Sosyal Bilimler Enstitüsü

T.C.

MARMARA ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

SAYISAL YÖNTEMLER BİLİM DALI

MODA KATEGORİSİNE YÖNELİK EVRİŞİMLİ SİNİR AĞLARI TABANLI ÖNERİ SİSTEMİ TASARIMI

Yüksek Lisans

MEHMET YİĞİT ÖZGENÇ

İSTANBUL,2022

T.C.

MARMARA ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

SAYISAL YÖNTEMLER BİLİM DALI

MODA KATEGORİSİNE YÖNELİK EVRİŞİMLİ SİNİR AĞLARI TABANLI ÖNERİ SİSTEMİ TASARIMI

Yüksek Lisans

MEHMET YİĞİT ÖZGENÇ

Danışman: Prof. Dr Ömer Önalan

İSTANBUL,2022

T.C.

MARMARA ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

SAYISAL YÖNTEMLER BİLİM DALI

TEZ ONAY BELGESİ

İşletme Anabilim Dalı Sayısal Yöntemler Bilim Dalı Yüksek Lisans öğrencisi Mehmet Yiğit Özgenç'in Moda Kategorisine Yönelik Evrişimli Sinir Ağları Tabanlı Öneri Sistemi Tasarımı adlı tez çalışması, Enstitümüz Yönetim Kurulunun ..... tarih ve .....sayılı kararıyla oluşturulan jüri tarafından oy birliği / oy çokluğu ile Yüksek Lisans olarak kabul edilmiştir.

Tez Savunma Tarihi ...../...../.....

Öğretim üyesi Adı Soyadı			**İmzası**
1	Tez Danışmanı		
2	Jüri Üyesi		
3	Jüri Üyesi		

ÖZET

MODA KATEGORİSİNE YÖNELİK EVRİŞİMLİ SİNİR AĞLARI TABANLI ÖNERİ SİSTEMİ TASARIMI

Kalabalık ve rekabetçi bir pazarda öne çıkmaya çalışan işletmeler için kişiselleştirme, son yıllarda önemli bir odak noktası haline gelmiştir. Yapay zeka ve makine öğrenimi tabanlı öneri sistemleri, işletmelerin son derece kişiselleştirilmiş müşteri deneyimleri sağlamaları için güçlü bir araç olarak kullanılmaktadır. İşletmeler, her müşterinin özel ihtiyaç ve ilgi alanlarına göre uyarlanmış öneriler sunarak müşteri deneyimini iyileştirmekte ve satış yapma olasılıklarını arttırmaktadır. Küresel ölçekte dijitalleşmenin hız kazanmasıyla birlikte e-ticaret sektörü de bu süreçten oldukça fayda sağlamaktadır. Bu da e-ticaret sektöründe kişiselleştirmeye yönelik rekabeti arttırmaktta ve kişiselleştirme teknolojilerinin değerinin artmasına olanak sağlamaktadır. Bu çalışma, herhangi bir ürün sayfasında bulunan görsel üzerinde, tüm ürün gruplarına öneriler sunabilecek evrişimli yapay sinir ağı tabanlı bir yapay öğrenme sistemi önermeyi amaçlamaktadır. Üç aşamadan oluşan yapay öğrenme sistemi, ilk aşamada nesne tespiti ikinci aşamada gömülü temsil öğrenim ve son aşamada ise yaklaşık en yakın komşu algoritmaları üzerinden ilerlemektedir. Deney süreci içerisinde, ilk aşama için MaskRCNN mimarisinin iMaterialist veri seti üzerinde farklı hiper parametreler üzerinden eğitilmiş ve farklı iou eşikleri için mAP değerleri karşılaştırıldığında en etkili sonuç resnet 101 omurgasına sahip ve tüm katmanları önceden eğitilmiş ağırlıklar üzerinden eğitime devam edilerek alınmıştır.İkinci aşama ve üçüncü aşama birlikte değerlendirilip ikinci aşama için Resnet16, Vgg18 ve VİT modelleri sokaktan dükkana veri seti ile eğitilip annoy en yakın komşu algoritması üzerinden benzerlik önerileri gerçekleştirilmiştir ve farklı k değerleri üzerinden precision@k ve recall@k metrikleri ile değerlendirilmiştir. Değerlerndirme sonucunda VİT mimarisine sahip model tüm metriklerde diğer mimarilere kıyasla daha iyi sonuçlar elde etmiştir.

Anahtar Kelimeler:Öneri Sistemleri,Görsel Öneri Sistemler,Evrişimli Yapay Sinir Ağlar

ABSTRACT

CONVOLUTIONAL NEURAL NETWORK

BASED RECOMMENDATION SYSTEM DESIGN FOR FASHION CATEGORY

Personalization has become a key focus for businesses in recent years as they try to stand out in a crowded and competitive market. AI and machine learning-based recommendation engines can be a powerful tool for businesses to provide highly personalized customer experiences. By providing recommendations that are tailored to the specific needs and interests of each customer, businesses improve the customer experience and increase the likelihood of making a sale.With the acceleration of digitalization on a global scale, the e-commerce sector is gaining more power than ever before. This increases the competition for personalization in the e-commerce sector and allows the value of personalization technologies to increase. This study will propose a convolutional artificial neural network-based artificial learning system that will be able to offer suggestions to all product groups on the image on any product page. The artificial learning system consists of three stages, in the first stage object detection, in the second stage embedded representation learning, and in the third stage approximate nearest neighbor algorithms are used. During the experimentation process, in the first stage, the MaskRCNN architecture was trained on the iMaterialist dataset with different hyperparameters and the mAP values were compared for different iou thresholds. The most effective result was obtained by continuing the training using pretrained resnet 101 backbone. In the second stage, Resnet16, Vgg18, and VIT models were trained on the street-to-shop data set and similarity recommendations were made using the annoy nearest neighbor algorithm and precision@k and recall@k metrics were evaluated using different k values. As a result of the evaluation, the VIT architecture model obtained better results in all metrics compared to the other architectures.

Keywords: Recommender Systems,Visual Recommender Systems,Convolutional Neural Network

ÖNSÖZ

Tez sürecinde, bilgisi birikimi ve tecrübesi ile beni aydınlatan ve bu tezin oluşmasında bana destek veren tez danışmanım Prof. Dr. Ömer Önalan'a, hayatımın her aşamasında, sağlığım mutluluğum ve iyiliğim için olağanüstü çaba içerisinde olan aileme, bu süreçte beni yalnız bırakmayan arkadaşlarıma, son olarak üniversite ve bilim özgürlüğü için mücadele veren tüm akademisyenlere sonsuz teşekkür ederim.

MEHMET YİĞİT ÖZGENÇ

İstanbul,202

İÇİNDEKİLER

ÖZET I

ABSTRACT II

ÖNSÖZ III

İÇİNDEKİLER IV

KISALTMALAR VI

TABLOLAR LİSTESİ VII

ŞEKİLLER LİSTESİ VII

GRAFİKLER LİSTESİ IX

1 - GİRİŞ **1**

2 - TEMEL KAVRAMLAR **3**

2.1 - (Convolutional Neural Networks) Evrişimli Yapay Sinir Ağları 3

2.1.1 - Evrişim katmanı 4

2.1.2 - Havuzlama Katmanı 6

2.1.3 - Tam bağlantı katmanları 7

2.2 - Transformer 7

2.2.1 - Kodlayıcı kısmı 8

2.2.2 - Kod Çözücü Kısmı 8

2.3 -Görüntü Sınıflandırma 10

2.3.1 - VGG Net - (Visual Geometry Group) 11

2.3.2 - Residual Ağlar (Resnet) 12

2.3.3 - Görüntü Transformer (Vision Transformer) 14

2.4 - Nesne Segmentasyonu ile Nesne Tespiti 15

2.4.1 - Mask RCNN 17

2.5 - Yaklaşık En Yakın Komşu (YSA) 19

2.6 - Öneri Sistemleri 20

2.6.1 - İşbirlikçi Filtreleme Öneri Sistemleri 20

2.6.1.1 Hafıza Temelli İşbirlikçi Filtreleme 21

2.6.1.1.1 Ürün Temelli İşbirlikçi Filtreleme 21

2.6.1.1.2 Kullanıcı Temelli İşbirlikçi Filtreleme 21

2.6.1.2 - Model Temelli İşbirlikçi Filtreleme 21

2.6.2 - İçerik Tabanlı Öneri Sistemleri 21
2.6.3 Hibrid Öneri Sistemleri 22
2.7 - Literatür İncelemesi 22
3 - METODOLOJİ 26
3.1 - Çıkarım Modülü 27
3.2 Gömülü Temsil Öğrenim Modülü 28
3.3 - Benzerlik Modülü 30
4 - UYGULAMA 32
4.1 - Veri Setleri 32
4.1.1 - iMaterialist (Fashion) 2019 at FGVC6 32
4.1.2 - Sokaktan Dükkana Veri Seti (Exact Street2 Shop) 35
4.2 - Değerlendirme Metrikleri 39
4.2.1 - Nesne Tespiti 39
4.2.2 - Öneri Sistemi 40
4.3 Deney Sonuçları 41
4.3.1 - Gömülü Temsil Öğrenim Performans Deneyi 41
4.3.1.1 Eğitim Veri Kümesi 44
4.3.1.2 - Test Veri Kümesi 46
4.3.2 Nesne Tespiti Performans Deneyi 48
5 - SONUÇ 52
6 - KAYNAKLAR 53

KISALTMALAR

CNN: Evrişimli Yapay Sinir Ağları - Convolutional Neural Networks

ANN: Yapay Sinir Ağları - Artificial Neural Networks

RNN: Tekrarlayan Yapay Sinir Ağları - Recurrent Neural Networks

LSTM: Uzun Kısa Dönemli Bellek - Long Short Term Memory

ReLU : Doğrultulmuş Doğrusal Birim - Rectified Linear Unit

ILSVRC : ImageNet Büyük Ölçekli Görsel Tanıma Yarışması

MSE : Ortalama Karesel Hata - Mean - Squared Error

VGG : Visual Geometry Group

RESNET : Artık Sinir Ağı - Residual neural network

R-CNN: Regional Convolutional Neural Network - Bölge Tabanlı Konvolüsyonel Sinir Ağı

RPN - Bölge Öneri Ağı - Region Proposal Network

YSA - Yaklaşık En Yakın Komşu

TABLOLAR LİSTESİ

Tablo 2.3.2.1 ImageNet için Mimariler (He et al. 5) 14
Tablo 4.1.1.1 iMaterialist Eğitim Verisi Segment Dağılımı 34
Tablo 4.1.1.2 iMaterialist Validasyon Verisi Segment Dağılımı 34
Tablo 4.1.1.3 iMaterialist Test Verisi Segment Dağılımı 35
Tablo 4.1.2.1 Eğitim Verisi İçerisinde Kullanılan Sokak Görsellerinin Kategori Bazında Dağılımı 37
Tablo 4.1.2.2 Eğitim Verisi İçerisinde Kullanılan Ürünlerin Kategori Bazında Dağılımı 37
Tablo 4.1.2.3 Eğitim Verisi İçerisinde Kullanılan Sokak Fotoğraflarının Kategori Bazında Dağılım 38
Tablo.4.1.2.4 Eğitim Verisi İçerisinde Kullanılan Sokak Görsellerinin Kategori Bazında Dağılımı 38
Tablo 4.1.2.5 Test Verisi İçerisinde Kullanılan Ürünlerin Kategori Bazında Dağılımı 39
Tablo 4.1.2.6 Test Verisi İçerisinde Kullanılan Sokak Fotoğraflarının Kategori Bazında Dağılım 39
Tablo 4.3.1.1 Model ve İterasyonlara Göre Kayıp Değerleri 42
Tablo 4.3.1.1.1 Eğitim Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Precision Değerleri 44
Tablo 4.3.1.1.2 Eğitim Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Recall Değerleri 44
Tablo 4.3.1.1.3 Eğitim Veri Kümesi İçin Kategori Bazında Modellerin Precision Değerleri 45
Tablo 4.3.1.1.4 Eğitim Veri Kümesi İçin Kategorilerin Precision Değerleri 45
Tablo 4.3.1.2.1 Test Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Precision Değerleri 46
Tablo 4.3.1.2.2 Test Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Recall Değerleri 46
Tablo 4.3.1.2.3 Test Veri Kümesi İçin Kategori Bazında Modellerin Precision Değerleri 47
Tablo 4.3.1.2.4 Test Veri Kümesi İçin Kategorilerin Precision Değerleri 48

