

**DERİN ÖĞRENME İLE GRUP HAREKETLERİNİN SABİT
RESİM ÜZERİNDEN TANINMASI**

**RECOGNITION OF GROUP ACTIVITIES ON STILL
IMAGES USING DEEP LEARNING**

ANIL ATVAR

Doç. Dr. NAZLI İKİZLER CİNBİŞ

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
Bilgisayar Mühendisliği Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2018

ANIL ATVAR'ın hazırladığı "**Derin Öğrenme ile Grup Hareketlerinin Sabit Resim Üzerinden Tanınması**" adlı bu çalışma aşağıdaki jüri tarafından **BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Pınar DUYGULU ŞAHİN

Başkan


.....

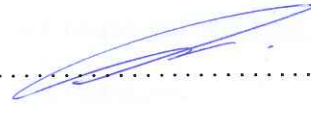
Doç. Dr. Nazlı İKİZLER CİNBİŞ

Danışman


.....

Doç. Dr. Ahmet Burak CAN

Üye


.....

Doktor Öğretim Üyesi Emre AKBAŞ

Üye


.....

Doç. Dr. İbrahim Aykut ERDEM

Üye


.....

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU
Fen Bilimleri Enstitüsü Müdürü

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

22/09/2018

ANIL ATVAR

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarda kullanıma açma iznini Hacettepe Üniversitesi'ne verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.
- Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ay ertelenmiştir.
- Tezim ile ilgili gizlilik kararı verilmiştir.

05./.../..2018

Anıl ATVAR

ÖZET

DERİN ÖĞRENME İLE GRUP HAREKETLERİNİN SABİT RESİM ÜZERİNDEN TANINMASI

Anıl ATVAR

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Danışmanı: Doçent Nazlı İKİZLER CİNBİŞ

Eylül 2018, 64 sayfa

Tez kapsamında odaklanılan temel problem, insana ait grup aktivite bilgilerinin sabit resimler üzerinden anlamlandırılması ve bu hareket bilgilerinin sınıflandırılmasıdır. Aktivite bilgilerinin çoğunlukla zaman çizelgesine bağlı olarak analiz edildiğinde anlamlı hale geliyor olması üstünde uğraşılan problemi zorlaştıran nedenlerin başında gelmektedir. Aynı aktivite içerisinde bulunmayan ancak görüntünün elde edildiği anda yan yana gelen iki insana ait durum bilgisi sabit görüntü üzerinden analiz edildiğinde büyük ihtimal ile aynı aktiviteye dahil olarak sınıflandırılacaktır. Üzerinde uğraşılan problemin bu ve benzer sebeplerden dolayı sahip olduğu zorlukların üstesinden gelebilmek adına sabit görüntüde bulunan tekil insan bilgilerinin yüksek doğrulukla görüntüden elde edilmesi ve bu bilgilerin yüksek doğrulukta sınıflandırılması büyük önem arz etmektedir.

Son yıllarda obje tespit ve sınıflandırma alanlarında başarılı sonuçlar üreten derin öğrenme teknikleri, tez kapsamında uğraşılan problemin başarısını artırmak için de tercih edilen yöntemlerin başında gelmektedir. Görüntü üzerinde bulunan insan ve insanların oluşturduğu grupların tespit ve sınıflandırılması aşamalarında insan ve grupları temsil eden özniteliklerin iyi seçilmiş olması önem arz etmektedir. Derin öğrenme tekniklerinin başarısı da bu aşamada diğer yöntemlere göre avantaj sağlamaktadır. Derin öğrenme modeline girdi olarak verilen eğitim

kümesi üzerinde öğrenme süreci boyunca eğitim kümesinde sınıflandırılması istenilen objelere yönelik öznitelikler model tarafından otomatik olarak üretilmektedir.

Problem çözümüne yönelik olarak derin öğrenme yöntemleri seçildiğinde de bu yöntemlere ait başka zorluklar önümüze çıkmaktadır. Bunların en başında probleme uygun olan bir derin öğrenme modeli seçimi gelmektedir. Tez kapsamında uğraşılan problem, sınıflandırma temelli bir problem olduğundan dolayı derin öğrenme yöntemi olarak Evrişimli Sinir Ağları (CNN) [1] tercih edilmiştir. Bu yöntemde kullanılacak olan model olarak da son yıllarda kompleks sınıflandırma problemleri için çokça tercih edilen ResNet [2] mimarisi tercih edilmiştir. Derin öğrenme alanında karşılaşılan diğer bir zorluk da eğitim aşaması için tercih edilecek görüntü denektaşı veri kümelerinin büyüklüğünün ve çeşitliliğinin yüksek olması gerekliliğidir. Tez kapsamında da en çok zorlanılan konu, tercih edilen SGD [3] denektaşı veri kümesinde, tekil insana ait yönelim ve grup aktivite sınıflarının yeterli sayı ve çeşitlilikte örnek içermemesinden kaynaklı sınıflandırma başarılarının daha yukarılara taşınamamasıdır. Bu zorluğun üstesinden gelebilmek adına grup aktivite bilgilerine ek olarak poz ve bölüt bilgileri de kullanılmıştır. Bu bilgilerin de derin öğrenme sürecine dahil olabilmeleri adına füzyon işlemleri gerçekleştirilmiş ve elde edilen sonuçlar gözlemlenmiştir.

Tez kapsamında derin öğrenme tekniklerinin tercih edilmesi ile derin olmayan diğer yöntemlere göre daha başarılı tespit ve sınıflandırma başarısının elde edildiği gözlemlenmiştir.

Anahtar Kelimeler: grup aktivitesi tanıma, derin öğrenme, ince ayarlama, sınıflandırma

ABSTRACT

RECOGNITION OF GROUP ACTIVITIES ON STILL IMAGES USING DEEP LEARNING

Anıl ATVAR

Master, Computer Engineering Department

Supervisor: Assoc. Prof. Dr. Nazlı İKİZLER CİNBİŞ

September 2018, 64 pages

The main problem focused in this thesis is inferring the group activity information from still images and classifying them. Activity information is often meaningful when analyzed based on the timeline. This is one of the reasons that complicates the problem. For example, if two people are not in the same activity but are standing side-by-side in a still image, they will most likely be classified to have the same activity. For reasons like this, classification of group activities in still images is a challenging problem. To overcome these difficulties, detection and classification of individual human from still image should be done with high accuracy. At the same time, this approach constitutes the first part of the thesis.

Deep learning techniques, which are yielding successful results in object detection and classification problems in recent years, are preferred methods to solve the problems dealt with in the scope of this thesis. It is important that the features should be well chosen to represent individual humans and groups in images. The success of deep learning techniques also provides more advantages than other methods at this point. These features can be automatically learned by the model in deep learning approaches.

There are additional challenges associated with choosing deep learning methods as a base problem solvers. At the top of these difficulties is deciding a deep learning model that is

suitable for problem. Since we are dealing with classification problems, Convolutional Neural Networks (CNN) [1] have been chosen to be adapted for group activity recognition. As a model to be used in this method, ResNet [2] architecture, which is preferred for complex classification problems in recent years, has been preferred. Another difficulty in the field of deep learning is that decide size and variety of dataset . SGD [3] was preferred as a dataset in the thesis. The most challenging issue in the thesis is that in the preferred dataset [3], the singular human orientation and group activity classes are not able to increase the classification performance due to the lack of sufficient numbers and diversity of samples. In addition to group activity information, joint and segment informations were also used to overcome these difficulties. In order to merge these informations into the deep learning process, fusion processes were performed and then results were observed.

Within the thesis, it can be observed that the success of the detection and classification is achieved by the choice of the deep learning techniques.

Keywords: group activity recognition, deep learning, fine-tuning, classification

TEŐEKKÜR

Tez danıřmanım Doç. Dr. Nazlı İvizler Cinbiř'e lisans/lisansüstü öğrenim hayatımdaki ve tez çalışmamdaki gösterdiği destek, bilgi ve anlayışından dolayı teşekkürlerimi sunuyorum.

Hiçbir zaman desteğini esirgemeyen ve hayatımı daima anlamlı kılan eşim Nihan Atvar'a en derin teşekkürlerimi sunuyorum.

Ailelerimizin bugüne kadar koşulsuz olarak gösterdiği tüm sevgi ve emeklerine sonsuz teşekkürlerimi sunuyorum.

Tez dönemim boyunca gösterdikleri desteklerinden dolayı abim Özhan Atvar ve yakın dostlarım Yılmaz Özuzun ile Sedat Kurt'a teşekkürlerimi sunuyorum.

Son olarak, lisans ve lisansüstü öğrenim hayatım boyunca çalışma fırsatı bulduğum tüm Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü akademik personeline teşekkürlerimi iletiyorum.

Bu çalışma kısmen TÜBİTAK 116E102 no'lu proje kapsamında desteklenmiştir.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER	vi
1. GİRİŞ	1
2. İLGİLİ ÇALIŞMALAR	6
3. TEMEL BİLGİLER	14
3.1. DERİN ÖĞRENME TEMELLERİ	14
3.2. DENEKTAŞI VERİ KÜMESİ TEMELLERİ	16
4. YÖNTEM	19
4.1. İNSAN DURUŞ ve YÖNELİM BİLGİSİNİN SINIFLANDIRILMASI	19
4.1.1. Sabit Resim Üzerinden Tekil İnsan Tespitinin Yapılması	22
4.1.2. Tekil İnsan Bilgisinin Duruş Sınıfına Göre Sınıflandırılması	22
4.1.3. Tekil İnsan Bilgisinin Yönelim Sınıfına Göre Sınıflandırılması	23
4.1.4. Deneysel Sonuçlar	24
4.1.5. Değerlendirme	27
4.2. GRUP BİLGİSİNİN SINIFLANDIRILMASI	27
4.2.1. Verilerin Hazırlanması	28
4.2.2. Sınıflandırma Süreci	33
4.2.3. Deneysel Sonuçlar	38
4.2.4. Değerlendirme	44
5. YAZILIM ve KÜTÜPHANELER	47
6. SONUÇ & GELECEK ÇALIŞMALARI	49
REFERANSLAR	50

1. GİRİŞ

Günümüzde güvenlik, analiz gibi bir çok nedenle her an çok sayıda görüntü, kameralar aracılığıyla elde edilmekte ve çeşitli yöntemlerle bu görüntülerden anlam çıkarılmaya çalışılmaktadır. Elde edilen bu görüntü miktarlarının çok yüksek boyutlara ulaşması insan dışında bir karar vericinin ortaya çıkmasını zorunlu hale getirmiştir. Trafik, sağlık, güvenlik, akıllı sistemler ve daha nice konu başlığı altında bilgisayarlar bu konuda doğrudan etkin rol oynamaktadır. Bilgisayar destekli sistemlerin varlığı hem insanların işlerine yardımcı olurken hem de bir insanın ulaşamayacağı sayıda işlemi aynı anda yaparak sonuçları hızlı bir şekilde analiz edebilmektedir.

Bilgisayarlı Görü alanının temel uğraş alanı da insan gözünün ve beyninin işlevinin bilgisayar destekli sistemlere ve algoritmalara aktarmaktır. Bu algoritmaların da görevi, insan gözünün insan beyni ile iletişimini taklit etmeye çalışarak algoritmalara girdi olarak verilen görüntüleri anlamlandırma çalışmasıdır. Bu anlamlandırma aşamasında girdi olarak gelen bir resmin kedi veya köpek resmi olmasına karar vermesi de çözdüğü problemler arasında yer alabilirken bir arabanın sürücüsüz bir şekilde çevreden elde ettiği görüntülere anlam yükleyerek sürüş işlemini gerçekleştirmesi de yer alabilmektedir. Diğer yandan bu algoritmaların trafik ve güvenlik alanlarında da kullanım bölgesi oldukça geniştir. Trafikte bulunan araçların kural ihlal tespitlerinin gerçekleşmesi de bu sistemlerle gerçekleşmektedir. Diğer yandan artan güvenlik ihtiyaçlarından dolayı kurulan güvenlik kameralarından elde edilen görüntülerin de akıllı bir sistem tarafından kontrol edilmesi gerekmektedir. Aksi takdirde bu görüntülerin manuel olarak teker teker izlenmesi ve bu görüntülerden sonuç çıkarmak için o kadar saat bir insanın masa başında kalması sağlıklı bir yaklaşım değildir.

Yukarıda bahsedilen sebeplerden dolayı şehirlerde bulunan güvenlik kameralarından elde edilen görüntülerin anlamlandırılması için de bir akıllı sistemin varlığı zorunlu hale gelmiştir. Bu bağlamda da insanların görüntüler üzerinde tespit edilmesi, hareketlerinin sınıflandırılması ve oluşturdukları grup hareketlerinin tanınabilmesi için de bilgisayar destekli algoritmalara ve bu algoritmaların da yüksek doğruluklu sonuçlar üretebilmesine ihtiyaç duyulmaktadır.

Tez konusu olarak da belirlenen görüntüler üzerinde bulunan insanların oluşturduğu grupların aktivitelerine yönelik olarak görüntüden anlam çıkarılması, bilgisayarlı görü alanının yoğun olarak uğraştığı problemler arasında yer almaktadır. Bu problemin çözüm odağında yer alan hareket ve aktivite bilgisinin tanınması, bir video girdisinden yapılabileceği gibi sabit bir görüntü üzerinden de yapılabilmektedir. Problem çözümüne yönelik uygulanacak yöntemler, girdinin video ya da sabit görüntü olmasına göre büyük değişiklikler gösterebilmektedir. Bu tez çalışması kapsamında problem çözümü için seçilen yöntemler, sabit görüntülerde bulunan grup aktivite bilgilerinin anlamlandırılması ve sınıflandırılmasına yönelik olarak seçilmiştir.

Grup aktivitelerinin durağan resimler üzerinde tanınması bilgisayarlı görü alanında zor problemler arasında yer almaktadır. İnsan hareketlerinin çoğunlukla durağan bir yapı sergilememesi ve zamansal çizelge incelendiğinde anlamlı hale gelmesi, mevcut uğraşılan aktivite tanıma problemini zorlaştırmaktadır. Video sekansından alınan tek bir çerçevelik görüntüde yürüyen iki insanın yan yana denk gelmiş olması o iki insanın aynı aktivite içinde olduklarını kesin olarak göstermemektedir. Bu örnek senaryo için video üzerinde sıralı çerçeveler incelendiğinde ancak hareket için anlamlı bir sonuca ulaşılabilir. Bu ve buna benzer sebeplerden dolayı grup aktivitelerinin sabit görüntüler üzerinden sınıflandırılması aşamasında görüntülerde bulunan tekil insan tespitlerinin yüksek hassasiyetle bulunması ve bu insanların duruş ve yönelim bilgilerinin olabildiğince detaylı sınıflandırılması sabit görüntüler üzerinde bulunan aktivitelerin yüksek başarımla belirlenebilmesi açısından önem arz etmektedir.

Yukarıda bahsedilen alt problemlerin ve alt problemleri kapsayan ana problemin çözümü adına son yıllarda sınıflandırma ve tespit problemlerinde yüksek başarı elde edilebilen derin öğrenme yöntemlerine başvurulmuştur. Derin öğrenme yöntemlerinin tercih edilmesindeki en büyük sebep son yıllarda yapılan birçok çalışmada derin olmayan yöntemlere göre daha yüksek doğruluk üretebilmeleri ve incelenen problemin de doğası gereği derin öğrenme yöntemlerine uygun olması yer almaktadır.

Derin öğrenme yöntemlerinden Evrişimsel Sinir Ağları (CNN) [1] bu tez kapsamında görüntülerden öznitelik elde etme ve sınıflandırma aşamaları için temel başvuru yöntemidir. Tez kapsamında probleme yaklaşım açısı olarak W. Choi ve arkadaşlarının [3] çalışması temel alınmıştır. Veri kümesi olarak da yine bu çalışmada yer alan Structured Group Dataset (SGD) veri kümesi kullanılmıştır. [3] çalışmasında, insan yönelimleri ve grup aktivite bilgilerinin elde edilmesi için derin olmayan metodlar kullanılmıştır. Bunun için de kullandıkları veri kümesinin içerdiği toplam görüntü sayısının ve sınıflara denk gelen görüntü sayıları fazla değildir. [3] çalışmasındaki metodların aksine tez kapsamında üzerinde çalışılan yöntemler derin öğrenme temelli yöntemler olduğu için veri sayısının çokluğu ve çeşitliliği son derece önem arz etmektedir. Veri kümesindeki bu problemin üstesinden gelmek adına da yine literatürde benzer problemler için kullanılan İnce Ayar (Fine Tune) [4] işlemine başvurulmuştur. Bu sayede çok sayıda görüntü ile eğitilmiş literatürde çoğunlukla kullanılan eğitilmiş bir model alınarak SGD veri kümesinden elde edilen yeni kümelerle tekrardan eğitilerek modelde bulunan ağırlıklar güncellenmiş ve üzerinde çalışılan problemi çözmeye uygun olarak evrilmesi sağlanmıştır.

