

Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma

Literatür Makalesi/Review Article

Merve Ayyüce KIZRAK¹, Bülent BOLAT²

¹Elektrik-Elektronik Mühendisliği, Haliç Üniversitesi, İstanbul, Türkiye

²Elektronik ve Haberleşme Mühendisliği, Yıldız Teknik Üniversitesi, İstanbul, Türkiye

ayyucekizrak@halic.edu.tr, bbolat@yildiz.edu.tr

(Geliş/Received:27.04.2018; Kabul/Accepted:30.07.2018)

DOI: 10.17671/gazibtd.419205

Özet— Yapay sinir ağları ve makine öğrenmesi, uzun yıllardır birçok problemin çözümünde kullanılmıştır. Problemlerin ve modellerin karmaşılaşması ve veri sayısındaki artış hesaplama yükünü de beraberinde getirmiştir. Bu çalışmada yapay sinir ağlarından derin öğrenmeye tüm geçiş süreci, modeller ve pratik uygulamalar kısa ve öz gösterilmiştir. Ayrıca donanım, yazılım ve kullanılan kütüphaneler hakkında da bilgiler verilmiştir. Özel olarak kalabalık analizi için kullanılan geleneksel yöntemler özetlenmiştir. Kalabalık analizi için literatürdeki derin öğrenme yaklaşımları detaylıca anlatılmış ve veri kümeleri tanıtılmıştır. Ayrıca son yıllarda yapılmış çalışmalar analiz edilmiş ve karşılaştırılmıştır. Sonuç olarak, kalabalık analizi, derin öğrenme yardımıyla başarılı sonuçlar alınan hem akademik hem de pratik bir çalışma alanıdır.

Anahtar Kelimeler— derin öğrenme, yapay sinir ağları, evrişimli sinir ağları, öz yinelemeli sinir ağları, kalabalık analizi

A Comprehensive Survey of Deep Learning in Crowd Analysis

Abstract— Artificial neural networks and machine learning have been used to solve many problems for decades. The complexity of the problems and models and the increase in the number of data also brought with it the computation burden. In this study, the whole transition process from artificial neural networks to deep learning, models and applications are briefly demonstrated. Additionally information about hardware, software, and used libraries is also provided. In particular, canonical methods for crowd analysis have been summarized. Deep learning approaches in the literature are pointed out in depth for crowd analysis and datasets are overviewed. Furthermore, studies done in recent years have been analyzed and compared. Consequently, crowd analysis is both an academic and a practical field of study where successful results evaluation. As a result, crowd analysis is both an academic and a practical field where fruitful results are achieved with the help of deep learning.

Keywords— deep learning, artificial neural networks, convolutional neural networks, recurrent neural networks, crowd analysis

1. GİRİŞ (INTRODUCTION)

Gündelik yaşamda karşılaşılan birçok problemin çözümünde hareketsiz görüntüler veya video görüntüleri akıllı algoritmalar ile işlenmesi gerekmektedir. Örneğin büyük şehirlerde güvenliğin sağlanması nüfusun yoğunluğuna bağlı olarak zorluğu artan bir sorundur. Bu yüzden şehirlerde sokaklar, caddeler ve kamuya açık ortak kullanım alanları güvenlik kameralarıyla donatılmıştır. Ayrıca askeri maksatlı olarak kullanılan insansız kara, hava ve deniz araçları veya benzeri güvenlik amaçlı uygulamalarda da görüntülerin akıllı algoritmalar ile otomatik işlenmesi önemlidir. Güvenlik konusunun yanı

sıra artan nüfusa bağlı olarak sosyal alanların yönetilmesi, park, bahçe, stadyum, alışveriş merkezlerinin kapasiteleri ve verimli kullanılması gibi kalabalık analizine ihtiyaç duyulan alanlar mevcuttur. Toplumun yaşam kalitesini artırmaya yönelik çalışmalara ışık tutması nedeniyle video ve fotoğraf görüntülerinden yola çıkarak problemlere çözümler getirilmesi oldukça önemli bir konudur.

Sosyal yaşamın modellenmesinde kullanılan aranmakta olan belirli kişilerin bulunması, kalabalık yoğunluğunun analizi, kalabalık benzetimi, anormal ve tehlikeli durumların saptanıp değerlendirilmesi, işaretlenen kişi ve/veya grupların takibi, kamuya açık alanların

tasarlanması, spor müsabakalarının değerlendirilmesi, müze ve kültür-sanat toplantı alanlarının tasarlanması ve giriş-çıkış yönetimi gibi konularda statik görüntü ve video görüntülerinin analiz edilmesi gibi problemlerin gerçek zamanlı olarak çözülebilmesi büyük önem arz etmektedir. Ayrıca arenalar, stadyumlar, sinema, tiyatro salonları ve alışveriş merkezleri gibi insan akışının yoğun olduğu yerlerde de durağan görüntü ve video görüntüleri kullanılarak istatistiksel ve veri analizine yönelik çalışmalar yapılmasına her zaman ihtiyaç duyulmaktadır [1].

1980'li yıllarda temelleri atılmış olan derin öğrenme, 1990'lı yıllarda donanım kaynaklı engeller nedeniyle bir duraklama devri geçirmiş olsa da günümüzde yapay zekanın en popüler alt dalı olma özelliğini taşımaktadır. Her geçen gün gelişmekte olan bilgisayarlı görü, doğal dil işleme, otonom araçlar, müzik, sanat, savunma sanayii, güvenlik ve finans gibi birçok çalışma alanında uygulamaları bulunan derin öğrenme alanında çalışma yaparken yüksek hızlı grafik işlem birimlerine (GPU), yüksek kapasiteli belleklere ve/veya bulut çalışma ortamlarına ihtiyaç duyulmaktadır.

Yapay zekâ kavramı ilk kez 1950'li yıllarda McCharthy tarafından ortaya atılmış ve 1957'de Rosenblatt yapay sinir ağlarının temel yapı taşı olan algılayıcı (perceptron) tanımını yapmıştır [2]. İnsan sinir sisteminden ilham alarak oluşturulan yapay sinir ağlarında kullanılmak üzere çok katmanlı yapıya sahip olan denetimli öğrenme algoritması, ilk kez Ivakhnenko ve Lapa tarafından 1965 yılında yayınlanmıştır [3]. 1980 yılında ise, Fukushma ilk derin öğrenme mimarisi sayılabilecek 'neocognitron' yapısını önermiştir. Bu yapı denetimsiz bir öğrenme yapısına sahip olduğu için çok önemli bir adım olarak nitelendirilebilir [4]. 1986 yılında Hinton ve ekibi bugün hala oldukça yaygın kullanıma sahip öğrenme algoritması olan 'geriye yayılım' algoritmasını geliştirmiştir [5,6].

60'lı yılların ikinci yarısından itibaren popülerliğini yitirmeye başlayan yapay zeka kavramı, 1997'de IBM Deep Blue yazılımının Dünya satranç şampiyonu Gary Kasparov'u mağlup etmesi ile birlikte yeniden popüler bir konu haline gelmiştir [7]. Bu başarının ardından IBM Watson, ABD'nin ünlü TV programı Jeopardy bilgi yarışmasında en başarılı iki yarışmacıyı büyük farkla yenmiştir [8]. 1998 yılında LeCun ve ekibi, rakamları tanımak ve sınıflandırmak için tasarlanmış oldukları LeNet-5 olarak bilinen modelle elde ettikleri başarılı sonuçlarla evrişimli sinir ağlarının tekrar gündeme gelmesini sağlamıştır [9]. Ancak o yıllardaki bilgisayar ve donanım koşulları bu tip görüntülerin çok katmanlı sinir ağı modelleriyle hesaplanabilmesine elverişli değildi. Donanım kısıtının yanı sıra modellerin uygulanabilmesi için gerekli olan miktarda etiketli verinin de olmayışı, bu alanda çalışan akademisyenlerin konuya olan ilgisinin azalmasına neden oldu ve "yapay zekâ kıışı" adı verilen bir döneme girilmiştir [10].

2009 yılında Li ve ekibi, 167 ülkeden yaklaşık 50 bin çevrimiçi çalışanla önışlemleri yapılarak etiketlenmiş 22

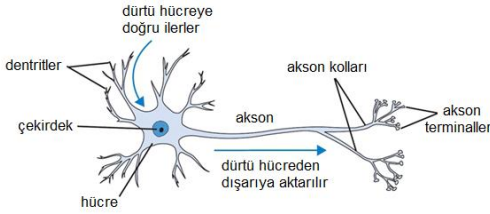
bin kategori ve 15 milyon görüntüden oluşan görüntü veri kümesi ImageNet'i oluşturarak sorunlardan birinin çözümüne yönelik önemli bir adım atmıştır. 2010 yılında düzenlenmeye başlanan ILSVRC (ImageNet Large Scale Visual Recognition Competition) yarışması ile birlikte büyük verilerle yapılan çalışmalar yeniden gündeme gelmiştir. Krizhevsky, Sutskever ve Hinton'ın 2012 yılında oluşturdukları evrişimli sinir ağı modeli AlexNet, ILSVRC yarışmasında birinci olarak bir önceki yılın hata oranını yaklaşık olarak yarıya düşürmüştür. 8 katmanlı bu evrişimli sinir ağı ile ilk-5 hata oranı %16,4'e kadar azaltılmıştır. Bu başarıyı takiben derin öğrenme yöntemleri popülerlik kazanmaya başlamış ve bu alanda yapılan akademik çalışmalar büyük bir ivme kazanmıştır [11]. 2014 yılında Szegedy ve ekibi, GoogLeNet modeli ile inception modülleri çok daha derin katmanlara sahip bir model ile çok daha başarılı sonuçlar elde etmiştir [12]. Aynı yıl Goodfellow çekışmeli üretici ağlar (Generative Adversarial Network-GAN) ile farklı problemlere çözüm sunmuştur [13]. 2016 yılında Microsoft'un geliştirdiği TAY isimli sohbet robotu insanlarla konuşurken öğrenmeye devam eden bir modele sahiptir. TAY zaman içinde insanların yazdıklarından öğrendikleri ile ırkçı söylemler üretmeye başlayınca yayından kaldırıldı [14]. 2017 yılında Google DeepMind şirketinin AlphaGo yazılımı çok kompleks bir oyun olan Go'da dünya şampiyonunu mağlup ederek büyük ses getirdi [15]. Yine 2017 yılının sonlarında yapay zekânın babası olarak anılan Hinton ve ekibinin dinamik yönlendirme algoritması ile kapsül ağı modelini sunması derin öğrenme modelleri içinde yaygın olarak kullanılan evrişimli sinir ağlarına yeni bir bakış kazandırmış oldu [16].

Yapay zekanın 80'li yıllardan günümüze çok önemli başarılar kaydederek gelmesi ve sayısız probleme çözüm üretmesi, konuyu çeşitli disiplinler için daha da önemli ve öncelikli bir hale getirmektedir. Bu çalışmanın konusu olan kalabalık analizi ile insanların sayılması ve takip edilmesi hem güvenlik zafiyetini önlemek hem de efektif sosyal ve kültürel alanlar tasarlamak için gerekli ve önemlidir. Derin öğrenme modelleri ile kalabalık analizi, 2012 yılından beri sürekli gelişen ve bu alanla ilgilenen araştırmacıların ilgisini çeken bir konudur.

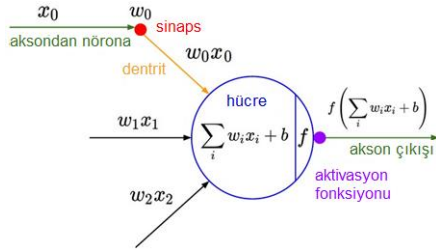
Bu çalışmanın ilk bölümünde yapay zeka tarihçesinden ve günümüze gelişinden söz edilmektedir. İkinci bölümde yapay sinir ağlarından türeyen derin sinir ağları konusunda detaylı bir özet sunulmaktadır. Özellikle derin öğrenme alanında kullanılan donanım, yazılım ve kütüphaneler hakkında güncel bilgiler sunulmaktadır. Üçüncü bölümde literatürde kalabalık analizine geleneksel yaklaşımlar özetlenmektedir. Dördüncü bölümde ise detaylı olarak kalabalık analizi problemine literatürdeki derin öğrenme yaklaşımları anlatılmakta ve başarımları tablolar halinde sunulmaktadır. Sonuç olarak, güncel, araştırılmaya ve geliştirilmeye açık bir konu olan kalabalık analizi konusunda ileri seviye çözüm yöntemleri araştırılarak kapsamlı bir inceleme yapılmıştır.

2. YAPAY SİNİR AĞLARI (ARTIFICIAL NEURAL NETWORKS)

Canlıların sinir sisteminin ve karar mekanizmalarının matematiksel olarak modellenmesi yapay sinir ağlarının temel konusudur. Eğitilebilir, kendi kendine organize olarak uyarlanabilir halde değerlendirme yapabilen yapay sinir ağları ile canlı beyninin öğrenme yapısı yapay olarak modellenmesi amaçlanmaktadır. Makinelerin bu vasıta ile insansı bir öğrenme ve karar verme mekanizmasına sahip olması beklenmektedir [17].



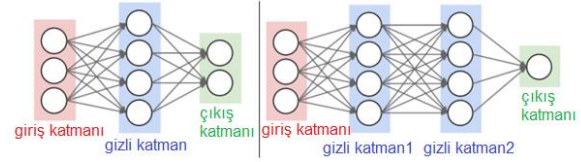
Şekil 1. İnsan nöronunun (sinir hücresi) yapısı [18]
(Human neural cell structure [18])



Şekil 2. Bir nöronun (sinir hücresi) matematiksel modeli [18]
(Mathematical model of a neural cell [18])

Dendritler sistemin girişi, aksonlar ise sistemin çıkışı olarak tanımlanmaktadır. Sinapslar, giriş ve çıkışların birbirine bağlanmasını sağlamaktadır. Çekirdek, elektriksel işaretin periyodik olarak üretilmesini sağlamaktadır. Miyelin tabaka ise yayılma hızına etki eden bir yalıtım malzemesidir. Aksonda taşınan işaret, sinapslara kimyasal taşıyıcılar yardımıyla iletilmektedir. Polarize sitoplazma -85mV değerindedir. Sitoplazma potansiyel değeri -40mV (Na⁺ iç) seviyesine vardığında iletim (+) değerli bir akımın oluşturur. Potansiyel -90mV (K⁺ dış) değerine ulaştığında ise bastırma (-) akımına yol açar. Sonuç olarak hücre uyarılması için belli bir eşik değerinin üzerinde gerilim uygulanmış olması gerekirken aksi halde sinirde iletim durdurulmuş olur. Bu yolla elde edilen çıkış değerine sinirsel hesaplama denmektedir. Yapay sinir ağının en küçük parçası olarak bilinen perseptron (algılayıcı), aşağıdaki gibi doğrusal bir fonksiyonla denklem (1)'deki gibi ifade edilmektedir. [2].