ŞEKİLLER LİSTESİ

Şekil 1.1 Evrişimli Yapay Sinir Ağı Katmanları 4
Şekil 1.1.1 Evrişim Operasyonu 5
Şekil 2.2 Transformer Mimarisi 8
Şekil 2.3.1 - VGG-Net (Davi Frossard) 12
Şekil 2.3.2 ImageNet için örnek ağ mimarileri (He et al. 4) 13
Şekil 2.3.3.1 Görüntü Transformer Mimarisi (Dosovitskiy et al. 3) 15
Şekil 2.4.0 Nesne Tespiti 16
Şekil 2.4.1.1 Mask RCNN Modeli Çalışma Prensibi 17
Şekil 2.6.1 En Yakın Komşu Çalışma Prensibi 19
Şekil 3.1.1 Önerilen Metodoloji 26
Şekil 3.2.1 CONVNet Mimarilerin Katman bölgelerinde öğrenimlerinin temsili 29
Şekil 3.2.2 İnce Ayar 30
Şekil 4.1.1 iMaterialist Veri Seti Girdi Ve Etiket Örnekleri 33
Şekil 4.1.2.1 Sokaktan Dükkana (Exact Street2 Shop) Gösterimi 36
Şekil 4.2.1.1 İşlem Oranı Metriği Görsel Anlatımı 40

GRAFİKLER LİSTESİ

Grafik 4.3.1.2 Model Kayıp Karşılaştırması Grafiği 43
Grafik 4.3.2.1 Model Kayıp Değeri Karşılaştırması 49
Grafik 4.3.2.2 mAP @ IoU=0.25 İçin Model Karşılaştırması 50
Grafik 4.3.2.3 mAP @ IoU=0.50 İçin Model Karşılaştırması 50
Grafik 4.3.2.4 mAP @ IoU=0.75 İçin Model Karşılaştırması 51

1 - GİRİŞ

"Doğuştan dijital" şirketler ve geleneksel şirketler arasındaki en önemli ayrım, insanlar, veri kümeleri veya hesaplama kaynakları değil, müşterilererine doğru ve eylemlerine göre öneri yapma konusunda gerçek zamanlı bir taahhüdü vermeleridir. Öneri sistemleri, kuruluşların verilerinden daha büyük değer elde etmeyi ve müşterilerine daha büyük değer yaratmayı nasıl sağlayabileceklerini düşünmelerini zorunlu hale getirir. Başka bir deyişle, platform perspektiflerine geçiş yapmak için mükemmel bir ortam ve mekanizmadır (Schrage 2017).

Araştırmacılar, öneri sistemlerinin bu ticaret platformlarında satışları artırdığını belirtmektedir öyle ki; Amazon'da tüketicilerin satın aldıklarının yüzde 35'inden fazlası ve Netflix'te izlediklerinin yüzde 60'ından fazlası önerilerden kaynaklanır (Magelssen 2021).

Öneri sistemleri yalnızca e-ticaret alanına sınırlamak bir hatadır. Youtube, Netflix ve Medium gibi hizmetler gösteriyor ki, öneri sistemleri müşterilere ürün ve hizmetler önermekle kalmaz, aynı zamanda içerik, tavsiye ve karar verme için bilgi sağlar. Öneri sistemleri, insanların öneri ve tavsiyeye ihtiyaç duydukları her alanda yararlı, kullanışlı ve kullanılabilir olabilir. Dijitalleşme ve verinin bol olduğu her alanda, öneri sistemlerinin gücü ve potansiyeli oldukça önem arz etmektedir.

Öneri sistemlerinin kullanım alanının genişliği ile birlikte, kullanılan teknolojilerin gelişmesi paralel doğrultuda ilerlemektedir. Günümüzde müşteri taleplerini karşılaması gereken her nokta, öneri sistemlerinin uygulanması için önemli bir potansiyel alan olarak işletmelerin karşısında durmaktadır.

Bu gelişmeler, öneri sistemlerinin en fazla etki gerçekleştirdiği e-ticaret gibi alanlarda ise, öneri sistemlerinin etkinliğinin daha farklı yönlere genişleyebilmesine olanak sağlamaktadır. Bu alanlardan birisi de ürün sayfasında bulunan fotoğraflara öneri yapabilme kabiliyetidir. İşletme tarafından müşteriye, ürün sayfasında gösterilen görseller sadece ilgili ürüne ait olmayabilir. Örneğin bir ceket ürün sayfasında, ürünün gösterildiği görselde bulunan manken farklı kıyafetler, örneğin ayakkabı, gömlek ve tişört giyebilir. Müşterinin bu noktada ilgisini, ilgili ürünün dışında farklı bir ürün çelebilir ve o ürüne veya benzerlerine ulaşmak isteyebilir. Müşteri gezintisine ve ürün özelliklerine dayalı klasik öneri sistemleri, müşterinin bu noktada ilgisini çekebilecek başka bir ürün grubuna öneri yapabilme kabiliyetine sahip değildir. Bu noktada ilgili görseli işleyebilecek,içerisindeki ürün gruplarını ayırabilecek ve her bir ürün grubuna benzer ürünleri önerebilecek bir sisteme ihtiyaç duyar.

Bu çalışma ilgili ihtiyacı karşılayacak ve benzer çalışmalarla kıyaslandığında oldukça az veri kümesinde eğitilip test edilecek, 3 aşamadan oluşan bir yapay öğrenme sistemi tasarlamayı

hedeflemektedir. Bu süreç içerisinde ilgili aşamalarda farklı mimariler veya kombinasyonlar denenerek, farklı mimarilerin ilgili görevi öğrenebilme kapasiteleri araştırılacaktır. Araştırma sonucunda en uygun modeller belirlenip, ilgili sistem için önerilecektir. Özellikle sadece evrişimli sinir ağları tabanlı yöntemleri değil, bununla birlikte nitekim yeni olarak adlandırılan Transformer mimarisine sahip bir yöntem de süreç içerisinde değerlendirilecektir.

2 - TEMEL KAVRAMLAR

2.1 - (Convolutional Neural Networks) Evrişimli Yapay Sinir Ağları

Evrişimli yapay sinir ağları, önemli bir yapay sinir ağı tekniğidir. Evrişimli yapay sinir ağları, diğer yapay sinir ağı mimarilerinin aksine problemin uzamsal yapısını (spatial information) korumak ile birlikte, yapısal olmayan verinin girdi olarak sağlandığı görevler için geliştirilmiştir. Özellikle görüntüyü tanıma, nesne algılama ve benzeri görevlerde oldukça etkili kullanılmaktadır.

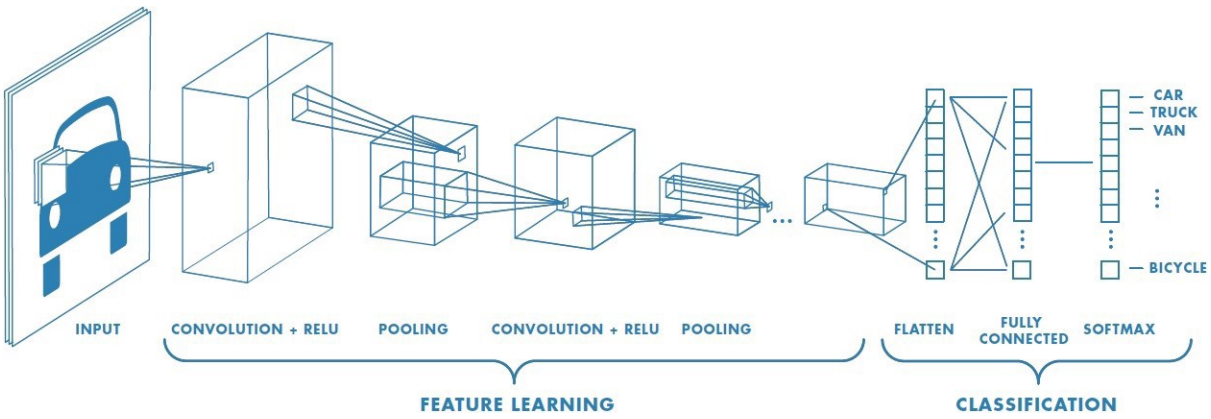
Evrişimli Yapay Sinir Ağları ile Geleneksel Yapay Sinir Ağları arasındaki en önemli dikkate değer fark, Evrişimli Yapay Sinir Ağlarının öncelikle görüntülerde örüntü tanıma alanında kullanılmasıdır. Bu, görüntüye özgü özellikleri mimariye kodlamamıza izin vererek ağı görüntü odaklı görevler için daha uygun hale getirirken modeli kurmak için gereken parametreleri daha da azaltır (O'Shea & Nash, 2015).

Bir diğer fark ise, CNN içindeki katmanların oluşturduğu nöronların, girdinin uzamsal boyutluluğu (yükseklik ve genişlik) ve derinlik olmak üzere üç boyutlu düzenlenmiş nöronlardan oluşmasıdır. Derinlik, yapay sinir ağı içindeki toplam katman sayısını değil, bir aktivasyon hacminin üçüncü boyutunu ifade eder. Standart yapay sinir ağlarından farklı olarak, herhangi bir katmandaki nöronlar, kendisinden önceki katmanın yalnızca küçük bir bölgesine bağlanır.

Evrişimli Yapay Sinir Ağları, görüntülerin özelliklerini keşfedebilmek için evrişim adı verilen bir işlemi kullanır. Bu işlem, girdi görüntüsünün özelliklerini bulmak için bir filtre kullanır. Filtre, girdi görüntüsünün bir parçasını alıp bu parçayı işler ve bu işlem sonucunda görüntünün bir özelliğini bulmaya yardımcı olur. Bu işlem tekrarlanarak görüntünün tüm özellikleri bulunur ve bu özellikler kullanılarak ilgili görevler gerçekleştirilir.

Evrişimli Yapay Sinir Ağları, boyutlandırılmış ve sıralı verileri işlemenin mimari olarak farklı bir yoludur. Girdideki verinin konumunun alakasız olduğunu varsaymak yerine (tamamen bağlı katmanların yaptığı gibi), evrişimli ve maksimum havuzlama katmanları ağırlık paylaşımını çevirisel olarak zorlar. Bu, insan görsel korteksinin çalışma şeklini modeller. Evrişimli yapay sinir ağlarının, nesne tanıma ve bir dizi başka görev için inanılmaz derecede iyi çalıştığı deneylerce kanıtlanmıştır.

CNN'ler üç tür katmandan oluşur. Bunlar evrişimli katmanlar, havuzlama katmanları ve tamamen bağlantılı katmanlardır. Bu katmanlar üst üste dizildiğinde bir CNN mimarisi oluşmuştur.



Şekil 1.1 Evrişimli Yapay Sinir Ağı Katmanları

Bir CNN mimarisi işlevi bakımından dört temel alana ayrılır.

1. Görüntünün piksel değerinin tutulduğu ve ANN'den farklı olarak uzlamsal boyut bilgisini de içeren girdi.
2. Evrişim katmanı, girdinin yerel bölgelerine bağlı olan filtre ağırlıkları ile girdi hacmine bağlı bölge arasındaki skaler çarpım, ve oluşan çıktının linear olmayan bir aktivasyon fonksiyona girdi olarak sağlanması.
3. Havuzlama katmanı, daha sonra verilen girdinin uzamsal boyutu boyunca basit bir şekilde altörnekleme gerçekleştirilmesi ve bu sayede aktivasyon içindeki parametre sayısının azaltılması.
4. Tam bağlantılı katmanlar, son aşamadaki özellik haritaları vektörel bir düzleme getirilir ve ANN katmanından geçirilerek sonuç katmanına iletilecek bilgi oluşturulur.

2.1.1 - Evrişim katmanı

Evrişimli yapay sinir ağının temelinde, evrişimli sinir ağlarına adını veren evrişim adı verilen katmanlar bulunur. Bu katman "konvolüsyon" veya evrişim adı verilen işlemi gerçekleştirir.

Konvolüsyon, parametreleri olan bir filtrenin, girdiye uygulanmasıyla birlikte aktivasyon işlemine girdi olarak sağlanmasıdır. Aynı filtrenin bir girdiye tekrar tekrar uygulanması, yapısal olmayan,örnek olarak görsel bir girdide, algılanan bir özelliğin konumunu ve önemini gösteren, özellik haritası adı verilen bir aktivasyon haritasıyla sonuçlanır. Öğrenilen filtre parametreleri görselin parçasına uygulanır.

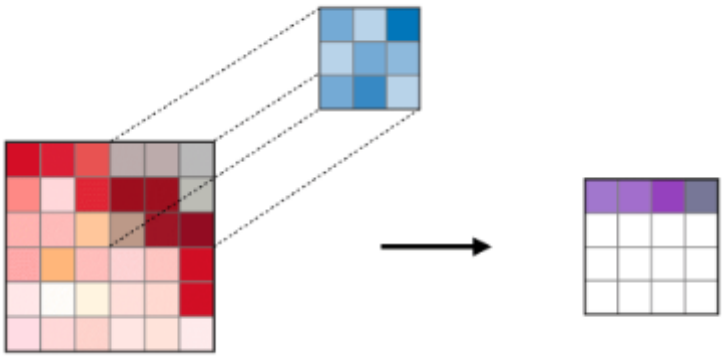
Evrişimli katmanımız tarafından takip edilen bir N×N kare nöron katmanında. ω bir m×m filtresi ise , evrişimli katman çıktısı (N−m+1)×(N−m+1) boyutunda olacaktır. Evrişim operasyon çıktısı X_{ij} şu şekilde hesaplanır;

$$x_{ij}^{l} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega_{ab} y_{(i+a)(j+b)}^{l-1}$$

Daha sonrasında linear olmayan bir fonksiyondan geçirilir;

$$y_{ij}^{l} = \sigma(x_{ij}^{l})$$

Filtre boyutu, girdi verisinin boyutundan daha küçüktür ve filtre görsel etrafında belirli bir adım alarak gezdirilir ve her işlem sürecinde girdi üzerinde bulunduğu kısım ile skaler çarpımı alınır. Spesifik olarak, filtre, soldan sağa, yukarıdan aşağıya, giriş verisinin her bir örtüşen parçasına uygulanır.