Tez çalışması kapsamında grup aktivitelerinin sabit resim üzerinden tanınması problemi için ilk aşama için tümevarım bir çözüm üzerinde durulmuştur. Bu çözümde tekil insan bilgisinin kullanılarak grup aktivite bilgisine ulaşılması hedeflenmiştir. Temel olarak tez kapsamında kullanılan denektaş veri kümesindeki resimler aktivite içerisindeki gruplardan, gruplar da tekil insan bilgilerinden oluşmaktadır. Bu sebepten dolayı tekil insan bilgisinin tüm resimden çıkarılması, duruş ve yönelim bilgisine göre sınıflandırılması genel problemin çözümü için önem arz etmektedir. Bu bağlamda da problemi 3 alt başlık altında incelemek mümkün olmaktadır. Bu başlıklar şu şekildedir: 1) Sabit Görüntü Üzerinden İnsan Tespiti, 2) Tekil İnsan Görüntülerinin Duruş Türünün Sınıflandırılması, 3) Tekil İnsan Görüntülerinin Yönelim Bilgisinin Sınıflandırılması. Eğer yukarıda bahsedilen tümevarım çözüm yöntemi kapsamında tekil insan bilgilerinin yönelim sınıfına göre sınıflandırma başarısı yüksek bir skor ile elde edilmiş olsaydı, grup bilgilerinin tanınması problemi için tümevarım yöntemi ile sonuca gidilecektir. Ancak elde edilen tekil insana ait yönelim bilgisinin sınıflandırma başarısı beklenenden az çıktığı için tümevarım yöntemi ile grup aktivitelerinin tanınması probleminin

çözümüne son verilmiş ve farklı bir yöntem üzerinden ilerlenilmesi kararı alınmıştır.

Tez kapsamında önerilen diğer yöntem ise girdi olarak elde bulunan resimlerdeki grup bilgilerinin doğrudan sınıflandırılmasına yönelik önerilmiş bir yöntemdir. Bu yöntem çerçevesinde resim içerisinde bulunan grup aktivite bilgilerinin sınırları alınarak sanki bir alt resim gibi sınıflandırma aşamalarına dahil edilmesi hedeflenmiştir. Bu bağlamda girdi olarak verilen resimlerde bulunan grup aktivite bilgilerinin sınıflandırma başarısının artırılması üzerine yönelik çalışmalar gerçekleştirilmiştir. Bu çalışmaları 3 başlık altında incelemek mümkündür: 1) Grup Aktivite Bilgilerinin Sınıflandırılması, 2) Geç Füzyon Yöntemi ile İki Modelin Birleştirilmesi, 3) Erken Füzyon Yöntemi ile İki Modelin Birleştirilmesi

Tezin literatüre olan katkıları da şu şekilde sıralanabilir:

- SGD [3] denektaşı veri kümesindeki tekil insan ve grup aktivite verilerinin derin öğrenme yöntemiyle sınıflandırılması ve derin olmayan yöntemlere göre daha başarılı sonuçlar elde edilmesi,
- Grup aktivite verilerinin sınıflandırma başarısını artırabilmek adına hareket tanımında önemli bir role sahip olan poz bilgilerinin de füzyon operasyonlarına dahil edilmesi ve sonuçlarının gözlemlenmesi,
- Grup aktivite verilerinin sınıflandırma başarısını artırabilmek adına görüntülerdeki insan ve insan dışı nesnelerin birbirinden ayrıldığı bölüt verilerinin de füzyon operasyonlarına dahil edilmesi ve sonuçlarının gözlemlenmesi.

Tez çalışmasının geri kalanı şu şekilde düzenlenmiştir: 2 numaralı başlık altında konu ile ilgili çalışmaların detayları verilmektedir. 3 numaralı başlık altında da geliştirilen yöntemlere girdi olarak verilen veri kümesine ait sınıf detayları bulunmaktadır. 4 numaralı başlık altında tez çalışması kapsamında üzerinde çalışılan problemin çözümüne yönelik geliştirilen derin öğrenme yöntemi anlatılmaktadır. 4 numaralı başlık tekil insan ve grup bilgileri için uygulanan yöntemlerin detaylandırılması için alt başlıklara bölünmüştür. Uygulanan yöntem ve

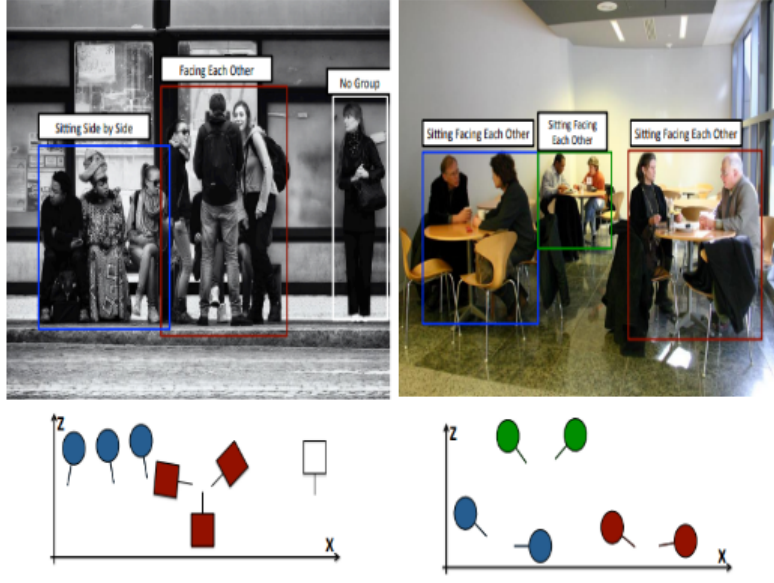
geliştirilen algoritmaların detaylarının aktarımının ardından da bu yöntem ve algoritmaların kullanılan veri kümesi ile birlikte yürütülen deneysel faaliyetlerin detayları ve sonuçları yer almaktadır. Son olarak da elde edilen algoritma performanslarının ve tez çalışma akışının değerlendirildiği ve tartışıldığı sonuç bölümü yer almaktadır.

2. İLGİLİ ÇALIŞMALAR

[3] çalışmasında grup aktivitelerinin tespit ve sınıflandırılması için sabit görüntüler üzerinde çalışan yöntemler önerilmiştir. Tez kapsamında kullanılan veri kümesi de yine [3] çalışmasında önerilen veri kümesidir. Tez konusu olarak belirlenen probleme yaklaşım noktasında temel alınan bu çalışmada uygulanan işlem adımları çalışmanın kendisinden farklı olarak derin öğrenme yöntemleri ile gerçekleştirilmiştir. [3] çalışmasında izlenen yöntemde tekil insan bilgisinden başlanılarak grup bilgisinin görüntüden elde edilmesine doğru uzanan bir işlem akışı izlenmektedir. Bu akışı özetlemek gerekirse: öncelikle tekil insanların sabit görüntü üzerinden tespiti daha sonra bu insanların duruş bilgisine göre sınıflandırılması daha sonra da yönelim bilgisine göre sınıflandırma aşaması gerçekleştirilmektedir. Elde edilen bu sınıflandırma bilgilerine göre ikili insan etkileşimleri bulunarak daha sonra da ikili etkileşimlerden yola çıkılarak grup bilgisi elde edilmektedir.

Tekil bir insanı, sabit görüntüden çıkarabilmek için Poselet detector [5] yönteminden elde edilen aktivasyon vektörü, MDP [6] aktivasyon vektörü ve HOG [7] belirteci birleştirilerek kullanılmıştır. Grup bilgisinin görüntüden elde edilmesi aşamasında da iteratif bir yöntem önerilmiştir. Tek bir görüntü için her bir iterasyonda kişilerin hangi gruplara dahil olacağına karar verilmektedir. Tez kapsamında ise tekil insan tespiti için Mask R-CNN [8] yöntemi tercih edilmiş ve bu yöntemden elde edilen çıktılardan insan etiketine sahip sonuçlar sonraki aşamalar için sınıflandırıcıya verilmiştir. [3] çalışmasında tekil insana ait duruş ile yönelim bilgisinin sınıflandırılması için SVM [9] yöntemi tercih edilirken mevcut önerilen çalışmada ise CNN [1] tabanlı bir model önerilerek sınıflandırma başarısı arttırılmaya çalışılmıştır. Aynı zamanda yine grup bilgisinin görüntüden elde edilmesi aşamasında da farklı CNN [1] tabanlı modellerinin bir arada kullanılmasıyla elde edilen yeni bir model önerilmiştir.

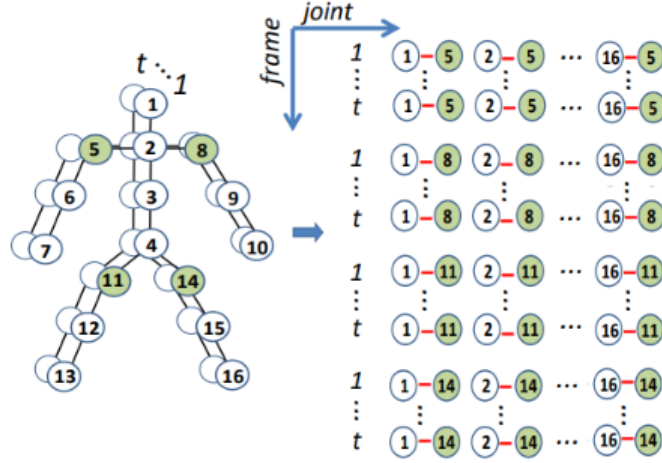
Q. Ke ve arkadaşlarının çalışmasında da [10] tekil insanı temsil etmek adına insan iskeletindeki eklem yeri bilgilerinin birbirine olan uzaklığı hesaplanmış ve bu bilgilerden yeni bir öznitelik matrisi elde edilmiştir. Oluşturulan öznitelik matrisinin CNN [1] tabanlı bir modele girdi



ŞEKİL 2.1.: [3] çalışmasından alınan grup bilgisi örneği.

oluşturabilmesi adına art arda gelen 3 çerçeve bilgisi, aynı işlemler uygulanarak 3 boyutlu bir matris haline getirilerek modele girdi olarak verilmiştir. Bu çalışma kapsamında LSTM [11] modelinin doğrudan zaman çizelgesine sahip bir problemde başarısının gösterilmesinden ziyade birkaç sıralı çerçeveden elde edilen bilginin CNN [1] ile öğrenilmesi hedeflenmiştir. Tez kapsamındaki girdiler bu çalışmanın aksine video değil sabit görüntü olmasına rağmen insan iskelet bilgisinin nasıl kullanılması konusunda ilham kaynağı olmuştur.

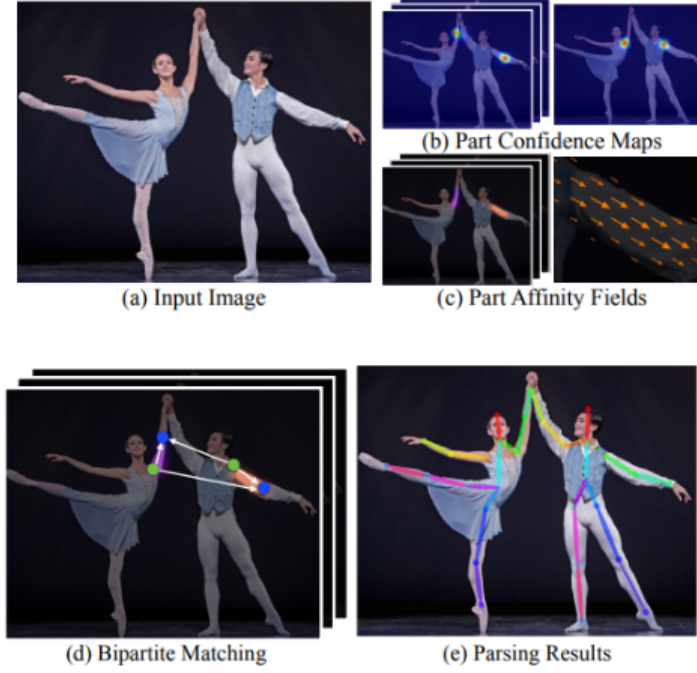
Z. Cao ve arkadaşlarının yaptığı çalışma [12] kapsamında yönelim bilgisi görüntü içerisinde bulunan insanların vücut parçalarının tespit edilmesi ve bu tespit edilen bölgelerin anlamlı bir şekilde birbiri ile birleştirilmesiyle elde edilmiştir. Girdi olarak verilen video olsa da çerçeveler arasındaki zamansal bilgidan faydalanmadan her bir tekil çerçeve için işlem yapılmaktadır. Gerçek zamanlı olarak çalışma süresine sahip tümevarım yaklaşımlı algoritmaları bulunmaktadır. Girdi olarak verilen sahnedeki her bir tekil insan için vücut parçaları bilgisi elde edilebilmektedir. Bu bilgiyi de insanın poz bilgisini tahmin etmekte kullanılmaktadır.



ŞEKİL 2.2.: Zaman çizelgesinde 3D insan iskeletinin gösterimi yapılmaktadır. Bu görüntü [10] çalışmasından alınmıştır.

Literatürdeki benzer çalışmaların aksine görüntüdeki insanları tespit edip daha sonra o insanlara ait eklem yerlerinin yerini tespit etmemektedir. Çünkü bu durumda insan tespitinin yapılmadığı durumlarda algoritmanın başarısız olma ihtimalinin olduğunu savunmaktadırlar. Bu yaklaşımları sayesinde de hız konusunda gerçek zamanlı çalışma süresi vaat etmektedirler. [12] çalışmasından farklı olarak tez kapsamında tekil insana ait insan öznitelikleri için herhangi bir eklem bilgisi kullanılmadan tekil insanın bulunduğu sahne, tüm içeriği ile birlikte CNN [1] modeline girdi oluşturmaktadır. Ancak bu eklem bilgileri ve eklem bilgilerinden elde edilen yoğunluk haritaları grup bilgisinin sınıflandırılması aşamasında kullanılmış ve elde edilen sınıflandırma başarısının artırılması hedeflenmiştir.

M. R. Amer ve arkadaşlarının yaptığı çalışmada [13] farklı ölçeklerdeki görüntülerde tekil insana ait veya bir gruba ait aktiviteyi tanınmasına yönelik bir yöntemler önerilmiştir. Bu çalışmadaki amaçlardan bir tanesi de farklı ölçekler için bu problemi çözerken çoklu belirteç kullanımlarından kaçınmaktır. Çalışma kapsamında öznitelik temsiliyeti olarak insan yüz yönelimlerinin bulunması için DPM [14] kullanılmıştır. Buradan elde edilen özniteliklerin sınıflandırılması için de SVM [9] yöntemine başvurulmuştur. Bu sayede öncelikle insan



ŞEKİL 2.3.: [12] çalışmasındaki eklem bölgelerinin gösterime yönelik üretilmiş olan çıktı.