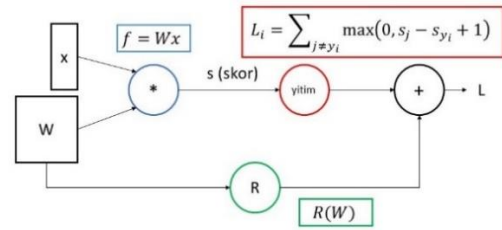
$$y = W \times x + b \quad (1)$$



Şekil 3. Tek ve çok gizli katmanlı yapay sinir ağı modelleri [18]

(Single and multi hidden layer artificial neural networks [18])

Her katmanın modele etkisi, her katmandaki nöronun da katmana, dolayısıyla modele etkisi bulunmaktadır. Klasik bir yapay sinir ağında bir katmanda bulunan nöronların birbirleriyle ilişkileri yoktur ve bilgi ancak bir önceki katmandan bir sonraki katmana ya da çıkışa doğru aktarılır. Peşpeşe gelen iki katmandaki nöronlar birbirlerini çeşitli aktivasyon değerleriyle etkilemektedir. Katmanlarda bulunan nöron sayıları sistemin performansını dolaylı olarak etkilemektedir. Bir yapay sinir ağı modeli tasarlanırken ağırlık vektörlerinin başlangıç değerlerinin nasıl belirlenmesi gerektiği de önemli konulardan biridir. Ağırlıklar başlangıçta rastgele atanabileceği gibi geçmişte eğitilmiş bir modelin ağırlıkları da başlangıç ağırlıkları olarak tanımlanabilir. Ancak bu değerler atanırken (+) ya da (-) değerlerden rastgele dağılımlı olarak belirlenmelidir. Başarılı şekilde tasarlanmış bir ağ modelinin başarımını belirleyen kayıp/yitim fonksiyonu (loss function) değeri çeşitli optimizasyon teknikleriyle 0'a yaklaşması istenen bir fonksiyondur [6].



Şekil 4. Temel bir yapay sinir ağı yapısında yitim fonksiyonuna düzenleme etkisi [6,18]

(Effects of regularization on loss function in a basic artificial neural network structure [6,18])

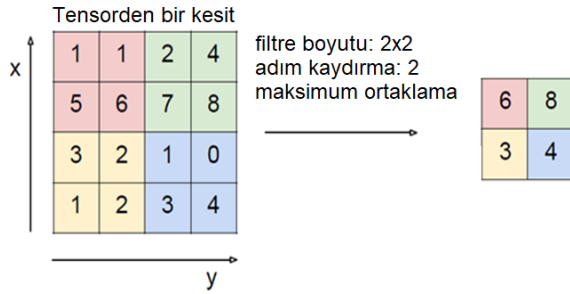
Yitim değerine eklenen düzenleme (regularization-R) değerinden dolayı yitim hiçbir zaman sıfır olmamaktadır. Düzenleme sayısı kullanılması zorunlu olmayan ancak pozitif modele pozitif etkisi dolayısıyla tavsiye edilen bir hiper parametredir. Başarım, yitim fonksiyonunun minimize edilmesiyle doğru orantılı olarak değerlendirilmektedir. Yitim fonksiyonunun parçalı-doğrusal bir matematiksel yapıda olduğu anlaşılmaktadır [18]. Yitim fonksiyonu genel olarak denklem (2) ve (3)'ten yola çıkarak denklem (4)'teki gibi elde edilmektedir.

$$R(W) = \sum_k \sum_l W_{k,l}^2 \quad (2)$$

$$L = \underbrace{\frac{1}{N} \sum_i L_i}_{\text{veri yitimi}} + \underbrace{\lambda R(W)}_{\text{düzenleme yitimi}} \quad (3)$$

belirlemek için $p=(f-1)/2$ denklemi kullanılmaktadır. Ayrıca evrişim işleminde gerçekleştirilen kaydırma işleminin adım sayısı (stride) da çıkış boyutunu etkilemektedir. Şekil 6'da gösterilen örnekte adım kaydırma sayısı $s=2$ olarak seçilmiştir. Pksel ekleme işleminde eklenen pikseller sıfır değerlerinden oluşabildiği gibi yanındaki pikselin değerinin kopyalanması ile de gerçekleştirilebilmektedir [18], [22].

Evrişimli sinir ağı modelinde boyut azaltmak için kullanılan bir diğer işlem de ortaklama (pooling) işlemidir. Bu katmanda maksimum ortaklama (max-pooling) yöntemi yaygın olarak kullanılmaktadır. Ağın maksimum ortaklama katmanında öğrenilen parametre bulunmamaktadır. Ortaklama işleminde giriş matrisinin kanal sayısı sabit tutulurken, yükseklik ve genişlik azaltılmaktadır. Genelde hesaplama karmaşıklığını azaltmak için ortaklama işlemi, Hinton'ın kapsül teorisine göre verideki önemli bazı bilgilerin de kaybolmasına sebep olduğu için başarımdan ödün verilmesine neden olmaktadır.

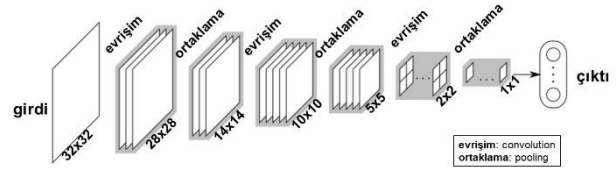


Şekil 6. Maksimum ortaklama işlemi [18]
(Maximum pooling [18])

3.1.1. Güncel Evrişimli Sinir Ağı Modelleri (The State of The Art Convolutional Neural Networks)

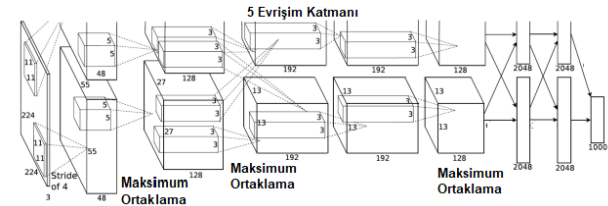
Literatürde yer alan güncel evrişimli sinir ağı modelleri şöyle sıralanabilir.

LeNet-5, 1998 yılında yayınlanmış ve ilk başarılı sonucu veren evrişimli sinir ağı modelidir. LeCun ve ekibi tarafından posta numaraları, banka çekleri üzerindeki sayıların okunması için geliştirilmiştir. LeNet-5 modeli MNIST (Modified National Institute of Standards and Technology) [9] veri kümesi ile eğitilmiştir. Bu modelde sonradan geliştirilecek diğer modellerden farklı olarak boyut azalma adımlarında maksimum ortaklama yerine ortalama ortaklama işlemi yapılmaktadır. Ayrıca aktivasyon fonksiyonu olarak da sigmoid ve hiperbolik tanjant kullanılmaktadır. Tüm bağlantı (fully connected-FC) katmanına giren parametre sayısı $5 \times 5 \times 16 = 400$ ve y çıkışında 0-9 arasındaki rakamları sınıflandırdığı için 10 sınıflı esnek eşikleyici (softmax) bulunmaktadır. Bu ağ modelinde, 60 bin parametre hesaplanmaktadır. Ağ boyunca matrisin yükseklik ve genişlik bilgisi azalırken derinlik (kanal sayısı) değeri artmaktadır [9].



Şekil 7. Temel bir evrişimli sinir ağı modeli
(A basic convolutional neural network model)

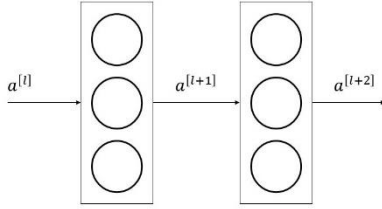
AlexNet, 2012 yılında evrişimli sinir ağı modellerinin ve derin öğrenmenin yeniden popüler hale gelmesini sağlayan ilk çalışmadır. Krizhevsky, Sutskever ve Hinton tarafından geliştirilmiştir. Birbirini takip eden evrişim ve ortaklama katmanları bulunmasından dolayı, temel olarak LeNet modeline benzemektedir. Aktivasyon fonksiyonu olarak ReLU (Rectified Linear Unit), ortaklama katmanlarında da maksimum ortaklama kullanılmaktadır. Daha büyük ve daha derin olan bu ağ modeli, paralel çift GPU (Grafik İşlem Birimi-Graphics Processing Unit) üzerinde çalışacak şekilde iki parçalı olarak tasarlanmıştır. Paralel GPU teknolojisini kullanan ilk çalışma olma özelliğini de taşımaktadır. Yaklaşık olarak 60 milyon parametrenin hesaplandığı AlexNet ile ImageNet ILSVRC yarışmasında sınıflandırma doğruluk oranı %74,3'ten %83,6'ya çıkarılmıştır [11].



Şekil 8. AlexNet mimarisi [11]
(AlexNet architecture [11])

VGG-16, basit bir ağ modeli olup daha önceki modellerden en önemli farkı ikili ya da üçlü evrişim katmanlarını ortaklama katmanları izlemektedir, VGG'den önceki modellerde evrişim ve ortaklama katmanları birbirini takip etmektedir. Tüm bağlantı katmanında $7 \times 7 \times 512 = 4096$ nöronlu bir öznelik vektörüne dönüştürülür. İki tüm bağlantı katmanı çıkışında 1000 sınıflı esnek eşikleyici (softmax) başarımları hesaplanır. Yaklaşık 138 milyon parametrenin hesabı yapılmaktadır [23].

Res-Net, kendinden önceki modellerden farklı olarak artık değerlerin (residual value) sonraki katmanlara besleyen blokların (residual block) modele eklenmesiyle oluşmaktadır. ResNet bu özelliği ile diğer klasik modellerden ayrılmaktadır. Doğrusal ve ReLU aktivasyon fonksiyonlarına sahip katmanların arasında iki katmanda bir eklenen bu değer aşağıdaki gibi sistem hesabını değiştirir. Önceden gelen $a^{[l]}$ değeri, $a^{[l+2]}$ hesabına eklenmektedir.



Şekil 9. ResNet blok modülü [24]
(ResNet block modüle [24])

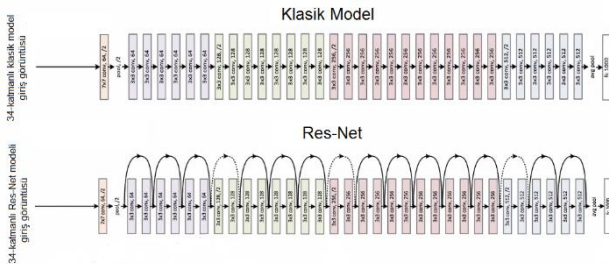
$$z^{[l]} = W^{[l+1]} a^{[l]} + b^{[l+1]} \quad (5)$$

$$a^{[l+1]} = g(z^{[l+1]}) \quad (6)$$

$$z^{[l+2]} = W^{[l+2]} a^{[l+1]} + b^{[l+2]} \quad (7)$$

$$a^{[l+2]} = g(z^{[l+2]}) \quad (8)$$

Teorik olarak, modelde katman sayısı arttıkça başarımın artacağı ön görülmektedir. Ancak pratikte durum böyle değildir. Böylece $w^{[l+2]} = 0$ olduğu durumda yeni teoriye göre $a^{[l+2]} = b^{[l+2]}$ olur. Bu türevin sıfır değeri üretmesi sorununu (vanishing gradient) oraya çıkarmaktadır. İstenmeyen bir durumdur [24]. Ancak artık değer (residual value) beslemesi yeni çıkış eşitliğini iki önceki katmandan gelen $a^{[l]}$ değeri o an ki ağırlık sıfır olsa bile öğrenme hatasını optimize etmektedir ve ağ daha hızlı eğitilmektedir.

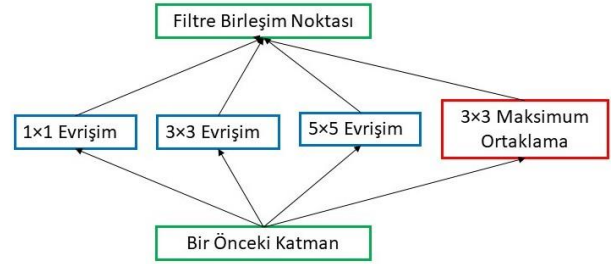


Şekil 10. Klasik derin bir sinir ağı modeli ve ResNet Modeli [24]
(The state of the art network model and ResNet model [24])

GoogLeNet, Lin ve arkadaşları 2014 yılında evrişimli sinir ağındaki doğrusal filtrelerin doğrusal olmayan aktivasyon fonksiyonları ile öğrenme yeteneklerinin artırılacağı fikrini ortaya atmaktadırlar. Tasarlamış oldukları modelde üç tane çok katmanlı algılayıcı yapısıyla evrişimli sinir ağı ve ortalama ortaklama işlemlerinden oluşmaktadır. Model MNIST, CIFAR10, CIFAR100 ve SVHN veri kümelerinde başarılı sonuçlar vermektedir. Veri artırma (data augmentation) ve seyreltme (dropout) uygulandıktan sonra başarımın tüm veri kümeleri için arttığı gözlemlenmiştir [25].

Temel Inception Modülü, ağ içinde ağ yapısıyla birlikte Google çalışanları, 2012 yılında gösterime giren 'Inception' filminden esinlenerek işlendirdikleri Inception modüllerinin birleşiminden oluşan GoogLeNet'i yayımladılar [12]. Inception modüllerindeki genişleme etkisi evrişim katmanlarındaki (1×1) , (3×3) , (5×5)

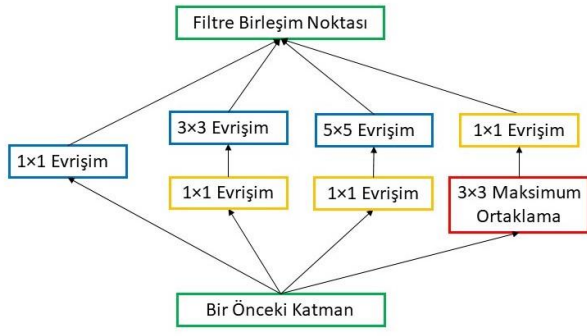
filtrelerin ve (3×3) maksimum ortaklama işleminin paralel olarak gerçekleştirilmesiyle oluşturulmaktadır. Bu yapı "naive Inception module" olarak isimlendirilmektedir. Temel bir Inception modülü Şekil 11'da gösterilen diyagramdaki gibidir ve önemli bir sorun vardır.



Şekil 11. Basit Inception modülü
(Basic inception modüle)

İşlem karmaşıklığı, çıkış boyutunun büyüklüğü ve parametre sayısı paralel işlemler yüzünden çok büyükmektedir. Bu problemin üstesinden gelebilmek için (1×1) evrişim katmanları paralel Naive Inception evrişim katmanlarının öncesine eklenerek boyut azaltma işlemi gerçekleştirilmektedir. Bu Inception modüllerinin ilk versiyonunu oluşturmaktadır. (1×1) evrişim derinliği olmayan x ve y düzlemine sahip bir matriste çarpma etkisine sahip olsa da bu derinlik bilgisi olan bir tensor haline geldiğinde evrişim işleminin parametre hesaplamada önemli derecede etkisi olmaktadır. Örneğin giriş matrisi 100 kanallıysa ve buna 30 kanallı bir (1×1) evrişim filtresi hacimsel olarak uygulandığında çıkış matrisinin kanal sayısı filtre sayısına eşit, yani 30 olur. O zaman (1×1) evrişim katmanı derinlikte boyut azaltmak demektir. Hesaplama karmaşıklığına ve büyüklüğüne bulunan çözüm beraberinde hız ve başarımı getirmektedir. Inception ağ modeli modüllerden oluşmaktadır. Her bir modül, farklı boyutlu evrişim ve maksimum ortaklama işlemlerinden meydana gelmektedir.