Şekil 1.1.1 Evrişim Operasyonu

Filtrenin, giriş verisinin her bir örtüşen parçasına uygulanması, görsel üzerinde var olan bir özelliğin, görselin her bir parçasında keşfedilebilmesine olanak sağlamaktadır, bu duruma "dönüşüm değişmezlik özelliği" adı verilir.

Bir özelliğin tam olarak nerede olduğundan çok var olup olmadığına daha çok önem veriliyorsa, dönüşüm değişmezlik özelliği çok yarar sağlamaktadır. Örneğin, bir görüntünün bir yüz içerip

içermediğini belirlerken, gözlerin konumunu mükemmel piksel doğruluğuyla bilmemiz gerekmez, sadece yüzün sol tarafında bir göz ve sağ tarafında bir göz olduğunu bilmemiz gerekir (Goodfellow et al., 2016).

Filtreyi giriş dizisiyle bir kez çarpmanın çıktısı tek bir değerdir. Filtre, giriş dizisine birden çok kez uygulandığında sonuç, girdinin filtrelenmesini temsil eden iki boyutlu bir çıktı değerleri dizisidir. Bu nedenle, bu işlemden elde edilen iki boyutlu çıktı dizisine "özellik haritası" denir.

2.1.2 - Havuzlama Katmanı

Havuzlama katmanları, temsilin boyutsallığını kademeli olarak azaltmayı ve böylece parametre sayısını ve modelin hesaplama karmaşıklığını daha da azaltmayı amaçlayan Evrişimli Sinir Ağları katmanıdır.

Evrişimli katmanların özellik haritası çıktısının bir eksi noktası, girdi özelliklerin kesin konumunu kaydetmeleridir. Bu, giriş görüntüsündeki özelliğin konumundaki küçük hareketlerin farklı bir özellik haritasıyla sonuçlanacağı anlamına gelir. Girdi görüntüsünde gerçekleşen kıpırdama, döndürme, kaydırma ve diğer küçük değişiklikler bu sonucu doğurabilmektedir.

Her durumda havuzlama, gösterimin girdinin küçük çevirilerine yaklaşık olarak değişmez hale gelmesine yardımcı olur. Çeviri değişmezliği, girdiyi küçük bir miktarda çevirirsek, havuzlanmış çıktıların çoğunun değerlerinin değişmeyeceği anlamına gelir (Goodfellow et al., 2016).

Havuzlama katmanı, evrişim katmanından sonra eklenen yeni bir katmandır. Spesifik olarak, bir evrişimli katman tarafından çıkarılan özellik haritalarına doğrusal olmayan bir fonksiyondan geçtikten sonra uygulanır. Havuzlama katmanı, aynı sayıda havuza alınmış özellik haritasından oluşan yeni bir set oluşturmak için her bir özellik haritası üzerinde ayrı ayrı çalışır. Havuzlama işleminin uygulandığı filtrenin boyutu, özellik haritasının boyutundan daha küçüktür.

Birkaç farklı türde havuzlama katmanı vardır, ancak en yaygın olanı, girdinin ilgili her bölgesinin maksimum değerini seçen maksimum havuzlama katmanıdır. Diğer havuzlama katmanı türleri arasında, ortalama değeri hesaplayan ortalama havuzlama ve girdi değerlerinin toplamını hesaplayan toplam havuzlama yer alır.

Havuzlama katmanı k×k bölgesini üzerinde işlem gerçekleştirip tek bir değer verir. Giriş katmanları bir N×N katmanıysa, her bir k×k bloğu örneğin maksimum işlevi aracılığıyla yalnızca tek bir değere indirgendiğinden, bir N/k×N/k katmanı çıkarırlar.

2.1.3 - Tam bağlantı katmanları

Tamamen bağlı bir katman, her nöronun bir ağırlık matrisi aracılığıyla giriş vektörüne doğrusal bir dönüşüm uyguladığı bir sinir ağını ifade eder. Sonuç olarak, katmandan katmana olası tüm bağlantılar mevcuttur, bu şekilde giriş vektörünün her bir düğümü, çıkış vektörünün her düğümünü etkiler.

Bir evrişimli sinir ağında (CNN), tamamen bağlı katmanlardan önce, evrişimli ve havuzlama katmanlarının yüksek boyutlu çıktısını, tamamen bağlı katmanlara girilebilecek tek boyutlu bir vektöre dönüştüren bir düzleştirme katmanı gelir. Bu vektör daha sonra, giriş verileri için nihai bir tahmin veya sınıflandırma çıktısı oluşturmak üzere tamamen bağlı katmanlar tarafından işlenir.

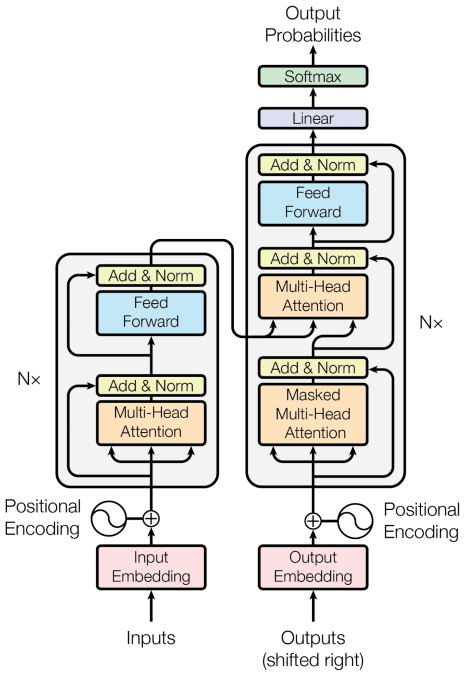
2.2 - Transformer

Bu çalışmada, tekrardan kaçınan ve bunun yerine girdi ve çıktı arasında küresel bağımlılıklar çizmek için tamamen bir dikkat mekanizmasına dayanan bir model mimarisi olan Transformer'ı öneriyoruz. Transformer, önemli ölçüde daha fazla paralelleştirmeye izin verir … Transformer, dizi hizalı RNN'ler veya evrişim kullanmadan girdi ve çıktılarının temsillerini hesaplamak için tamamen kendi dikkatine dayanan ilk transdüksiyon modelidir (Vaswani et al., 2017).

Dikkat mekanizmalarını kullanan transformer mimarisi, Uzun Kısa Süreli Bellek ağları (LSTM) ve diğer Tekrarlayan Sinir Ağları(RNN) modellerine alternatif olarak önerilmiştir. Başlangıçta dil modelleri üzerine önerilen mimari, sonrasında zaman serisi, görüntü işleme vs gibi alanlarda da aktif olarak kullanılmaya başlanmış ve etkili sonuçlar alınmıştır. Mimari içerisindeki dikkat mekanizması, bir girdi dizisinde modelin hangi girdi maddesine ne kadar ve nasıl odaklanması gerektiğine karar verir.

Transformer mimarisi kodlayıcı(Encoder) ve kod çözücü(Decoder) olmak üzere iki kısma ayrılır. Transformer mimarisinin ilk yarısındaki kodlayıcının görevi, bir girdi dizisini, daha sonra bir kod çözücüye beslenen sürekli temsiller dizisine eşlemektir. Mimarinin diğer yarısında bulunan kod çözücü, bir çıkış dizisi oluşturmak için kodlayıcının önceki zaman adımındaki tahminini kod çözücünün temsiliyle birlikte alır.

Her adımda model, bir sonrakini oluştururken ek girdi olarak önceden oluşturulmuş sembolleri tüketerek otomatik gerilemelidir (Vaswani et al., 2017).



Şekil 2.2 Transformer Mimarisi

2.2.1 - Kodlayıcı kısmı

Kodlayıcı, her katmanın iki alt katmandan oluştuğu 6 katman yığınından oluşur.

İlk alt katman, çok başlı bir öz-dikkat mekanizması(Multi-Head Attention) uygular. Çok başlı öz-dikkat mekanizması, birden fazla dikkat mekanizmasının paralel olarak çalışmasıyla oluşur,her bir dikkat mekanizmasının çıktısı lineer olarak birleştirilir ve istenen boyuta getirilir. Sezgisel olarak, birden fazla dikkat başlığı, sekansın bölümlerine farklı şekilde odaklanmaya izin verir.

İkinci alt katman, aşağıdakiler arasında Doğrultulmuş Doğrusal Birim (ReLU) aktivasyonu ile iki doğrusal dönüşümden oluşan tamamen bağlantılı bir ileri beslemeli ağdır.

Model herhangi bir sıralama mekanizmasına sahip olmaması nedeniyle konumsal bilgiyi kullanabilmek için konumsal kodlama mekanizmasını kullanır. Konumsal kodlama vektörleri, giriş katıştırmalarıyla aynı boyuttadır ve farklı frekansların sinüs ve kosinüs fonksiyonları kullanılarak üretilir. Daha sonra, konum bilgisini enjekte etmek için basitçe giriş katıştırmalarına toplanırlar.

2.2.2 - Kod Çözücü Kısmı

Kod çözücü her biri üç alt katmandan oluşan 6 aynı katman yığınından oluşur:

Ilk alt katman, önceki çıkışını kod çözücüsü yığınından alır, pozisyonel bilgilerle birlikte geliştirir ve üzerinde çok başlı bir öz-dikkat mekanizması uygular. Kodlayıcı, dizinin hangi pozisyonunda olursa olsun tüm girdilere dikkat etmeyi hedeflerken, çözücü sadece önceki girdilere dikkat etmek üzerine

odaklandırılmıştır. Bu nedenle, i. konumdaki bir girdinin tahmini sadece dizide ondan önce gelen girdilerin bilinen çıktılarına bağlı olabilir. Çok kafa dikkat mekanizmasında (paralel olarak birden fazla tek dikkat işlevini gerçekleştirir), bu, Q ve K matrislerinin ölçeklendirilmiş çarpımının ürettiği değerleri maskeleyerek sağlanır.

Maskeleme, çözücüyü tek yönlü yapar (çift yönlü kodlayıcıya göre değil) (Vasilev, 2017).

İkinci katman, birinci kodlayıcı alt katmanında gerçekleştirilen benzer bir çok başlı bir öz-dikkat mekanizması gerçekleştirir. Ancak, çözücü tarafında, bu mekanizma, önceki çözücü alt katmanından sorguları ve kodlayıcının çıktısından anahtarları ve değerleri alır. Bu, çözücünün bir kez bir girdiden oluşan çıktı dizisini oluşturmaya yönelik olmasına rağmen, tüm girdi dizisindeki girdilere dikkat etmesine izin verir.

Bu çok başlı bir öz-dikkat mekanizması kullanarak, çözücü tahminlerine girdi dizisinin tamamını dahil edebilir ve daha doğru ve akıcı çıktı dizileri oluşturabilir. Bu, Transformer modelinin önemli bir bileşenidir, uzun girdi dizileri için bile yüksek kaliteli çıktılar oluşturmaya yardımcı olur.

Bir Transformer modelindeki kod çözücünün üçüncü katmanı, kodlayıcının ikinci alt katmanında kullanılan katmana benzer şekilde tamamen bağlı bir ileri beslemeli ağ uygular. Bu ağ, önceki kod çözücü alt katmanının çıktısını girdi olarak alır ve katmanın nihai çıktısını üretmek için bir dizi doğrusal dönüşüm, doğrusal olmayan aktivasyon fonksiyonu uygular.

Bu tamamen bağlantılı ileri beslemeli ağ, önceki alt katmanların çıktısını daha da iyileştirmeye ve işlemeye yardımcı olarak kod çözücünün daha doğru ve akıcı çıktı dizileri oluşturmasına olanak tanır. İkinci katmandaki çok kafalı öz dikkat mekanizması gibi, bu ağ da Transformer modelinin model çıktısının kalitesini artırmaya yardımcı olan önemli bir bileşenidir.

Yukarıda açıklanan üç alt katmana ek olarak, bir Transformer modelindeki kod çözücü, bu alt katmanların çevresinde, modelin giriş ve çıkış dizilerindeki uzun menzilli bağımlılıkları daha iyi yakalamasına izin veren artık bağlantılara da sahiptir. Bu artık bağlantıları, eğitim sürecini stabilize etmeye ve modelin performansını iyileştirmeye yardımcı olan bir normalleştirme katmanı takip eder.