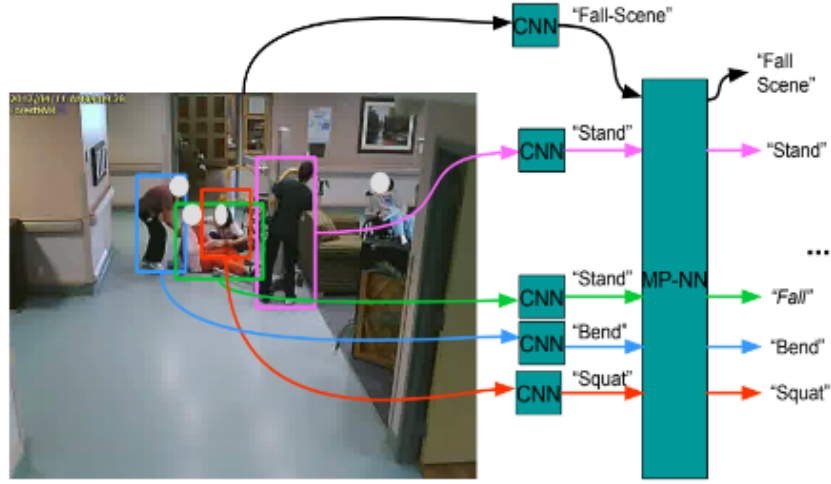
tespiti yapıp sonrasında ise 8 sınıflı sınıflandırıcı ile yüzün yönü bulunmaktadır. Sahnede bulunan obje tespitleri için de DPM [14] kullanılmıştır. Bu objelerin yönelimleri için de SVM [9] sınıflandırıcı kullanılmıştır. Çalışma kapsamında temel hareket ve eylemlerin tanımlanması için hareket bazlı STIP [15] öznitelikleri kullanılmıştır. Bu öznitelikler HOG [7] belirteçleri kullanılarak tanımlanmıştır. Tez kapsamında ise öznitelik çıkarma işlemi için ayrıca algoritma kullanılmamıştır. Bu işlem için derin öğrenme modellerinden elde edilen öznitelikler doğrudan sınıflandırıcıya verilmiştir.

Z. Deng ve arkadaşlarının yaptığı çalışmada [16] kompleks video görüntülerinde bulunan sahnelerin ve sahnelerde bulunan grup aktivitelerinin sınıflandırılması üzerine çalışılmıştır. Bu sınıflandırma işlemlerinin gerçekleştirilmesi aşamasında da derin öğrenme yöntemlerine başvurulmuştur. Bu çalışmanın temelinde sahnede bulunan tekil aktivitelerinin öğrenmesi ve eş zamanlı olarak da mevcut sahnenin açıklayıcı bir şekilde tanımlanması hedeflenmiştir. Bu hedefin yerine getirilmesi amacıyla da sahnede bulunan tekil aktivite bilgilerinin diğer aktiviteler ve sahnenin geneliyle olan ilişkilerinin çıkarılması yolu izlenmiştir. Sahne ve tekil

insanların ilişkilerini anlamlandırabilmek adına da CNN [1] yöntemine olasılıksal grafiksel modelin yeni bir katman olarak eklendi bir model kullanılmıştır.

Yeni oluşturulan bu model tekil hareket bilgisi, insanın vücut pozunu ve grup aktivite bilgilerinin arasındaki bağıllığı göz önünde bulundurarak sahne sınıflarını tahmin edebilmektedir. CNN [1] modeline katman olarak eklenen bu olasılıksal grafiksel model, çok adımlı mesaj aktarım yapay sinir ağları olarak modellenmiştir. Bu sayede sahne içerisinde bulunan tekil insan aktivitelerinin diğerleri ile olan ve sahnenin geneli ile olan bağı çıkartılabilmektedir. Çalışma kapsamında 2 tane model kullanılmıştır. Bunlardan ilki sahnenin genel bilgisini gösteren sınıfı, tekil insan aktivite bilgisini ve poz bilgisini çıkaran fine-tune [4] edilmiş CNN [1] modelidir. İkincisi ise yukarıda anlatılan mesaj aktarım yöntemi ile çalışan yapay sinir ağı modelidir. Fine-tune [4] edilen ilk CNN modeli AlexNet mimarisinin [1] ImageNet [17] verileri üzerinde eğitilerek ağırlıklarının elde edildiği modeldir. İkinci modelde her bir insan için o insanın hangi sahne, hangi aktivite ve hangi poz sınıfına ait olduğunu gösteren ağırlıklar mesaj aktarım yöntemiyle atanmıştır. Çalışma kapsamında uygulanan derin öğrenme ve yapay sinir ağı modellerine yönelik çalışmalar Caffe kütüphanesi [18] ile gerçekleştirilmiştir. Diğer yandan çalışma kapsamında iki farklı veri kümesi üzerinde performans sonuçları elde edilmiştir. Bunlar: Collective Activity [19] ve nursing home dataset [2] olarak seçilmiştir. Tez kapsamında odaklanılan alandan farklı olarak bu çalışma kapsamında sabit resimler değil videolar girdi olarak seçilmiştir. Diğer yandan bu çalışma kapsamında sınıflandırıcı olarak SVM [9] tercih edilmiştir.

A. Toshev ve arkadaşlarının yaptığı DeepPose [20] ismini verdikleri çalışmalarında üstündeki çalıştıkları problem insan poz tahmini için insan eklem yerlerinin tespit edilme ve bu bilgilerin kullanılmasıdır. Önerilen bu çalışmada önceden denenmiş olan bir sahnedeki olası tüm pozları bulmak için kullanılan Part-Based modellere göre daha farklı bir yöntem önerilmektedir. Çünkü part-based modellerin daha sınırlı bir çözüm ürettiği düşünülmektedir. Bu çalışma kapsamında önerilen yöntem DNN - derin sinir ağları olarak tanımlanmaktadır. Bu model



ŞEKİL 2.4.: [16] çalışmasında gerçekleştirilen aktivite tanıma yönelik akış gösterilmektedir.

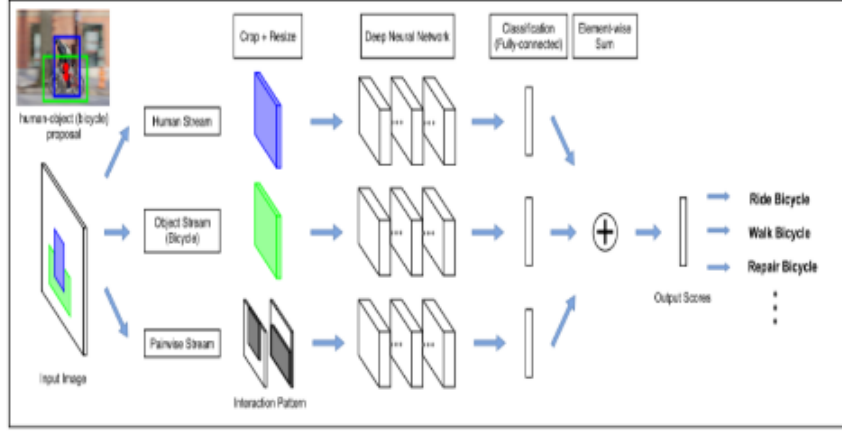
hem sahnede bulunan objelerin görsel sınıflandırılması hem de objelerin sahnedeki lokalizasyonunun yapılması için kullanılmaktadır. Çalışmanın iki katkısı bulunmaktadır. Bunlardan ilki her bir vücut eklem bilgisinin tespit edilebilmesi için tüm resmin girdi olarak verilmesi yeterli olmaktadır. Diğer katkısı ise vücut parçalarının tespit edilmesi için ayrı bir tespit edici (detector) kullanmaya gerek kalmamasıdır. Önerilen DNN modeli bu iki adımda bahsedilen görevleri yerine getirebilecek kabiliyettir. DNN modelinde öncelikle tüm görüntü kullanılarak poz tahmini yapılır ardından da tahmini yapılan bölge alt bir görüntü olarak ele alınarak eklem yeri tahminlemesi yapılır. İlk aşama haricinde diğer aşamalarda eklem yerlerinin tespiti önceki eklem yeri bilgisi kullanılarak bulunmaktadır. Çalışmada önerilen DNN modeli 7 katmandan oluşan bir mimariye sahiptir. Bu mimarinin eğitim ve test aşamaları için de iki farklı veri kümesi kullanılmıştır. Bunlardan ilki Hollywood filmlerinden elde edilmiş video kesitlerinden oluşan FLIC [21] veri kümesidir. Bu veri kümesinde çalışırken öncelikle yüz bölgesinin bulunup daha sonra insan vücudunun tamamının bulunması yönelik bir yöntem kullanılmıştır. Diğeri ise spor aktivitelerine ait video kesitleri içeren ve tüm bir vücut için 14 eklem bölgesinin etiket bilgisine sahip olan LSP [22] veri kümesidir.



ŞEKİL 2.5.: [20] çalışmasından örnek çıktılar gösterilmektedir.

Y. Chao ve arkadaşlarının çalışmasında [23] insan obje etkileşiminin bulunmasına yönelik olarak yeni bir yöntem ve bu problem kapsamında kullanılabilir geniş kapsamlı bir veri kümesi önerilmiştir. HO-RCNN ismini verdikleri yöntemde ilk aşama olarak literatürde popüler olarak tercih edilen insan ve obje tespit edicileri kullanılmıştır. İkinci aşamada ise çok akışlı bir yöntem ile insan ve obje arasındaki ilişki kurulmaya çalışılmıştır. HO-RCNN yönteminin ilk aşamasında obje ve insan tespiti için Fast R-CNN [24] mimarisi kullanılır ve hangi modelden dönen sonuç daha yüksek skor dönüyorsa o kullanılır. HO-RCNN yönteminin ikinci aşamasında ise insan akışı, obje akışı ve ikili akış ayrı olarak derin sinir ağı modeline girdi olarak verilir. Ayrı akışların derin sinir ağları mimarisinin sonucunda elde edilmiş olan öznitelikleri element bazında toplanarak sonuç skoru oluşturulmaktadır. Çalışmanın ikinci katkısı olarak da HICO-DET isimli insan obje etkileşimine yönelik olarak hazırlanmış olan veri kümesi bulunmaktadır. Bu veri kümesinin hazırlanması aşamasında öncelikle görüntülerde bulunan ve herhangi bir obje ile etkileşimde olan insanlar çerçeveye alınarak eylemi tanımlayacak şekilde etiketlenmiştir. Diğer yandan da görüntülerde bulunan objeler çerçeve içine alınarak etiketlenmiştir. En son aşamada ise insan ile objeler arasında

etkileşimden doğan bağlar tanımlanmıştır. Eğer bir insan birden çok obje ile etkileşim halinde ise hepsi birden tanımlanmıştır.



ŞEKİL 2.6.: [23] çalışmasında önerilen model akış şeması gösterilmektedir.

3. TEMEL BİLGİLER

Tezin bu bölümünde tez kapsamında kullanılan iki ana bileşenin detaylarına yer verilmiştir. İlk olarak problemin çözümüne yönelik olarak tercih edilen yöntem olan derin öğrenmenin temellerine ilişkin açıklamaya yer verilmiştir. İkinci olarak da tez kapsamında uğraşılan problemin temel taşı olan denektaşı veri kümesinin detaylarına yer verilmiştir.

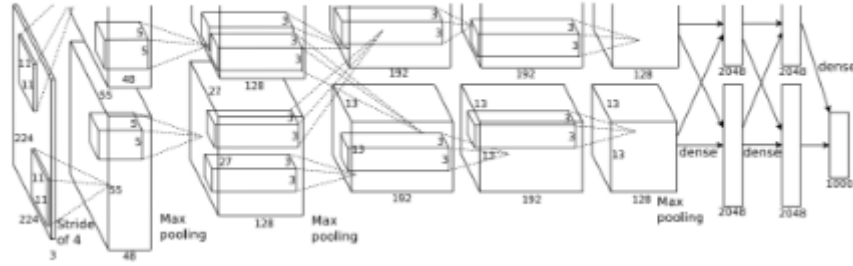
3.1. DERİN ÖĞRENME TEMELLERİ

Son yıllarda bilgisayarlı görü alanındaki problemlerin çözümüne ilişkin olarak sıklıkla derin öğrenme yöntemlerine başvurulmaktadır. Bunun sebepleri arasında derin öğrenme yöntemlerinin daha karmaşık ve zor problemlere yönelik çözüm üretebiliyor olması gelmektedir. Gün geçtikçe problem çeşitliliğinin artması da problemlerin çözümüne yönelik yeni derin öğrenme modellerinin üretilmesine neden olmaktadır. Örnek olarak üzerinde çalışılan problemin zaman bazlı bir problem oluşunda veya sınıflandırma bazlı bir problem oluşunda tercih edilen derin öğrenme yöntem ve modelleri birbirinden farklı olmaktadır.

Evrişimli Sinir Ağları (CNN) [1], tez kapsamında uğraşılan problem sınıflandırma temelli bir problem olduğundan dolayı, derin öğrenme yöntemi olarak tercih edilmiştir. CNN [1] mimarisi katmanlardan oluşan ve her bir katmandan girdi olarak verilen görüntüye ait farklı özniteliklerin elde edildiği bir mimaridir. Konvolüsyon operasyonu, CNN [1] mimarisinde çok önemli bir role sahiptir. Bu operasyonda daha önceden büyüklüğü konfigürasyonla belirlenen filtreler, işlem yapılacak olan vektör veya matris üzerinde dolaştırılarak o girdiden yeni bilgiler elde edilmekte ve sonraki katmanlara iletilmektedir. Ek olarak, konvolüsyon operasyonundan sonra doğrusallığı bozmak adına 'ReLU' aktivasyon vektörü uygulanabilmektedir. Diğer yandan başka bir katman türü de 'Pooling' katmanıdır. Bu katmanın amacı da operasyonda bulunan vektörün boyutunu azaltmaktır. En çok tercih edilen 'Pooling' katmanı türü arasında da 'Max Pooling' katmanı yer almaktadır. Bir diğer katman türü de

tam bağlantılı katmandır. Buradaki amaç da girdi ile çıktı arasında bulunan her bir düğüm arasındaki bağlantıyı kurmaktır. Genel olarak tam bağlantılı katman, tahmin katmanından hemen önce kullanılmaktadır.

Literatürde çok fazla CNN [1] modeli bulunmaktadır. Bu modellerin başarısı da "ImageNet Large Scale Visual Recognition Challenge (ILSVRC)" isminde düzenlenen yarışma ile ölçülmektedir. AlexNet modeli de [1] 2012 yılında bu yarışmada birincilik kazanmış popüler olan modellerden ilkidir ve en temel modern CNN yapısı bu çalışmada ortaya konmuştur. Şekil 3.1.'de de katman şeması gösterilen mimariye göre tez kapsamında tercih edilen model daha fazla katmana sahiptir.

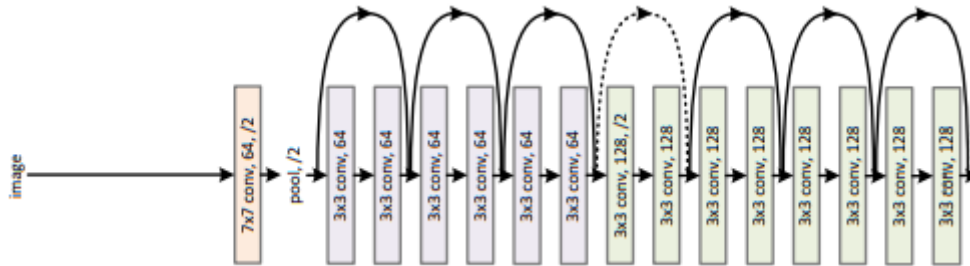


ŞEKİL 3.1.: AlexNet mimarisinin katman detayları gösterilmektedir. Bu görüntü [1] çalışmasından alınmıştır.

Tez kapsamında tercih edilen CNN [1] modeli olan ResNet [2] modeli de aynı AlexNet mimarisi [1] gibi ILSVRC yarışmasında birincilik elde etmiş bir modeldir. Bununla birlikte daha güncel bir model olduğu için ve daha karmaşık problemlerin çözümü adına tercih edildiği tez kapsamında da bu modelin kullanılmasına karar verilmiştir.

ResNet [2] mimarisi daha derin ve kompleks sınıflandırma problemlerini çözmek için önerilmiş olan bir CNN [1] mimarisidir. Aynı zamanda ILSVRC yarışmasında 2015 yılında birinci olarak seçilmiş olan modeldir. ResNet [2] mimarisinde, alt katmanlardan elde edilen öznetelik

bilgileri sıralı olarak üst katmanlara taşınarak alt ve üst katmanlar arasında bir köprü kurulması sağlanmaktadır. Bu mimarinin bir kısmına ait görsel Şekil 3.2.'de gösterilmektedir. Tez kapsamında da sınıflandırma aşamalarındaki problemlerin çözümü adına ResNet50 [2] mimarisi tercih edilmiştir. Toplamda 177 katmandan oluşan bu mimaride 'batch normalization' işlemleri de gerçekleştirilmektedir. Mimari 177 katmandan oluşmasına rağmen ResNet50 [2] olarak ifade edilmesinin sebebi, mimarinin birden çok katmanı bünyesinde bulunduran kolların oluşuyor olması ve bu kolların sayısının da 50 oluyor olmasındandır. İnce-ayar [4] işlemi adına, modelde ilk 80 katmanın ağırlıkları dondurulmuş ve yeni denektaşı veri kümesi girdi olarak alınarak ilk 80 katmandan sonraki katmanlarda bulunan ağırlıkların güncellenmesi sağlanmıştır. Öğrenme aşamasında tam bağlantılı katmanın boyutu 1000 olarak güncellenmiş ve 'batch size' değeri 16 olarak atanmıştır. 'stochastic gradient descent' algoritması da geri yayılım algoritması olarak tercih edilmiştir. Ek olarak modelde kullanılan öğrenme oranı parametresinin değeri de 0.01 olarak atanmıştır.