Inception modülü içindeki (5×5) evrişim işlem için $(28 \times 28 \times 32) \times (5 \times 5 \times 192) = 120$ milyon parametre hesabı gerekmektedir. Bunun gibi diğer evrişim ve maksimum ortaklama katmanlarını da aynı şekilde hesaplamak gerekmektedir. Bu da beraberinde büyük bir işlem yükü getirmektedir. Szegedy ve ekibi her evrişim katmanından önce (1×1) evrişim katmanı kullanılarak işlem yükünü optimize etmeye odaklanmıştır. Böylece daha karmaşık bir ağ modelin ile daha az hesap ve daha hızlı bir tasarım yapılmaktadır. Bu koşulda; (1×1) evrişim katmanında: $(28 \times 28 \times 16) \times (1 \times 1 \times 192) = 2,4$ milyon parametre ve (5×5) evrişim katmanında: $(28 \times 28 \times 32) \times (5 \times 5 \times 16) = 10$ milyon parametre olmak üzere toplamda 12,4 milyon parametre hesaplanmaktadır. İlk duruma göre yaklaşık 10 kat daha az parametre hesabı son derece çarpıcı bir etkiye sahiptir. Bu (1×1) evrişim işlemini 'darboğaz' (bottleneck) olarak tanımlanmaktadır [12].

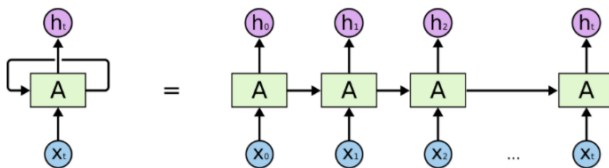


Şekil 12. GoogLeNet modelinin içindeki bir Inception modülü
(One of the inception module in the GoogLeNet model)

Bu işleme öznetelik ortaklama (pooling of feature) denir. Çünkü tensor derinliği yani katmanın kanal sayısı azaltılmaktadır. Aktivasyon fonksiyonu olarak ReLU kullanılmaktadır. Tüm ağ, 9 inception modülünden oluşmakta ve toplamda 100'den fazla katman bulunmaktadır. Tüm bağlantı katmanlarında ortalama ortaklama kullanılmaktadır. Böylece $(7 \times 7 \times 1024)$ adet tensor $(1 \times 1 \times 1024)$ boyutlu bir vektöre dönüştürülmektedir. Parametre maliyeti bakımından büyük avantaj sağlanmaktadır. GoogLeNet 5 milyon parametre ile AlexNet'e göre 12 kat daha az işlem yüküne sahiptir [12,26-27]. R-CNN (Region with Convolutional Neural Networks) yöntemi, nesne tanıma modelinde Inception v1 modelinden faydalanmaktadır [28-29].

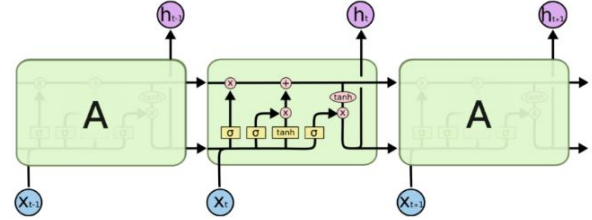
3.2. Özyinelemeli Sinir Ağı ve Uzun Kısa Vadeli Bellek Ağları (Recurrent Neural Networks and Long Short Term Memory Networks)

1980'li yıllarda temelleri atılan bir yapay sinir ağı modeli olan özyinelemeli sinir ağları (Recurrent Neural Network-RNN) özellikle doğal dil işleme (Natural Language Processing-NLP) alanında dizi işleme özelliği olmasından dolayı sıklıkla tercih edilmektedir. Konuşma, metin, müzik, video gibi sıralı akışa sahip verilerde başarılı sonuçlar üreten RNN yapısında klasik sinir ağı yapısından farklı olarak ilerideki katmanlardan girişlere geri besleme yapılmaktadır. Ağ bu sayede önceki bilgilere bağlı bir çıkış üretmektedir. Özyinelemeli ağların da önemli kısıtları vardır. Örneğin bir metin işleme probleminde ilk cümlede yer alan "ben Fransa'da büyüdüm." bilgisinden sayfalarca kelime ve cümleden sonra "ben akıcı bir şekilde ... konuşurum." cümlesindeki boşluğu "Fransızca" diye tamamlaması zor olmaktadır. Çünkü yazı boyunca Almanya'dan, İngilizceden bahseden cümleler varsa modelin uygun bilgiyi seçebilmesi önemli bir problemdir.



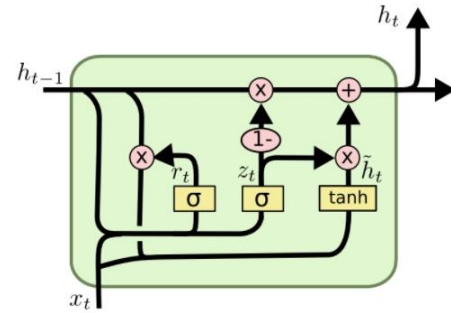
Şekil 13. Temel özyinelemeli sinir ağı modeli [31]
(Basic recurrent neural network model [31])

Buna çözüm olarak, Hochreiter 1997 yılında, uzun kısa vadeli bellek ağları (Long Short Term Memory-LSTM) önermiştir. Giriş katmanı, unutmama katmanı ve çıkış katmanı olmak üzere üç kısımdan oluşmaktadır. Klasik özyinelemeli sinir ağı modelinde unutmama katmanım bulunmamaktadır [30-31].



Şekil 14. LSTM modeli blok diyagramı [31]
(Block diagram of LSTM model [31])

Sabit hata döngüsü, çıkış aktivasyon fonksiyonu ve gözetleme (peephole) bağlantıları ile LSTM son yıllarda çok kullanılan modellerdendir.



Şekil 15. LSTM iç yapısı [31]
(LSTM internal structure [31])

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (9)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(W[\tilde{r}_t * h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (12)$$

Son yıllarda özyinelemeli sinir ağları video görüntülerinde nesne ve insan takibi konularında da kullanılmakta ve hatta çok yeni modellerden üretici çekişmeli ağlar (Generative Adversarial Networks-GANs) ile birlikte kullanımı da literatürde yerini almıştır [13].

3.3. Yazılım ve Donanım (Software and Hardware)

Derin öğrenme modellerinin 2010'lu yıllarda gelişmesinin en önemli sebebi donanımsal kısıtların aşılmış olmasıdır. 90'lı yıllarda bilgisayarlarda yeterli işlemci gücü olmadığından, tasarlanmış olan ağ modellerinden verimli sonuçlar alınamamıştır. Günümüzde grafik işlemci birimleri (GPU) sayesinde özellikle vektör ve matris işlemleri çok daha hızlı yapılabilmektedir. Böylece karmaşık derin öğrenme ağ modellerinin de

uygulanabilirliği sağlanmaktadır. Bununla birlikte dünyaca bilinen büyük servis sağlayıcıları bulut (cloud) servislerini de son kullanıcıya ücretli/ücretsiz şekillerde açmaktadır. Böylece yüksek kapasiteli işlem yapabilen makinelerle uzaktan bağlantı sağlanarak derin öğrenme modelleri tasarlayıp eğitmek mümkün olmaktadır.

Makine öğrenmesi ve derin öğrenme konularında en çok kullanılan 4 programlama dili şöyle sıralanabilir: Python (%76,3), R (%59,2), C/C++ (%19,2), Java (%18,3). Özellikle Python birçok derin öğrenme çatısının ve

kütüphanesinde kullanılmasından dolayı en çok tercih edilen programlama diliyken istatistik alanında en tercih edilen dil R'dır. Python programlama dilinin bir diğer avantajı da hem akademik hem ticari uygulamalar için kullanımının elverişli olmasıdır [32].

Derin öğrenme için çeşitli üniversiteler ve şirketlerce geliştirilmiş farklı özelliklere sahip birçok çatı ve kütüphane bulunmaktadır [33,34]. Bu kütüphaneler ve özellikleri Tablo 1'de özetlenmiştir.

Tablo 1. Derin öğrenme kütüphaneleri ve özellikleri
(Deep learning frameworks and properties)

Kütüphane	Geliştirici	Programlandığı Dil	Özellikleri
TensorFlow [35]	Google	Python	-Hızlı derleme yapabilmektedir. -TensorBoard ile görselleştirme yapabilmektedir. -Veri ve model paralellliği sağlar. -GPU veya CPU'da paralel çalışabilmektedir.
Caffe [36]	Berkeley Vision and Learning Center (BVLC)	Python	-İleri beslemeli ağlar ve görüntü işleme konularında hızlıdır. -Hassas ayarlama (finetuning) için önceden eğitilmiş modelleri vardır. -Hiçbir kod yazmadan model eğitilebilir. -Python arayüzü oldukça kullanışlıdır. GPU desteği vardır.
Caffe2 [36]	Facebook	Python	-Python API ile C++ desteği sağlar. Berkeley yazılım dağıtım lisansı vardır. -GPU desteği vardır.
Torch/ PyTorch [37]	Google/ Facebook	Lua/ Python	-Birçok modül parçayı birleştirmek kolaydır. -Yeni katmanları yazıp GPU üzerinde çalıştırması kolaydır. -Çokça önceden eğitilmiş model vardır.
Keras [38]	Francois Chollet-Google	Python	-Torch kütüphanesinden esinlenilmiş sezgisel bir API'dir. -Theano, TensorFlow, DeepLearning4j ve CNTK arkaplanda kullanılmaktadır. -Hızlı büyüyen bir yapısı vardır. -GPU veya CPU'da paralel çalışabilmektedir.
MxNet [39]	Pedro Domingos, Amazon AWS	R, Python ve Julia	-Çoklu GPU desteği vardır. -Eğitim hızı ve verimliliği yüksektir. -Yeni katmanlar eklemek kolaydır.
CNTK [40]	Microsoft	C++	-Geri planda Python API kullanımına izin verir.
DeepLearning4j [41]	Adam Gibson	Java, Scala	-Java sanal makinesi (Java Virtual Environment-JVM) tabanlı olmasıdır.
KNet [42]	Deniz Yüret, Koç Üniversitesi	Julia	-Kolay anlaşılır olması ve ifade gücü yüksektir. GPU Desteği vardır.
Theano [43]	Montreal Institute for Learning Algorithms (MILA) Lab.	Python	-Eylül 2017'de resmi olarak Theano kütüphanesinin geliştirilmeye devam edilmeyeceği duyuruldu.

4. KALABALIK ANALİZİ (CROWD ANALYSIS)

Video tabanlı görüntü işleme ve analiz yöntemleri son yıllarda önem kazanmaya başlayan geniş bir konudur. Günümüz dünyasında açık ve kapalı alan gözetmeksizin güvenlik en önemli olgulardan biri haline almıştır. Bu yüzden dijital kamera ve görüntüleme sistemlerinin kullanım yaygınlığı artmakta ve değişen gereksinimler nedeniyle çözülmesi gereken yeni problemler tanımlanmaktadır. Örneğin, izlenen görüntüdeki anormallikleri, güvenliğe, ekonomiye etki edecek etkinlikleri ve hayati önem taşıyan insan aktivitelerini saptamak öncelikli çalışma konuları arasındadır [44].

Akıllı kamera sistemleri ve kullanılan yöntemler görüntüdeki insan davranışlarının analiz edilmesi üzerinde yoğunlaşmaktadır. Bunun başlıca sebebi etken ve edilgenin çoğunlukla insan oluşudur. Tüm bu görüntüleme sistemleri, dönüşümler ve analizlerin yanında insan davranışlarının temel olarak araştırılmasını söz konusu haline gelmektedir. İnsan davranışı üzerine çok sayıda bilimsel çalışma yapılmakta ve bitip tükenmez bir araştırma konusudur.

Bilgisayarlı görüntüleme tekniklerinin gelişmesiyle birlikte insan davranışlarının anlaşılması ve/veya atletik performansların ölçülmesi problemlerinin üstesinden gelmek için otonom ya da yarı otonom teknikler kullanılmaya başlanmıştır. Bu gelişmelere rağmen insan ve

kalabalık görüntülerin analizine ilişkin karmaşık zorluklarla karşılaşmaktadır [44-45].

Tanıma işlemlerine göre, görüntüleme ve insan davranışlarının analiz edilmesi gerektiğinde spesifik kişilerin saptanabilmesi ya da bir kalabalığın grup olarak değerlendirilmesi önemli bir bilimsel araştırma konusu olmanın yanında yüksek seviyeli analiz teknikleri gerektirmektedir. Bu konuyla ilgili temel olarak uygulamalar aşağıdaki gibi sıralanabilir:

Kalabalık Yönetimi: Kalabalık analizi, kalabalığın yönetilmesi örneğin; popüler etkinlikler, spor müsabakaları, konser organizasyonları ve halk gösterilerinde oluşabilecek felaketleri önlemek ve güvenliği sağlamak için kullanılabilir.

Kamusal Alan Tasarımı: Kalabalık analizi, alışveriş merkezleri, şehir, çevre ve park düzenlemeleri, çalışma ve ofis alanları gibi kamusal hizmet vermesi planlanan bölgelerin tasarlanmasında kullanılabilir. Bu konu içeriğinde kalabalık simülasyonu konusu da belli ölçüde kapsamaktadır.

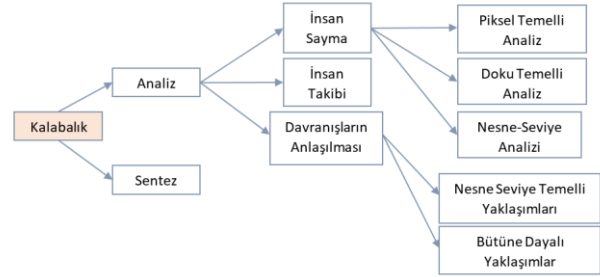
Sanal Çevre Tasarımı: Kalabalık analizi, kalabalığın matematiksel modeli oluşturularak simülasyonlar üzerinden tasarımların performansını artırmak amaçlı kullanılabilir. Örneğin; bir bina içi veya dışı alandaki boğaz noktalarında kalabalığın yoğunlaşması, acil çıkış noktalarının felaket anındaki kapasitesinin hesaplanabilmesi için kullanılabilir.

Görsel Gözetim: Kalabalık analizi, anormallikler ve alarm durumlarının otomatik olarak saptanması amacıyla kullanılabilir. Ayrıca polislerin şüphelileri kalabalık içinden seçip takip edebilmesi amaçlı da kullanılması hedeflenen çalışmalar arasındadır [46].

Buna örnek olarak 2013 yılında Amerika Birleşik Devletleri'nde gerçekleşen Boston Maratonu'ndaki bombalı saldırıda saldırganın tespit edilmesi için polislerin 400'ü aşkın kez aynı görüntüleri izlemesi gerekmiştir. Ancak bölgedeki kameralardan alınan veriler üzerine çalışan Hwang ve arkadaşları suçlu kişinin tespitinde hem başarılı hem de oldukça hızlı bir sistemin başarısını test etmişlerdir [47].

