilen sonuçlarda UR Düşme Veri Kümesi’nden alınan görseller incelenmiştir. İlk görsel sinir ağının doğru etiketlediği bir baş objesini simgelemektedir. İkinci görsel baş objesinin bulunmadığı bir görselin baş olarak etiketlenmesidir. İkinci imgeyi sinir ağının kafa olarak etiketlenmesinin sebebi imgenin bir bölümünü kaplayan ve aynı uzaklık değerlerine sahip bir grup pikselden kaynaklı olabilir. Üçüncü görsel, içinde baş objesinin bulunduğu resmin yanlış etiketlendiği bir durumu göstermektedir. Arka planda bulunan gürültü ya da eğitim sırasında aynı pozisyonda bulunan insan sayısının az olması

sebepler olarak gösterilebilir. Dördüncü görsel ise içinde baş objesinin bulunmadığı imgenin doğru etiketlenmesidir. İnsan gözü ile bakıldığında imgede baş var denebilirken sinir ağı sayesinde imgede insan olmadığı anlaşılmaktadır.

Duyarlılık-doğruluk (precision-recall) eğrisi sınıflandırma tabanlı yöntemlerin başarılarının ölçülmesinde kullanılan hassasiyet ve geri çağırma değerlerinin birbirine oranının kullanıldığı bir yöntemdir. Bu iki değer için hata matrisi kullanılmaktadır.

Çizelge 4.16. Hata matrisi

|                                |                     |                             |                                |
|--------------------------------|---------------------|-----------------------------|--------------------------------|
|                                | <b>Gerçek Değer</b> |                             |                                |
| <b>Tahmini Değer</b>           |                     | <b>Pozitif Değer (Head)</b> | <b>Negatif Değer (NotHead)</b> |
| <b>Pozitif Değer (Head)</b>    |                     | Gerçek Pozitif (TP)         | Yanlış Pozitif (FP)            |
| <b>Negatif Değer (NotHead)</b> |                     | Yanlış Negatif (FN)         | Gerçek Negatif (TN)            |

Çizelge 4.16.'da gösterilen hata matrisi gerçek olan imge etiketleri ile yapay sinir ağının çıktısı olarak tahmin edilen etiketleri karşılaştırmaktadır.

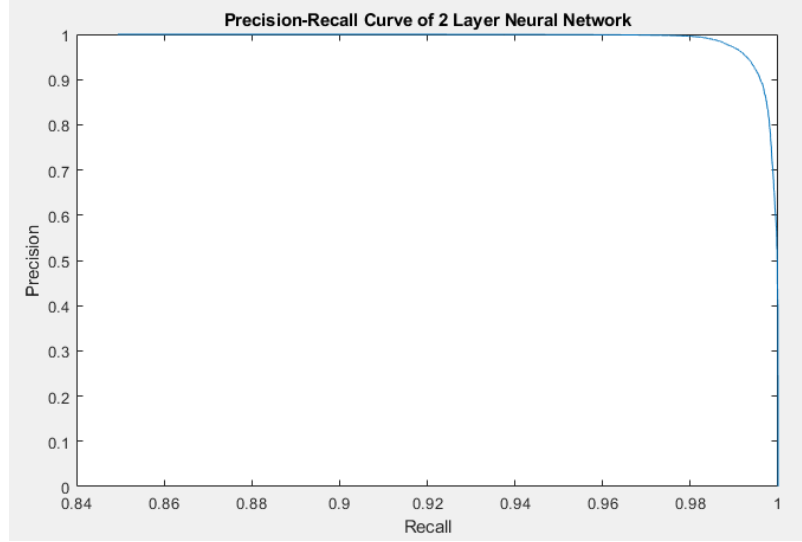
Duyarlılık-doğruluk eğrisi hesaplanırken duyarlılık pozitif olarak tahmin edilen bir durumdaki başarıyı, doğruluk ise pozitif durumların ne kadar başarılı tahmin edildiğini gösterir. Hata matrisine göre bu değerleri matematiksel olarak yazarsak