Kodlayıcı gibi, kod çözücü de giriş yerleştirmelerinin bir parçası olarak konumsal kodlamaları alır. Bu konumsal kodlamalar, modele girdi ve çıktı dizilerindeki her kelimenin göreli konumu hakkında ek bilgi sağlayarak, modelin daha doğru ve akıcı çıktılar üretmesini sağlar.

Artık bağlantılar ve normalleştirme katmanıyla birlikte kod çözücünün üç alt katmanı, Transformer modelinin giriş dizisine dayalı olarak yüksek kaliteli çıktı dizileri üretmesini sağlamak için birlikte çalışır. Bu bileşenler, modelin girdi ve çıktı dizileri arasındaki ilişkiyi etkili bir şekilde öğrenmesi ve doğru ve akıcı çıktılar üretmesi için gereklidir.

2.3 -Görüntü Sınıflandırma

Görüntü sınıflandırma, bir görüntüye içeriğine göre bir etiket veya sınıf atama işlemidir. Örneğin, köpekler, kediler ve kuşlar gibi farklı hayvan türlerini tanımak için bir görüntü sınıflandırma algoritması eğitilebilir. Algoritmaya bir dizi etiketli hayvan resmi verilir ve bunları her bir hayvan sınıfının hangi özelliklerinin karakteristik olduğunu öğrenmek için kullanılır. Ardından, yeni bir görüntü sunulduğunda, algoritma, görüntünün hangi sınıfa ait olduğunu belirlemek için eğitilmiş bilgisini kullanır.

ILSVRC (ImageNet Büyük Ölçekli Görsel Tanıma Yarışması), dünyanın dört bir yanından ekiplerin en doğru görüntü sınıflandırıcıları geliştirmek için yarıştığı yıllık bir yarışmadır. Yarışma, görüntü sınıflandırma algoritmalarının eğitiminde ve değerlendirilmesinde kullanılmak üzere geniş bir etiketli görüntüler veritabanı tutan ImageNet konsorsiyumu tarafından organize edilmektedir.

ILSVRC'de, katılımcılara bir görüntü veri seti verilir ve görüntüleri 1000 farklı sınıftan birine doğru bir şekilde sınıflandırma kabiliyetine sahip algoritmalar geliştirme görevi verilir. Sınıflar, hayvanlar, binalar ve doğal afetler gibi çeşitli nesneleri, sahneleri ve olayları içerir. Yarışmanın amacı, veri setindeki görüntüleri sınıflandırmada yüksek düzeyde doğruluk elde edebilen algoritmalar geliştirmektir.

ILSVRC' yarışmaları içerisinde, Evrişimli Sinir Ağları oldukça popülerdir. Özellikle yarışmanın ilk döneminde klasik yapay öğrenme algoritmaları yarışmanın popüler yöntemi iken, Evrişimli Sinir Ağlarının kullanımı ile çok ciddi başarılar elde edilmiştir.

ILSVRC, görüntü sınıflandırmasında en son teknolojinin ilerlemesinde önemli bir rol oynamış ve bu görev için etkili evrişimli sinir ağı mimarilerinin geliştirilmesine yardımcı olmuştur.

Görüntü sınıflandırma görevleri için kullanılabilecek birkaç kayıp(Loss) fonksiyonu vardır. Bu tür görevler için bazı yaygın kayıp(Loss) fonksiyonları, çapraz entropi kaybı, hinge kaybı ve ortalama karesel hata (MSE) kaybını içerir. Çapraz entropi kaybı, sınıf etiketleri için tahmin edilen olasılık dağılımı ile gerçek olasılık dağılımı arasındaki farkı ölçtüğü için görüntü sınıflandırma görevleri için yaygın bir seçimdir. Hinge kaybı ikili sınıflandırma görevleri için kullanılır ve tahmin edilen sınıf etiketi ile gerçek sınıf etiketi arasındaki farkı ölçer. MSE kaybı, tahmin edilen sınıf etiketi ile gerçek sınıf etiketi arasındaki farkın bir ölçüsüdür ve regresyon görevleri için kullanılır. Görüntü sınıflandırması için kayıp fonksiyonunun seçimi, görevin belirli özelliklerine ve kullanılan modele bağlıdır.

Üçlü kayıp(Triplet Loss), yüz tanıma ve kişi yeniden tanımlama gibi görevler için derin öğrenme modellerinin eğitiminde yaygın olarak kullanılan bir tür kayıp işlevidir. Üçlü kayıp işlevi, benzer veri

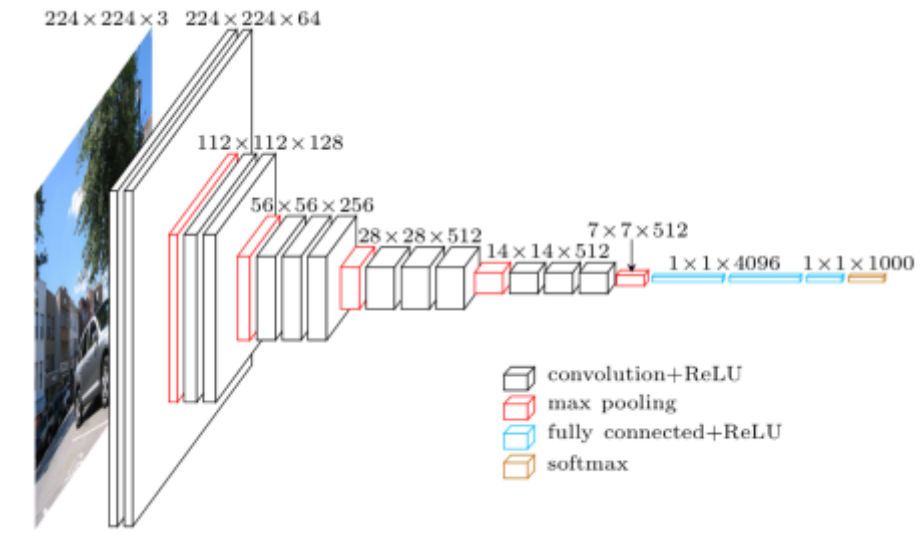
noktalarının birbirine daha yakın olduğu, benzer olmayan veri noktalarının ise daha uzak olduğu bir özellik alanına (görüntüler gibi) giriş veri noktalarını eşleyen bir model öğrenme fikrine dayanır.

Üçlü kayıp fonksiyonu girdi olarak, bir referans veri noktası, bir pozitif veri noktası (referans veri noktasına benzer) ve bir negatif veri noktasından (referans veri noktasına benzemeyen) oluşan bir üçlü veri noktası alır. Üçlü kayıp fonksiyonunun amacı, referans veri noktasını ve pozitif veri noktasını özellik uzayında benzer bir konuma eşleyen ve aynı zamanda referans veri noktasını ve negatif veri noktasını daha uzak bir konuma eşleyen bir model öğrenmektir. Bu, modelin veri noktaları arasındaki benzerlik ve benzemezlik ilişkilerini koruyan verilerin bir temsilini öğrenmesini sağlar. Üçlü kayıp işlevi, yüz tanıma ve kişi yeniden tanımlama gibi görevler için derin öğrenme modellerini eğitmek üzere genellikle diğer kayıp işlevleriyle birlikte kullanılır.

2.3.1 - VGG Net - (Visual Geometry Group)

VGG, Oxford'daki Visual Geometry Group tarafından geliştirilen, görüntü sınıflandırması için evrişimli bir sinir ağı mimarisidir. Basitliği ve küçük (3x3) evrişim filtrelerinin kullanımı ile karakterize edilir. VGG, 2014 yılında ImageNet Büyük Ölçekli Görsel Tanıma Yarışmasını (ILSVRC) kazanmak için geliştirimiştir ve geliştirildiği dönem içerisinde, nesne tanıma görevinde in iyi performansı elde etmiştir. Basitliğine rağmen, VGG'nin çeşitli görüntü sınıflandırma görevleri için etkili olduğu gösterilmiştir. (Simonyan and Zisserman)

VGG mimarisi ile LeNet-5 ve AlexNet gibi önceki evrişimli sinir ağı mimarileri arasındaki temel farklardan biri, çok sayıda küçük filtrenin kullanılmasıdır. VGG, 1 adımlı 3x3 filtreler kullanırken, LeNet-5 daha büyük filtreler kullanır ve AlexNet, 4 adımlı daha küçük ama yine de nispeten büyük filtreler kullanır.



Şekil 2.3.1 - VGG-Net (Davi Frossard)

Küçük filtrelerin kullanılması, VGG mimarisinin, görsel sınıflandırma gibi görevler için önemli olan girdi görüntülerinden daha ayrıntılı uzamsal bilgi yakalamasına olanak tanır. Küçük filtreler ayrıca ağda daha fazla filtre kullanılmasına izin vererek modelin kapasitesini ve performansını artırabilir.

Genel olarak, küçük filtrelerin kullanılması, VGG ile önceki evrişimli sinir ağı mimarileri arasındaki temel farklardan biridir ve görüntü sınıflandırma görevleri için etkili bir tasarım seçeneği olduğu gösterilmiştir.

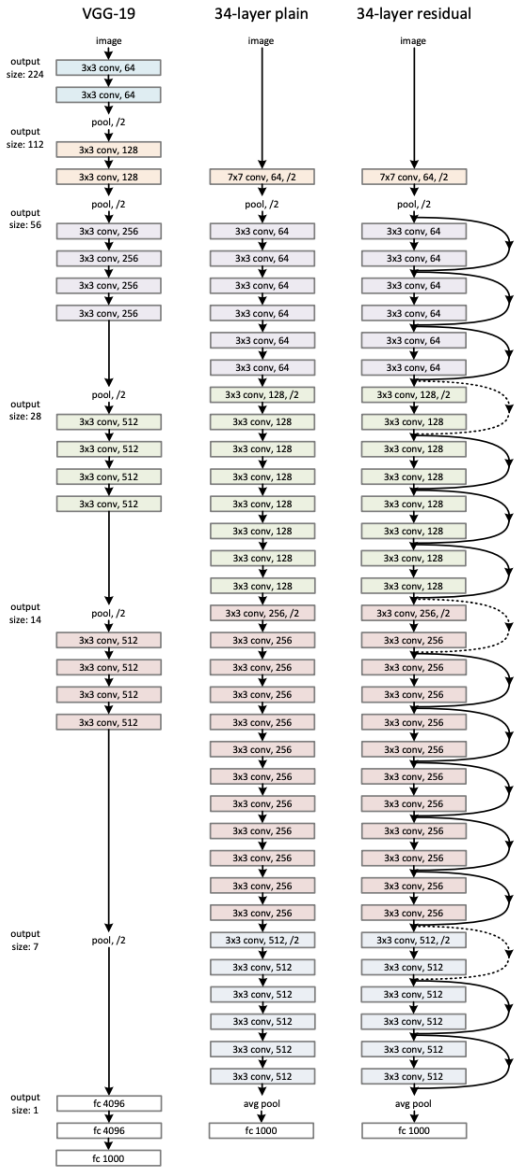
VGG ağı, evrişimli katmanlarında çok sayıda filtre kullanması ile bilinir. Filtre sayısı modelin derinliği ile birlikte artmakta, 64'ten başlayıp modelin öznitelik çıkarma kısmının sonunda 128, 256 ve 512'ye kadar çıkmaktadır. Çok sayıda filtrenin kullanılması, VGG ağının giriş görüntüsünden daha çeşitli özellikler öğrenmesini sağlar. Bu, modelin görüntüdeki nüansları ve ayrıntıları daha fazla yakalamasına izin verdiği için görüntü sınıflandırma görevinde modelin performansını arttırmaktadır.

VGG mimarisinin, katman sayısı ve ağ tasarımının belirli ayrıntıları bakımından farklılık gösteren bir dizi çeşidi vardır. En çok atıfta bulunulan iki varyant, sırasıyla 16 ve 19 öğrenilmiş katmana sahip olan VGG-16 ve VGG-19'dur.

VGG mimarisinin bu varyantları, performansları ve derinliklerine göre geliştirilip ve değerlendirilmiştir. VGG-16 ve VGG-19, VGG mimarisinin en iyi performans gösteren çeşitleri arasında kabul edilir ve çeşitli görüntü sınıflandırma görevlerinde yaygın olarak kullanılmaktadır.

2.3.2 - Residual Ağlar (Resnet)

Referanssız fonksiyonları öğrenmek yerine girdi katmanına referansla artık fonksiyonları öğrenmek için eğitilmiş bir evrişimli sinir ağı (CNN) türüdür. Bu, ResNet modellerinin geleneksel CNN'ler kullanılarak mümkün olandan çok daha fazla katmana sahip olmasını sağlar ve onları görüntü tanıma gibi görevler için çok daha etkili hale getirir.



Şekil 2.3.2 ImageNet için örnek ağ mimarileri (He et al. 4)

ResNet modellerinin çeşitli görevlerde diğer son teknoloji CNN modellerinden daha iyi performans gösterdiği ve birçok makine öğrenimi araştırmacısı ve uygulayıcısı için popüler bir seçim haline geldiği görülmüştür.