Zhan ve arkadaşlarının çalışmalarında öne sürdüğü görüşlerden biri insan ve/veya nesne görüntülerinde başarılı olarak takip, tanıma yoğunluk bilgisi veren algoritmalar insan yoğunluğunun arttığı kalabalık görüntülerinde elverişli sonuçlar vermemektedirler. Kalabalık, bir bireyin tanımının ötesinde yeni bir tanımlamaya ihtiyaç duymaktadır. Kalabalık, her bir bireyin beklenen, farklı ve karmaşık davranışlarını kapsayan ortaklaşa yeni bir tanım getirmektedir. Örneğin, kişinin değil kalabalığın grup olarak davranışının agresif veya huzurlu olup olmadığının kararının verilmesi esastır [48].

Genel olarak kalabalık analizi, otomatik insan davranışı saptama ve tanımlama sistemlerinin önemli ve temel aşamasıdır.



Şekil 16. Kalabalık analizi/sentezi [44]
(Crowd analysis/syntesis [44])

Kalabalık video görüntülerinin otomatik analizi için üç temel ve önemli probleme göre sınıflandırmak mümkündür. Bu temel sınıflar; piksel seviyeli analizler, doku seviyeli analizler ve nesne seviyeli analizlerdir. Kalabalık analizinde insanların sayısının/yoğunluğunun ölçülmesi ve/veya kestirilmesi önemli problemlerden biridir. Örneğin kamu konfor seviyesi ya da potansiyel tehlike durumları için kalabalık yoğunluğunun kestirimi ve analizi önemlilik arz etmektedir. Bu alanda yapılan çalışmalarda birçok farklı yaklaşım sunulmaktadır. Bu yaklaşımla üç temel başlık altında toplanabilir [49].

- Kalabalık Yoğunluğu Kestirimi
- Kalabalık İçinden Alt Grupların veya Kişilerin Takibi
- Kalabalık Video Görüntülerinde Anormal Durumların Kestirilmesi/Saptanması

Kronolojik olarak literatür incelendiğinde bu alanda yapılan çalışmalarda birbirinden farklı yaklaşımlarla çözüm önerileri sunulduğu ancak hala konuyla ilgili karmaşık problemlerin çözülmediği görülmektedir.

3.1. Kalabalık Yoğunluğu Kestirim, Takibi, Davranışların Tanımlanması (Density Estimation, Tracking and Behavior Recognition of Crowd)

Kalabalık video görüntülerinde yoğunluğun saptanabilmesi için görüntülerde gerekirse boyut azaltma ve ön işlemlerin gerçekleştirilmesi, özniteliklerin belirlenip uygun bulunanların çıkarılması ve sonuçta karar vermesi için istatistiksel ve/veya akıllı modellere ihtiyaç vardır.

2008'de, Ali ve Shah çalışmalarında yüksek yoğunluklu kalabalık görüntülerinde akıştaki çıkış ve giriş noktaları gibi statik alanları, hareketli bireylerin bulunduğu dinamik alanları ve duvar gibi sınır alanlarını belirleyerek problemin üstesinden gelmeye çalışmıştır. Ortalama kaydırma yöntemi ile yaya takip işlemlerini gerçekleştirmişlerdir [50].

2010'da Mao ve arkadaşları görüntü çerçeveleri arasında fark olarak hareketli bölgeleri tespit etmişlerdir. Sekiz adet

alt çerçeve üzerinden öznitelik çıkarma işlemi gerçekleştirmişlerdir. Harris köşe detektörü ile görüntü üzerindeki köşe bölgelerini belirlemişlerdir. Böylece KLT (Kanade-Lucas-Tomasi) öznitelik noktaları üzerinden kalabalığın yoğunluğunu 3 farklı veri kümesi üzerinden analiz etmeyi başarmışlardır [51-52].

Guo ve arkadaşları 2010 yılında yaptıkları çalışmada insan güvenliğini sağlamak amacıyla kalabalık analizini rastgele Markov süreçler yönteminden faydalanarak gerçekleştirmişlerdir. Bunun için optik akış (Optical Flow), önalanın ayrılması ve kenar detektörü işlemlerinin sonrasında sınıflandırma yapmışlardır. Tasarladıkları sistem gerçek zamanlıdır [53].

Li ve çalışma arkadaşları 2010'da yapmış oldukları çalışmalarda görüntülerden optik akış yöntemi kullanarak hareketli bölgeleri tespit etmiş ve arka planı silerek işe başlamışlardır. Elde ettikleri ön plan görüntüsünde doku analizi için öznitelikleri çıkarmışlardır. Kendi kendine organize olabilen haritalama yöntemi ile farklı kalabalık gruplarını tespit etmişlerdir. Bunu 500 görüntülük bir veri kümesi için gerçekleştirmişlerdir [54-55].

2010 yılında Ge ve Collins, kalabalık içinden insan sayısının kestirimi üzerine çalışmışlardır. Bunun için direkt ve dolaylı olarak iki yöntem öne sürmüşlerdir. Literatürde kalabalık içindeki insanları işaretleyip sayma yöntemi direkt yöntem olarak adlandırılmıştır. Dolaylı yöntemde ise çerçeveler arasındaki ilişkiden faydalanarak kalabalık tespiti ve belirtilen boyuta düşebilecek insan sayısını görüntünün yoğunluk haritasına bağlı olarak kestirilmesi sağlanmıştır. Çalışmada da direkt olmayan yöntem ele alınmıştır. SURF (Speeded Up Robust Features) noktaları belirlenmiş ve kümeleme için K-ortalama, Haar Dalgacık dönüşümü, Gaussian işleme yöntemleri kullanılmıştır [56].

Kim ve arkadaşları 2010'da otomatik bir kalabalık yoğunluğu hesaplama sistemi tasarlamışlardır. Kalabalığın grup hareketini belirlemek için optik akış algoritmasını kullanarak, kaç kişilik bir grubun hareketi olduğunu da kenar piksel sayılarına bağlı olarak hesaplamayı başarmışlardır. Farklı yoğunluklara sahip 5 tür seviyeyi kalabalık bakımından sınıflandırmayı başarmışlardır [57].

Hsu ve arkadaşları 2011'de yapmış oldukları çalışmada ayırık kosinüs dönüşümü ve frekans dağılımı bilgilerinden faydalanarak farklı frekans bantlarındaki öznitelikleri belirleyip kalabalığı beş farklı yoğunluk seviyesine ayırmışlardır. Bu sınıflandırmayı yaparken destek vektör makineleri (Support Vector Machines-SVM) yöntemine başvurmuşlardır [58].

Xiong ve çalışma arkadaşları 2011 yılında iki farklı tür anormal davranışı kalabalık görüntülerinden tespit etmek üzerine çalışmışlardır. Bu iki sınıfı gruplaşma ve koşma olarak tanımlamışlardır. Bunun için görüntünün potansiyel ve kinetik enerji bilgilerinden faydalanmışlardır. Uyarlamalı Gauss bileşim modeli ile elde edilen ön plan

görüntüsünden potansiyel enerji, histogram ve optik akışı hesaplanmıştır. Optik akıştan elde edilen bilgi içinden kinetik enerjileri ile kalabalığın yoğunluğu kestirilmektedir [59].

2011'de Yu ve diğerleri yoğun ve çok yönlü kalabalık akışını dinamik doku, optik akış ve ROI öznitelikleri kullanarak gerçek zamanlı GRNN (Gaussian regresyon sinir ağları) ile analiz etmeyi başarmışlardır [60].

Yang ve Zhoa 2012 yılında arka plandan Lucas-Kanade optik akış yöntemi ve Gauss yöntemi ile kullandıkları görüntüler üzerinde doku analizi ile öznitelikleri çıkarmışlardır. Bunun üzerine farklı kalabalık gruplarını kendi kendine organize olabilen haritalama (Self-Organizing Map) yöntemi ile sınıflandırmışlardır [61].

2012'de Subburaman ve diğerlerinin yapmış olduğu çalışmada kalabalıktaki kişi sayısını ve kalabalık yoğunluğunu işaretledikleri ilgili noktalara göre hesaplamışlardır. Bunun için insanların kafa bölgelerini ilgili nokta olarak belirleyip bunları da gradyan oryantasyonu ile bulmuş ve yerel ikili örüntüler yöntemini kullanmışlardır [62].

Yine 2012 yılında Chan ve Vasconcelos'un çalışmaları da dinamik ve doku hareket modellerinin kombinasyonu üzerine inşa edilmiştir. Bu şekilde öznitelikleri çıkarmışlar ve Bayes, Gauss ve Poisson regresyon yöntemlerini kullanarak kalabalık analizi gerçekleştirmişlerdir. İstatistiksel yaklaşımlarla başarılı sonuçlar elde etmişlerdir [63-65].

Wang ve çalışma arkadaşları 2012'de diğer çalışmalardan farklı olarak görüntünün genelindeki yoğunluğunun kestirilmesi değil yerel bölgelerde değişen yoğunlukların kestirilmesi üzerine çalışmışlardır. Bunun için doku tanımlama işlemlerinde yerel ikili örüntü modeli (Local Binary Pattern-LBP) ve gri düzeyli eş oluşum matrisi (co-occurrence matrix) birlikte tanımlanmıştır. LBPCM (Local Binary Pattern Co-occurrence Matrix) istatistiksel özellikleri ve uzamsal bilgileri kullanılarak, pencere kaydırma tekniği ile doku öznitelikleri tanımlanmaktadır. Öznitelikler destek vektör makineleri yöntemi ile eğitilip test edilmiş ve olası kalabalık bölgelerin kestirimi gerçekleştirilmiştir [66].

Fradı ve Dugelay'ın 2012 ve 2013 yıllarındaki çalışmalarında çok kalabalık ve perspektifi zorlayıcı görüntüler üzerinde analiz yapmışlardır. Öncelikle hareketli noktaları çerçeve farkları yöntemiyle belirlemişlerdir. Ardından perspektif normalizasyonu ve yoğunluk normalizasyonu gerçekleştirmişlerdir. Bunun üzerine Gauss regresyon modeli ile işaretlenen ilgili noktaların sayımı ile kestirilen kişi sayısına ulaşmışlardır [67-68].

2013 yılında Tehranipour ve arkadaşları istatistiksel ve bulanık mantık karar modelleri ile kalabalık yoğunluğunun kestirimi üzerine çalışmışlardır. Kalabalığın yoğunlaştığı

bölgeleri riskli bölgeler olarak belirlemişlerdir. Bu bölgeleri dalgacık dönüşümü yöntemi ve modifiye piksel tabanlı işleme yöntemleri ile çıkarmışlardır. Destek vektör makineleri yöntemi kullanarak modelledikleri sistemi eğitmiş ve olası risk bölgelerini bulanık karar verme modeli ile kestirmişlerdir [69].

2013'te Fradi diğer çalışma arkadaşları lokal ikili örüntülerinin alt uzay öğrenme yöntemiyle kombine ederek bir kalabalık analiz modeli oluşturmuşlardır. Alt uzaylara doğrusal diskriminant analizi ve temel bileşenler analizi ile hesaplamışlardır. Daha sonra elde ettikleri öznelikler K-en yakın komşu (K-Nearest Neighbourhood) yöntemi ve destek vektör modeli ile sınıflandırıp bunları karşılaştırmışlardır. Destek vektör makineleri yöntemi ile başarımları daha yüksek sonuçlar elde etmişlerdir [70].

Rao ve arkadaşları 2013'te optik akış yaklaşımını görüntü üzerinde bloklar ayrıştırarak uygulamışlardır. Hiyerarşik bir kümeleme yaptıktan sonra Öklid uzaklığı ile blokların ve görüntülerin birbirleriyle ilişkilerini incelemişlerdir. Buna göre elde ettikleri bölgeleri sayarak yoğunluğun kestirimini yapmışlardır [71].

Yuan ve çalışma arkadaşları 2013 yılında kalabalık analizine yeni bir bakış açısı kazandırmışlardır. "Günümüz dünyasında herkesin mobil telefon sahibi olduğu bilgisinden yola çıkarak yoğunluğu kestirilmek istenen bölgedeki internet ağı ve radyo frekansı yoğunluğuna bakarak kişilerin sayısı hakkında bilgi sahibi olmak mümkündür" savını ortaya atmışlardır. Diğer klasik yöntemlerle kıyaslanabilecek sonuçlar elde etmişlerdir [72].

2013'te Karpagavalli ve Ramprasad çalışmalarında kalabalık çevrenin insan güvenliği için yönetilmesi gereken bir unsur olduğuna dikkat çekmişlerdir. Daha özel bir çalışma yapıp kalabalığın genel yoğunluğunu değil kalabalığı oluşturan kişi sayısını kestirmeye yönelik çalışmışlardır. Geliştirilmiş uyarlamalı K- Gauss model ile görüntülerden arka plan çıkarılmıştır. Kalabalığın boyutunu anlamak için sınır bölgelerin alanı belirlenmiştir [73].

Wu ve arkadaşları 2013 yılındaki çalışmalarında Lucas-Kanade optik akış ile kalabalığın hareketli olduğu bölge tespit edilmiştir. Destek vektör makineleri modeli ile yoğunluklu bölgeler sınıflandırılmıştır [74].

2013 yılında Ping ve çalışma arkadaşları görüntüyü 8 bloka bölünmüş ve üzerinden istatistiksel yaklaşımlarla incelemeler yapmışlardır. En küçük kareler regresyonu ve gri seviye matris bileşimi ile doku analizi gerçekleştirmişlerdir. Destek vektör makineleri modeli ile kalabalık yoğunluğunun kestirimi için çözüm önermişlerdir [75].

2014 yılında Yuan mobil telefon kullanımı sayesinde kalabalık yoğunluğunun analizi üzerinde çalışmıştır.

Kalabalığın içindeki gruplaşmaları da tespit etmek adına hem kablosuz internet sinyallerinden hem de bluetooth sinyallerinden faydalanmıştır. Bluetooth kullanarak haberleşme yapan cihazları kendi içlerinde gruplar, kablosuz ağ kullananları da tek başlarına gruplardan uzakta olan bireyler olarak sınıflandırır. Böylece kalabalığın yoğunlaştığı ve seyrekleştiği bölgeleri, alt grupları belirlemekte herkesi kullandığı mobil telefondan faydalanmıştır [76].

3.2. Kalabalık İçinden Alt Grupların veya Kişilerin Takibi (Tracking Pedestrians or Group of People from Crowd)

Bir video görüntüsünden kalabalık içinden bir kişinin veya grubun takip edilmesi literatürde bu alandaki bir başka konudur. Takip işlemi için video görüntüsündeki çerçeveler arasındaki ilişkinin önemi artmaktadır. Sürekli hareket halinde olan bir kalabalık söz konusuysen işler daha karmaşık olmaktadır.

2008 yılında Khansari ve arkadaşları kalabalık video görüntülerindeki bir nesnenin takibini dalgacık öznelikler ve doku analizi ile gerçekleştirmişlerdir. Dalgacık öznelikler (wavelet features) ile görüntü akışındaki lokasyonu, doku analizi sayesinde de nesnenin hareket yönü ve hızını belirleyerek referans görüntü belirlendikten sonra kalabalık görüntüsünden takibi işlemini başarmışlardır [77].