ŞEKİL 3.2.: ResNet [2] mimarisinin bir parçası gösterilmektedir. Bu görsel [2] çalışmasından alınmıştır.

3.2. DENEKTAŞI VERİ KÜMESİ TEMELLERİ

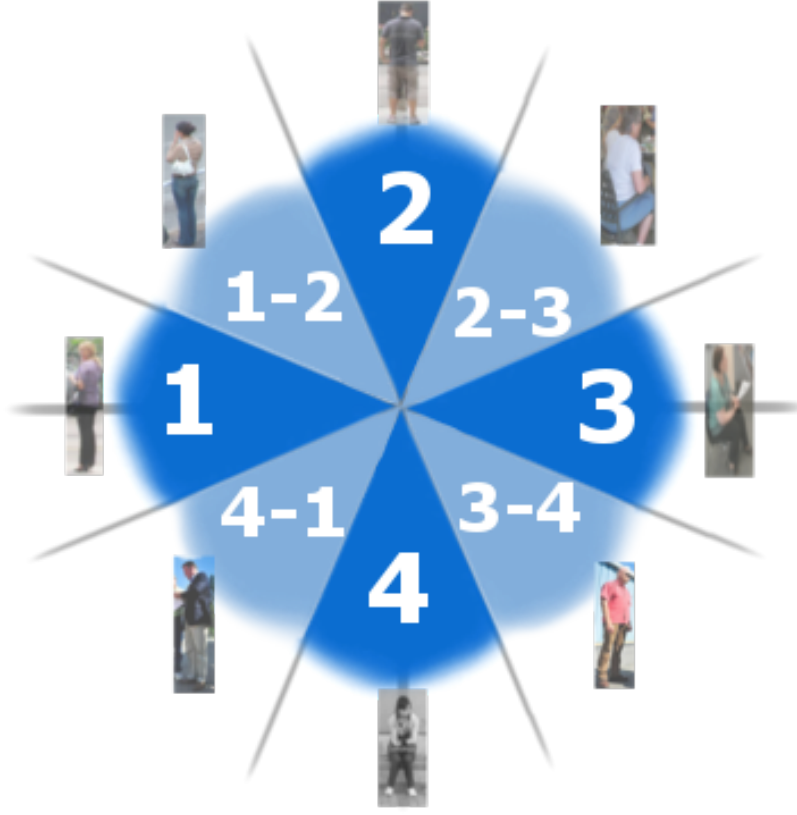
Tez kapsamında, [3] çalışmasında önerilen denektaşı veri kümesi kullanılmıştır. SGD (Structured Group Dataset) isimli bu denektaşı veri kümesi, sabit resimler üzerinden grup aktivite tanınması problemine yönelik olarak önerilmiş zorlu bir kümedir. Denektaşı veri kümesi

içerisinde grup aktivite tanıma problemine aşamalı yaklaşabilmek adına tekil insan bilgilerine yönelik etiketlemeler de bulunmaktadır. Bu bilgilere ek olarak hangi tekil insan bilgisinin hangi grup aktivitesine dahil olduğu bilgisi de yer almaktadır.

SGD [3] denektaşı veri kümesinde 588 tane görüntü bulunmaktadır. Bu görüntüler içerisinde toplamda 5415 adet tekil insan bilgisine yönelik etiketleme, 1719 tane de grup bilgisine yönelik etiket bulunmaktadır. Tekil insan bilgileri veri kümesi içerisinde duruş sınıf bilgisine göre “ayakta (standing)”, “obje üzerinde (sitting on an object)” ve “yerde (sitting on the floor)” olmak üzere 3 tanedir.

Ek olarak tekil insanlar için bu sınıflamanın dışında yönelim bilgisine ait olarak da “azimuth” değerleri verilmiştir. Bu değeri kullanarak her bir tekil insan için yönelim açısını hesaplamak mümkün olmaktadır. Tez kapsamında ve [3] çalışmasında da bu yönelim bilgisi için 8 bölge yeterli olmuştur. Bu 8 bölge Şekil 3.3’te gösterildiği üzere koordinat düzleminin merkezine insan yerleştirilerek ve 360 derecenin eşit bir şekilde bölünmesiyle elde edilmiştir. İnsan, koordinat düzleminin temeline yerleştirildiğinde 2 numaralı bölge sırtı dönük bireyi temsil ederken, 4 numaralı bölge yüzü dönük bireyi temsil etmektedir. Diğer yandan 1 numaralı bölge yüzün sola döndüğü bireyi temsil ederken, 3 numaralı bölge yüzün sağa dönük olduğu bireyi temsil etmektedir. Ayrıca koordinat düzleminde eşit olarak bölünen parçalar düzlem üzerinde -22.5 dereceden başlayarak eşit olarak dağılım göstermiştir. Bunun sebebi olarak da denektaşı veri kümesinde verilen resimlerde bulunan tekil insan bilgilerinin yönelim sınıfına ait bilgilerinin yüksek hassasiyette olmaması gösterilebilir.

SGD [3] denektaşı veri kümesinin en büyük zorluklarından biri tekil insanlarına yönelik yapılan etiketleme durumlarında bireyin önünde bireyin vücudunun herhangi bir bölgesinin görünmesini engelleyen bir durum olsa dahi vücudun tamamının etiket bölgesine dahil olarak kabul etmesidir. Böyle bir durumda karşılaşıldığı zaman vücut bölgesinin görünen kısmı tespit edilse dahil veri kümesinde belirtilen o bireye ait etiketleme bilgisi ile tam örtüşmemektedir.



ŞEKİL 3.3.: Yönelim sınıfları gösterilmektedir. Her bir yönelim sınıfı koordinat düzleminin 8 eş parçaya bölünmesiyle elde edilmiştir.

Diğer yandan vücudunun sadece küçük bir kısmı gözükken tekil birey bilgisinin da ayakta, obje üzerinde veya yerde sınıflarından birine dahil edilmesinin beklenmesi, sınıf temsiliyetinin iyi yapılamamasına sebebiyet vermektedir. SGD [3] denektaşları veri kümesi içerisindeki görüntülerde bulunan grup bilgileri için de etiketleme yer almaktadır. Bu etiketleme içerisinde hangi insanın hangi grup aktivitesine dahil olduğu bilgisi ve insana ait sınır çerçeve bilgisi bulunmaktadır. Grup aktiviteleri 7 sınıf altında toplanmıştır. Bu sınıflar: “sırada(queueing)”, “yüz yüze ayakta (standing facing each other)”, “yüz yüze oturarak (sitting facing each other)”, “yüz yüze yerde oturarak (sitting on the ground facing each other)”, “yan yana ayakta (standing side by side)”, “yan yana oturarak (sitting side by side)” ve “yan yana yerde oturarak (sitting on the ground side by side)” şeklindedir.

4. YÖNTEM

Tez kapsamında odaklanılan temel problem, grup aktivitelerin sabit görüntüler üzerinden tanınmasının derin öğrenme yöntemleri ile çözülmesi olmuştur. Bu bağlamda [3] çalışmasındaki denektaşı veri kümesi kullanılarak yine o çalışmadaki işlem adımları probleme yaklaşım açısından uygulanmaya çalışılmıştır. Ancak problemin çözümüne ilişkin olarak tez kapsamında önerilen yöntemler, baz alınan çalışmadan farklı olarak derin öğrenme temelli yöntemlerdir. Tekil bireylerin görüntü içerisindeki lokalizasyonunun yapılması, tespit edilmesi, tekil bireye ait sınıflandırma aşamaları, görüntüde bulunan grupların hangi aktivite sınıfa dahil oldukları gibi tüm alt problemler derin öğrenme yöntemleri ile çözülmüştür.

Tez çalışması kapsamında uygulanan yöntemi 2 adımda incelemek mümkündür. Bunlardan ilki, tekil insan duruş ve yönelim bilgisini sınıflandırarak grup aktivitesini tanımaya yönelik yapılan yöntemler olup, [3] çalışmasındaki işlem akışı uygulanmaya çalışılmıştır. Diğer yöntem ise baz alınan çalışmasındaki [3] işlem akışından bağımsız olarak, doğrudan görüntüden grup bilgisinin elde edilmeye çalışılarak bu grupların hangi aktivite sınıfına ait olduğunun bulunmasına yönelik geliştirilmiş yöntemdir. Tez kapsamında uygulanan ikinci yöntem, birinci yöntemde hedeflenen tümevarım yönteminde başarı elde edilememesi ve tekil insan bilgisinin sınıflandırma sonuçlarının grup aktivite bilgisini temsil edecek yeterli seviyede olmamasından tercih edilmiştir. Detayları alt başlıklarda verilmiştir.

4.1. İNSAN DURUŞ ve YÖNELİM BİLGİSİNİN SINIFLANDIRILMASI

Bu başlık altında odaklanılan yöntem, grup aktivitelerinin belirlenmesi için ön koşul olan ve grup aktivitelerinin belirlenmesine zemin kabul edilen tekil insan tespiti ve bu tespitten elde edilen duruş ile yönelim bilgilerinin çıkartılmasıdır. Bu bilgiler çıkartıldıktan sonra da elde edilen bilgiler birleştirilerek grup aktivite bilgisinin bulunması hedeflenmiştir. Bu

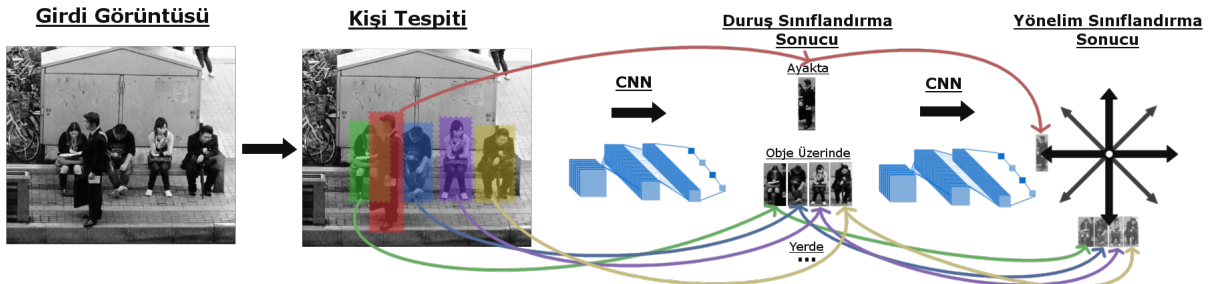


ŞEKİL 4.1.: SGD [3] denektaşı veri kümesine ait örnekler gösterilmektedir.

bağlamda da tekil insana yönelik problemi 3 ana başlık altında incelemek mümkün olmaktadır. Bu başlıklar, şu şekildedir: 1) Sabit Görüntü Üzerinden İnsan Tespiti, 2) Tekil İnsan Görüntülerinin Duruş Türünün Sınıflandırılması, 3) Tekil İnsan Görüntülerinin Yönelim Bilgisinin Sınıflandırılması.

Tez kapsamında, bahsi geçen bu problemler için kademeli bir yaklaşım kullanılmaktadır. Duruş ve yönelim sınıflandırma problemleri için uçtan uca eğitilebilen Evrişimli Sinir Ağı (CNN) [1] modellerinin kullanılması önerilmektedir. İlgili denektaşı veri kümesinde (SGD [3]) deneyler yapılmış, kullanılması önerilen Evrişimli Sinir Ağı modelinin başarılı sonuçlar üretebildiği değerlendirilmiştir.

Yukarıda da belirtildiği gibi, bu problem için üç ana adım bulunmaktadır. Bunlardan ilki grup aktivite bilgisi içeren görüntülerden tekil insan sınırlarının çıkarılması, ikincisi elde edilen tekil insan görüntülerinin insan duruş etiketlerine göre sınıflandırılması, üçüncüsü ise tekil insan görüntülerinin insan yönelim etiketlerine göre sınıflandırılmasıdır. Üzerinde çalışılan yöntemin işleyiş akışı ŞEKİL 4.2 üzerinde gösterilmiştir.



ŞEKİL 4.2.: Önerilen yöntem işleyişi gösterilmektedir. Grup aktivite tespitine doğru uzanan akış içerisinde sırasıyla tekil insan tespitinin gerçekleşmesi, elde edilen tekil insan bilgisinin duruş ve yönelim sınıflarına göre sınıflandırılması yer almaktadır.

4.1.1. Sabit Resim Üzerinden Tekil İnsan Tespitinin Yapılması

Grup aktivite görüntülerinden tekil insan görüntülerinin çıkartılması yakın zamanda önerilmiş olan Mask R-CNN yöntemi [8] ile elde edilmektedir. Mask R-CNN yöntemi, Faster R-CNN [25] yönteminin daha genişletilmiş halidir. Faster R-CNN yönteminin aksine Mask R-CNN yönteminde, girdi ve çıktı arasında piksel bazında eşleşme sağlanması hedeflenmektedir. Bu yöntemde, temel olarak Faster R-CNN [25] yöntemindeki modele yeni bir kol eklenerek bulunan bölgedeki pikselin ilgili obje sınıfına ait olup olmadığı belirlenerek piksel bazında sonuç üretilmektedir.

Bu başlık altında, veri kümesindeki her bir görüntü, Mask R-CNN [8] yöntemine girdi olarak verilmiştir. Elde edilen çıktılardan, önceden belirlenen eşik değerinin üzerinde kesin referans (GT) bilgisi ile kesişen tekil insan görüntüleri, duruş/yönelim açısından sınıflandırılmak üzere kesin referans verisinde yer alan ilgili duruş/yönelim sınıflarına dahil edilirken eşik değerinin altındaki görüntüler yanlış pozitif (false positive) sınıfına dahil olmaktadır.

4.1.2. Tekil İnsan Bilgisinin Duruş Sınıfına Göre Sınıflandırılması

Grup aktivite çalışmalarındaki denektaşları veri kümelerinde, üç farklı poz sınıfı göz önüne alınmaktadır. Bu sınıflar "ayakta (*standing*)", "obje üzerinde (*sitting on an object*)", "yerde (*sitting on the ground*)" şeklindedir. Ayrıca bu sınıflara ek olarak tespit işlemi sonucunda yanlış pozitif (*false positive*) sınıfına dahil olan görüntüler de ek bir sınıf olarak eklenmiştir ve negatif küme olarak etiketlenmiştir. Oluşturulan sınıflara ait örnekler Şekil 4.3'te gösterilmiştir.

Tekil insan duruşuna yönelik belirlenen bu 3 sınıf ve Bölüm 4.1.1.'de bahsedilen yanlış pozitif (*false positive*) sınıfı tespit aşamasının ardından, Evrişimli Sinir Ağı (*CNN*) [1] mimarisi temel alınarak sınıflandırılma aşamasına geçilmiştir. Burada uygulanan Evrişimli



ŞEKİL 4.3.: Duruş sınıflarına ait örnekler gösterilmektedir. Duruş sınıfları 3 tanedir: "ayakta", "obje üzerinde", "yerde".

Sinir Ağı (*CNN*) [1] yaklaşımında temel alınan *ResNet* [2] modelidir. Bu model, ilk eğitim aşamasında *ImageNet* [17] denektaşı veri kümesi ile eğitilmiş, ve sonraki aşamalarda bu ilk eğitim sürecinde elde edilmiş olan ağırlıklar kullanılmıştır. Önceden eğitilmiş olan bu model üzerinde yeni oluşturulan veri kümesi ile ince ayar (*fine-tune*) [4] işlemi uygulanmıştır ve ağırlıklar yeni veri kümesini sınıflandıracak şekilde güncellenmiştir.

4.1.3. Tekil İnsan Bilgisinin Yönelim Sınıfına Göre Sınıflandırılması

Duruş bilgisinin sınıflandırılmasına oranla insan yönelim bilgisinin sınıflandırılması daha zor bir problemdir. Bunun sebebi sınıflar arasındaki benzerliklerin duruş bilgisine göre yönelimde çok daha fazla olmasıdır. Diğer yandan yönelime ait sınıf sayısı da duruş sınıf sayısına göre daha fazladır.