ResNet modelinde artık blokların ve kısayol bağlantılarının kullanılması, geleneksel bir CNN ile mümkün olandan çok daha derin bir ağ mimarisine sahip olmasını sağlar. Bu derinlik, modelin verilerdeki daha karmaşık kalıpları öğrenmesine ve daha doğru tahminler yapmasına olanak tanır. ReLU aktivasyonunun ve 1x1 kıvrımların kullanılması da modelin performansını iyileştirmeye yardımcı olur.

Tüm ResNet modellerinde, havuzlama boyutu 3x3 olan ve ilk katmandan sonra uygulanan adım 2 olan yalnızca bir maksimum havuzlama katmanı vardır. Bu, girişin çözünürlüğünü azaltmaya yardımcı olur, ancak diğer CNN mimarilerine kıyasla sınırlıdır. Tek bir maksimum havuzlama katmanının ve sınırlı bir adımın kullanılması, görüntü tanıma gibi görevler için faydalı olabilecek girdi verilerinde daha fazla uzamsal bilginin korunmasını sağlar. Ayrıca model karmaşıklığını azaltır ve performansı arttırır.

ResNet modellerinde, ortalama havuzlama katmanı, ağın sonundaki tamamen bağlı katmanların yerini almak için kullanılır. Bunun birkaç avantajı vardır. Birincisi, bu katmanda optimize edilecek herhangi bir parametre olmadığı için model karmaşıklığını azaltır. İkinci olarak, ortalama havuzlama katmanı, özellik haritaları ile kategoriler arasındaki ilişkileri daha iyi uygulamakta ve bu da modelin performansını arttırmaktadır.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Tablo 2.3.2.1 ImageNet için Mimariler (He et al. 5)

Yukarıdaki tablo, ImageNet veri setini sınıflandırmak için kullanılan farklı ResNet mimarilerini göstermektedir. Katman sayısı 18 ile 152 arasında değişmektedir. Kalan bloklar iki veya üç katmana sahiptir.

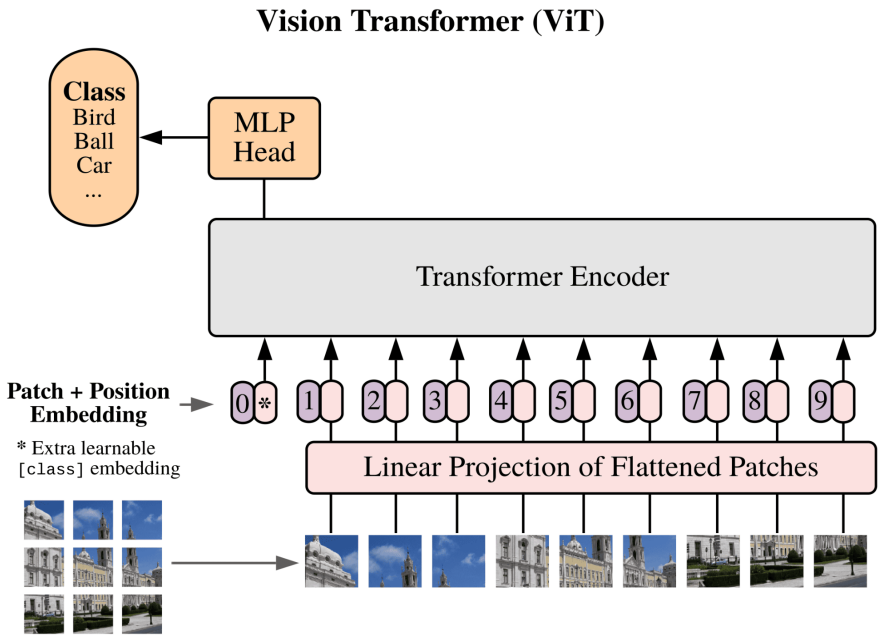
2.3.3 - Görüntü Transformer (Vision Transformer)

Vision Transformer, görüntü tanıma ve nesne algılama gibi görsel verileri içeren görevler için tasarlanmış bir tür derin öğrenme modelidir. Sıralı verileri işlemek için kendi kendine dikkat mekanizmalarını kullanan bir tür sinir ağı olan transformer mimarisine dayanır. Bir Vision Transformer'da, görsel verilere kişisel dikkat mekanizması uygulanarak modelin girdi görüntüsünün farklı bölgeleri arasındaki ilişkileri öğrenmesine olanak sağlanır. Bu, Vision Transformers'ı bir görüntüdeki nesneler arasındaki uzamsal ilişkileri anlama yeteneği gerektiren görevler için çok uygun hale getirir.

Standart Transformer modeli başlangıçta doğal dil işleme görevleri için tasarlanmıştır ve bu nedenle girdi olarak tek boyutlu bir kelime gömme dizisi alır. Bunun aksine, görüntü sınıflandırma görevine uygulandığında, Transformer modeline girdi verileri iki boyutlu görüntüler biçiminde sağlanır. Bu, giriş görüntüsünün farklı bölgeleri arasındaki uzamsal ilişkileri öğrenmek için kendi kendine dikkat mekanizmalarının kullanılması gibi, görsel verileri işlemeye uygun hale getirmek için standart Transformer mimarisinde bazı değişiklikler gerektirir.

Sınıflandırma kafası, ön eğitim zamanında bir gizli katmana sahip bir MLP tarafından ve ince ayar zamanında tek bir doğrusal katman tarafından uygulanır (Dosovitskiy et al., 2021).

Vision Transformer (ViT) modeli, giriş verilerini işlemek için orijinal Transformer mimarisinin kodlayıcı kısmını kullanır. Kodlayıcının girdisi, konum bilgisi ve öğrenilebilir bir sınıf yerleştirme ile zenginleştirilmiş bir dizi gömülü görüntü yamalarıdır. Kodlayıcının çıktısına eklenen sınıflandırma kafası, öğrenilebilir sınıf gömme değerini alır ve durumuna göre bir sınıflandırma çıktısı üretir. Bu, ViT modelinin girdi görüntüsünün farklı bölgeleri arasındaki ilişkileri öğrenmesine ve görüntüdeki nesnelerin uzamsal düzenlemesine dayalı tahminler yapmasına olanak tanır.

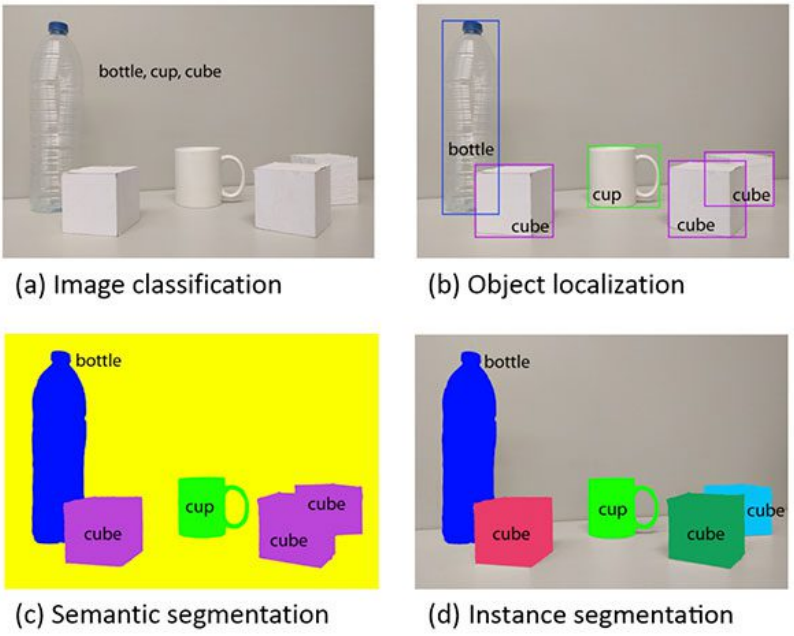


Şekil 2.3.3.1 Görüntü Transformer Mimarisi (Dosovitskiy et al. 3)

2.4 - Nesne Segmentasyonu ile Nesne Tespiti

Nesne bölütleme ile nesne algılama, bir görüntüdeki nesneleri tanımlamayı ve yerleştirmeyi ve ayrıca sınırlarını vurgulamak için nesneleri bölümlere ayırmayı içeren bir bilgisayar görüşü tekniğidir. Bu,

bir görüntüdeki nesnelerin daha doğru bir şekilde tanımlanmasını ve konumlandırılmasını sağlar ve görüntü sınıflandırması ve nesne tanıma gibi görevler için yararlı olur. Nesne bölümleme ile nesne algılama, giriş görüntülerini işlemek ve nesne algılama ve bölümleme sonuçlarını oluşturmak için tipik olarak evrişimli sinir ağları (CNN'ler) gibi derin öğrenme algoritmalarının kullanımını içerir.



Şekil 2.4.0 Nesne Tespiti

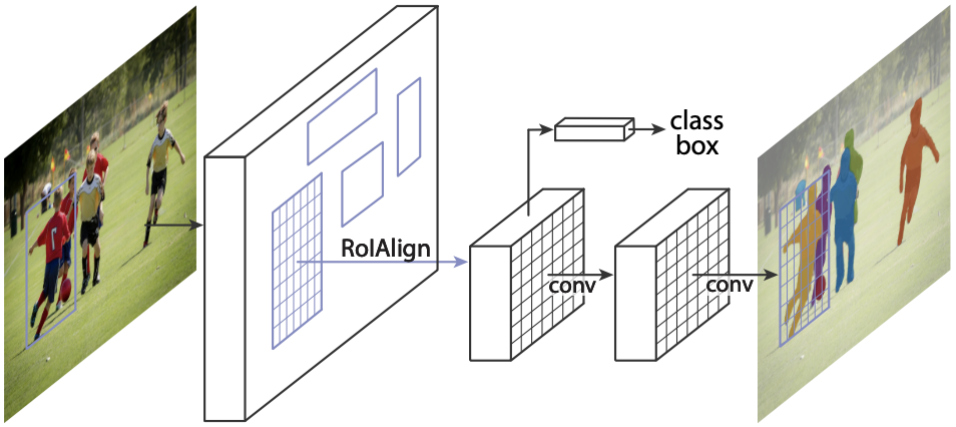
Görüntü sınıflandırma, bir görüntüdeki bir nesnenin sınıfını tahmin etmeyi içerir. Nesne bölütleme, bir görüntüdeki bir veya daha fazla nesnenin konumunun belirlenmesi ve kapsamları etrafında çok sayıda kutunun çizilmesi anlamına gelir. Nesne algılama, bu iki görevi birleştirir ve bir görüntüdeki bir veya daha fazla nesneyi bulur ve sınıflandırır.

Nesne algılamanın bir uzantısı, nesne segmentasyonu sırasında kaba sınırlayıcı kutular kullanmak yerine görüntüde algılanan her nesneye ait belirli piksellerin işaretlenmesini içerir. Problemin bu daha zor versiyonuna nesne bölütleme veya semantik bölütleme denir.

... hem görüntü sınıflandırmasını (görüntüde hangi nesne sınıflarının bulunduğunu belirlemek için bir algoritma gerektiren bir görev) hem de nesne algılamayı (görüntüde bulunan tüm nesneleri yerelleştirmek için bir algoritma gerektiren bir görev) kapsayacak şekilde geniş bir şekilde nesne tanıma terimini kullanacağız (Russakovsky et al., 2015).

2.4.1 - Mask RCNN

Mask R-CNN (Bölge Tabanlı Konvolüsyonel Sinir Ağı), bir görüntüdeki tek tek nesneleri tanımlamayı ve bölümlere ayırmayı içeren, bölümleme görevi için tasarlanmış bir derin öğrenme modelidir. Son teknoloji nesne algılama modeli olan Faster R-CNN modelinin bir uzantısıdır. Mask R-CNN modeli, Faster R-CNN mimarisine, nesne bölütleme kutusuna ek olarak bir nesne maskesini tahmin etmekten sorumlu olan paralel bir dal ekler. Bu, modelin bir görüntüdeki nesneleri bölümlere ayırmasına ve sınırlarını vurgulamasına olanak tanır; bu, görüntü sınıflandırma ve nesne tanıma gibi görevler için etkin olarak kullanılır.



Şekil 2.4.1.1 Mask RCNN Modeli Çalışma Prensibi

Sınırlayıcı kutu nesne algılamaya yönelik Bölge tabanlı CNN (R-CNN) yaklaşımı, yönetilebilir sayıda aday nesne bölgesine katılmak ve evrişimli ağları her bir RoI'de bağımsız olarak değerlendirmektir. R-CNN, RoIPool kullanarak özellik haritalarında RoI'lere katılmaya izin verecek şekilde genişletildi, bu da yüksek hız ve daha iyi doğruluk sağlıyor. Daha hızlı R-CNN, bir Bölge Öneri Ağı (RPN) ile dikkat mekanizmasını öğrenerek bu akışı geliştirdi. Daha hızlı R-CNN esnektir ve takip eden pek çok iyileştirmeye karşı dayanıklıdır ve birçok kıyaslamada mevcut lider çerçevedir (He et al., 2016).