Rodriguez ve çalışma arkadaşlarının 2011 yılında oluşturdukları sistem, görüntüdeki insanların kafalarını bulması ve bunların sayılmasını esas almaktadır. Ancak öncelikle görüntü üzerindeki enerji yoğunluğu ile insanların olduğu bölgeleri bulup bunun üzerinden kafa bölgelerinin bulunması işlemleri yapılmaktadır. Böylece video görüntüsünün getirdiği işlem yükü probleminde biraz kurtulmuşlardır [78-79].

Ayrıca kalabalık video görüntülerindeki kişilerin, alt grupların ya da nesnelere takip edilmesinde Kalman filtreler gibi uyarlamalı filtre modelleri de başarılı olmaktadır. Kalabalık sayımında geliştirilen yeni çözümler üç gruba ayrılmıştır; izleme tabanlı, algılama tabanlı ve regresyon tabanlı yöntemlerdir. Algılama tabanlı yöntemler, kafa, yüz ve diğer vücut kısımlarını içerebilen sahnelerden ilgi bölgesini algılamayı öğrenir.

Subburaman ve arkadaşları kafa detektörü olay yerindeki insanın sayısını saymak için kullanmışlardır. Algılama yöntemleri, ilgi alanındaki renk ve eğim özelliklerinin yönlendirilmesi ile sınıflandırıcı eğitilmektedir. Regresyon tabanlı model bir resimdeki kişilerin özelliklerinden doğrusal veya doğrusal olmayan bir işlev olduğunu öğrenir ve öğrenilen modeli uygulayarak insan sayısını hesaplar [62].

Coneto ve ark. modeli ilgi noktalarına ve kümedeki kişi sayısına göre regresyon temelli yöntemi uygular. Bu yöntemlerin genelliği, farklı kamera ayarlarıyla farklı eğitime veri kümesi gereksinimi nedeniyle sınırlı olmasıdır.

İzleme tabanlı yöntemler, benzer yörüngeden bölümleri kümelemeye çalışmaktadır [80].

Antonini ve ark. video diziliminde kişi sayımı yapan bir sistem önermişlerdir. Benzer yörüngeler kombine edilmiş ve insan vücuduna ait olanlar yayaları saymak için kullanılmıştır [81].

Kişi sayma, daha önce de belirtildiği gibi bilimsel birçok bir araştırmanın ilgi odağı olmuştur. Bunlardan birkaçını saymak gerekirse; güvenlik kamera sisteminde yoğun taleplerin olması, özellikle trafik kontrolü ve metropol güvenliği gibi kalabalık yönetimidir. Kalabalık analizi, bilgisayarlı görü alanında en zorlu görev olma ününü kazanmıştır. Güvenlik ve gözetim talepleri nedeniyle dikkat çeken bir çalışma alanıdır. Kalabalık analizi, kalabalık yoğunluk tahmini, kişi sayma, kalabalık davranış analizi, kalabalık izleme, kalabalık bölütleme, vb. alt konulardan oluşur [82-98].

Kalabalık analizi için çoğu sistem, sahneleri açıklamak için genelleme konusunda zayıftır çünkü bunlar sahneye özgü çözümlerden oluşmaktadır; yani kalabalık ayrımı yalnızca aynı sahnelerden öğrenildiğinde uygulanabilir. Son çalışmalar, yüzlerin, nesnelerin, eylemlerin ve sahne görüntülerinin nitelik temelli gösterimlerini içermektedir [99-109].

3.3. Kalabalık Video Görüntülerinde Anormal Durumların Kestirilmesi/Saptanması (Estimation of Abnormality from Crowded Video Scenes)

MKalabalık video görüntülerinde popüler olan konulardan biri de anormal durumların kestirilmesi, kişilerin ve/veya grupların saptanmasıdır. Bunun için yoğunluk kestirimi yöntemlerine ek olarak literatürde aşağıdaki yöntemler üzerine çalışmalar yapılmıştır.

2009'da Kratz ve Nishino çalışmalarında uzamsal-zamansal hareket örüntü modeli ile zamanla olan mekân değişimlerini görüntüler üzerinde yakalamayı saklı Markov modellerinden (Hidden Markov Model) faydalanarak başarmışlardır. Böylece belirlemiş oldukları çok hızlı ve ani davranışları kestirmişlerdir [110].

Jiang ve çalışma arkadaşları 2009 yılında modellerini eğitici olarak tasarlamış ve uzamsal değişimlerden elde ettikleri öznitelikler ile birlikte doku analizi de yaparak görüntü içindeki anormallikleri kestirmeye çalışmışlardır [109].

2010 yılında Mahadevan ve arkadaşları dinamik dokuların birleşimi modeli oluşturmuşlardır. İlk adımlarında görüntüdeki dinamik dokuların değişen bölgelerini Gauss bileşim modeli kullanarak bulmuşlardır. Sonra görüntüde uzamsal analiz yaparak anormal değişikliklerin kestirimini gerçekleştirmişlerdir [111].

Reddy ve arkadaşları 2011'de kalabalık video görüntüsünden yola çıkarak gelişmiş bir anormal bulan

model tasarlamışlardır. Bu modele ön plan görüntüsünün hızını, büyüklüğünü ve dokusunu hücre temelli analiz etme adını vermişlerdir. Dinamik doku modellerinin bileşimi ile başarılı sonuçlar elde etmişlerdir [112].

2012'de Saligrama ve Chen, Neyman-Pearson karakterizasyonu yaklaşımını temel alarak yerel öznitelikleri her çerçeve için tanımlamışlardır. Uzamsal değişimler filtresi uygulamış ve yerel K-en yakın komşuluklarını (K-Nearest Neighbors) hesaplamışlardır. 5 farklı video veri kümesi üzerinde sistemlerini test etmiş ve başarılı sonuçlar elde etmişlerdir [113].

Cao ve Huang 2013 yılında alışılmış olmayan hareketler yapan kişi ve grupları kalabalığın içinden ayırtmak üzerine bir çalışma yapmışlardır. Gauss birleşim modeli (Gauss mixture model), diğer çalışmalardan farklı olarak birikmiş mozaik görüntü farkları ve perspektif bozukluklarını düzeltme işlemleri sonrasında ani hareketleri görüntüler üzerinde tespit etmeyi başarmışlardır [114].

Ayrıca kalabalık grupların bulunduğu video görüntülerinin analizi ve takibinde insan gücünün kullanılması başarımlar için hız bakımından yetersiz kalmaktadır. Bu yüzden her geçen gün artan bu büyük verilerin başarılı analizi için çalışmalar sürdürülmektedir. Bilgisayarlı görü konusunda kalabalık analizi gibi karmaşık bir probleme bahsi geçen geleneksel yöntemlerin yanında derin öğrenme (deep learning) modelleri ile başarılı çözümler üretilmeye 2012 yılından bu yana devam etmektedir.

5. DERİN ÖĞRENME MİMARİLERİ İLE KALABALIK ANALİZİ (CROWD ANALYSIS WITH ARCHITECTURE OF DEEP LEARNING)

Görüntüler üzerinden kalabalık analizi ve kişi sayısı tespiti için derin öğrenme alanındaki son çalışmalar ve uygulamalar araştırmacılara standart yöntemlerle karşılaşılan birçok zorluğun aşılmasına destek olmaktadır. Özellikle probleme uygun özniteliklerin seçilmesi önemli miktarda deney ve tecrübe gerektirmekteydi. Bilgisayarlı görü sorunları için insan görme sistemini modellemesinden dolayı evrişimli sinir ağları ve evrişimli sinir ağlarıyla birlikte özyinelemeli sinir ağları kullanılmaktadır. Bu derin öğrenme modelleri diğer tekniklerden daha iyi performans göstermektedir. Bu yapının daha başarılı sonuçlar vermesi hem öznitelikleri seçmek için bir işlem yapmak gerekmemesi hem de daha derin katmanlarda görüntülere ait üst düzey özelliklerin öğrenilmesinden kaynaklanmaktadır. Böylece araştırmacılar kalabalık sahnelerin analizi için farklı evrişim ve ortaklama katmanı konfigürasyonları ile modeller tasarlayarak derin öğrenmeden faydalanmaktadır.

Bu bölüm kalabalık sayımında derin öğrenme, evrişimli sinir ağı, HOG (histogram orientated gradient) dönüşümü, SIFT (scale-invariant feature transform) karakteristik özellikleri kullanan ve diğer bazı tekniklerdeki son gelişmeleri kapsamaktadır.

Derin evrişimli sinir ağı, öznelik haritalarını verilen girdiden çıkaran evrişim ve ortaklama katmanlarından oluşan özel yapısı nedeniyle, video ve görüntü işleme sistemlerinde yaygın olarak kullanılmaya başlanmıştır [115-116]. Sınıflandırma ve regresyon görevleri için el yapımı (hand-engineering) öznelik çıkarımı yöntemlerini kapsamamaktadır. Evrişimli sinir ağının bir diğer avantajı ise, tüm bağlantılı derin sinir ağıyla karşılaştırıldığında parametrelerin sayısının önemli ölçüde azalmasıdır. Çünkü tüm bağlı katmanlar ile ağırlıkları paylaşılan örneğin (3×3), (5×5) boyutlu filtreleri olan evrişim katmanları ile değiştirilir. Sınıflandırma ve regresyon amacıyla test edilen evrişimli sinir ağının çeşitli yapıları vardır. Çoğu literatür, evrişim katmanlarının yığınlarını tanımlar, ardından ortaklama katmanı ve yapının sonunda tüm bağlı katmanı takip eder. Fakat görüntü ve video sınıflandırma problemleri için izlenen tek bir standart yapı bulunmamaktadır. Bu yüzden, sınıflandırma problemleri için en iyi evrişimli sinir ağı yapısı türetmek bile açık bir araştırma problemidir.

Video işleme üç ana aşamadan oluşur: İlk aşamasında yoğun veya seyrek ilgi noktaları kümesini tanımlayan yerel görsel özelliklerdir. İkinci aşamasında, özellikleri sabit boyutlu video düzeyi tanımına birleştirmek ve son olarak, farklı sınıfları ayıran ayıklanan özellikler üzerine eğitilmiş bir sınıflandırıcı oluşturulmasıdır.

2015 yılında Zhang ve çalışma arkadaşları derin evrişimli sinir ağı modelini kullanarak çapraz sahne kalabalığı saymayı önermiştir. Eğitim verisi, kalabalık yoğunluğu ve kalabalık sayımı olmak üzere iki alternatif hedef içermektedir. Bu öneri dört tip problemi ele alıyor [117]:

- Kalabalığı tanımlayan etkili özelliklerin elde edilmesi ve kalabalık saymada daha fazla dayanıklılık yardımı. Bu problemin üstesinden gelmek için evrişimli sinir ağı, iki öğrenme hedefi, kalabalık sayısı ve kalabalık yoğunluk haritası ile kullanılmıştır.
- Farklı aydınlatma, kalabalık yoğunluğu ve dağılımlara bağlı olarak sahne bozulma (distortion) sorunlarını çözmek amacıyla kullanılmıştır. Bu, farklı sahnelerin ek eğitim verilerini gerektirir. Önceden eğitilmiş evrişimli sinir ağı modeli, bu sorunun üstesinden gelebilmek için hedef verilerle ince ayarlanmıştır.
- Kalabalık bölütleme, kalabalık analizi ve saymada çok zor ve vazgeçilmez bir görevdir. Derin öğrenme modelleri, kalabalığın hareket halindeyken kalabalık dokusunu yakalar.
- Kalabalık sayımı için daha az sayıda eğitim örneği mevcuttur. En geniş eğitim verisi UCF_CC_50 seti, farklı kalabalık sahnelerdeki 50 titreşim görüntüsünden oluşur. Bu nedenle, yazarlar, sahnede kalabalık sayım için yeni veri kümesi sunmuşlardır [118].

Evrişimli sinir ağının yapısı, birinci ve ikinci evrişim katmanlarından ve üç tamamen bağlı tabakadan sonra birleştirme katmanı olmak üzere üç evrişim katmanından oluşur. Kayıp katman, yoğunluk haritası ve genel sayı olmak üzere iki çeşit sonuçtan oluşur. Yöntemler, yaya UCF_CC_50 veri kümesi WorldExpo10 kalabalık veri kümesi ve UCF_CC_50 veri kümesi olmak üzere üç veri kümesi üzerinde değerlendirilmiştir [118].

Wang ve arkadaşları, eğitim örnekleri, perspektif, karmaşık sahneler ve ciddi yoğunluk ve aşırı yoğun kalabalıkta insanların sayılması sorunu üzerinde durmuşlardır. Mevcut geleneksel özellikleri kalabalığın yüksek yoğunluğunda başarısız olan SIFT, HOG vb. yöntemlerden elde etmektedirler. Yazarlar, çok yoğun kalabalıktaki insanları saymak için evrişimli sinir ağı modelini önermişlerdir. Sistemin yapısı evrişimli katmanlardan oluşur; bunlardan bazıları ortaklama katmanıdır ve sonunda iki tüm bağlantı katmanı takip eder [119]. Veri kümesi, UCF_CC_50 halka açık insanlardan ve Google ve Flickr gibi kamuya açık olan web sitelerinden toplanmıştır [118]. Evrişimli sinir ağı modeli, yoğun kalabalığa sahip pozitif örnekler ve sistemi daha dayanıklı hale getirmek için sıfır sayımlı negatif örnekler üzerine eğitilmiştir. Olumsuz örnekler, sıfır sayımlı binalar ve diğer bölgelerden oluşur. Bu tür örnekler, modelin kalabalık sayımı için daha sağlam ve güçlü özellikler elde etmesine olanak sağlar. Negatif örnekler de dahil edildiğinde, sonuçları, kalabalık sayımının doğruluğunun mutlak fark hesabında önemli ölçüde artmıştır.

Li ve çalışma arkadaşları 2011'de öncelikle ilgi bölgesi kalabalık sayım amacı için seçilmiştir. Kişi sayısını doğru bir şekilde hesaplamak için kendi kendine öğrenme stratejisi uygulanmıştır. Bu sistemde, bir yayayı tanımlamak için, özelliklerin çevredeki bölgelerden elde edildiği bir özellikler yaklaşımı kullanılmıştır. Özellikler, K ortalamalı kümeleme ile üretilen kod dizini kullanılarak oluşturulmuştur. Yöntem ayrıca arka planda yaya tespit ederken ön planda yaya olmayı kaldırır. Yöntemler, evrişimli kayan çerçeve yönteminden daha iyi sonuçlar ortaya koymaktadır. Sistem, görünmeyen bir sahnede bir kalabalık tahmin etmek zorunda kalırsa, bu yöntemler iyi performans göstermez [120]. Derin evrişimli sinir ağının bu eksikliği giderme potansiyeli vardır. [119]'de iki öğrenme hedefi ile derin bir evrişimli sinir ağı önerilmektedir. İlk kalabalık yoğunluğu diğeri ise kalabalıktaki insan sayısıdır.