Sınıflandırma işleminden önce SGD [3] denektaşı veri kümesinde bulunan verilerin tekil insanlar için yönelim sınıflarına göre gruplanması ve etiketlenmesi gerekmektedir. [3] çalışmasında her bir duruş sınıfındaki tekil insan yönelimi için 8 alt kategori oluşturulduğu bilgisi yer almaktadır. Ancak SGD [3] veri kümesi içerisinde yönelime ait bilgisi doğrudan paylaşılmamaktadır. Bunun yerine her bir tekil insan için azimut (*azimuth*) değeri verilmiştir. Bu değer kullanılarak her bir tekil insan için kosinüs değeri hesaplanarak yönelim bilgisi koordinat düzlemine aktarılmaktadır.

Azimet deęerinden elde edilen kosinüs ve sinüs deęerlerinin iřaret bilgisi yönelimin, koordinat düzleminde hangi bölgeye düřtüęünü belirlerken kosinüsün iřaretsiz açısai deęeri kullanılarak da yönelimin hangi sınıfa dahil olduęu bilgisi belirlenmektedir. Őekil 4'te de gösterildięi üzere koordinat düzlemi yönelim bilgisi için 8 farklı bölgeye bölünmüřtür. 1, 2, 3 ve 4 numaralı bölgeler temel yönleri temsil ederken, 1-2, 2-3, 3-4 ve 4-1 numaralı bölgeler ise ara yönleri temsil etmektedir. Ana yönler ile ara yönler açısai olarak eř parçalara bölünmüřtür. Ancak veri kümesinde bulunan tekil insanlardan ara yön sınıflarına dahil olanların sayısı ana yön sınıflarına dahil olanlardan daha az sayıdadır. Bu durumun sınıflandırma performansına olan etkisi Bölüm 4.1.5'te açıklanmıřtır.

Tekil insanı koordinat düzleminin ortasına konumlandırdığımızda Őekil 3.3'te gösterilen 2 numaralı sınıfa dahil olan insan görüntüsü sırtı dönük bireyi temsil ederken 4 numaralı bölgede bulunan insan görüntüsü yüzü dönük bireyi temsil etmektedir.

Yukarıda açıklanan kural ile veriler gruplandırıldığında her bir duruř için 8 sınıflık yönelim bilgisi elde edilmektedir ve 3 duruř sınıfı için toplamda 24 yönelim sınıfı ve bu sınıfa ek olarak negatif sınıf bulunmaktadır. Toplamda 25 sınıflık veri içeren veri kümesi için ResNet [2] mimarisi kullanılmıř, yukarıda anlatılmıř olan iliřkin yöntem ve parametreler kullanılarak eęitim ařaması gerçekteřtirilmiřtir. Sınıflandırma başarısına iliřkin detaylar Bölüm 4.1.4'te açıklanmıřtır.

4.1.4. Deneysel Sonuçlar

Deney sonuçları, mevcut çalıřma için temel alınan [3] çalıřmasındaki poz ve duruř sonuçlarıyla karřılařtırılabilecek řekilde üretilmiřtir. SGD [3] denektařı veri kümesi içerisinde 588 tane görüntü ve bu görüntüler içerisinde de etiketlenmiř olarak 5415 adet insan bulunmaktadır.

TABLO 4.1.: Tekil insan sınıflandırma sonuçları gösterilmektedir. Değerler, Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	Tekil İnsan Tespiti Başarı Sonuçları	
	[3] Çalışma Sonuçları	MaskRCNN [8] Sonuçları
Eğitim Seti	83.1/83.6	90.8/88.5
Test Seti	79.6/80.5	84.8/88.4

TABLO 4.2.: Duruş sınıflandırma sonuçları gösterilmektedir. Değerler, Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	Duruş Bilgisinin Sınıflandırma Sonuçları	
	[3] Çalışma Sonuçları	Tez Çalışması
Eğitim Seti	78.0/77.7	91.7/93.3
Test Seti	71.0/66.9	70.2/76.8

TABLO 4.3.: Yönelim sınıflandırma sonuçları gösterilmektedir. Değerler, Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	Yönelim Bilgisinin Sınıflandırma Sonuçları	
	[3] Çalışma Sonuçları	Tez Çalışması
Eğitim Seti	42.5/47.6	72.3/84.5
Test Seti	41.7/33.6	22.4/41.7

Hesaplama yapılırken tekil insana ait kesin referans (ground truth (GT)) bilgileri ile Mask R-CNN [8] çıktılarının kesişme oranları dikkate alınmıştır. Mevcut çalışma kapsamında kesişme için eşik değeri %75 olarak belirlenmiştir. Eğer kesişme miktarı belirlenen orandan küçük ise Mask R-CNN [8] tarafından bulunan kişi tespit sınırlı çerçevesi, yanlış pozitif (false positive) sınıfına, büyük ise duruş/yönelim sınıfına ait olmaktadır.

Tablolarda gösterilen sonuçlarda her sınıf için 2 ayrı doğruluk değeri hesaplanmıştır. Bunlardan ilki sınıf bazında ortalama doğruluk değerini (per class accuracy) ikincisi ise sınıflandırmadaki genel doğruluk değerini (overall accuracy) temsil etmektedir. Bu iki hesaplama yöntemi ile elde edilen doğruluk sonuçlarının birbirinden farklı çıkmasındaki temel neden özellikle 25 sınıflık değerlendirme için sınıfların içerisinde bulunan tekil insan görüntü sayılarının dengeli dağılmamış olmasından kaynaklı tekil sınıf doğruluk değerlerinin düşük gelebilmesinden kaynaklanmaktadır.

Deneyleer sırasında veriler, %60 eđitim, %20 dođrulama, %20 test ve kumesi olmak üzere b÷l÷nmüştür. Bu bölümlenmenin ardından 2, 4 ve 25 sınıf için ResNet50 [2] modelleri eđitilmiştir. Eđitim aşamasının tamamlanmasının ardından da yeni üretilen ince ayar (fine-tune) [4] edilmiş modeller her 3 grup için hazırlanmış test kümeleri ile test edilerek başarı sonuçları listelenmiştir. Sonuçlar Tablo 4.1., 4.2. ve 4.3. üzerinde gösterilmiştir.

Tablo 4.1., 4.2. ve 4.3.'te de görüldüğü üzere mevcut çalışma kapsamında önerilen yöntemlerin üretmiş olduđu sonuçlar [3] çalışmasına göre daha yüksek sonuçlar üretebilmiştir. Bunun altında yatan temel sebep olarak derin öğrenme yaklaşımının sınıflandırmadaki başarısı ve önceden öğrenilmiş Evrişimli Sinir Ađı (CNN) [1] modellerinin eldeki veri, sayıca az olsa dahi öğrenme durumuna imkan tanıyor olması gösterilebilir. Tablolardaki sonuçlar incelendiğinde 2 ve 4 sınıf için üretilmiş olan sınıflandırma modelinin dođruluk oranının temel alınan çalışmaya göre daha yüksek olduđu görülmektedir. Tablo 3, incelendiğinde de 25 sınıflık sınıflandırma sonucunun ortalama dođruluk deđerinin düşük çıktığı görülmektedir. Bunun asıl sebebi olarak veri sayısının azlığı gösterilebilir. SGD [3] denektaşı veri kümesinde özellikle 25 sınıflık tekil insan yönelimi için sınıflara bölümlene işleme gerçekleştirildiğinde sınıf başına düşen veri miktarı oldukça azalmaktadır. Bu sebepten dolayı bu kadar az sayıdaki verinin derin öğrenme ile o sınıfı temsil edecek ađırlıkları bulması oldukça zorlaşmaktadır. Görüntü sayıları az olan sınıfların başarısı düşük geldiğinde de genel ortalamaya dođrudan etki etmekte ve bu da ortalama dođruluđun düşmesine sebep olmaktadır.

Dođruluđu etkileyen diđer bir sebep olarak da, Mask R-CNN [8] yönteminin insan tespiti için ürettiği sonuçların yüksek hassasiyette olması ve her durumda mutlaka insanın görünen kısmını bulması gösterilebilir. SGD [3] denektaşı veri kümesinde bulunan tekil insan görüntüleri için paylaşılmış kesin referans (GT) bilgileri ise görüntüdeki insan bir masanın altında oturuyor olsa dahi o insanın tüm vücudunu kapsayacak şekilde verilmiştir. Bu sebepten dolayı Mask R-CNN [8] sonuçları insan tespiti için dođru olsa dahi kesin referans (GT) ile yüksek oranda eşleşmemektedir.

4.1.5. Değerlendirme

Mevcut çalışma kapsamında derin öğrenme teknikleri kullanılarak, grup aktivite görüntülerinde insan duruş ve yönelimlerinin sınıflandırılması üzerinde çalışılmıştır. Bu kapsamda ele alınan uçtan-uca eğitilmiş olan Evrişimli Sinir Ağı [1] modellerinin hem duruş tanıma, hem de yönelim tanıma problemleri için başarılı sonuçlar alabildiği gözlemlenmiştir. Ancak tabloda da sonuçlar incelendiğinde 25 sınıflık veri için önerdiğimiz yöntem yüksek bir doğrulukla üstünlük elde edemediği görülmektedir. Bundan dolayı tekil insan duruş ve yönelim bilgileri kullanılarak grup aktivitelerinin tespitine doğru uzanan yaklaşım durdurulmuştur. Bunun en büyük sebebi, SGD [3] denektaşı veri kümesinin 25 sınıflık sınıflandırma problemi için yeterli sayıda veri içermemesidir. Tez kapsamındaki ana hedef, grup aktivitelerinin tanınması problemi olduğundan dolayı bu yaklaşımda tekil insanların yöneliminin yüksek hassasiyet ile bulunması ön koşuldur. Ancak denektaşı veri kümesi, 25 sınıflık yönelim bilgisi için klasörlendiğinde sınıf başına düşen resim miktarı sınıf temsiliyetini sağlayacak sayı ve çeşitlilikte olmadığından dolayı sınıflandırma başarısı beklenen düzeyden daha düşük çıkmıştır. Bu başlık altında yapılan faaliyetler neticesinde de bu yöntemin grup aktivite sınıflandırma çalışması için uygun olmadığı görülmüştür.

4.2. GRUP BİLGİSİNİN SINIFLANDIRILMASI

İlk metodun aksine bu başlık altında ele alınan yöntemde tümevarım bir yaklaşım izlenmemektedir. Bu başlık altında önerilen çözümde doğrudan grup bilgilerinin kullanılması ve sınıflandırılması hedeflenmiştir. Önerilen yöntemi 3 alt başlık altında incelemek mümkündür: "Grup Bilgisinin Sınıflandırılması", "Geç Füzyon Yöntemi ile İki Akışlı Model" ve "Erken Füzyon Yöntemi ile İki Akışlı Model".

4.2.1. Verilerin Hazırlanması

[3] çalışmasında önerilen SGD denektaşında doğrudan gruplara ait sınır bir bilgisi bulunmamaktadır. Çünkü o çalışmada uyguladıkları yöntem tümevarım şeklinde olduğundan dolayı tek bir insanın tespit ve yöneliminden sonra insanları önce ikili olarak gruplamak sonra da ikili grupları iterasyon ile bir grup olarak bulunması hedeflenmiştir. Ancak tez kapsamındaki bu başlık altındaki yöntemde doğrudan bir gruba dahil olan insanların tümünün o grup aktivite sınıfına dahil edilerek sınıflandırılması planlanmıştır. Bunu gerçekleştirebilmek adına da SGD [3] denektaşı veri kümesinde her bir görüntüde bulunan tekil insan etiketlerinin hangi grup aktivitesine dahil olduğu bulunmuş ve aynı grup aktivitesine dahil olan insan sınırlarının tüm grubu kapsayacak şekilde birleştirilmesiyle bir görüntüdeki tek bir grup aktivitesinin sınırları elde edilmiştir. Bunun için figürde de görüldüğü üzere $\min(x,y)$ değerleri ile $\max(x,y)$ değerleri alınarak grup aktivite sınır bilgisi elde edilmiştir. Şekil 4.4'te de görüldüğü üzere sarı, yeşil ve mavi renkler farklı grupları temsil etmektedir. Her bir grup içerisinde de tekil insana ait sınır bilgisi ve tekil insanın hangi gruba ait olduğu bilgisi bulunmaktadır.

Veriler şekildeki gibi SGD [3] denektaşı veri kümesinde bulunan her bir görüntü için aynı şekilde ayıklandıktan sonra 7 sınıf için aşağıdaki tabloda verildiği sayı adedince örneklem toplanmış olundu. Tabloda da görüldüğü üzere eğitim ve test kümesi sayıları ince ayar (fine tune) [4] yöntemi ile sınıflandırma için yeterli sayıdadır. Ancak sınıflara ait örneklem sayısının homojen olmaması test aşamasında sınıf bazında doğruluğu etkileyen temel faktörler arasında yer almaktadır. Diğer yandan sayısal değerlere ek olarak sınıflandırma başarısını etkileyen temel nedenler arasında verilerin çeşitliliği ve temsil ettiği sınıfın dışında başka sınıflara ait sahne içermemesi de önemli olmaktadır. Sınıflara ait dağılım ve sınıflarda bulunan örnek sayılarına ilişkin detaylar Tablo 4.4.'te gösterilmektedir. Şekil 4.5.'te verilen denektaşı veri kümesinden örnekler incelendiğinde de aynı sınıfa ait aktivitelerin bile aslında çok farklı formlarda olabildiği görülmektedir.



ŞEKİL 4.4.: SGD [3] denektaşı veri kümesinde bulunan tekil insanların oluşturduğu grup aktivite örnekleri gösterilmektedir.

TABLO 4.4.: SGD [3] denektaşı veri kümesinden elde edilen grup aktivite bilgilerine yönelik sınıfların dağılımı gösterilmektedir.

Sınıf İsmi	Test Set Örneklem Sayısı	Eğitim Seti Örneklem Sayısı
sırada	27	107
yüzyüze ve ayakta	79	314
yüzyüze ve oturuyor	76	301
yüzyüze ve yerde	20	80
yanyana ve ayakta	45	176
yanyana ve oturuyor	81	323
yanyana ve yerde	24	92

Aynı denektaşı veri kümesi için iki farklı alt küme, SGD [3] denektaşı veri kümesinden elde edilen grup aktiviteleri kullanılarak oluşturulmuştur. Bu alt kümelerin oluşturulmasındaki temel amaç, grup aktivite bilgisi kullanılarak doğrudan elde edilen sınıflandırma performansını arttırmaktır. Farklı sınıflandırma bilgilerinin mevcut sınıflandırma sürecine dahil edilmesi ve bu durumun etkilerinin gözlemlenmesi amaçlanmıştır. Bu amaç doğrultusunda grup aktivite görüntüleri [12] çalışmasındaki yöntem girdi olarak verilmiş ve grup bilgilerindeki insanların poz bilgileri elde edilmiştir. Buradaki temel amaç elde edilen poz bilgilerinin grup aktivitelerinin sınıflandırma olan etkisinin gözlemlenmesidir. Bu yöntemin

detayları Bölüm 4.2.3.'te verilmiştir. Bu yöntemden elde edilen 7 sınıfa ait poz görsel örnekleri de Şekil 4.6'da gösterilmektedir. Örnek poz görsellerinde de görebileceğiniz gibi, bazı çıktılar eylemi açıkça tanımlarken, bazı çıktılar ise eylemleri temsil etmez. Bu durumun nedenleri bölüm 4.2.3'te detaylandırılmıştır.



ŞEKİL 4.5.: Aynı grup aktivite görüntülerine ait farklı formlardaki örnekler gösterilmektedir.

Eğer insana ait bir aktivite tanımlı yapıyorsa, insan vücudunda yer alan eklem bilgilerinin konumu ve aldığı açılar o eylemi tanımlamak için önemli ipuçları verebilmektedir. Örnek olarak ayakta sabit olarak duran bir insanın kol ve bacak eklemleri arasında 180 derecelik bir açı oluşurken oturan veya çömelmiş bir insanın kol ve bacak eklemleri arasındaki açı

180 dereceden daha az olmaktadır. Eklem bilgisinin sınıflandırma ve füzyon operasyonlarına olan etkisinin incelenmesi adına yapılan çalışmanın detayları Bölüm 4.2.3 ve 4.2.2’de verilmiştir.