Mask R-CNN modeli dört ana bileşenden oluşur: bir öznitelik çıkarıcı, bir bölge önerme ağı, bir sınıflandırma ve regresyon ağı ve bir maske tahmin ağı.

Öznitelik çıkarıcı, girdi görüntüsünü işleyen ve ondan üst düzey özellikleri çıkaran evrişimli bir sinir ağıdır.

Bölge öneri ağı (RPN), özellik çıkarıcının çıktısını alır ve girdi görüntüsündeki nesneler içerebilecek bölgeler için bir dizi teklif oluşturur. Daha sonra sınıflandırma ve regresyon ağı, her nesnenin sınıfını ve her nesne için sınırlayıcı kutu koordinatlarını tahmin etmek için bu önerileri kullanır.

Son olarak, maske tahmin ağı, bölge önerme ağının çıktılarını alır ve giriş görüntüsündeki nesneleri bölümlere ayırmak için kullanılan her nesne için maskeyi tahmin eder.

Mask R-CNN modelinde kayıp fonksiyonu, modelin performansını ölçmek için kullanılır. Mask R-CNN için, kayıp fonksiyonu üç farklı fonksiyonun birleşimi olacaktır: kutu regresyon kaybı, maske segmentasyon kaybı ve sınıflandırma kaybı.

Kutu regresyon kaybı, modelin nesneleri için sıkı kutular çizme konusunda ne kadar iyi olduğunu ölçerken, maske segmentasyon kaybı, bu kutular içindeki nesnelerin şeklinin nasıl belirlendiğini ölçer.

Resmi olarak, eğitim sırasında çoklu görev kaybı tanımlarız.Kayıp fonksiyonu L = Lcls + Lbox + Lmask olarak tanımlanır (He et al., 2017).

Bölge Öneri Ağı Sınıf Kaybı (Rpn Class Loss): RPN (Bölge Öneri Ağı) için sınıflandırma kaybını ifade eder, bu da bir nesnenin bulunabileceği aday bölgeleri önerir. RPN için sınıflandırma kaybı, RPN'nin bir görüntüdeki bir bölgede bir nesnenin olup olmadığını belirleme konusundaki başarısını değerlendirme için kullanılır. Bölge Öneri Ağının, arka planı nesnelerle ne kadar iyi ayırdığını ifade eder.

Bölge Öneri Ağı Tespit Kutusu Kaybı (Rpn Bbox Loss):RPN (Bölge Öneri Ağı) için kutu regresyon kaybını ifade eder. RPN, bir nesnenin bulunabileceği aday bölgeleri önerir ve bu kayıp, RPN'nin bir görüntüdeki bir bölgede bir nesnenin sınırlarını nasıl çizdiğini değerlendirme için kullanılır.

Mask R-CNN Tespit Kutusu Kaybı (Mrcnn Bbox Loss):Mask R-CNN için kutu regresyon kaybını ifade eder. Bu kayıp, modelin bir görüntüdeki nesneleri nasıl sınırlandırdığını değerlendirme için kullanılır

Mask R-CNN Sınıf Kaybı (Mrcnn Class Loss):Mask R-CNN için sınıflandırma kaybını ifade eder. Bu kayıp, modelin bir görüntüdeki nesneleri nasıl sınıflandırdığını değerlendirme için kullanılır.Mask RCNN'nin her bir nesne sınıfını ne kadar iyi tanıdığı belirtir.

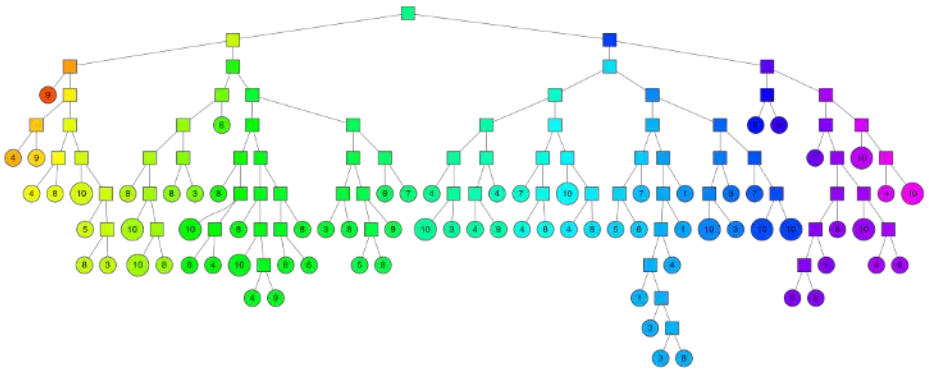
Mask R-CNN Maske Kaybı(Mrcnn Mask Loss) : Mask R-CNN modeli için maske segmentasyon kaybını ifade eder. Bu kayıp, modelin kutular içindeki nesnelerin şeklini nasıl belirlediğini değerlendirme için kullanılır. Mask RCNN nesneleri ne kadar iyi segmente ediyor sorusunun cevabını verir.

kayıp : Tüm küçük kayıpların lineer kombinasyonu

2.5 - Yaklaşık En Yakın Komşu (YSA)

Yaklaşık en yakın komşu (YSA), yüksek boyutlu uzaylarda verim sağlayan en yakın komşu arama tekniğidir. Bir sorgulama noktasının tam olarak en yakın komşusunun gerekli olmadığı, bunun yerine gerçek en yakın komşuya yeterince yakın olan başka bir en yakın komşunun gerekli olduğu uygulamalarda yaygın olarak kullanılır.

YSA algoritmaları, en yakın komşuya yaklaşmak için yerelliğe duyarlı karma (LSH), k-d ağaçları ve rastgele projeksiyonlar gibi çeşitli teknikler kullanır. Bu teknikler, YSA algoritmalarının yaklaşık en yakın komşuları, yüksek boyutlu uzaylarda hesaplama açısından pahalı olabilen tam en yakın komşu arama algoritmalarından daha verimli bir şekilde aramasına izin verir. YSA algoritmaları, görüntü arama, belge benzerlik arama ve öneri sistemleri gibi uygulamalarda yaygın olarak kullanılmaktadır.



Şekil 2.6.1 En Yakın Komşu Çalışma Prensibi

Yaklaşık En Yakın Komşu teknikleri, verileri verimli bir dizine önceden işleyerek aramayı hızlandırır. Farklı çeşitleri olmasına karşın genellikle 3 aşamadan oluşur.

1. Vektör Dönüşümü — vektörler indekslenmeden önce uygulanır, bunların arasında boyut küçültme ve çeşitli vektör transform teknikleri vardır.
2. Vektör Kodlama — arama için gerçek dizini oluşturmak amacıyla vektörlere uygulanır, bunların arasında Ağaçlar, LSH gibi veri yapısına dayalı teknikler mevcuttur.
3. Kapsamlı Olmayan Arama Bileşeni — Kapsamlı aramadan kaçınmak için vektörler üzerinde uygulanır.

2.6 - Öneri Sistemleri

Tavsiye sistemi olarak da bilinen öneri sistemi, bir kullanıcının bir öğeye vereceği derecelendirmeyi veya tercihi tahmin etmek için kullanılan bir algoritma türüdür. Öneri sistemleri yaygın olarak internet yayın hizmetlerinde film ve müzik önerileri, e-ticaret sitelerinde ürün önerileri ve haber sitelerinde haber makalesi önerileri gibi uygulamalarda kullanılmaktadır. Üç farklı türde öneri sistemi mevcuttur.

1. İşbirlikçi Filtreleme Öneri Sistemleri
2. İçerik Temelli Öneri Sistemleri
3. Hibrit Öneri Sistemleri

2.6.1 - İşbirlikçi Filtreleme Öneri Sistemleri

Collaborative Filtering yaklaşımı, kullanıcıların site üzerindeki geçmiş davranış verisi üzerinden modellenmektedir. Yaklaşım, geçmiş etkileşim verileri üzerinde benzerlik bulunan kullanıcıların benzer ürünlere ilgi duyabileceği varsayımı üzerinden hareket etmektedir (Resnick et al., 1994).

Kullanıcıların etkileşim verileri iki farklı şekilde tedarik edilebilir,sırasıyla;

Açık veri toplama: kullanıcılardan favori öğelerin bir listesini derlemelerini veya daha önce satın alınan ürünleri bir ölçekte en çok sevilenden en az sevilene doğru derecelendirmeleri üzerinden tedarik edilen etkileşim.

Örtülü veri toplama :Kullanıcıların site üzerinde gerçekleştirdikleri gezinme,beğeni,alışveriş, tıklama gibi verileri bir çerez yardımı ile toplanması üzerinden tedarik edilen etkileşim.

Kullanıcıların, ürünlerle gerçekleştirdiği davranış "kullanıcı-ürün" etkileşimleri matrisinde yer alır. Bu yaklaşıma göre elde edilen bu matris, birbirlerine yakınsayan ve uzaklaşan kullanıcıları belirlemekte kullanılabilme ile birlikte, benzer kullanıcılar arasında ortak olarak henüz ilişki kurulmamış ürünlerin önerilmesi ile sağlar. Kullanıcıların temsili matris üzerinde bulundukları satırın vektörel ifadesi olarak alınırken, kullanıcıların yakınlığı da farklı uzaklık metrikleri üzerinden ifade edilir.

Kullanıcılar arasındaki uzaklığı ifade eden fazlasıyla uzaklık metriği olmasına karşın en sık kullanılan yöntemlerden biri kosinüs benzerliğidir. Kosinüs benzerliği, boyutlarından bağımsız olarak veri nesnelerinin ne kadar benzer olduğunu belirlemede yardımcı olan bir ölçüdür. Kosinüs benzerliği, matematiksel olarak aralarında benzerlik hesaplanacak iki vektörün nokta çarpımının iki vektörün öklid normlarının çarpımına bölümüdür.

$$similarity = cos(\theta) = \frac{A.B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

İşbirlikçi filtreleme hesaplanma yöntemi bakımından ikiye ayrılır,bunlar;

1. Hafıza Temelli İşbirlikçi filtreleme
2. Model Temelli İşbirlikçi filtreleme

2.6.1.1 Hafıza Temelli İşbirlikçi Filtreleme

Hafıza Temelli işbirlikçi filtreleme yöntemi, yalnızca kullanıcı-öğe etkileşim matrisinden gelen bilgiler kullanılmaktadır, yeni öneriler üretmek için herhangi bir model varsaymamalarıdır.

Hafıza Temelli işbirlikçi filtreleme yöntemi, hesaplanma amacı doğrultusunda ikiye ayrılır. Bunlar sırasıyla ürün temelli işbirlikçi filtreleme ve kullanıcı temelli işbirlikçi filtreleme yöntemidir.

2.6.1.1.1 Ürün Temelli İşbirlikçi Filtreleme

Ürün temelli işbirlikçi filtreleme i ve j ürünleri arasındaki benzerlik hesaplamasının temel fikri, Fig2 de gösterildiği gibi bu ürünlerin her ikisini de derecelendiren kullanıcıları ayırmak ve ardından si,j benzerliğini belirlemek için bir benzerlik hesaplama yöntemi uygulamaktır (Yao and Cai, 2015).

2.6.1.1.2 Kullanıcı Temelli İşbirlikçi Filtreleme

Kullanıcı temelli işbirlikçi filtreleme mentalitesi ise öneri sürecinde kullanıcının önceden "olumlu" etkileşimde bulunduğu ürünlere benzer ürünler önermektir. İki öğe arasındaki benzerliği hesaplamak yerine, iki kullanıcı arasındaki benzerliğe odaklanılır. Yöntem içerisinde kullanıcıları etkileşim verisi üzerinden belirli yakınlık algoritmaları kullanarak kullanıcılar arasında yakınlık ilişkisi kurulur, sonrasında kullanıcıların etkileşime geçmeyip yakınlık kurduğu kullanıcıların tercih ettiği ürünler, önem sırasına göre önerilir. Benzer kullanıcıların benzer ürünleri beğenebileceği varsayımından yola çıkılır.

2.6.1.2 - Model Temelli İşbirlikçi Filtreleme

Model Temelli İşbirlikçi Filtreleme yönteminde ise yalnızca kullanıcı-ürün etkileşim bilgilerine dayanır ve bu etkileşimleri açıklayabilecek bir model üzerinden öneri sistemini kurar. Burada öngörülen süreç kullanıcı ürün etkileşimlerinin girdi olarak verildiği bir model oluşturmak ve öneri sistemini bir optimizasyon problemi olarak ele almaktır. Matris çarpanlara ayırma, karar ağaçları, destek vektör makinesi modelleri aktif olarak kullanılmaktadır.