Shao ve arkadaşları 2015 yılında derin öğrenme mimarisi ile kalabalık sahnelerinin analizi üzerine çalışmalarını sunmuşlardır. Yapısı, görünüş ve hareket özelliklerini öğrenmeye ve birleştirmeye çalışan bir sistem geliştirmişlerdir. Yazarlar, mimarilerinde görünüm ve hareketler için iki kanal oluşturmuştur. Hareket özelliklerini öğrenmek için video çerçevelerini doğrudan derin mimariye giriş olarak vermek yerine, hareket kanalı modele girdi olarak iletilmektedir. Derin mimari, iki dal oluşturmak suretiyle iki kanallı girişi kapsar ve her dal, evrişim, ortaklama ve normalizasyon katmanlarından oluşan üç evrişim katmanı, bir ortaklama ve bir tüm

bağlantı katmanı olmak üzere iki parçadan oluşur. Son olarak, benzer yapıya sahip iki kol tam olarak bağlanmış bir katmanla birleştirilmiştir. Bu çalışma, derin mimarinin, nitelik tabanlı öğrenme ile yalnızca tek bir davranış bulmayı amaçlayan tek görevli yapıdan daha iyi performans göstermektedir. Yazarlar, sistemin performansını artırmak için mevcut veri setlerine, yeni video verilerini de eklemiştir. Toplamda 8.257 kalabalık sahneden 10.000 videodan oluşan veri kümesi ile çalışmışlardır [97].

Boominathan ve çalışma arkadaşları 2016'da kalabalık sayımı için yine evrişimli bir sinir ağı modeli olan CrownNet sistemini sunmuştur. Yazarlar kalabalık görüntünün yoğunluk haritasını tahmin etmek için hem derin hem de sığ (shallow) evrişimli sinir ağından yararlanmışlardır. Hem beden, yüz gibi semantik bilgi hem de düşük seviye özellikleri etkili bir şekilde yakalamıştır. Eğitim örneklerinin eksikliği nedeniyle, evrişimli sinir ağı ölçeğin değişmez temsillerini öğrenmesi için veri artırma (data augmentation) kullanmıştır. İki çeşit artırma gerçekleştirilmiştir, birincisi kalabalığın ölçeğinde değişmez sorunları çözmüştür. İkincisi, hataları azaltmak için oldukça yoğun olan kalabalığın model performansını iyileştirmeye çalışmıştır. Derin öğrenme yapısı, semantik seviye özellikleri öğrenen VGG-16 ağına benzer [97]. Sığ ağ her biri ortalama ortaklama katmanının izlediği üç evrişim katmanından oluşur. Sığ ağ modelinde her katmanda boyutu 5×5 olan 24 filtre bulunmaktadır. Bu ağ, düşük seviyeli bilgilerin çıkarılmasında ve algılanmasında yardımcı olur. Son olarak, her iki ağdan gelen tahminler birleştirilmiştir [121]. Çıkışın giriş görüntüsü boyut yoğunluğu haritasını oluşturmak için ara değer olarak eklenmiştir. Kişi sayımı, tahmin edilen yoğunluk haritasını ile hesaplanmıştır. Sistem, UCF_CC_50 veri kümesinde değerlendirilmiş ve sonuçlar, ortalama mutlak hata (mean absolute error-MAE) yöntemiyle elde edilmiştir.

Xu ve arkadaşları, yoğun Yönlü Gradyan Histogram (histogram orientated gradient-HOG) özelliklerini kullanarak yoğun kalabalığın içinde insanların sayılmasını önermektedir [122]. Önerilen yaklaşım, kalabalığın yapısının mekânsal farkını yakalar ve çok sayıda regresörü eğitmeye gerek duymamaktadır. Özellikler, yalnızca bir regresyon modelinin yetiştirilmesi için görüntülerin bölünmüş bölgelerinden çıkarılmıştır. Her bölge için, görüntünün birikmiş HOG özelliği çıkarılmıştır. Mevcut regresyon tabanlı modellerde HOG özellikleri, bölüt, kenar, doku gibi diğer düşük seviye özelliklerine kıyasla iyi performans göstermiştir. Bu çalışmada, Gauss Süreç Regresyonu (Gaussian Process Regression-GPR), HOG özelliklerini, kişi sayısı ile eşleştirmek için kullanılmıştır. Tasarlanan sistemin performansı, kullanılan çekirdek türüne bağlıdır. Böylece, yazarlar doğrusal ve üstel (Radial Basis Function-RBF) çekirdeğin kombinasyonunu kullanmışlardır [123-124].

Zhang ve çalışma arkadaşları 2016 yılında, kalabalık sayısını, kalabalık yoğunluğu ve değişken perspektiflerden oluşan görüntülerden tahmin etmek için evrişimli sinir ağı modeli geliştirilmiştir. Çok sütunlu evrişimli sinir ağı

yapısı, görüntüyü kalabalık yoğunluk haritasına eşleştirmektedir. Önerilen mimarinin belirgin özelliği, girilen görüntünün rasgele bir boyut veya çözünürlükte olmasını sağlar. Küçük, orta ve büyük alıcı alan boyutlarına sahip filtreler görüntüye dair özelliklerin öğrenilmesine karşılık gelir ve kalabalık yoğunluğundaki değişime uyarlanır. Bu özelliğinden dolayı önerilen yöntem bozulmayı önlemek için herhangi bir boyuttaki görüntüyle çalışabilme imkânı sağlar [125].

Kafa veya yüz detektörünü içeren kalabalığın geleneksel sayım yöntemi, yoğun bir sahnede çok iyi çalışmamaktadır. İnsanı bir doku olarak, dolayısıyla her insan için bir kafa ve omuz şeklini işleme almak daha uygun olabilmektedir. Bu nedenle yoğun bir öğrenme, yoğun bir kalabalığın insan sayımını daha iyi tahmin etmek için kullanılabilir [126]. Bazı kalabalık sayma yöntemlerinde, kalabalık bölge bölütlemesi için kalabalığın yeri önceden bilinmelidir. Kalabalık bölütleme probleminin kendisi de zor bir sorundur ancak gözetim kameralarının hareket analizi için statik olduğunu varsayarsak problemi kolayca çözülebilir hale getirir. Ancak uygulamada statik kalabalık trafik akışını büyük ölçüde etkilediği için analiz edilmelidir. Yukarıdaki zorluklarla başa çıkmak ve kalabalık bölütlemesi için bir evrişimli sinir ağı modeli önerilmiştir [126]. Yazarlar; görüntülere ait, hareket ve yapı bilgilerini ve önerilen füzyon şemalarını birleştirmiştir. Kalabalık sayımında mevcut veri kümesi, kamera açısı ve yüksekliği gibi yan bilgiler içermez. Kang ve arkadaşlarının çalışmalarında ise, kalabalık sayımı için uyarlamalı bir evrişimli sinir ağı kullanılır. Evrişimli ağlardaki filtre ağırlıkları yan bilgiyi kullanarak mevcut sahneye uyum sağlar. Yazarlar, yan bilgisine sahip uyarlanabilir ağırlıkları kullanarak daha iyi sonuçlar verdiğini öne sürmüşlerdir. Kalabalık gözetimini etkileyen, örneğin hava durumu, kalabalık yoğunluğu gibi farklı durumlar vardır. Dolayısıyla, dayanıklı bir yöntemdir [127]. Cao'nun çalışmasında kalabalığın sayılması için evrişim temelli bir sinir ağı önerilmiştir. Yöntem, evrişimli sinir ağı ile sınıflamasını, evrişimli sinir ağı regresyon ile birleştirmek için dolaylı strateji kullandığından mütevellit daha sağlam olduğu kanıtlanmıştır. Yoğunluk ve ölçek değişiklikleri nedeniyle bir hedefin büyük görünüm değişiklikleri nedeniyle kalabalık sayımı daha zorlaşmaktadır. Geleneksel kalabalık sayımı, büyük görünüm değişikliği hedeflerini saymada daha iyi sonuçlar vermeyen bir öngörücü kullanır [128].

Marsden ve arkadaşları çalışmalarında, modeli optimize edebilmek için hesap karmaşıklığından dolayısıyla parametre sayısının azaltılması konusuna yoğunlaşmışlardır. Derin öğrenme modellerinde öğrenilen ve hesaplanan parametre sayısı sistem performansını etkileyen önemli faktörlerdendir. Model her katmanda bir kestirim yapmakta ve bu kestirimlerin ortalaması sonuç kalabalık yoğunluk kestirimini vermektedir. Ayrıca bu çalışmada eğitim seti için veri artırma yapılmaktadır. Veri artırma da sistemin performansını pozitif yönde etkileyen yöntemlerdendir. Önceki çalışmalarda veri artırma için örtüşmeli rastgele kırpmaya işlemi yapılırken bu çalışmada,

örtüşmesiz dikey ve yatay yönlendirme tercih edilmektedir. Bu sayede modelin aşırı uydurulmasının (overfitting) önüne geçilmektedir. Buna ek olarak, önerilen yöntemin genelleme performansını çapraz veri kümesi performans ölçümü ile incelenmiştir [129].

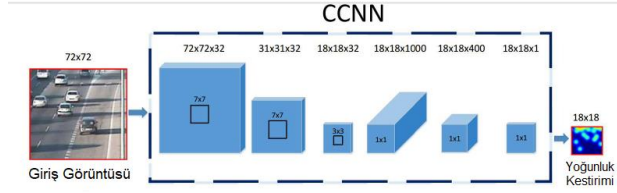


Şekil 17. Model diyagramı [129]
(Model diagram [129])

Shang ve arkadaşları 2016'da evrişimli sinir ağını uçtan uca kalabalık sayan girdi olarak tüm görüntüyü alıp sonuç tahmini sağlamak üzere önerilmiştir. Bağlamsal bilgiler hem yerel hem de global sayımı tahmin etmek için bir araç olarak kullanılır [130].

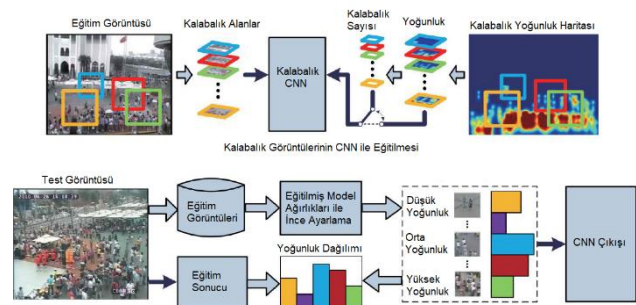
Hu ve ekibi çalışmalarında kalabalık sayısını tahmin etmek için derin öğrenme yaklaşımı kullanmaktadırlar. Bu çalışmada kullanılan yöntem, aşırı kalabalıklar için kullanılabilir ve hareketsiz görüntüler için de başarılı sonuçlar vermektedir. CovNet adı verilen evrişimli sinir ağı ile özellikler elde edilmektedir. Kalabalık yoğunluğu ve kalabalık sayısı da kalabalık özelliklerini elde etmek için kullanılır. Yazarlar ayrıca farklı sahneleri kapsayan 107 görüntü içeren AHU-CROWD adlı kendi veri kümesini oluşturup, bu veri kümesi ile çalışmışlardır [131].

Onoro vd., ölçek temelli sayma modeli ve Hidra evrişimli sinir ağı (Hydra CNN) adında bir model geliştirmişlerdir. Nesne yoğunluğu hesaplayan bu model derin tüm evrişimli sinir ağı yapılarını içinde barındırmaktadır. CCNN (counting convolutional neural network-CCNN) modeli 6 katmanlı bir yapıdan meydana gelir ve paralel olarak işlenmesiyle hedeflenen sonuca ulaşılmaktadır. Nesne ve/veya kişi sayma konusunda başarılı bir modeldir. Modelin blok şeması aşağıdaki gibidir [132].



Şekil 18. Hidra CNN modeli [132]
(Hydra-CNN model [132])

Zhang ve arkadaşları var olan yöntemleri kendi yöntemlerinde performansı iyileştirmek için ön eğitim (pre-training) işlemi için kullanmışlardır. Kalabalık yoğunluk haritası oluştururken çapraz görüntü sayımı işlemi uygulamışlardır. Tasarlanan model tamamen evrişimli sinir ağı temellidir. Bu çalışma, model eğitiminde iki temel üzerine yoğunlaşmaktadır: kalabalık yoğunluk haritası ve ilişkili nesnelere. Ön eğitim sayesinde tasarlanan model ekstra etikete ihtiyaç duymadan test görüntüleri üzerinde başarılı sonuçlar vermiştir. Çalışmada tek görüntü ve çapraz görüntüler üzerinde model test edilmiştir [133].



Şekil 19. Önerdikleri modelin diyagramı [134]
(Their proposed model diagram [134])

Kumagai ve arkadaşları çalışmalarında birden fazla önerilen öngörücü, her biri belirli bir görünüm için eğitilmiştir. Tahmini kalabalık sayısı bakımından önerinin görünüm değişikliklerinde daha güçlü olduğu gösterilmiştir. [119]'de yoğun bir çevrede ve geniş bir eğitime veri kümesi olmaksızın kalabalık tahmini sağlamak için uçtan uca derin evrişimli sinir ağı önerilmektedir. Geleneksel yöntemler, daha az eğitim örneği ve yoğun kalabalıkla baş etmekte başarısız olmuştur [134]. Önerilen yöntem, etkili özellikleri öğrenen bir modeldir ve adına evrişimli sinir ağı karışım (mixture of CNN-MoCNN) demektirler. Evrişim ve maksimum ortaklama katmanlarından oluşan klasik bir yapıya sahiptir. Ayrıca, yazarlar, negatif örneklerle önerilen yöntemlerin daha dayanıklı ve verimli olduğunu gösteren kanıtlar da sunmaktadır.

Farklı bir yaklaşımda Sindagi ve arkadaşlarının önerisi olan çoklu görev için kaskat CNN mimarisidir. Burada

Tablo 3. Kalabalık veri setlerinden örnekler: UCSD (a), Mall (b), UCF CC 50 (c), WorlExpo'10 (d), ShanghaiTech Part-A (e), ShanghaiTech Part-B (f) [1]

(Examples of crowd datasets UCSD (a), Mall (b), UCF CC 50 (c), WorlExpo'10 (d), ShanghaiTech Part-A (e), ShanghaiTech Part-B (f) [1])



İncelenen çalışmalarda çoğunlukla performans kriteri olarak ortalama karesel hata (OKH) ve ortalama mutlak hata (OMH) kullanıldığı görülmektedir.

$$OKH = \sqrt{\frac{1}{N} \sum_{i=1}^N |y_i - y_i^*|^2} \quad (13)$$

$$OMH = \frac{1}{N} \sum_{i=1}^N |y_i - y_i^*| \quad (14)$$

Burana N test örnek sayısı, y_i görüntülerdeki gerçek kişi sayısı, y_i^* ise görüntülerden kestirilen kişi sayısıdır.

Tablo 2'de gösterilen, çeşitli kalabalık seviyeleri için hazırlanan farklı veri kümeleri vardır. Bu veri kümeleri ile yapılan çalışmalarda genellikle OKH ve OMH performans kriterleri kullanılmaktadır. Literatürdeki evrişimli sinir ağı temelli derin öğrenme modeli kullanan çalışmalar Tablo 4'te incelenmektedir. Spesifik olarak yöntem ve ilgili yöntemin kullanıldığı veri kümesine ilişkin oluşturulan tabloda elde edilen hata sonuçlarını karşılaştırmak mümkündür. Aynı anda birkaç farklı kalabalık veri kümesiyle yapılan çalışmalar olduğu gibi yalnızca bir veri kümesiyle yapılan çalışmalar da mevcuttur. Çalışmaların başarımları görüntüdeki kalabalık yoğunluğunun artmasıyla düştüğü incelenilmektedir. Bu da problemin ne kadar zorlaşabileceğiyle ilgili önemli bir çıkarımdır.