ŞEKİL 4.6.: [12] çalışmasındaki metodun SGD [3] denektaşı veri kümesi üzerindeki çıktıları gösterilmektedir.

Grup aktivite görüntüleri için üretilen diğer küme ise MaskRCNN [8] yönteminin bölütleme çıktılarıdır. Bölütleme çıktıları insan ve insan olmayan diğer bilgiler şeklinde görüntüde



ŞEKİL 4.7.: Mask R-CNN [8] yönteminin SGD [3] denektaşı veri kümesindeki bölütleme çıktıları gösterilmektedir.

sadece 2 bilgi bulundurmaktadır. Bu kümenin üretilmesindeki temel amaç grup aktivite bilgilerinin sınıflandırılma aşamasında insan dışında kalan sahnedeki diğer bileşenlerin olumsuz etkilerini aza indirmektir. SGD [3] denektaşı veri kümesindeki bazı görüntülerde insanların önünde başka nesnelere bulunmaktadır. Bu da o insanın dahil olduğu grup aktivitesinin tanınmasını zorlaştırmaktadır. Bölütleme çıktısı bu ve benzeri etkilerin giderilmesi amacıyla sınıflandırma ve füzyon operasyonlarına dahil edilmektedir. Şekil 4.7.'deki çıktılar incelendiğinde insana ait bazı vücut detaylarının ortadan kalktığı ve hareket tanımı için yeterli bilgi içermeyen görüntülerin olduğu görülmektedir. Bu ve benzeri durumların sınıflandırma başarısı üzerindeki etkileri Bölüm 4.2.3'te ve 4.2.4'te detaylandırılmıştır.

4.2.2. Sınıflandırma Süreci

Bölüm 4.2.1'de açıklanan yöntemleri kullanarak, SGD [3] denektaşı veri kümesi için üç farklı alt küme elde edilmiştir. Bu kümelerin her birinde, aynı görüntülerin işlenmemiş ve işlenmiş durumları vardır. Bu kümelerdeki test ve eğitim klasörleri aynı sayıda örnek içerir. Bu 3 alt küme sırasıyla, grup bilgilerinin doğrudan kullanıldığı RGB görüntüleri, grup bilgilerinden elde edilen poz bilgilerine ait görüntüleri ve grup bilgilerinden elde edilen bölütleme bilgilerine ait görüntüleri içeren kümelerdir.

3 farklı kümede bulunan 7 adet grup aktivite sınıfına ait görüntüler Imagenet [17] denektaşı veri kümesi ile eğitilerek ağırlıkları güncellenmiş ResNet [2] mimarisi ile ince-ayar(fine-tune) [4] işlemi uygulanarak eğitilmiştir. Bu eğitim aşamasında ResNet [2] mimarisine ait bazı katmanlardaki ağırlıklar dondurulurken bazı ağırlıklar ise yeni denektaşı veri kümesinden gelecek bilgiler ile güncellenmeye açık bırakılmıştır. İlk 80 katmandaki ağırlıklar dondurulurken 80'den sonra gelen katmanlardaki ağırlıklar ince-ayar(fine-tune) [4] yöntemiyle güncellenmek üzere serbest bırakılmıştır. İlk katmanlardaki görüntülere ait daha temel özellikleri tespit etmeye yarayan ağırlıklar olduğu gibi kullanılmıştır. Burada Imagenet gibi büyük bir veri kümesinden elde edilen temel nesnelere ait temsiliyet gücünün kullanılması hedeflenmiştir.

[3] çalışmasında grup aktivite bilgileri üzerinde doğrudan sınıflandırma işlemi uygulanmamıştır. Onun yerine tümevarım yöntemi ile tekil insandan başlayarak grupların temsiliyeti ve aktivite sınıflarının bulunması işlemleri uygulanmıştır. Tez kapsamında ise grup aktivite bilgileri doğrudan aktivite bilgilerine göre sınıflandırılmıştır. Bu iki yöntem başarısı karşılaştırıldığında ise tez kapsamında elde edilen başarının daha yüksek olduğu görülmektedir. Tez kapsamında ek olarak elde edilen grup aktivite sınıflandırma başarısının daha ileriye taşınabilmesi adına çeşitli füzyon teknikleri uygulanmıştır. Bu füzyon tekniklerinin kullanılmasıyla derin öğrenme tekniklerinin sayıca az olan veri kümesi üzerindeki negatif etkilerinin azaltılması amaçlanmıştır.

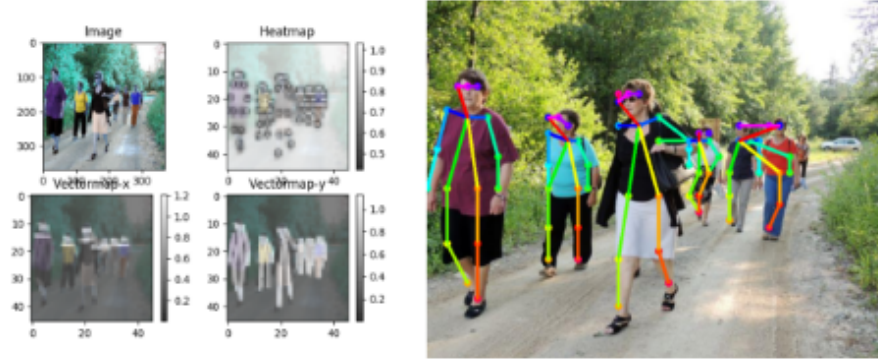
Eğitim aşamasında tabloda gösterildiği üzere her sınıftaki toplam verinin yaklaşık olarak yüzde 20'lik kısmı test için ayrılmıştır. Doğrulama (validation) adımı için veri ayrılmamasının sebebi de veri sayısının daha fazla azalmasını önlemektir.

4.2.2..1 Geç Füzyon Yöntemi ile İki Akışlı Model

Füzyon işlemi, farklı derin öğrenme modellerindeki ağırlıkların çeşitli yöntemlerle birleştirilip tek bir model olarak kullanılması ve sonuç üretmesi aşaması olarak tanımlanabilir. Füzyon işlemleri farklı şekilde yapılabilmektedir. Bu füzyon işlemlerini birbirinden ayıran özellik, farklı modellerin hangi aşamalarda birleştirilmesine karar verildiğidir. Eğer erken katmanlarda modellerin ağırlıkları birleştirilirse buna erken füzyon işlemi, eğer geç katmanlarda modellerin ağırlıkları birleştirilirse buna da geç füzyon işlemi denilmektedir.

Tez kapsamında füzyon işlemi için 2 farklı model kullanılmıştır. Bu modellerden ilki bir önceki başlıkta detayları verilen SGD [3] denektaşı veri kümesinde bulunan grup bilgilerinin doğrudan sınıflandırılması yönelik gerçekleştirilmiş olan ince-ayar (fine-tune) [4] işlemi ile elde edilmiş olan modeldir. Diğer model ise [12] çalışmasında önerilen yöntem kullanılarak vücuttaki eklem yerlerinin her bir görüntü için çıkartılıp görüntünün o halinin eğitildiği

modeldir. [12] çalışmasında sabit resimler üzerinden eklem bölgelerinin yeri tayin edilerek bu bölgelerin birbiri ile ilişkisi kurularak tekil bir insan için poz bilgisi gerçek zamanlı olarak bulunabilmektedir. Tez kapsamında da ilgilenilen konular arasında bireylerin yaptığı aktivitenin tanınması olduğundan dolayı bir bireyin eklem yerlerinin doğru tespit edilmesi o bireyin içerisinde bulunduğu aktivitenin tanınabilmesine yardımcı olmaktadır. [12] çalışmasının SGD [3] veri kümesindeki örnek bir görüntü üzerinde çalışmış hali aşağıdaki Şekil 4.8’de gösterilmiştir.



ŞEKİL 4.8.: [12] çalışmasındaki yöntem ile elde edilen SGD [3] denektaşı veri kümesindeki bir görüntünün örnek çıktısı gösterilmektedir..

Şekil 4.8.’de görüldüğü üzere resim bütün olarak verildiği halde de içerisindeki tekil bireyler için eklem bilgisinin çıkartılması işlemi gerçekleştirilebilmektedir. Tez kapsamında ise yukarıdaki çıktılardan “Vectormap-x” ve “Vectormap-y” olarak işaretlenmiş çıktılar kullanılarak grup resim bilgisinin temsiliyetinin eklem bölgeleri ile yapılması hedeflenmiştir. Bir önceki başlıkta anlatılan yöntemin birebir aynısı yine aynı veri kümesi ve aynı görüntülerin “vectormap-x” ve “vectormap-y” bilgileri kullanılarak ince-ayar(fine tune) işlemine tabii tutulmuştur. Her görüntü için ince ayar işlemine uygun hale gelmesi amacıyla “vectormap-x” ve “vectormap-y” değeri toplanmış ve elde edilen tek boyutlu gri resim 3 bant için kopyalanmıştır. Bu işlemin sonunda elimizde x ve y koordinatları için eklem bölgeleri bulunan grup resimleri elde edilmiştir.

Diğer geç füzyon operasyonu için de Mask R-CNN [8] yönteminden elde edile bölütleme sonuçları kullanılmıştır.

Yukarıda açıklanan süreçler uygulandıktan sonra yeni bir model elde edilmiştir. Bu sürecin amacı, iki farklı kümeden gelen bilgilerin birleştirilerek grup etkinliklerini daha iyi temsil eden yeni bir model elde etmektir. Sınıflandırma katmanı, geç füzyon operasyonunda birleştirme katmanı olarak kullanılmıştır. Son sınıflandırma katmanının, iki farklı veri kümesinden gelen sınıflandırma sonucu üzerindeki etkisini gözlemlemek için, bu katmanlardaki ağırlıklar, eleman bazında çarpılarak modele eklenmiştir. Başka bir yöntem olarak da, son katmanlardaki ağırlıklar, model bazında aritmetik ortalama alınarak modele eklenmiştir. Geç füzyon için uygulanan yöntemin şematik özeti ayrıntılı olarak Şekil 4.9.'da verilmiştir. Şekil 4.9.'da görülebileceği gibi, iki akışlı olarak bağımsız olarak gelen farklı kümelere ait olan ResNet [2] mimarisi, sınıflandırma aşamasından önce füzyona tabi tutulmuştur. Bu füzyon işlemi sırasında elemanların çarpımı ve aritmetik ortalaması alınmaktadır. Şekil 4.9.'da gösterilen mimariye, RGB grup bilgisi ile poz bilgisi ve RGB grup bilgisi ile bölütleme bilgisi ayrı ayrı girdi olarak verilmiştir. Burada, füzyon işleminin gerçekleştiği katmanın uzunluğu 1×7 boyutunda bir vektördür. Bu vektördeki değerler, eğitilmiş model için 7 sınıfın ağırlıklarını temsil eder. İki farklı ağırlığın sınıf temsil kabiliyetlerini üzerindeki etkilerini görebilmek için 1×7 boyutundaki vektörler eleman bazında çarpılarak ve aritmetik ortalama alınarak performans değerleri ölçülmüştür.

4.2.2..2 Erken Füzyon Yöntemi ile İki Akışlı Model

Geç füzyon yönteminin detayları Bölüm 4.2.2.1'de açıklanmıştır. Bir başka füzyon yöntemi de, erken füzyondur. Bu yöntemde, öğrenme süreci, sınıflandırma aşamasından önce 2 farklı yoldan ilerleyen öğrenme akışını birleştirilerek ilerlemektedir. Bu noktada, 2 ayrı yoldan ilerleyen akış önceden belirlenen katman seviyesinde birleşir ve ağırlıklar 2 koldan gelen ağırlıklar ile güncellenir. Bu yöntemde temel amaç, iki farklı kümeden gelen bilgileri tek bir

modelde birleřtirerek sınıflandırma başarısını arttırmaktır. Tez kapsamında uygulanan erken füzyon tekniğinde modele, aynı görüntülere ait farklı özellikler içeren iki küme verilmekte ve tek bir sınıflandırma sonucu üretilmektedir.

Erken füzyon tekniğinde hangi katman seviyesinin birleřtirileceğine karar vermek önemlidir. Tez kapsamında birleřtirilecek olan bu katmanın kararlařtırılması için çeřitli denemelerde bulunulmuřtur. Tezde kullanılan ResNet50 [2] modelinde dal yapıları mevcuttur. Her dal kendi içinde 3 veya 4 katmandan oluřmaktadır. Tez kapsamında, füzyon iřlemi için hangi katmanın seçileceğine karar verilmek adına son katmandan bařlayarak alt katmanlara inilerek en yüksek sınıflandırma sonucunu üreten katman bulunmaya çalıřılmıřtır.

Erken füzyon iřleminde, iki farklı koldan gelen ağırlıkların belirlenen katman seviyesindeki hangi operasyonlarla birleřtirileceğı de önem arz etmektedir. Bu operasyonlar element bazında çarpma ve element bazında aritmetik ortalama alma řeklinindedir. Bu bölümde, ilk olarak, hangi katmanlarda erken füzyon iřleminin yapılması gerektiğine karar vermek için farklı katmanlarda birleřtirme iřlemi uygulanarak performans ölçülmüřtür. Daha sonra, birleřtirilmeye karar verilen katmanda, birleřtirme yönteminin eleman tabanlı bir çarpım mı yoksa aritmetik ortalama mı olduđunu belirlemek için deneyler yapılmıřtır. Elde edilen performans sonuçları Bölüm 4.2.3'te detaylandırılmıřtır.

Erken füzyon iřlemi řematik akıř řekil 4.10. üzerinde gösterilmektedir. Model üzerindeki birleřme noktasının řekil 4.10.'da görüldüğü üzere seçilmesinin sebebi Bölüm 4.2.3'te açıkladıđı gibi daha yüksek sınıflandırma başarısına sahip olmasıdır. řekil 4.10.'da gösterildiğı gibi, erken füzyon iřlemi "res5c_branch2c" katmanından hemen önce gerçekteřtirilmiřtir. Bu katmandan sonra, 2 ayrı koldan ilerleyen mimari, tek bir kola indirilir ve 2 girdi için tek bir çıktı veren bir model haline gelir. Birleřtirme yöntemi olarak Keras [26] kütüphanesinde yer alan çarpma ve ortalama alma fonksiyonları kullanılmıřtır. İřlem yapılan matrisin uzunluđu 512x2048'dir. Füzyon için bu matris üzerindeki tüm iřlemler eleman bazında yapılmıřtır.

Tüm bu yapılarla ek olarak farklı bir katman birleřtirme yöntemi daha uygulanmıřtır. Bu

yöntemde iki farklı koldan gelen mimari FC katmanına kadar ayrı akışlarına devam etmiştir, FC katmanına gelindiğinde ise eleman bazlı bir işlem yerine mevcut vektörlerin uç uca eklenmesi yöntemi uygulanmıştır. Bu sayede iki ayrı koldan elde edilen 1x1000 boyutlu FC katmanı uç uca eklenme işleminin ardından 1x2000 boyutunda bir vektöre dönüşmüştür. Bu modelin de performans sonuçları hesaplanmıştır..