$$\text{Hassasiyet} = \frac{TP}{TP + FN} \quad (4.4)$$

$$\text{Geri çağırma} = \frac{TP}{TP + FN} \quad (4.5)$$

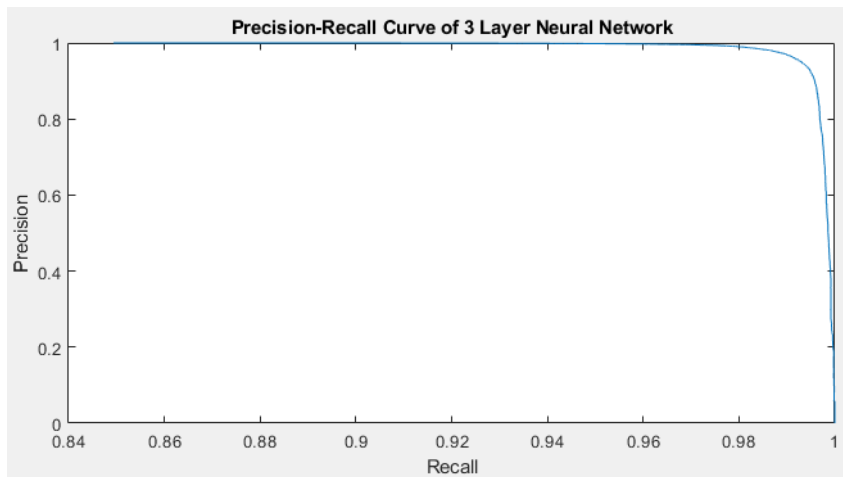
Yapay sinir ağında varsayılan olarak bir imgenin “baş” ya da “baş-değil” olarak etiketlenmesi için 0.5 değeri eşik değeri olarak kullanılır. Duyarlılık doğruluk eğrisini hesaplamak için ağ çıktısında alınan doğrulama kümesinin skor değerleri 0 ile 1 değerleri arasında sıralanmış ve her bir farklı değer aynı doğrulama kümesi için eşik değeri olarak kullanılmıştır

Duyarlılık doğruluk eğrisi 2 katmanlı, 3 katmanlı, 4 katmanlı ve 5 katmanlı mimariler için hesaplanmıştır.



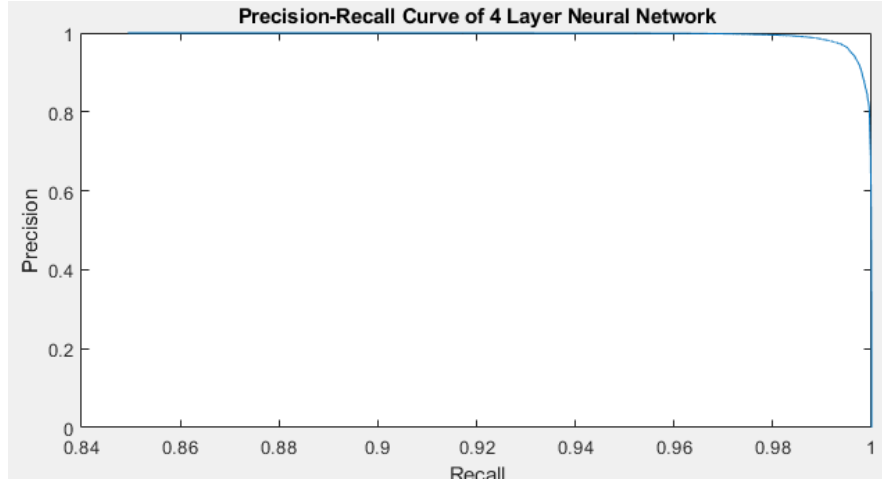
Şekil 4.11. 2 katmanlı sinir ağı mimarisi için duyarlılık-doğruluk eğrisi

Şekil 4.11.'de gösterilen duyarlılık doğruluk eğrisinde 124889 adet etiketlenmiş imge ve 90033 adet eşik değeri vardır. Her eşik değeri için tp, fp ve fn değerleri hesaplanmıştır. 4.4. denklemleri ve 4.5. denklemlerine göre hassasiyet ve geri çağırma değerleri hesaplanıp duyarlılık-doğruluk eğrisi elde edilmiştir.