Tablo 4. Çeşitli veri setleri ve literatürdeki derin öğrenme temelli kalabalık analizi çalışmalarının başarımlarının karşılaştırması

(Comparison of the performances based on crowd analysis using deep learning in literature and various datasets)

	Veri Kümesi	UCSD		MALL		UCF CC 50		WorldExpo'10		Shanghai Tech-A		Shanghai Tech-B	
		OKH	OMH	OKH	OMH	OKH	OMH	OKH	OMH	OKH	OMH	OKH	OMH
Evrişimli Sinir Ağı Temelli Derin Öğrenme Modeli	Yöntem												
	Çapraz sahne [117]	1.60	3.31			467.0	498.5	12.9		181.8	277.7	32.0	49.8
	Derin+Sığ Ağ [121]					452.5							
	M-CNN [125]	1.07	1.35			377.6	509.1	11.6		110.2	173.2	26.4	41.3
	CNN-artırma [143]	1.10		2.01		364.4							

Veri Kümesi	UCSD		MALL		UCF CC 50		WorldExpo'10		Shanghai Tech-A		Shanghai Tech-B	
Hidra-CNN [132]					333.7	425.2						
Ortak yerel ve global sayım [130]					270.3		11.7					
MoCNN [134]			2.75	13.4	361.7	493.3						
FCN [129]					338.6	424.5			126.5	173.5	23.76	33.12
CNN-piksel [126]	1.12	2.06			406.2	404.0	13.4					
Ağırlıklı V-LAD [144]	2.86	13.0	2.41	9.12								
Kaskat-MTL [135]					322.8	341.4			101.3	152.4	20.0	31.1
Anahtarlamalı-CNN [136]	1.62	2.10			318.1	439.2	9.4		90.4	135.0	21.6	33.4

6. SONUÇ (CONCLUSION)

Bu literatür incelemesinde, derin öğrenmenin tarihçesinden günümüze gelişi ve elde edilen başarımların aktarmanın yanında temel olarak bilgisayarlı görü alanında tercih edilen derin öğrenme mimarileri teorik ve detaylı olarak da anlatılmaktadır.

Spesifik olarak, görüntüler üzerinden kalabalık analizi konusunda geleneksel yöntemlere değinilmekte birlikte, esasen derin öğrenme mimarileri ve bunların kalabalık analizi konusuna katkıları araştırılmıştır. Bu çalışma, kalabalık yoğunluk hesaplama, kişi sayma ve takip etme, kalabalık davranışı ve anormalliklerin tespit edilmesiyle ilgili çeşitli çalışmaların literatürde klasik yöntemler kısaca özetlenirken, yapay zeka alanının özel bir alanı olan derin öğrenmenin kalabalık analizine katkıları kapsamlı bir şekilde özetlenmektedir. Özellikle evrişimli sinir ağı modeli bilgisayarlı görü alanında başarılı sonuçlar vermesi dolayısıyla tercih edilmektedir. Ayrıca kullanılan veri setleri ve yapılan çalışmaların başarımları tablolar halinde belirtilmektedir.

Sonuç olarak, literatürdeki çalışmalar başarımlar bakımından da karşılaştırılmıştır. Makine öğrenmesi ve derin öğrenme ile kalabalık analizi çalışmaları yapmak isteyen araştırmacılar için yol gösterici bir içerik olması hedeflenmiştir. Son yılların en büyük problemlerinden biri olan metropollerdeki insan davranışlarının modellenmediği veya olası güvenlik problemlerinin önüne geçilebilmesi için kalabalık analizi üzerinde çalışması gereken ve derin öğrenme ile başarılı değerlendirmeler alınan hem akademik hem pratik bir çalışma alanıdır.

KAYNAKLAR (REFERENCES)

- [1] V. D. Sindagi, V. M. Patel, "A Survey of Recent Advances in CNN-Based Single Image Crowd Counting and Density Estimation", *Pattern Recognition Letters*, 107, 3-16, 2018.
- [2] F. Rosenblatt, **The Perceptron a Perceiving and Recognizing Automaton**, Cornell Aeronautical Laboratory, 1957.
- [3] A. G. Ivakhnenko, V. G. Lapa, **Cybernetic Predicting Devices**, Purdue University School of Electrical Engineering, 1965.
- [4] K. Fukushima, "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", **Biological Cybernetics by Springer-Verlag**, 36, 193-202, 1980.
- [5] G. E. Hinton, "Learning Distributed Representations of Concepts", **Proceedings of the Eighth Annual Conference of the Cognitive Science Society, Amherst, Mass**, Parallel Distributed Processing: Implications for Psychology and Neurobiology, Editör: R. G. M. Morris, Oxford University Press, Oxford, UK, 46-61, 1986.
- [6] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning Representations by Back-Propagating Errors", *Nature*, 323, 533-536, 1986.
- [7] M. Newborn, "Deep Blue's Contribution to AI", *Annals of Mathematics and Artificial Intelligence*, 28(1-4), 27-30, 2000.
- [8] D. Ferrucci, A. Levas, S. Bagchi, D. Gondek, E. Mueller, "Watson: Beyond Jeopardy!", *Artificial Intelligence*, 199-200, 93-105, 2013.
- [9] Y. LeCun, L. Bottou, Y. Bengio ve P. Haffner, "Gradient Based Learning Applied to Document Recognition", *Proceeding of the IEEE*, 86(11), 2278-2324, 1998.
- [10] Internet: W. Knight, AI Winter Isn't Coming, Intelligent Machines, MIT Technology Review, <https://www.technologyreview.com/s/603062/ai-winter-isnt-coming/>, 07.11. 2016.
- [11] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", **Advances in Neural Information Processing Systems25 (NIPS'12)**, 2012.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed,

- D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going Deeper with Convolutions", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)**, 1-9, 2015.
- [13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Nets", **Advances in Neural Information Processing Systems**, 2672-2680, 2014.
- [14] Internet: F. Ferreira, How Tay "Machine Learned" Her Way to Become a Twitter Troll, Harvard University, Graduate School of Arts and Science, SITN, Science in the News, 12 Nisan 2016, <http://sitn.hms.harvard.edu/flash/2016/how-tay-machine-learned-her-way-to-become-a-twitter-troll/>, 20.01.2018.
- [15] Internet: D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the game of Go with deep neural networks and tree search, doi:10.1038/nature16961, Nature | Vol 529 | 28 Ocak 2016 <https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf>, 21.11.2017.
- [16] S. Sabour, N. Frosst, G. E. Hinton, "Dynamic Routing Between Capsules", **31st Conference on Neural Information Processing Systems (NIPS'17)**, Long Beach, CA, USA, 2017.
- [17] E. Alpaydın, **Yapay Öğrenme**, Boğaziçi Üniversitesi Yayınları, Türkiye, 2011.
- [18] Internet: A. Karpathy, Stanford University, Stanford CS class CS231n: Convolutional Neural Networks for Visual Recognition, Course Notes, 20.03.2018.
- [19] B. Widrow ve M. E. Hoff, "Associative Storage and Retrieval of Digital Information in Networks of Adaptive 'Neurons'", *Biological Prototypes and Synthetic Systems*, 1, 160, 1962.
- [20] C. Cortes, V. Vapnik, "Support-Vector Networks", *Journal of Machine Learning*, 20(3), 273-297, 1995.
- [21] R. C. Gonzalez, R. E. Woods, **Digital Image Processing**, Pearson Publication, 1977.
- [22] Internet: M. Nielsen, Y. Bengio, I. Goodfellow, A. Courville, Deep Learning Book, <http://neuralnetworksanddeeplearning.com/>, 2016, 10.2017.
- [23] M. D. Zeiler, R. Fergus, "Visualizing and Understanding Convolutional Networks", **European Conference on Computer Vision (ECCV'14)**, 818-833, 2013.
- [24] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)**, 770-778, 2015.
- [25] M. Lin, Q. Chen, S. Yan, "Network in Network", *arXiv:1312.4400*, 2014.
- [26] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", **Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)**, 4278-4284, 2016.
- [27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16)**, 2818-2826, 2016.
- [28] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich Feature Hierarchies for Accurate Object Detection And Semantic Segmentation", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)**, 580-587, 2014.
- [29] Internet: A. Ng, Y. B. Mourri, K. Katanforoosh, Deep Learning Specialization, Convolutional Neural Networks, Coursera, <https://www.coursera.org/learn/convolutional-neural-networks>, 02.01.2017.
- [30] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, 9(8), 1735-1780, 1997.
- [31] Internet: A. Karpathy, The Unreasonable Effectiveness of Recurrent Neural Networks, <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>, 21 Mayıs 2015, 09.01.2018.
- [32] Internet: Kaggle Survey, The State of Data Science & Machine Learning, <https://www.kaggle.com/surveys/2017>, 20.09.2017.
- [33] A. Şeker, B. Diri, H. H. Balık, "Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme", *Gazi Mühendislik Bilimleri Dergisi*, 3(3), 47-64, 2017.
- [34] Y. Poyraz, S. Sevgen, "GPU Programlama Tekniği ile Yüksek Performanslı Araç Takibi", *Bilişim Teknolojileri Dergisi*, 10(3), 255-261, 2017.
- [35] Internet: M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, <https://www.tensorflow.org/>, 01.05.2017.

- [36] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding", **Proceeding MM '14 Proceedings of the 22nd ACM international conference on Multimedia**, 675-678, 2014.
- [37] Internet: R. Collobert, C. Farabet, K. Kavukcuoğlu, Torch | Scientific computing for LuaJIT, NIPS Workshop on Machine Learning Open Source Software, <http://torch.ch/>, 01.05.2017.
- [38] Internet: F. Chollet, Keras, <https://keras.io/>, 10.10.2017.
- [39] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, Z. Zhang, "MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems", **Workshop on Machine Learning Systems In Neural Information Processing Systems**, 2016.
- [40] F. Seide, A. Agarwal, "CNTK: Microsoft's Open-Source Deep-Learning Toolkit", **Proceeding KDD'16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, 2135-2135, 2016.
- [41] Internet: Skymind, Deeplearning4j: Open-source, Distributed Deep Learning for the JVM, <https://deeplearning4j.org/>, 10.10.2017.
- [42] Internet: D. Yuret, Welcome to Knet.jl's documentation!, <http://denizyuret.github.io/Knet.jl/latest/>, 10.08.2016.
- [43] Theano Development Team, "Theano: A {Python} framework for fast computation of mathematical expressions," *ArXiv e-prints*, abs/1605.02688, 2016.
- [44] J. C. S. Jacques Junior, S. R. Musse, C. R. Jung, "Crowd Analysis using Computer Vision Techniques," *In IEEE Signal Processing Magazine*, 27, 66-77, 2010.
- [45] M. A. Kızrak, B. Bolat, "A Novel Approach for People Counting and Tracking from Crowd Video," **In IEEE International Conference on Innovations in Intelligent SysTems and Applications (INISTA)**, 2017.
- [46] A. K. Abdulrahman, S. Öztürk, "Çoklu Görüntü Damgalama Yönteminde Farklı Frekans Bölgelerinin Değerlendirilmesi", *Bilişim Teknolojileri Dergisi*, 11(1), 75-88, 2018.
- [47] J. Hwang, C. Chu, H. Pai, K. Lan, "Tracking Human Under Occlusion Based On Adaptive Multiple Kernels With Projected Gradients", *In IEEE Transaction on Multimedia*, 15(7), 1602-1615, 2013.
- [48] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, L. Q. Xu, "Crowd Analysis: A Survey", *Machine Vision Application*, 19(2), 345-357, 2008.
- [49] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, S. Yan, "Crowded Scene Analysis: A Survey", *IEEE Transactions on Circuits and Systems for Video Technology*, 25(3), 367-386, 2015.
- [50] S. Ali, M. Shah, "Floor Fields for Tracking in High Density Crowd Scenes", *10th European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science*, 5303, 1-14, 2008.
- [51] Y. Mao, J. Tong, W. Xiang, "Estimation of Crowd Density using Multi-Local Features and Regression", **Proceedings of the 8th World Congress on Intelligent Control and Automotion**, 6295-6300, 2010.
- [52] W. Ma, L. Huang, C. Liu, "Crowd Density Analysis using Co-Occurrence Texture Features", **In 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT'10)**, 170-175, 2010.
- [53] J. Guo, X. Wu, T. Cao, S. Yu, Y. Xu, "Crowd Density Estimation via Markov Random Field (MRF)", **Proceedings of 8th World Congress on Intelligent Control and Automation**, 258-263, 2010.
- [54] W. Li, X. Wu, K. Matsumoto, H. Zhao, "A New Approach of Crowd Density Estimation", **IEEE Region 10 Conference TENCN**, 200-203, 2010.
- [55] W. Li, X. Wu, K. Matsumoto, H. Zhao, "Crowd Density Estimation: An Improved Approach", **IEEE 10th International Conference on Signal Processing (ICSP'10)**, 1213-1216, 2010.
- [56] W. Ge, R. T. Collins, "Crowd Density Analysis with Marked Point Processes," *In IEEE Signal Processing Magazine*, 27, 107-123, 2010.
- [57] G. Kim, K. Eom, M. Kim, J. Jung, "Automated Measurement of Crowd Density Based on Edge Detection and Optical Flow", **In IEEE 2nd International Conference on Industrial Mechatronics and Automation, Volume 2**, 553-556, 2010.
- [58] W. Hsu, K. Lin, C. Tsai, "Crowd Density Estimation Based on Frequency Analysis," **7th International Conference on Intelligent Information Hiding and Multimedia Signal Processing**, 348-351, 2011.
- [59] G. Xiong, X. Wu, J. Cheng, Y. Chen, Y. Ou, Y. Liu, "Crowd Density Estimation Based on Image Potential Energy Model", **Proceedings of the IEEE International Conference on Robotics and Biometrics (ROBIO)**, 538-543, 2011.
- [60] H. Yu, Z. He, Y. Liu, L. Zhang, "A Crowd Flow Estimation Method Based on Dynamic Texture and GRNN", **7th IEEE Conference on Industrial Electronics and Applications (ICIEA)**, 79-84, 2012.