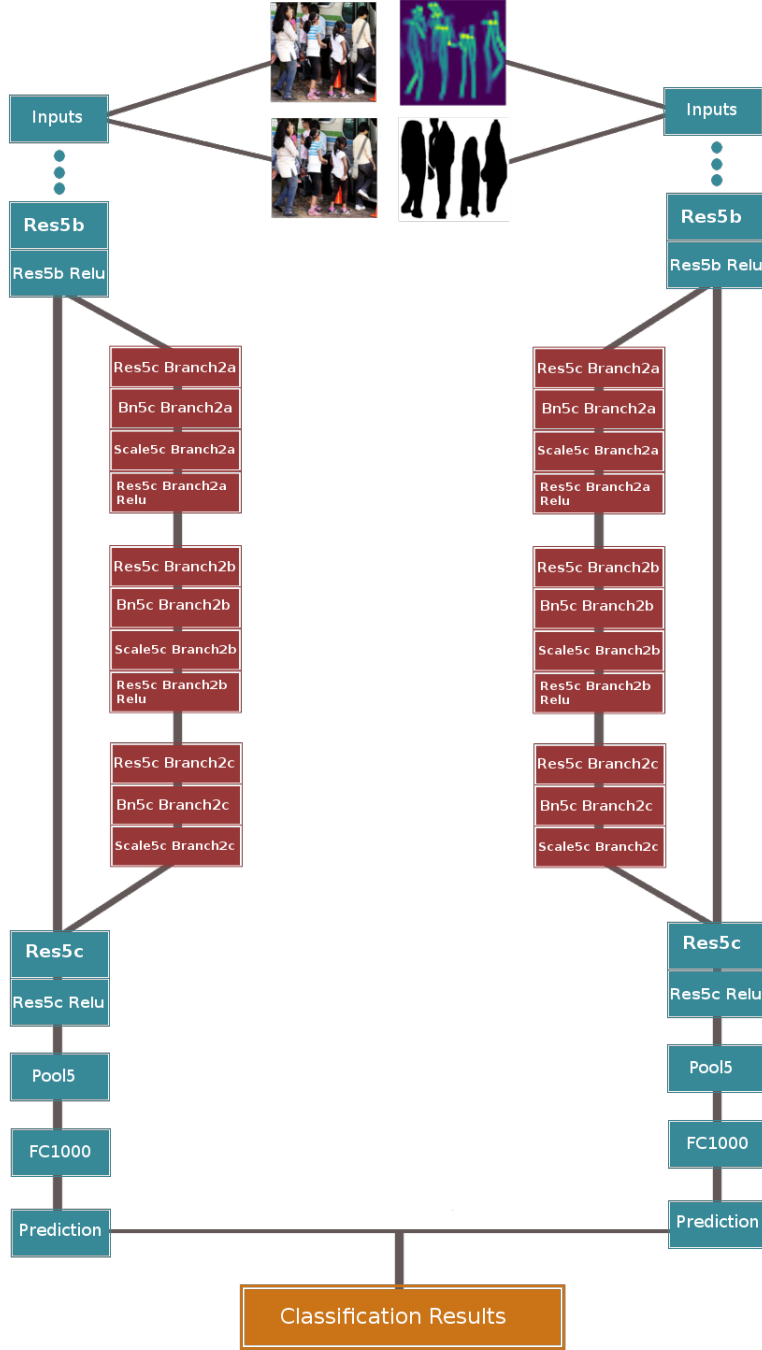
4.2.3. Deneysel Sonuçlar

Tezin bu kısmında grup aktivite görüntülerinin sınıflandırılması için kullanılan tekniklerin avantaj ve dezavantajlarından bahsedilmektedir. SGD [3] veri kümesinden elde edilen grup aktive bilgilerinin doğrudan sınıflandırılması ile bu görüntülerden elde edilen poz bilgileri ve bölütleme bilgilerinin yer aldığı kümelerin sınıflandırma başarıları detaylandırılmaktadır. Bunlara ek olarak bu kümelerin birbirileri ile olan etkileşimlerinden doğan modellere ilişkin sınıflandırma başarıları da hesaplanmıştır.

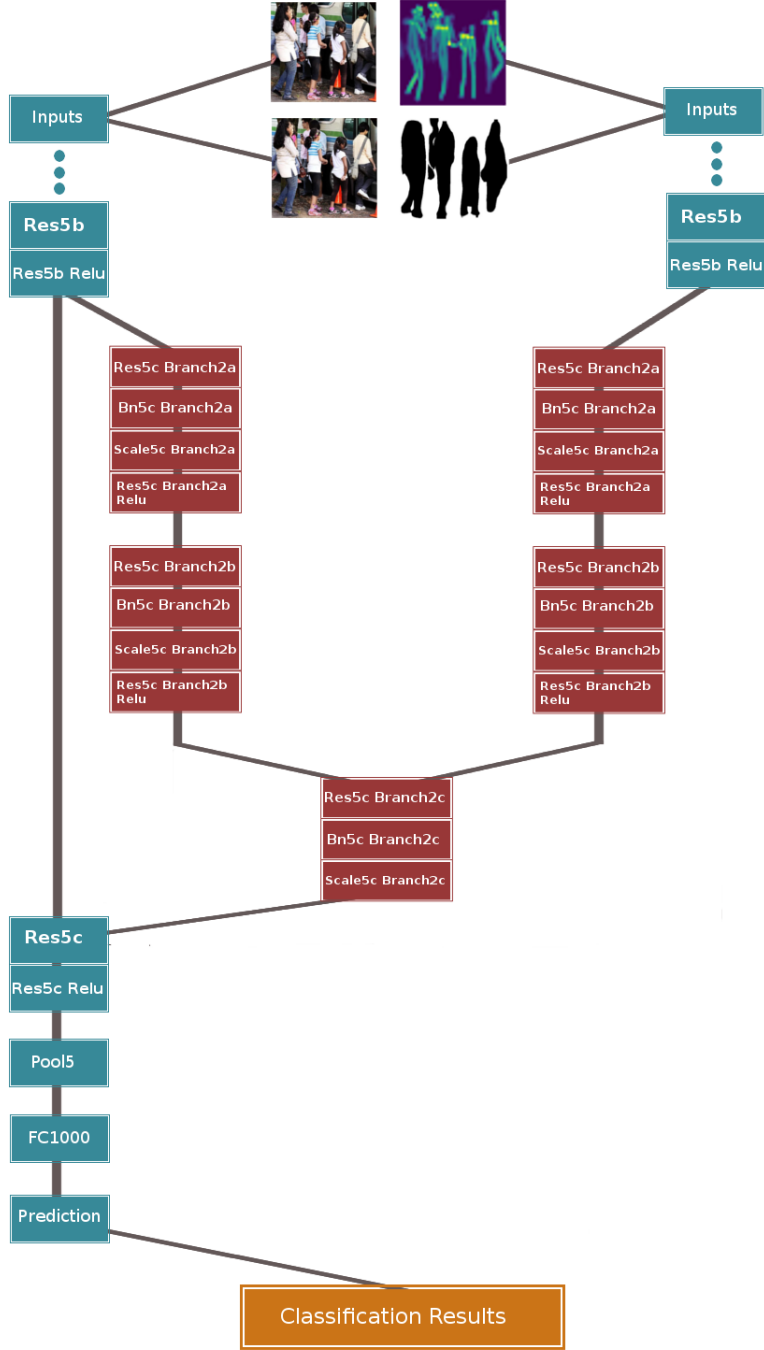
Performans ölçümleri için 4.2.2.1. ve 4.2.2.2. başlıklarındaki yöntemlerle elde edilen 3 kümedeki görüntüler de SGD [3] denektaşı veri kümesinden elde edilmiştir. Test ve eğitim setleri bu kümeler için tamamen birbirinden farklı olarak seçilmiştir. Tablolardaki ilk performans değeri, her bir sınıf için ortalama doğruluğu gösterirken, ikinci performans değeri kümedeki toplam doğruluk değerini gösterir.

Tablo 4.5.'te de görüldüğü üzere, SGD [3] denektaşı veri kümesindeki grup bilgilerinin RGB olarak doğrudan sınıflandırılması ile elde edilen başarı oranı, [3] çalışmasında yer alan sınıflandırma performansından daha yüksek olmuştur.

Tablo 4.6.'da SGD [3] veri kümesinin [12] çalışmasındaki yöntemle elde edilen kümesinin ResNet50 [2] ince-ayar(fine-tune) [4] işlemi sonucunda eğitim aşamasını tamamlayabildiği



ŞEKİL 4.9.: Geç füzyon işleminin şematik akışı gösterilmektedir.



ŞEKİL 4.10.: Erken füzyon işleminin şematik akışı gösterilmektedir.

TABLO 4.5.: Grup bilgilerinin RGB olarak sınıflandırma sonuçları gösterilmektedir. Değerler Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	Grup Aktivite Bilgisi Başarı Sonuçları	
	[3] Çalışması Sonuçları	Tez Çalışması Sonuçları
Eğitim Seti	52.6/66.7	88.5/88.8
Test Seti	43.0/52.7	54.3/59.5

TABLO 4.6.: Poz verisinin sınıflandırma sonucu gösterilmektedir.

Sonuç Türü	Başarı Sonuçları
Sınıf Bazlı Ortalama Başarısı	21.9
Genel Başarı	24.7

TABLO 4.7.: Bölüt verisinin sınıflandırma sonucu gösterilmektedir.

Sonuç Türü	Başarı Sonuçları
Sınıf Bazlı Ortalama Başarısı	31.5
Genel Başarı	33.5

görülmektedir. Bu sonuçlar aynı zamanda elde edilen poz bilgilerinin tek başına grup aktivitesini temsil edecek bir sınıflandırma başarısı oluşturamadığını göstermektedir.

Tablo 4.7.'de SGD [3] veri kümesinin MaskRCNN [8] çalışmasındaki bölütleme çıktılarından elde edilen kümesinin ResNet50 [2] ince-ayar(fine-tune) [4] işlemi sonucunda eğitim aşamasını tamamlayabildiği görülmektedir. Bu sonuçlar aynı zamanda elde edilen bölütleme tek başına grup aktivitesini temsil edecek bir sınıflandırma başarısı oluşturamadığını göstermektedir.

Tablo 4.6. ve 4.7.'de görüldüğü üzere poz bilgisinin ve bölütleme bilgisinin yer aldığı kümelerin doğrudan sınıflandırılması ile elde edilen başarılar yüksek değildir. Doğrudan grup bilgilerinin RGB olarak sınıflandırmaya dahil edilmesiyle elde edilen başarılar diğerlerine göre daha yüksektir. Genel olarak sınıflandırma başarılarını arttırabilmek adına füzyon işlemleri Bölüm 4.2.2.1. ve 4.2.2.2.'de açıklandığı şekliyle uygulanmış ve performans değerleri ölçülmüştür.

Tez kapsamında uygulanan iki farklı füzyon tekniğinden biri olan geç füzyon tekniği uygulanarak elde edilmiş olan sonuçlar Tablo 4.8. ve 4.9'da gösterilmiştir.

TABLO 4.8.: RGB-Poz ikili akış modelinin geç füzyon ile birleştirilmesiyle elde edilen başarı değerleri gösterilmektedir. Değerler Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

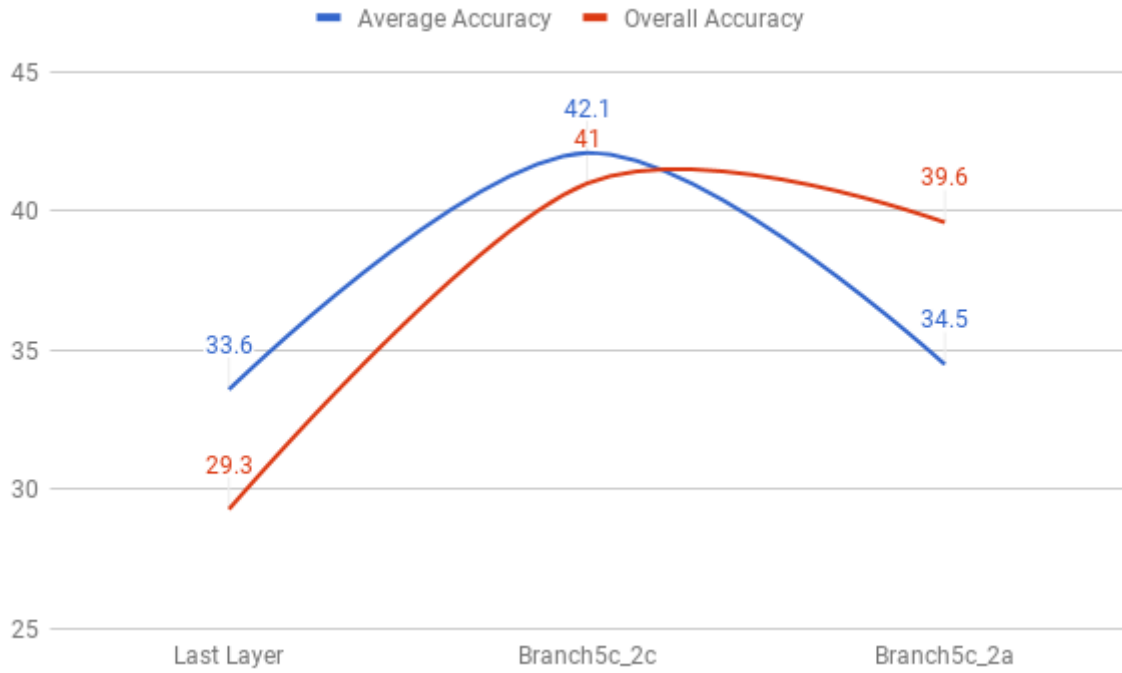
Sınıflar	RGB-Poz İkili Akış Sonuçları	
	Skaler Çarpım	Ortalama
Test Seti	12.4/7.9	53.1/58.4

TABLO 4.9.: RGB-Bölüt ikili akış modelinin geç füzyon ile birleştirilmesiyle elde edilen başarı değerleri gösterilmektedir. Değerler Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	RGB-Bölüt İkili Akış Sonuçları	
	Skaler Çarpım	Ortalama
Test Seti	8.6/7.12	52.8/59.2

Tezde geç füzyon tekniği sınıflandırmasının başarılı bir şekilde değerlendirilmesinden sonra, aynı görüntü kümeleri için erken füzyon tekniği için sınıflandırma başarısı hesaplanmıştır. Geç füzyon tekniğinden farklı olarak, erken füzyon tekniğinde füzyon sürecinin hangi katmandan itibaren yapılacağına karar verilmesi gerekmektedir. Bu kararı almak için çeşitli deneyler gerçekleştirilmiştir. Son katmandan başlanarak daha alt katmanlara inilerek her defasında sınıflandırma başarısı ölçülmüştür. Bu performans ölçümlerinin en yükseği, erken füzyon için birleşmenin yapılacağı katman olarak seçilmiştir. Şekil 4.11.'de, grup bilgisi ve bölütleme bilgisi kullanılarak erken füzyon için birleştirilecek katmanı seçmek için yapılan performans ölçüm grafiği vardır.

Şekil 4.11.'de gösterildiği üzere "Branch5c_2c" katmanında elde edilen sınıflandırma başarısı diğer katmanlara göre daha yüksektir. Öncesindeki ve sonrasındaki katmanlar incelendiğinde "Branch5c_2c" katmanının erken füzyon aşaması için birleşme noktası olarak seçilmesine karar verilmiştir. Tablo 4.10. ve 4.11'de erken füzyon operasyonlarından elde edilen sonuçlar görülmektedir.



ŞEKİL 4.11.: Erken füzyon işleminin farklı katmanlardaki birleşme sonucu sınıflandırma başarısı gösterilmektedir.

TABLO 4.10.: RGB-Bölüt ikili akış modelinin erken füzyon ile birleştirilmesiyle elde edilen başarı değerleri gösterilmektedir. Değerler Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	RGB-Bölüt İkili Akış Sonuçları	
	Skaler Çarpım	Ortalama
Test Seti	42.1/41.0	29.5/30.7

Tablo 4.12.'de ise FC katmanı seviyesinde iki ayrı koldan gelen bilgilerin uç uca eklenmesi yöntemiyle elde edilmiş performans sonuçları görülmektedir.

Tablolarda gösterilen önerilen yöntemlerden elde edilen sonuçların değerlendirilmesi Bölüm 4.2.4.'te açıklanmıştır.

TABLO 4.11.: RGB-Poz ikili akış modelinin erken füzyon ile birleştirilmesiyle elde edilen başarı değerleri gösterilmektedir. Değerler Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	RGB-Poz İkili Akış Sonuçları	
	Skaler Çarpım	Ortalama
Test Seti	26.3/31.6	30.3/32.6

TABLO 4.12.: Uç uca ekleme yöntemiyle elde edilen sınıflandırma sonucu. Değerler Sınıf Bazlı Doğruluk / Genel Doğruluk olarak gösterilmektedir.

Sınıflar	Model Sonuçları	
	RGB-Bölüt Sonuçları	RGB-Poz Sonuçları
Test Seti	28.4/30.7	21.8/20.7

TABLO 4.13.: Alt kümelerden elde edilen en yüksek başarı değerleri gösterilmektedir.