2.6.2 - İçerik Tabanlı Öneri Sistemleri

İçerik tabanlı öneri sistemleri, bir kullanıcıya benzer öğeleri önermek için öğelerin özelliklerini kullanan bir öneri sistemi yöntemidir. Bu yaklaşım, öğenin açıklaması ve kullanıcının tercih ettiği seçeneklerin içeriğine dayanır. İçerik tabanlı bir öneri sisteminde, ürünü tanımlamak için ürün

tanımlama içeriği kullanılır. Algoritmalar, geçmişte beğendiği ürünlerle benzer ürünleri önermeye odaklanır. İçerik tabanlı filtreleme bir kullanıcının bir ürünü beğenmesi durumunda, aynı kullanıcının o ürünün bir benzerini beğeneceği varsayımıyla çalışır

İçerik tabanlı öneri sistemlerinin en temel sorunu, kullanıcı tercihlerini öğelerin içeriğinden öğrenme yeteneğidir. Bu, kullanıcının tercihlerine ilişkin tüm bilginin öğelerin içeriğinde yansıtılmaması nedeniyle zor olabilir. Sonuç olarak, içerik tabanlı öneri sistemleri, birden fazla kullanıcının tercihlerini dikkate alan diğer yaklaşımlar gibi etkili olmayabilir.

2.6.3 Hibrid Öneri Sistemleri

Kullanıcıların etkileşim verisi üzerinden uygulanan işbirlikçi öneri sistemleri ile öğelerin özellik verisine dayalı olan içerik tabanlı öneri sistemleri birleştirilerek, hibrit öneri sistemleri oluşturulur. Bu yaklaşım, kullanıcıların tercihlerini ve önerilen öğelerin özelliklerini dikkate alarak daha kişiselleştirilmiş öneriler sunar.

Hibrit öneri sistemlerini uygulamak için farklı yöntemler mevcuttur. Bu yöntemlerden bir tanesi iki ayrı öneri sistemi oluşturmak ve oluşturulan iki ayrı öneri sistemini birleştirmektir. Bir diğer yöntem ise, içerik tabanlı bir öneri sistemine işbirlikçi özellikler eklemek veya işbirlikçi bir öneri sistemine içerik tabanlı özellikler eklemektir..

2.7 - Literatür İncelemesi

Görünüm Alma: Günlük Fotoğraflarda Otomatik Ürün Önerileri için Kıyafet Tanıma ve Segmentasyon (Kalantidis et al., 2013) çalışması bir görsel üzerinden meta veriler olmadan ilgili moda ürünlerini otomatik olarak önerebilen bir yöntemi önermektedir. Bu yöntem, görüntüdeki kişiyi ayırarak ve görüntüdeki moda sınıflarını belirleyen makine öğrenimi tekniklerini kullanarak, sonra da görüntü benzerliği tekniklerini kullanarak her bir sınıftan benzer ürünleri almaktadır. Makale yazarları, yöntemlerinin son zamanlarda işaretlenmiş bir veri kümesinde en son teknolojiye kıyasla benzer bir performans gösterdiğini ve çok daha hızlı olduğunu iddia etmektedir. Ayrıca, ürün veritabanında bir milyondan fazla ürün bulunan büyük ölçekli bir moda öneri senaryosu sunmaktadır. Bu yaklaşım, bir görüntüdeki giysi bölgelerini algılamaya ve sınıflandırmaya ilişkin birkaç temel adım içermektedir. Öncelikle, pozlandırma kullanılarak görüntüdeki insan vücudunun konumu belirlenir. Bu bilgi, giysi açısından en umut verici görüntü bölgelerini ayrıştırmak için kullanılır. Sonra, bu bölgeler görsel olarak tutarlı parçalara bölünür. Bu parçalar, açık bir ceket gibi uzak mekânsal parçaları birleştirmeye olanak sağlamak için birleştirilir.Her bölgenin belirlenen insan pozuya göre konumunu ve şeklini açıklamak için, yepyeni bir ikili temsili olan mekânsal görünüm maskesi kullanılır. Bu temsil, eğitim görüntülerinden alınan etiketlenmiş örneklerle olan yakınlığa

göre tüm sorgu bölgelerinin sınıflandırılmasına olanak sağlar. Eğitim kümesi bölgeleri mekânsal görünüm maskesi olarak temsil edilir ve LSH dizininde indekslenir, böylece gerçek bir öğrenme gerektirmeksizin etkili ve hızlı sınıflandırma yapılabilir. Bu yaklaşım verimlidir ve hızlı bir şekilde uygulanabilir, gerçek zamanlı giysi algılama ve sınıflandırma için yararlıdır. Bir görüntüdeki giysi sınıfları belirlendikten sonra, ikinci adım büyük bir ürün veritabanından görsel olarak benzer ürünleri almaktır. Benzer görünümlü öğeler, örtüşme benzerliklerini toplayarak alınmıştır.

Transformer mimarileri, gömülü görüntü temsili gibi birçok bilgisayar görüş görevlerinde başarılı olduğu için. Evrişimli Yapay Sinir Ağları ve diğer modeller gibi, Transformer'lar tarafından yapılan tahminlerin açıklanabilirliği önemli bir konudur. Ancak, Transformer modeli için görselleştirme yaklaşımları genellikle mimari üzerinden değişkenliğe sahiptir.Transformer Ağlarında Eşleştirilmiş Görüntü Benzerliğinin Görselleştirilmesi (Black et al., 2022) çalışmasında, bir Transformer ile gömülü temsile sahip iki resim verildiğinde benzerliğe katkıda bulunan bölgeleri gösteren yeni bir yorumlanabilir görselleştirme yöntemi sunulmuştur. Bu çalışma gömülü görüntü temsili elde etme görevinde Evrişimli yapay sinir ağları ile Transformer mimarisini karşılaştırmıştır. Transformer mimarileri, daha büyük bölgeleri ve nesnelerin tamamını kodlamaya eğilimlidir, ancak konvolüsyonel yerleştirme ağları daha küçük dokulu bölgeleri kodlamaya eğilimlidir. Ayrıca, Transformasyonlar tarafından oluşturulan kodlamalarda kamera konumu bir etken olarak görünüyor. Bu bilgi, çeşitli görüntü alanları ve görüntü yerleştirme dışındaki görevler için hangi modelin seçileceğini belirlemede faydalı olabilir.

Moda Fotoğraflarında Giysileri Ayrıştırma çalışması (Yamaguchi et al., 2012), bir kişinin kıyafetini ayrıntılı ve doğru bir şekilde ayrıştırmak için etkili bir yöntem önermektedir. İki senaryo incelenmektedir: meta-veri sağlanan giysi etiketleri kullanarak ayrıştırma ve sınırsız etiket kümeleri kullanarak ayrıştırma. Araştırma ile birlikte büyük bir veri kümesi ve etiketleme araçları da sunulmaktadır. Giyim ayrıştırma problemi, görüntü içindeki kişisel kıyafet öğelerini tanımlamak için bir görüntünün bireysel piksellerini veya bölgelerini etiketlemekle ilgilidir. Bu, genel görüntü ayrıştırma problemine benzer ancak görüntüdeki diğer nesneler yerine kişisel kıyafet öğelerini tanımlamaya odaklanır. Önceki araştırmalarda önerildiği gibi, uniform görünüm bölgelerinin aynı giyim öğesi olduğu varsayımıyla, bu problem görüntü içinde bir dizi superpixel için bir etiketleme tahmini yapmaya indirgenmiştir. Bu yaklaşım, giyim ayrıştırma problemini daha yönetilebilir hale getirmeyi ve daha doğru tahminler yapmayı mümkün kılar.

Nereden Alınır: Çevrimiçi Mağazalarda Eşleşen Sokak Kıyafeti Fotoğrafları (Kiapour et al., 2015) çalışması, gerçek dünyadaki giyilebilir öğeleri çevrimiçi karşılıklarına eşleştirmeyi içeren yeni bir görev olan Exact Street to Shop'u tanımlar. Görevin zor olduğu, cadde fotoğrafları ile online mağaza

fotoğrafları arasındaki görsel farklılıklardır. Yazarlar, 404,683 mağaza fotoğrafı ve 20,357 cadde fotoğrafından oluşan bir veri kümesi toplar ve Exact Street to Shop görevi için üç yöntem geliştirir. Yöntemlerden ikisi derin öğrenme tabanlıdır, üçüncü yöntem ise cadde ve mağaza alanları arasındaki benzerlik öğrenmesini içerir. Sonuçlar, öğrenilmiş benzerliğin derin öğrenme tabanını geçtiğini göstermektedir. Caddeden dükkana eşleştirme problemi için birkaç farklı bilgiye erişim yöntemi uygulamışlardır. Yöntemlerin girdileri, sokak sorgusu görüntüsü, ilgilenilen öğe kategorisi ve sorgu görüntüsündeki öğenin etrafındaki nesne belirleme kutusudur. Mağaza tarafında, çok sayıda görüntü olduğu için, öğelerin elle etiketlenmiş herhangi bir nesne belirleme bölgelerini varsayılmamıştır, bunun yerine algoritmanın görüntünün tamamında veya nesne önerme bölgelerinde hesaplanan özelliklere dayanmasına izin verilmiştir.

1-Görüntü özellikleri yalnızca kırpılan öğe bölgesinde hesaplanırken, mağaza görüntüleri için, görüntünün tamamında CNN özelliklerini hesaplanır. Daha sonra, sorgu özellikleri ile tüm mağaza görseli özellikleri arasındaki kosinüs benzerliğini karşılaştırır ve bu benzerliğe göre benzerlikler sıralanır.

2-Spesifik olarak, seçici arama algoritmasını kullanılır ve genişliği resim genişliğinin ⅛ 'inden küçük olan teklifler filtrelenir. Tüm görüntü alma yöntemine benzer şekilde, sokak öğesi sınırlama kutusunda ve her bir mağaza görüntüsü için en güvenilir 100 nesne önerisi üzerinde CNN özelliklerini hesaplanır. Daha sonra kosinüs benzerliğini kullanarak benzerlikler sıralanır.

3-Benzerlik öğrenme görevi, amacın iki öğenin ne kadar benzer veya farklı olduğunu belirlemek olduğu bir tür makine öğrenimi problemidir. Bu durumda, öğeler, esas olarak görüntülerin bir bilgisayarın anlayabileceği ve değiştirebileceği bir biçimdeki temsilleri olan CNN özelliği çiftleridir. Görev, bir ikili sınıflandırma problemi olarak çerçevelenmiştir, yani amaç, her bir özellik çiftini pozitif (iki özelliğin aynı öğeyi temsil ettiğini gösterir) veya negatif (farklı öğeleri temsil ettiğini gösterir) olarak sınıflandırmaktır.

Bu sorunu çözmek için çapraz entropi hatası optimize edilir. Çapraz entropi, bir dizi tahmin edilen olasılığın, bir dizi olayın gerçek olasılıklarıyla ne kadar iyi eşleştiğinin bir ölçüsüdür. Bu durumda, tahmin edilen olasılıklar, belirli bir özellik çiftinin aynı veya farklı öğeleri temsil etme olasılıkları olacaktır ve gerçek olasılıklar, her bir çiftin gerçek etiketleri olacaktır (yani, pozitif veya negatif). Model, çapraz entropi hatasını optimize ederek, belirli bir özellik çiftinin aynı veya farklı öğeleri temsil edip etmediğini doğru bir şekilde tahmin etmeyi öğrenir.

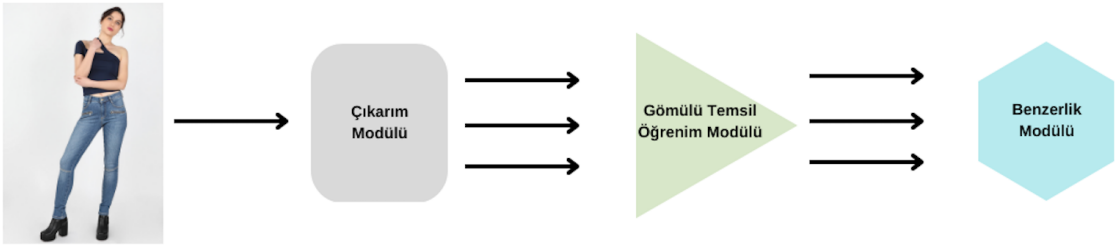
Bana O Görünümü Satın Al (Ravi et al., 2021) çalışması, bir e-ticaret platformunda kullanıcılara benzer moda ürünlerini önerebilen bir bilgisayar görüşü tekniği olan ShopLook'u önermektedir. Kullanıcının sorgusu ve ilgili ürün görüntüleme sayfası verildiğinde, yöntemin amacı, görüntüleme

sayfasındaki model tarafından giyilen tüm moda ürünlerine benzer ürünler önerebilmektir. Bu, satışı artırmak ve müşteri deneyimini geliştirmek için önemlidir. Yöntem, insan nokta tespiti, poz sınıflandırma, makale yerelleştirme ve nesne tespiti gibi birkaç adımı içermektedir, ayrıca aktif öğrenme geribildirimi de bulunmaktadır. Bu, sistemin model tarafından giyilen moda ürünlerini doğru bir şekilde tanıyarak kullanıcıya benzer ürünler önerebilmesine olanak verir ve 4 aşamadan oluşur.