- [61] H. Yang, H. Zhao, "A Novel Method for Crowd Density Estimations", **IET International Conference on Information Science and Control Engineering (ICISCE)**, 1-4, 2012.
- [62] V. B. Subburaman, A. Descamps, C. Carincotte, "Counting People in the Crowd using a Generic Head Detector", **IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance**, 470-475, 2012.
- [63] A. Chan, N. Vasconcelos, "Counting People with Low-Level Features and Bayesian Regression", *IEEE Transactions on Image Processing*, 21(4), 2160-2177, 2012.
- [64] A. B. Chan, N. Vasconcelos, "Modeling, Clustering and Segmenting Video with Mixtures of Dynamic Textures", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5), 909-926, 2008.
- [65] A. B. Chan, Z. J. Liang, N. Vasconcelos, "Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)**, 1-7, 2008.
- [66] Z. Wang, H. Liu, Y. Qian, T. Xu, "Crowd Density Estimation Based on Local Binary Pattern Co-Occurrence Matrix", **IEEE International Conference on Multimedia and Expo Workshops (ICMEW)**, 372-377, 2012.
- [67] H. Fradi, J. Dugelay, "People Counting System in Crowded Scenes Based on Feature Regression", **Proceedings of the 20th European Signal Processing Conference (EUSIPCO)**, 136-140, 2012.
- [68] H. Fradi, J. Dugelay, "Crowd Density Map Estimation Based on Features Tracks", **In IEEE 15th International Workshop on Multimedia Signal Processing**, 40-45, 2013.
- [69] F. Tehranipour, R. Shishegar, S. Tehrenipour, S. Seterehdan, "Attention Control Using Fuzzy Inference System in Monitoring CCTV Based on Crowd Density Estimation", **IEEE 8th Iranian Conference on Machine Vision and Image Processing (MVIP)**, 204-209, 2013.
- [70] H. Fradi, X. Zhao, J. Dugelay, "Crowd Density Analysis using Subspace Learning on Local Binary Pattern," **In IEEE International Conference on Multimedia and Expo Workshops (ICMEW)**, 1-6, 2013.
- [71] A. S. Rao, J. Gubbi, S. Marusic, P. Stanley, M. Palaniswami, "Crowd Density Estimation Based on Optical Flow and Hierarchical Clustering," **IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)**, 494-499, 2013.
- [72] Y. Yuan, J. Zhao, C. Qui, "Estimating Crowd Density in an RF-Based Dynamic Environment", *IEEE Sensors Journal*, 13(10), 3837-3845, 2013.
- [73] P. Karpagavalli, A. V. Ramprasad, "Estimating the Density of the People and Counting the Number of People in a Crowd Environment for Human Safety", **In IEEE International Conference on Communication and Signal Processing**, 663-667, 2013.
- [74] Z. Wu, H. Zheng, J. Wang, "Pedestrian Counting Based on Crowd Density Estimation and Lucas-Kanade Optical Flow", **IEEE 7th International Conference on Image and Graphics (ICIG)**, 471-476, 2013.
- [75] K. Ping, P. Bo, Z. Wenying, L. Shuai, "Research on Central Issues of Crowd Density Estimation", **10th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)**, 143-145, 2013.
- [76] Y. Yuan, "Crowd Monitoring using Mobile Phones", **IEEE 6th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Volume 1**, 261-264, 2014.
- [77] M. Khansari, H. R. Rabiee, M. Asadi, M. Ghanbari, "Object Tracking in Crowded Video Scenes Based on the Undecimated Wavelet Features and Texture Analysis", *EURASIP Journal on Advances in Signal Processing*, 2008(1), 1-18, 2008.
- [78] M. Rodriguez, I. Laptev, J. Sivic, J. Audibert, "Density-Aware Person Detection and Tracking in Crowds", **IEEE International Conference on Computer Vision**, 2423-2430, 2011.
- [79] M. Rodriguez, J. Sivic, I. Laptev, J. Audibert, "Data-Driven Crowd Analysis in Videos", **In IEEE International Conference on Computer Vision**, 1235-1242, 2011.
- [80] D. Conte, P. Foggia, G. Percannella, F. Tufano, M. Vento, "A Method for Counting Moving People in Video Surveillance Videos", *EURASIP Journal on Advances in Signal Processing*, 2010, 1-10, 2010.
- [81] G. Antonini, J-P. Thiran, "Counting Pedestrians in Video Sequences Using Trajectory Clustering", *IEEE Transactions on Circuits and Systems for Video Technology*, 16(8), 1008-1020, 2006.
- [82] E. L. Andrade, S. Blunsden, R. B. Fisher, "Modelling Crowd Scenes for Event Detection", **In Pattern Recognition (ICPR'06) 18th International Conference on, Volume 1**, 175-178, 2006.
- [83] X. Wang, X. Ma, W. E. L. Grimson, "Unsupervised Activity Perception in Crowded and Complicated Scenes using Hierarchical Bayesian Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3), 539-555, 2009.
- [84] C. C. Loy, T. Xiang, S. Gong, "Multi-Camera Activity Correlation Analysis", **In IEEE**

- Conference on Computer Vision and Pattern Recognition (CVPR'09)**, 1988-1995, 2009.
- [85] R. Mehran, A. Oyama, M. Shah, "Abnormal Crowd Behavior Detection using Social Force Model", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)**, 935-942, 2009.
- [86] T. Hospedales, S. Gong, T. Xiang, "A Markov Clustering Topic Model for Mining Behaviour in Video", **In IEEE 12th International Conference on Computer Vision**, 1165-1172, 2009.
- [87] S. Ali, M. Shah, "A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)**, 1-6, 2007.
- [88] J. Liu, B. Kuipers, S. Savarese, "Recognizing Human Actions by Attributes", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)**, 3337-3344, 2011.
- [89] B. Zhou, X. Tang, X. Wang, "Coherent Filtering: Detecting Coherent Motions from Crowd Clutters", **Computer Vision (ECCV'12)**, 857-871, 2012.
- [90] B. Zhou, X. Tang, X. Wang, "Measuring Crowd Collectiveness", **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13)**, 3049-3056, 2013.
- [91] J. Shao, C. C. Loy, X. Wang, "Scene-Independent Group Profiling in Crowd", **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)**, 2219-2226, 2014.
- [92] K. Kang, X. Wang, "Fully Convolutional Neural Networks for Crowd Segmentation", *ArXiv preprint*, arXiv:1411.4464, 2014.
- [93] S. Yi, X. Wang, C. Lu, J. Jia, "L0 Regularized Stationary Time Estimation for Crowd Group Analysis", **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)**, 2211-2218, 2014.
- [94] F. Zhu, X. Wang, N. Yu, "Crowd Tracking with Dynamic Evolution of Group Structures", **In European Conference on Computer Vision**, Springer, 139-154, 2014.
- [95] B. Zhou, X. Tang, X. Wang, "Learning Collective Crowd Behaviors with Dynamic Pedestrian-Agents", *International Journal of Computer Vision*, 111(1), 50-68, 2015.
- [96] S. Yi, H. Li, X. Wang, "Understanding Pedestrian Behaviors from Stationary Crowd Groups", **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, 3488-3496, 2015.
- [97] J. Shao, K. Kang, C. C. Loy, X. Wang, "Deeply Learned Attributes for Crowded Scene Understanding", **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)**, 4657-4666, 2015.
- [98] B. Zhou, X. Wang, X. Tang, "Understanding Collective Crowd Behaviors: Learning a Mixture Model of Dynamic Pedestrian-Agents", **In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)**, 2871-2878, 2012.
- [99] N. Kumar, A. C. Berg, P. N. Belhumeur, S. K. Nayar, "Attribute and Simile Classifiers for Face Verification", **In IEEE 12th International Conference on Computer Vision**, 365-372, 2009.
- [100] A. Farhadi, I. Endres, D. Hoiem, D. Forsyth, "Describing Objects by Their Attributes", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)**, 1778-1785, 2009.
- [101] C. H. Lampert, H. Nickisch, S. Harmeling, "Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)**, 951-958, 2009.
- [102] T. L. Berg, A. C. Berg, J. Shih, "Automatic Attribute Discovery and Characterization from Noisy Web Data", **In European Conference on Computer Vision**, Springer, 663-676, 2010.
- [103] Y. Fu, T. Hospedales, T. Xiang, S. Gong, "Attribute Learning for Understanding Unstructured Social Activity", **Computer Vision (ECCV'12)**, 530-543, 2012.
- [104] G. Patterson, J. Hays, "Sun Attribute Database: Discovering, Annotating, and Recognizing Scene Attributes", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)**, 2751-2758, 2012.
- [105] A. Oliva, A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope", *International Journal of Computer Vision*, 42(3), 145-175, 2001.
- [106] F-F. Li, I. Asha, K. Christof, P. Pietro, "What Do We Perceive in a Glance of a Real-World Scene?", *Journal of Vision*, 7(1), 10-10, 2007.
- [107] D. Parikh, K. Grauman, "Interactively Building a Discriminative Vocabulary of Nameable Attributes", **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)**, 1681-1688, 2011.
- [108] F. Jiang, Y. Wu, A. K. Katsaggelos, "Detecting Contextual Anomalies of Crowd Motion in Surveillance Video," **16th IEEE International Conference on Image Processing (ICIP'09)**, 1117-1120, 2009.
- [109] Ö. Akşehirli, H. Ankaralı, Ş. Cangür, M. A. Sungur, "Breiman Algoritması Kullanılarak

- Homojen Alt Grupların Belirlenmesi: Bir Uygulama”, *Bilişim Teknolojileri Dergisi*, 7(1), 19-24, 2014.
- [110] L. Kratz, K. Nishino, “Anomaly Detection in Extremely Crowded Scenes using Spatio-Temporal Motion Pattern Models”, **In IEEE Conference on Computer Vision and Pattern Recognition (CVPR’09)**, 1446-1453, 2009.
- [111] V. Mahadevan, W. Li, V. Bhalodia, N. Vasconcelos, “Anomaly Detection in Crowded Scenes”, **IEEE Conference on Computer Vision and Pattern Recognition (CVPR’10)**, 1975-1981, 2010.
- [112] V. Reddy, C. Sanderson, B. C. Lovell, “Improved Anomaly Detection in Crowded Scenes via Cell-Based Analysis of Foreground Speed, Size and Textures”, **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)**, 55-61, 2011.
- [113] V. Saligrama, Z. Chen, “Video Anomaly Detection Based on Local Statistical Aggregates”, **IEEE Conference on Computer Vision and Pattern Recognition (CVPR’12)**, 2112-2119, 2012.
- [114] L. Cao, K. Huang, “Video-Based Crowd Density Estimation and Prediction System for Wide-Area Surveillance”, *China Communications*, 10(5), 79-88, 2013.
- [115] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, F-F. Li, “Large-Scale Video Classification with Convolutional Neural Networks”, **In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR’14)**, 1725-1732, 2014.
- [116] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *ArXiv preprint*, arXiv:1409.1556, 2014.
- [117] C. Zhang, H. Li, X. Wang, X. Yang, “Cross-Scene Crowd Counting via Deep Convolutional Neural Networks”, **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’15)**, 833-841, 2015.
- [118] H. Idrees, I. Saleemi, C. Seibert, M. Shah, “Multi-Source Multi-Scale Counting in Extremely Dense Crowd Images”, **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, 2547-2554, 2013.
- [119] C. Wang, H. Zhang, L. Yang, S. Liu, X. Cao, “Deep People Counting in Extremely Dense Crowds”, **In Proceedings of the 23rd ACM international conference on Multimedia**, 1299-1302, 2015.
- [120] J. Li, L. Huang, C. Liu, “An Efficient Self-Learning People Counting System”, **In First Asian Conference on Pattern Recognition (ACPR’11)**, 125-129, 2011.
- [121] L. Boominathan, SS. S. Kruthiventi, R. V. Babu, “Crowdnet: A Deep Convolutional Network for Dense Crowd Counting”, **In Proceedings of the 2016 ACM on Multimedia Conference**, 640-644, 2016.
- [122] N. Dalal, B. Triggs, “Histograms of Oriented Gradients for Human Detection”, **In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), Volume 1**, 886-893, 2005.
- [123] C. E. Rasmussen, C. K. I. Williams, **Gaussian Processes for Machine Learning**, University Press Group Limited, 2006.
- [124] T. Xu, X. Chen, G. Wei, W. Wang, “Crowd Counting using Accumulated HOG”, **In IEEE 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC FSKD)**, 1877-1881, 2016.
- [125] Y. Zhang, D. Zhou, S. Chen, S. Gao, Y. Ma, “Single Image Crowd Counting via Multi-Column Convolutional Neural Network”, **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’16)**, 589-597, 2016.
- [126] L. Lebanoff, H. Idrees, **Counting in Dense Crowds using Deep Learning, REU Participants & Projects Final Report**, University of Central California, 2015.
- [127] D. Kang, D. Dhar, A. B. Chan, “Crowd Counting by Adapting Convolutional Neural Networks with Side Information”, *ArXiv preprint*, arXiv:1611.06748, 2016.
- [128] L. Cao, X. Zhang, W. Ren, K. Huang, “Large Scale Crowd Analysis Based on Convolutional Neural Network”, *Pattern Recognition*, 48(10), 3016-3024, 2015.
- [129] M. Marsden, K. McGuinness, S. Little, N. E. O’Connor, “ResnetCrowd: A Residual Deep Learning Architecture for Crowd Counting, Violent Behaviour Detection and Crowd Density Level Classification”, **13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)**, 1-7, 2016.
- [130] C. Shang, H. Ai, B. Bai, “End-to-End Crowd Counting via Joint Learning Local and Global Count”, **In IEEE International Conference on Image Processing (ICIP’16)**, 1215-1219, 2016.
- [131] Y. Hu, H. Chang, F. Nian, Y. Wang, T. Li T, “Dense Crowd Counting from Still Images with Convolutional Neural Networks”, *Journal of Visual Communication and Image Representation*, 38, 530-539, 2016.
- [132] D. O’norro, R. R. Lopez-Sastre, “Towards Perspective-Free Object Counting with Deep Learning”, **Computer Vision-ECCV 2016: 14th European Conference**, Amsterdam, The Netherlands, 2016.

- [133] C. Zhang, K. Kang, H. Li, X. Wang, R. Xie, X. Yang, "Datadriven Crowd Understanding: a Baseline for a Large-Scale Crowd Dataset", *IEEE Transactions on Multimedia*, 18, 1048-1061, 2016.
- [134] S. Kumagai, K. Hotta, T. Kurita, "Mixture of Counting CNNs: Adaptive Integration of CNNs Specialized to Specific Appearance for Crowd Counting", *ArXiv preprint*, arXiv:1703.09393, 2017.
- [135] V. D. Sindagi, V. M. Patel, "CNN-Based Cascaded Multi-Task Learning of High-Level Prior and Density Estimation for Crowd Counting", **IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)**, 1-6, 2017.
- [136] D. B. Sam, S. Surya, R. V. Babu, "Switching Convolutional Neural Network for Crowd Counting", **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)**, 4031-4019, 2017.
- [137] Internet: UCSD Anomaly Detection Dataset, <http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm>, 01.10.2017.
- [138] Internet: Mall Dataset, Crowd Counting Dataset, http://personal.ie.cuhk.edu.hk/~ccloy/downloads/mall_dataset.html, 01.10.2017.
- [139] Internet: UCF_CC_50 Dataset, University of Central Florida, Center for Research in Computer Vision, http://crcv.ucf.edu/data/crowd_counting.php, 01.10.2017.
- [140] Internet: WorldExpo'10, http://cs-chan.com/downloads_crowd_dataset.html, 18.02.2018.
- [141] Internet: ShanghaiTech Part A, <https://github.com/svishwa/crowdcount-mcnn>, 01.10.2017.
- [142] Internet: ShanghaiTech Part B, <https://github.com/svishwa/crowdcount-mcnn>, 01.10.2017.
- [143] E. Walach, L. Wolf, "Learning to Count with CNN Boosting", **European Conference on Computer Vision**, Springer, 660-676, 2016.
- [144] B. Sheng, C. Shen, G. Lin, J. Li, W. Yang, C. Sun, "Crowd Counting via Weighted Vlad on Dense Attribute Feature Maps", *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.