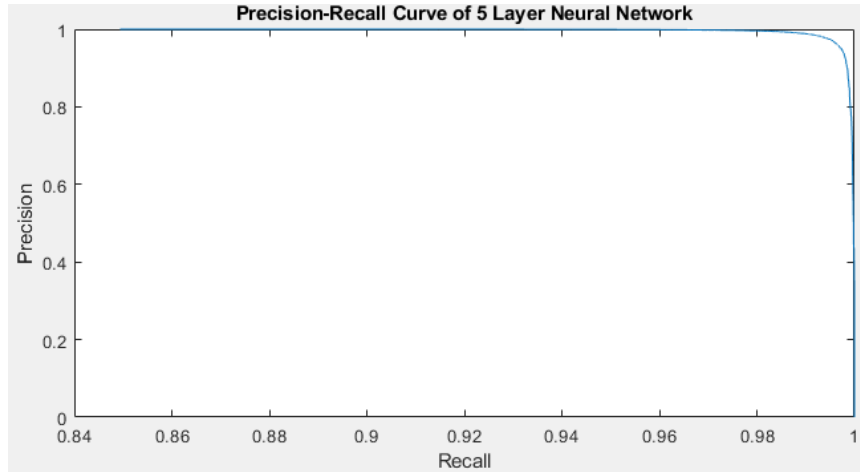
Şekil 4.12. 3 katmanlı sinir ağı mimarisi için duyarlılık-doğruluk eğrisi

Şekil 4.12.'de gösterilen duyarlılık doğruluk eğrisinde 124889 adet etiketlenmiş imge ve 71728 adet eşik değeri vardır. Diğer katmanlara göre en düşük doğruluk oranına sahip 3 katmanlı mimaride hassasiyet oranının diğerlerine oranla düşük olduğu görülmüştür.



Şekil 4.13. 4 katmanlı sinir ağı mimarisi için duyarlılık-doğruluk eğrisi

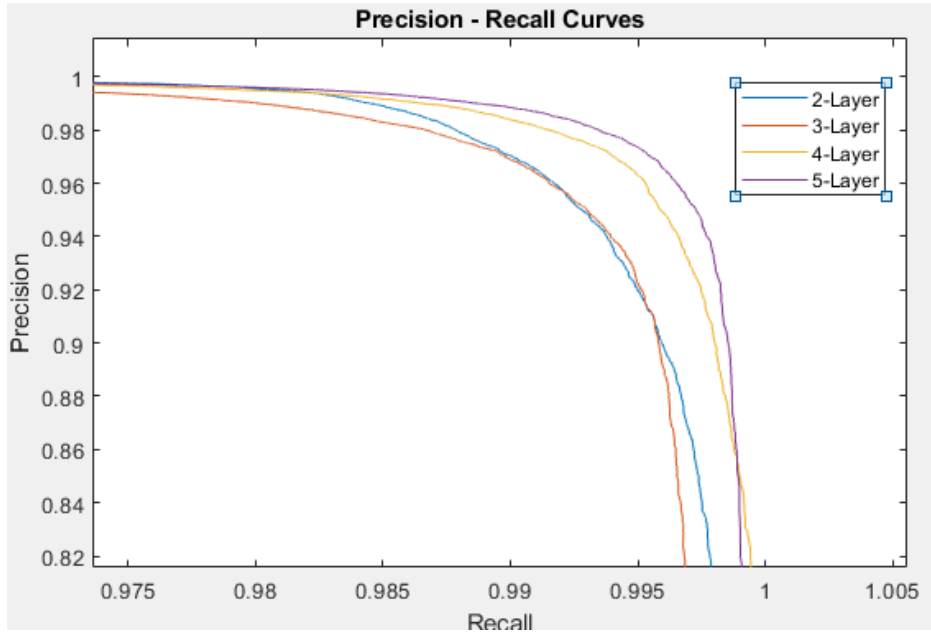
Şekil 4.13.'de gösterilen duyarlılık doğruluk eğrisinde 124889 adet etiketlenmiş imge ve 87719 adet eşik değeri vardır.



Şekil 4.14. 5 katmanlı sinir ağı mimarisi için duyarlılık-doğruluk eğrisi

Şekil 4.14.'de gösterilen duyarlılık doğruluk eğrisinde 124889 adet etiketlenmiş imge ve 79796 adet eşik değeri vardır. Diğer katmanlara göre en yüksek doğruluk oranına sahip 5 katmanlı mimaride hassasiyet oranının diğerlerine oranla daha yüksek olduğu görülmüştür.