Alt Küme Türü	Elde Edilen En Yüksek Başarı Sonuçları
RGB	59.5
RGB-Poz	58.4
RGB-Bölüt	59.2

4.2.4. Değerlendirme

Çalışmaların sonunda elde edilen sonuçlar incelendiğinde, poz ve bölütleme verilerinden elde edilen bilgilerin genel olarak sınıflandırma hedefine olumlu bir katkı sağlamadığı görülmektedir. Bu kümelermelerin katkılarını gözlemlemek için ilk önce bireysel olarak eğitilmiş ve test edilmiştir. Daha sonra, bu kümeler, grup bilgileri ile geç ve erken füzyon işlemleri uygulanarak birleştirildi ve farklı katmanlarda performans ölçümleri yapıldı. Sonuçlar incelendiğinde grup bilgisinin doğrudan sınıflandırılmasıyla elde edilen başarı oranının Tablo 4.5.'te görüldüğü gibi daha yüksek olduğu görülmektedir. Aynı zamanda elde edilen bu başarı derin öğrenme yöntemlerinin derin olmayan yöntemlere göre daha iyi sınıflandırma sonucu üretebildiğini göstermektedir. Sınıftaki grup bilgisi için kullanılan kümedeki veri sayısı çok büyük olmasına rağmen, sınıflandırma sonucunun daha iyi olduğu görülmektedir. Bu da ön eğitilmiş olan bir model kullanılarak sınıflandırma işleminin gerçekleştirilmesinden kaynaklanmaktadır. Öte yandan, grup aktivite bilgilerinin doğrudan sınıflandırılması problemi de zor bir problem olarak düşünebilir. Çünkü aynı sınıfta yer alan bir grup aktivitesi, birçok farklı form ve kombinasyonda alabilir. Örneğin, insan yüzünün veya izleyicinin konumuna bağlı olarak,

iki kiři yan yana birbirlerinden bağımsız olarak duruyor veya birbirleriyle sohbet ediyor olabilirler. Tablo 4.5.'te gösterildiđi gibi, grup bilgisinin eğitim kümesindeki sınıflandırma başarıları % 88.5 / % 88.8 ve test sınıflandırma başarıları % 54.3 / % 59.5 olarak hesaplanmıştır. Buradaki eğitim ve test başarıları arasındaki farkın nedeninin, eğitim kümeleneşlerinin sayı ve çeşitlilik açısından yetersiz olması görülebilir.

Tablo 4.6. ve Tablo 4.7.'de gösterilen sonuçlar incelendiđinde, poz ve bölütleme bilgilerinin tek başına başarılı bir sınıflandırma sonucu üretmediđi görülmektedir. Tablo 4.6. ve Tablo 4.7.'deki sonuçlar incelendiđinde, elde edilen sınıflandırma performansının, grup bilgilerinin doğrudan sınıflandırılmasıyla elde edilen Tablo 4.5.'te elde edilen sonuçlardan daha düşük olduđu görülmektedir. Örneđin, problemi poz bilginin detayında incelersek, eklem bilgilerinin görüntülerde her birey için aynı hassasiyetle çalışmadıđı görülür. Şekil 4.12'de görüldüđu gibi, resimdeki her birey için tüm eklem bölgeleri çıkartılamaz. İlk bireyi temsil eden bir özellik olsa da, diđer bireyler için bir sonuç üretilememiştir. Bu nedenle, grubu temsil eden özellikler kaybolur ve sınıflandırma, sınıfı başarıyla temsil edemeyen ağırlıklardan oluşan bir model ile gerçekleştirilmiş olur. Bu ve benzeri örnekler kümedeki birçok görüntü için kullanılabilir. Bu durum da, tez kapsamında uygulanan yöntemlerin sınıflandırma başarıları ve füzyon süreçlerini doğrudan etkiler.



ŞEKİL 4.12.: Örnek bir görüntü üzerinden poz bilgisi gösterilmektedir.

Poz ve bölütleme bilgilerinin bağımsız olarak eğitilmesi neticesinde yukarıda belirtilen sebeplerden dolayı başarılı sonuçlar üretilememiştir. Bu durumun önüne geçebilmek adına

füzyon işlemine başvurulmuştur. Amaç doğrudan grup bilgisinin kullanılmasıyla elde edilen sınıflandırma başarısına olumlu bir katkı yapabilmektir.

İlk olarak poz ve bölütleme bilgileri grup bilgileri ile sırasıyla geç füzyon işlemine tabi tutulmuştur. Bilgilerden gelen ağırlıklar element bazında çarpma ve aritmetik ortalama ile birleştirilmiştir. Tablo 4.8. ve 4.9.'daki değerler incelendiğinde element bazlı çarpmanın farklı kollardan gelen bilgilerinin etkilerini birbirleri üzerine daha fazla yansıttığı görülmektedir. Aritmetik ortalama ile elde edilen doğruluk değerleri element bazlı çarpma işlemine göre daha yüksek olmasına rağmen grup bilgilerinin doğrudan sınıflandırılmasıyla elde edilen sınıflandırma başarısına göre daha düşük olduğu görülmektedir.

Son olarak poz ve bölütleme bilgileri grup bilgileri ile sırasıyla erken füzyon işlemine tabi tutulmuştur. Tablo 4.10. ve 4.11. incelendiğinde elde edilen sınıflandırma başarılarının düşük olduğu ancak geç füzyon işlemine göre biraz daha yüksek olduğu görülmektedir. Bunun sebebi olarak erken katmanda gerçekleşen birleşme işleminden sonra ağırlıkların sınıflandırma katmanına kadar güncellenmeye devam etmesi gösterilebilir.

Genel olarak sonuçlar incelendiğinde poz ve bölütleme bilgilerinin de grup bilgilerinin sınıflandırma başarılarına pozitif etki yapamadığı görülmektedir. Bunun sebebi olarak işlem yapılan denektaşı veri kümesinde bulunan grup görüntülerinin yeterli çeşitlilik ve sayıda olmaması gösterilebilir.

5. YAZILIM ve KÜTÜPHANELER

Tez kapsamında kullanılmış olan kütüphaneler ile geliştirilmiş olan yazılım parçalarının büyük çoğunluğu Python dili ile düzenlenmiştir. Derin öğrenme içeren kod parçalarının hepsi Tensorflow [27] altyapısını kullanarak geliştirilmiştir. MaskRCNN [8] çalışmasının Tensorflow için geliştirilmiş olan versiyonun kullanılması ile tez kapsamında geliştirilen diğer kod parçalarına daha kolay entegrasyon için yardımcı olmuştur. Diğer yandan bireylerin eklem bilgilerinin çıkarıldığı tf-openpose [28] isim çalışmanın da Tensorflow versiyonu kullanılmış ve mevcut probleme uygun hale getirilmiştir.

[3] çalışmasında bulunan SGD veri kümesinin etiket bilgileri Matlab verisi ve o verilerin kullanımına yönelik olarak bir betik verilmiştir. Bu veri ve betik sayesinde SGD veri kümesi içerisindeki tekil birey bilgileri ile grup bilgilerine ulaşılabilmektedir.

İnce-ayar (fine-tune) [4] işlemi için de yine Tensorflow [27] kütüphanesi kullanılmıştır. Bu kütüphaneye ek olarak Keras kütüphanesi kullanılarak daha modüler bir tasarım uygulanmıştır. İnce-ayar (fine-tune) [4] işlemi için ResNet [8] mimarisi tercih edilmiştir. ResNet [8] mimarisinin imageNet veri kümesi ile eğitilmesinden elde edilen ağırlıklar kullanılmıştır. Bu sayede ince ayar işleminde dondurulan katmanlar imageNet veri kümesi ile eğitilmiş ağırlıkları içermiş, dondurulmayan katmanlar da dondurulan katmanlardaki ağırlıklar ile SGD [3] denektaşı veri kümesindeki verilerin CNN [1] altyapısındaki katmanlardaki işlemler sonucunda oluşturulmuştur.

Tüm teknik altyapının birbiri ile aynı dili konuşacak şekilde planlanmış olması uçtan uca çalışabilecek bir sistemin doğmasına yardımcı olmuştur. Sistemin çalışabilmesi adına en büyük ihtiyacı GPU ihtiyacıdır. Derin öğrenme işlemlerinin GPU ile birlikte etkili ve hızlı çalışabilme gereksiniminden dolayı GPU barındıran bir sistem ile ancak geliştirilen kodlar çalışabilmektedir. Eğitim aşamasındaki tercih edilen küme büyüklüğü değeri (batch size) GPU kartında kullanılacak olan bellek miktarını doğrudan etkileyen bir faktördür. Aynı zamanda küme büyüklüğü (batch size) değeri eğitim başarı oranını da etkileyen bir parametredir. Ancak tez kapsamında kullanılan SGD [3] denektaşı veri kümesinde sınıf başına düşen

örneklem sayısı çok fazla olmadığından dolayı küme büyüklüğü (Batch size) değerini de çok büyük seçmek başarıyı yükselten değil düşüren faktörler arasında yer almaktadır. Parametre değerlerine göre performansın değişimini gösteren tabloda da görüldüğü üzere SGD [3] denektaşı veri kümesi üzerinden yapılan mevcut sınıflandırma işlemleri için 8 ve 16 değeri yeterli olmaktadır.

6. SONUÇ & GELECEK ÇALIŞMALARI

Tez kapsamında, sabit görüntüler üzerinden grup aktivite bilgilerinin elde edilmesi için çalışmalar yapılmıştır. Bu bağlamda, ana odak alanı olarak tespit ve sınıflandırma aşamalarının derin öğrenme teknikleri ile gerçekleştirilmesi hedeflenmiştir. Tezde karşılaşılan temel zorluk, denektaşı veri kümesindeki örneklerin ve çeşitliliğin derin öğrenme için yeterli olmasıdır. Bu durumun üstesinden gelmek için ince-ayar(fine-tune) [4] yöntemiyle önceden eğitilmiş bir CNN [1] modeli kullanılmıştır. Bu şekilde, denektaşı veri kümesindeki az sayıda örnek nedeniyle ortaya çıkacak olumsuz etkilerin ortadan kaldırılması amaçlanmaktadır. Ayrıca, aynı kümeden elde edilen farklı özniteliklerin çeşitli operasyonlarla birleştirilerek sınıflandırma başarısının artırılması amaçlanmıştır. Ancak, çalışılan denektaşı veri kümesinin sayısı ve çeşitlilik açısından yeterli düzeyde olmaması nedeniyle elde edilen başarılar belli bir seviyeye getirilmiş, daha ileriye taşınamamıştır.

Problemin çözümü için planlanan çalışma ise, farklı denektaşları kullanılarak deneylerin tekrarlanmasıdır. Hedeflenen denektaşı veri kümelerinin çok fazla örnek içereceği ve sınıfları temsil eden görüntü çeşitliliğinin fazla olması beklenen özelliklerden biridir. Ayrıca, derin öğrenme modellerinin performans değerlerinin birbirleriyle karşılaştırılması için farklı modellerde eğitim aşamasının tamamlanması ve elde edilen modellerin füzyon süreçlerine tabi tutulması da uygulanabilecek hedefler arasındadır. Yapılan deneyler sonucunda, derin öğrenme ve evrimsel sinir ağı tabanlı yöntemlerin, grup aktivite tanıma problemi için umut verici olduğu görülmektedir. Her ne kadar incelenmiş olan poz ve bölüt yapılarının katkıları belirlenen düzeyde olmasa da, farklı birleşim yöntemleri ve daha büyük veri kümelerinin kullanılması ile bu bilgilerden de ek çıkarımlar yapılabilir.

REFERANSLAR

- [1] Krizhevsky A. Sutskever I. & Hinton G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105. **2012**.
- [2] He K. Zhang X. Ren S. & Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778. **2016**.
- [3] Choi W. Chao Y. W. Pantofaru C. & Savarese S. Discovering groups of people in images. In *European conference on computer vision*, pages 417–433. **2014**.
- [4] Y. Wang J. Song Y. Leung T. Rosenberg C. Wang J. Philbin J. & Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1386–1393. **2014**.
- [5] Bourdev L. & Malik J. Poselets: Body part detectors trained using 3d human pose annotations. In *IEEE 12th International Conference*, pages 1365–1372. **2009**.
- [6] Singh S. Gupta A. & Efros A. A. Unsupervised discovery of mid-level discriminative patches. In *Computer Vision–ECCV*, pages 73–86. **2012**.
- [7] Dalal N. & Triggs B. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference*, pages 886–893. **2005**.
- [8] He K. Gkioxari G. Dollár P. & Girshick R. Mask r-cnn. In *Computer Vision (ICCV)*, pages 2980–2988. **2017**.
- [9] Chang C. C. & Lin C. J. Libsvm: a library for support vector machines. In *ACM transactions on intelligent systems and technology*, pages –. **2011**.
- [10] Ke Q. Bennamoun M. An S. Sohel F. & Boussaid F. A new representation of skeleton sequences for 3d action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4570–4579. **2017**.

- [11] Gers F. A. Schmidhuber J. & Cummins F. Learning to forget: Continual prediction with lstm. In -, pages -. **1999**.
- [12] Cao Z. Simon T. Wei S. E. & Sheikh Y. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, pages Vol. 1, No. 2, p. 7. **2017**.
- [13] Amer M. R. Xie D. Zhao M. Todorovic S. & Zhu S. C. Cost-sensitive top-down/bottom-up inference for multiscale activity recognition. In *European Conference on Computer Vision*, pages 187–200. **2012**.
- [14] Felzenszwalb P.F. Girshick R.B. McAllester D. Ramanan D. Object detection with discriminatively trained part-based models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1627–1645. **2010**.
- [15] Laptev I. Marszalek M. Schmid C. Rozenfeld B. Learning realistic human actions from movies. In *CVPR*, pages -. **2008**.
- [16] Deng Z. Zhai M. Chen L. Liu Y. Muralidharan S. Roshtkhari M. J. & Mori G. Deep structured models for group activity recognition. In *arXiv*, page 1506.04191. **2015**.
- [17] Deng J. Dong W. Socher R. Li L. J. Li K. & Fei-Fei L. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition(CVPR)*, pages 248–255. **2009**.
- [18] Yangqing Jia Evan Shelhamer Jeff Donahue Sergey Karayev Jonathan Long Ross Girshick Sergio Guadarrama and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *arXiv*, page 1408.5093. **2014**.
- [19] Choi W. Shahid K. & Savarese S. Learning context for collective activity recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3273–3280. **2011**.

- [20] Toshev A. & Szegedy C. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660. **2014**.
- [21] B. Sapp and B. Taskar. Modec: Multimodal decomposable models for human pose estimation. In *CVPR*, pages –. **2013**.
- [22] S. Johnson and M. Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *BMVC*, pages –. **2010**.
- [23] Chao Y. W. Liu Y. Liu X. Zeng H. & Deng J. Learning to detect human-object interactions. In *arXiv*, page 1702.05448. **2017**.
- [24] R. Girshick. Fast r-cnn. In *arXiv*, page 1504.08083. **(2015)**.
- [25] Ren S. He K. Girshick R. & Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99. **2015**.
- [26] <https://keras.io/>. In -, pages –. -.
- [27] <https://www.tensorflow.org/>. In -, pages –. -.
- [28] <https://github.com/ildoonet/tf-pose-estimation>. In -, pages –. -.

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı, Soyadı : Anıl ATVAR
Doğum Yeri : Ankara
Medeni Hali : Evli
E-posta : n14320602@cs.hacettepe.edu.tr
Adres : Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü

Eğitim

Lisans : Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü
Yüksek Lisans : Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü
Doktora : -

Yabancı Dil ve Düzeyi

İngilizce (Üst Düzey)

İş Deneyimi

Turkcell İletişim Hizmetleri A.Ş., Havelsan A.Ş.

Deneyim Alanları

Yazılım Geliştirme

Tezden Üretilmiş Projeler ve Bütçesi

-

Tezden Üretilmiş Yayınlar

Atvar, A., İkizler-Cinbis, N. Classification of Human Poses and Orientations with Deep Learning.
In Signal Processing and Communication Application Conference (SIU), 2018 IEEE.

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar

In Signal Processing and Communication Application Conference (SIU), 2018 IEEE



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 05/11/2018

Tez Başlığı / Konusu: DERİN ÖĞRENME İLE GRUP HAREKETLERİNİN SABİT RESİM ÜZERİNDEN TANINMASI

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 64 sayfalık kısmına ilişkin, 01/10/2018 tarihinde tez danışmanım tarafından *Turnitin* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 3 'tür.

Uygulanan filtrelemeler:

- 1- Kaynakça dâhil
- 2- Alıntılar dâhil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Tarih ve İmza

Adı Soyadı: ANIL ATVAR

Öğrenci No: N14320602

Anabilim Dalı: BİLGİSAYAR MÜHENDİSLİĞİ

Programı:

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

05/11/2018

DANIŞMAN ONAYI

UYGUNDUR.

Doç. Dr. Nazlı İKİZLER CİNBİÇ

(Unvan, Ad Soyad, İmza)