Şekil 4.15. Katmanlara ait duyarlılık-doğruluk eğrileri

Sınıflandırmanın kullanıldığı probleme göre duyarlılık-doğruluk eğrisinden eşik değeri seçilebilir. İnsan algılamanın hayati olduğu ve imgedeki insanı algılayamamanın kötü sonuçlar doğuracağı durumlarda hassiyetten ödün verilip geri çağırma değerinin yüksek olduğu eşik değerleri kullanılabilir.

## 5. SONUÇ VE ÖNERİLER

İmge işleme ile insan sezimi üzerine literatürde bir çok çalışma yapılmıştır. İnsan sezimine bir çok farklı alanda ihtiyaç duyulduğu ve bazı durumlarda gerçek zamanlı insan seziminin hayati bir önem kazandığı görülmektedir.

İmge işleme ile insan sezimi çalışmalarında genellikle filtre bazlı sınıflandırma yöntemleri kullanılmaktadır. Gerçek zamanlı işlemler için derin sinir ağları kullanılmasının daha hızlı sonuçlar verdiği görülmektedir. En sık kullanılan görüntü yapısı ise daha fazla özellik içermesinden dolayı renkli imgelerdir. Renkli imgelere erişilemeyeceği durumlarda derinlik imgelerinin önemi artmaya başlamıştır. Bu tez çalışmasında yalnızca derinlik imgeleri kullanılarak derin sinir ağları ile insan sezimi üzerine çalışılmıştır. Çalışma kapsamında dört açık kaynaklı veri kümesi ve MATLAB Sinir Ağı Paketi (Neural Network Toolbox) kullanılmıştır. Üç farklı veri kümesine ait imgeler üzerinden ilgili alanlar el ile etiketlenip kesilmiştir. Kesilen imgeler kaydırma, döndürme, aynalama gibi işlemler ile çoğaltılmıştır. Testler sırasında veri kümelerine ait pozitif ve negatif imgelere ait doğruluk oranları farklı değişkenlerle elde edilmiş ve incelenmiştir.

Hazırlanan farklı katmanlara sahip yapay sinir ağları ile belirlenen parametreler değiştirilip sonuçlar elde edilmiştir. Elde edilen sonuçlar incelendiğinde genellikle pozitif doğruluk oranı arttıkça negatif doğruluk oranının düştüğü gözlemlenmiştir. Katman artırımının bazı veri kümelerinde, pozitif doğruluk oranlarında artışa bazılarında ise düşüşe sebep olduğu gözlemlenmiştir. Katman sayıları sabit tutulduğunda eğitimde kullanılan imge sayısını arttırdığında parametrelerin güncellenmesi için kullanılan başlangıç parametrelerinin de artırılması gerektiği görülmüştür.

İleride yapılacak çalışmalar için, girdi olarak kullanılan imgelerin boyutlarının büyütülmesi ile sinir ağı eğitimi için daha iyi sonuçlar elde edilmesi planlanmaktadır. Ayrıca katmanlarda kullanılan filtre sayılarının artırılması ve sonuçların incelenmesi planlanmaktadır.

## KAYNAKLAR DİZİNİ

- Ballotta D., Borghi G., & Vezzani R., Cucchiara R., 2018a. Head detection with depth images in the wild .p. 56-63.
- Ballotta D. Borghi G., Vezzani R., Cucchiara R., 2018b, Fully convolutional network for head detection with depth images, 24th International Conference on Pattern Recognition (ICPR)
- Borghi G., Venturelli M., Vezzani R., Cucchiara R., 2017, Poseidon: Face-from-depth for driver pose estimation, Computer Vision and Pattern Recognition (CVPR)
- Chen S., Bremond F., Nguyen H., Thomas H., 2016 , Exploring depth information for head detection with depth images, Advanced Video and Signal Based Surveillance (AVSS), pp. 228-234.
- Eitel A., Springberg J. T., Spinello L., Riedmiller M. A., Burgard W, 2015, Multimodal deep learning for robust {rgb-d} object recognition, IROS, p. 681-687
- Ikemura, S., Fujiyoshi, H., 2010 Real-time human detection using relational depth similarity features. Computer Vision–ACCV, p. 25–38.
- Karpathy A., 2015a, CS231n Convolutional Neural Networks for Visual Recognition Introduction to Neural Networks <http://cs231n.github.io/neural-networks-1/> erişim tarihi: 22.07.2018
- Karpathy A., 2015b, CS231n Convolutional Neural Networks for Visual Recognition Linear Classification <http://cs231n.github.io/linear-classify/> erişim tarihi: 16.07.2018
- Karpathy A., 2015c, CS231n Convolutional Neural Networks for Visual Recognition Setting Up The Data And The Model, <http://cs231n.github.io/neural-networks-2/> erişim tarihi: 22.07.2018
- Khan M. H., Shirahama K., Farid M. S., Grzegorzec M., 2016, Multiple human detection in depth images, Multimedia Signal Processing (MMSP), p. 1-6.
- Kwolek B., Kepski M., 2014, Human fall detection on embedded platform using depth maps and wireless accelerometer, Computer Methods and Programs in Biomedicine, 117, 3,p. 489-501
- Lecun Y., Bottou L., Bengio Y., Haffner P., 1998 Gradient-based learning applied to document recognition, IEEE, 86, 11, p. 2278-2324.
- Li F., Johnson J., Yeung S., 2018a, Backpropagation and neural networks [http://cs231n.stanford.edu/slides/2018/cs231n\\_2018\\_lecture04.pdf](http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture04.pdf) erişim tarihi: 22.07.2018

### KAYNAKLAR DİZİNİ (devam)

- Li F., Johnson J., Yeung S., 2018b, Convolutional Neural Networks, [http://cs231n.stanford.edu/slides/2018/cs231n\\_2018\\_lecture05.pdf](http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture05.pdf) erişim tarihi: 25.07.2018
- Lison P., 2012, An introduction to machine learning <http://folk.uio.no/plison/pdfs/talks/machinelearning.pdf> erişim tarihi: 4.05.2018
- Munaro M., Basso F., Manegatti. E., 2012, People tracking within groups with RGB-D data, Intelligent Robots and Systems, p. 2101-2107
- Nghiem A. T., Auvinet E., Meunier J., 2012 Head detection using Kinect camera and its application to fall detection, Signal Processing and their Applications (ISSPA), p. 164-169.
- Spinello L., Arras K. O., Triebel R., Siegwart R., 2010, A layered approach to people detection in 3d range data, Conference on Artificial Intelligence, AAAI
- Spinello L., Arras K., 2011, People detection in RGB-D data, Intelligent Robots and Systems, p. 3838-3843
- Xia L., Chen C., Aggarwal J. K., 2011, Human detection using depth information by kinect, Computer Vision and Pattern Recognition (CVPR), p. 15-22.
- Xu F., Fujimura K., 2003, Human detection using depth and gray images, IEEE, p. 115-121
- Wang A., Lu J., Cai J., Cham T., Wang G., 2015, Large-margin multi-modal deep learning for rgb-d object recognition, IEEE, 17, 11, p. 1887-1898
- MathWorks, 2016a, trainingOptions <https://www.mathworks.com/help/nnet/ref/trainingoptions.html> erişim tarihi: 20.04.2018
- MathWorks, 2016b, imageInputLayer <https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.imageinputlayer.html> erişim tarihi: 19.04.2018
- MathWorks, 2016c, convolution2dLayer, <https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.convolution2dlayer.html> erişim tarihi: 18.04.2018
- MathWorks, 2016d, reluLayer, <https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.reluLayer.html> erişim tarihi: 18.04.2018

**KAYNAKLAR DİZİNİ (devam)**

MathWorks, 2016e, maxPooling2dLayer,

<https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.maxpooling2dlayer.html>

erişim tarihi: 18.04.2018

MathWorks, 2016f, fullyConnectedLayer,

<https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.fullyconnectedlayer.html>

erişim tarihi: 18.04.2018

MathWorks, 2016g, DropoutLayer

<https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.dropoutlayer.html> erişim

tarihi: 19.04.2018

MathWorks, 2016h, softmaxLayer

<https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.softmaxlayer.html> erişim

tarihi: 18.04.2018

MathWorks, 2016i, classificationLayer

[https://www.mathworks.com/help/nnet/ref/classificationlayer.html?s\\_tid=doc\\_ta](https://www.mathworks.com/help/nnet/ref/classificationlayer.html?s_tid=doc_ta)

erişim tarihi: 20.04.2018

MathWorks, 2017, batchNormalizationLayer

[https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.batchnormalizationlayer.h](https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.batchnormalizationlayer.html)

[tml](https://www.mathworks.com/help/nnet/ref/nnet.cnn.layer.batchnormalizationlayer.html) erişim tarihi: 18.04.2018