1. Öne Taraf Tam Çekim Görüntü Algılama. Tekniğin ilk adımı, ürün görüntüleme sayfasındaki görüntülerden tam kare görüntüyü otomatik olarak tanımlamaktadır. Bunu yapmak için, sistem insan anahtar noktası tespiti gerçekleştirir, bu da insan vücudundaki belirli noktaların tanımlanmasını içerir, örneğin eklemler ve önemli vücut bölgeleri. Bu sayede sistem, görüntüdeki insanın pozunu anlayarak, tam kare görüntü olup olmadığını belirleyebilir. Tam kare görüntü tanımlandıktan sonra, sistem işlemin bir sonraki adımına geçebilir, bu da poz sınıflandırmasıdır.
2. Moda makalesi Algılama ve Yerelleştirme,Tam kare görüntünün tanımlanmasından sonra, tekniğin bir sonraki adımı, görüntüdeki model tarafından giyilen tek tek moda makalelerini tanımlamak ve ayırmaktır. Bunu yapmak için, sistem bir özel moda makale veri kümesinde eğitilmiş bir Mask RCNN modelini kullanır. Bu, modelin görüntüdeki farklı moda makalelerini doğru bir şekilde algılayıp yerelleştirebilmesine olanak verir. Model ayrıca aktif öğrenme kullanır, bu da içerideki etiketleyicilerin yanlış sınıflandırılmış örnekleri tanımlayarak ve bunları kullanarak modeli tekrar eğitmesine olanak verir, bu performansı geliştirir. Moda makaleleri tanımlanıp yerelleştirildikten sonra, sistem kullanıcıya benzer ürünler önerebilir.
3. Türler için gömme oluşturma, tam kare görüntüden önemli moda makaleleri çıkarıldıktan sonra, tekniğin bir sonraki adımı, katalog veritabanından benzer moda ürünlerini almaktır. Bunu yapmak için, sistem çıkarılan makale türlerini ve veritabanındaki ürünleri benzer makaleleri bir araya getiren ortak bir gömme alanına temsil eder ve benzer olmayanları oradan uzaklaştırır. Bu, üçlü tabanlı bir ağ mimarisi kullanarak gerçekleştirilir, bu mimari üç ağırlıkları paylaşan eşdeğer Evrişimli Yapay Sinir Ağından oluşur. Ağ, ikinci ve üçüncü resimler arasında semantik olarak benzer olan üçlü resimler kullanarak eğitilir. Bu sayede sistem gömmeyi öğrenir ve kullanıcıya doğru bir şekilde benzer ürünler önerebilir.

3 - METODOLOJİ

Oluşturulmak istenen sistemin amacı, bir fotoğraf üzerinden,fotoğrafta bulunan kişinin üzerinde olan farklı moda ürünlerinin benzerlerini önerebilecek bir uygulama kurmaktır. Bu bölümde bahsedilen uygulamanın kurulumu ve bölümleri anlatılacaktır. Önerilen uygulama (şekil 4.1.1'de görüldüğü gibi) bir görsel üzerinden fotoğraf anlamlandıracak ve içerisindeki istenen moda kategorilerini segmente edebilecek ve sonrasında her segment özelinde benzerlik üzerinden ürün önerisi yapabilecek 3 ana kısımdan oluşmaktadır. Her bir kısım, bu başlık altında ilerleyen kısımlarda anlatılacaktır.



Şekil 3.1.1 Önerilen Metodoloji

- **Çıkarım Modülü** (kullanıcının moda görseli)

  Amacı: Görsel içerisinde bulunan kişinin veya mankenin, üzerinde bulunan ve önceden belirlenmiş kategoriler içerisinde olan ürünlerin segmente edilip ayrı görseller olarak ayrılması.

  Dönüş: Görsel üzerinde bulunan manken veya kişinin üzerinde bulunan ve önceden belirlenmiş olan kategoriler içerisinde yer alan kiyafetlerinin farklı görsel dosyaları şeklinde ayrılmış durumu .

- **Gömülü Temsil Öğrenim Modülü** (Nesne tespiti ile ayrılmış ürün görseli)

  Amacı: ayrılan her bir görselin vektörel temsili, Yapay Sinir Ağına sahip bir mimari üzerinden çıkarılması.

  Dönüş: Ayrılan her bir görselin vektörel temsili

- **Benzerlik Modülü** (Vektörel Temsil)

  Amacı: Vektörel temsili alınan her bir görselin vektörel betimleme havuzu içerisinde bulunan tüm ürünler için kosinüs uzaklığı metriği üzerinden uzaklık değeri belirlenmesi ve benzerlik değeri en yüksek 10 ürünün belirlenmesi.

  Dönüş: Vektörlük uzaklık değeri en yüksek olan 10 ürün.

3.1 - Çıkarım Modülü

Önerilen sistem içerisinde Gömülü temsil öğrenim dizini modeline girecek veri setinin belirli moda kategorileri üzerinden ayrılmış görseller şeklinde olması esastır. Burada amaç önceden belirli kategoriye ait olan görsellerin vektörel dizimini öğrenmektir. Çıkarım dizini sisteme girecek olan ve birden fazla kategoriyi içinde barındıran manken veya model görselin önceden belirlenmiş kategoriler düzleminde kesilerek tek bir görselden farklı görseller oluşturulmasını sağlamaktır. Bu görevi sağlaması için obje tespiti gerçekleştirebilen evrişimli yapay sinir ağı temelli mimariler kullanılmaktadır.

Yapay sinir ağı içinde bulunan parametre adedi kadar komplekstir, kompleks olması ise modelin eğitilme süresi ile doğru orantıdadır. Özellikle zaman ve donanım gibi imkanlar model seçimi ve değerlendirmesi noktasında önem arz etmektedir.

Bu nedenle bir modeli en baştan eğitmek yerine, transfer öğrenme metodu tercih edilir.

Transfer öğreniminde, önce bir temel ağı bir temel veri kümesi ve görev üzerinde eğitiriz ve sonra öğrenilen özellikleri yeniden kullanırız veya bunları, bir hedef veri kümesi ve görev üzerinde eğitilmek üzere ikinci bir hedef ağa aktarırız. Bu süreç, özellikler genelse, yani temel göreve özgü yerine hem temel hem de hedef görevlere uygunsa işe yarayacaktır (Yosinski et al., 2014).

En yaygın transfer öğrenimi yöntemlerinden birisi olan ince-ayar (fine-tune) yöntemi, eğitimli bir sinir ağının parametre değerlerini almakla birlikte, bu değerleri benzer yapıya sahip veriler üzerinde eğitilen yeni bir model için ilk parametre değerleri olarak kullanılmasıdır. Bu sayede gerçekleştirilmek istenen görev zaten benzer bir görevde eğitilmiş parametre değerlerini kullanacağı için daha çabuk ve daha iyi öğrenebilme potansiyeli taşır.

Çok çeşitli mimariler içerisinde Ayrıca, çok çeşitli esnek mimari tasarımlarını kolaylaştıran obje tespiti mimarileri mevcuttur. İnce-ayar gerçekleştirilmesi kolay ve daha hızlı olmakla birlikte önceki eğitilmiş parametre değerleri farklı veri seti özelliklerini daha kapsamlı yakınsaması bakımından Mask R-CNN mimarisi tercih edilmiştir.

3.2 Gömülü Temsil Öğrenim Modülü

Çıkarım dizini belirli moda kategorilerine ait görsel parçalarını, tek bir fotoğraftan çıkarmak için önemli bir işlev görmektedir, ancak sistemin bütünü için elde edilen görsel parçalarının bir veri kümesi içerisinde tutulup,daha sonrasında farklı moda görselleri verildiğinde, veri kümesi içerisinde bulunan fotoğrafların her biri ile benzerliği belirlenip, en benzer ürünleri listelenmesi gerekmektedir. Bu bütünde ihtiyaç, görsellerin birbirleri ile karşılaştırılabilir temsillerini oluşturmaktır. Bu noktada ise Gömülü temsil öğrenim metodu kullanılmaktadır. Gömülü öğrenme, her bir görselin vektörel düzlemde bir temsilini oluşturmakla birlikte benzer görsellerin daha yakın farklı görsellerin ise daha uzakta olduğu bir uzay yaratır.

Bu çerçevede görüntü gömme, görüntünün vektörel bir temsilinin çıkarılmasıdır.Gömülü yöntemin seçilmesinin bir başka nedeni ise özellikle son yapılan bilimsel çalışmalar sonucunda, görsel verinin en etkili temsil yöntemlerinden birisi olduğunun saptanmasıdır (Veit et al., 2017).

Gömülü vektörel temsil için çok farklı yöntemler mevcuttur. Ancak görsel verisi ile en etkili sonuçlar üretilen çalışmaların büyük çoğunluğu evrişimli yapay sinir ağı tabanlı sınıflandırma görevi gerçekleştirebilecek mimariler üzerinden gerçekleşmektedir.

Evrişimli sinir ağlarına ismini veren evrişim operasyonu girdi verisinin ağırlık kümeleriyle çarpılmasını içeren doğrusal bir operasyonun linear olmayan fonksiyonlardan geçirilmesiyle geleneksel sinir ağlarına çok benzeyen bir yapıya sahipken özellikle yerel bağlantı özelliği ile içerik bilgisini taşımasıyla görsel girdisi olan görevlerde geleneksel sinir ağı tabanlı mimarilerden daha etkili olabilmektedir.

Evrişimli sinir ağları belirli bir görev üzerine eğitilir ve bu görev sinir ağının son katmanında nasıl bir katman kullanılacağı noktasında belirleyici olur. Hem regresyon hem de sınıflandırma görevleri için aktif olarak kullanılır. Regresyon görevinde tek nöronlu bir katman belirlenirken sınıflandırma problemlerinde sınıf sayısı kadar nörona sahip olan katman belirlenir. Ancak gömülü vektörel temsil öğrenme görevinde sinir ağının son katmanı bir önem teşkil etmemektedir. Gömülü vektörel temsil öğreniminde daha önceden eğitilmiş parametre değerlerine sahip bir sinir ağının son katmanı çıkartılır ve sondan bir önceki lineer katmanın nöronlar gömülü vektörel temsilin bileşenlerini oluşturacak şekilde kullanılır, burada amaç gömülü vektörel temsil öğrenilecek verinin kapsandığı benzer bir veri üzerinde eğitilen yapay sinir ağının parametrelerini kullanabilmektir. Bu nedenle gömülü vektörel temsil öğrenimi aynı zamanda öznitelik çıkarımı anlamınada gelmektedir.

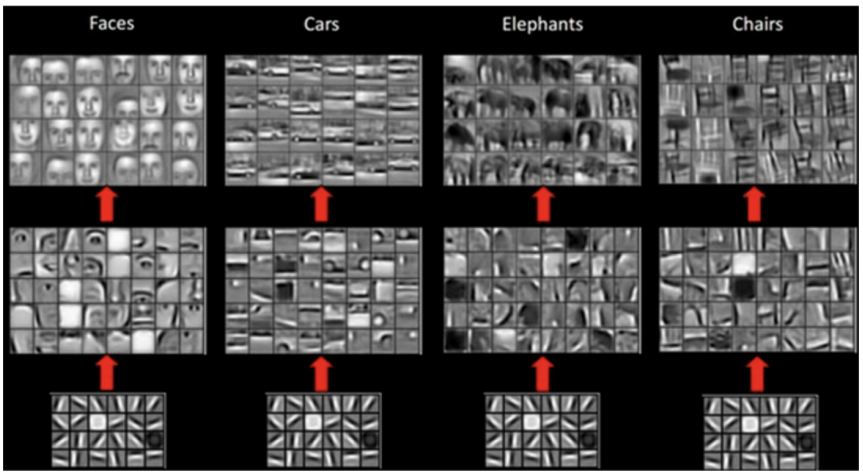
Bir diğer sık kullanılan ince-ayar yöntemlerinden birisi de eğitimli bir sinir ağının parametre değerlerini alarak ilk katmanlarda bulunan parametre ağırlıklarının doldurulmasıdır. Dondurulan katmanların üzerine görev temelli sınıflandırma katmanları eklenmesi durumunda sadece

sınıflandırma katmanları eğitilecektir,dondurulmuş katmanlar sadece özellik çıkarma görevi görecektir. Bu sayede sadece dondurulmayan katmanlar içerisindeki parametreler üzerinden eğitim gerçekleşecektir. Belirlenmiş görev üzerinden ince-ayar eğitimi gerçekleştikten sonra model bir üst paragrafta belirlenen şekilde düzenlenerek gömülü vektörel temsil öğrenime hazır hale getirilmektedir.

İlk katmanların dondurulması, ilk aşamalarda ağın temel özellikleri öğrenmesine bağlıdır. İnce ayarı uygulandığında tam olarak çıkarılmak istenen budur. (“CS231n Convolutional Neural Networks for Visual Recognition”)

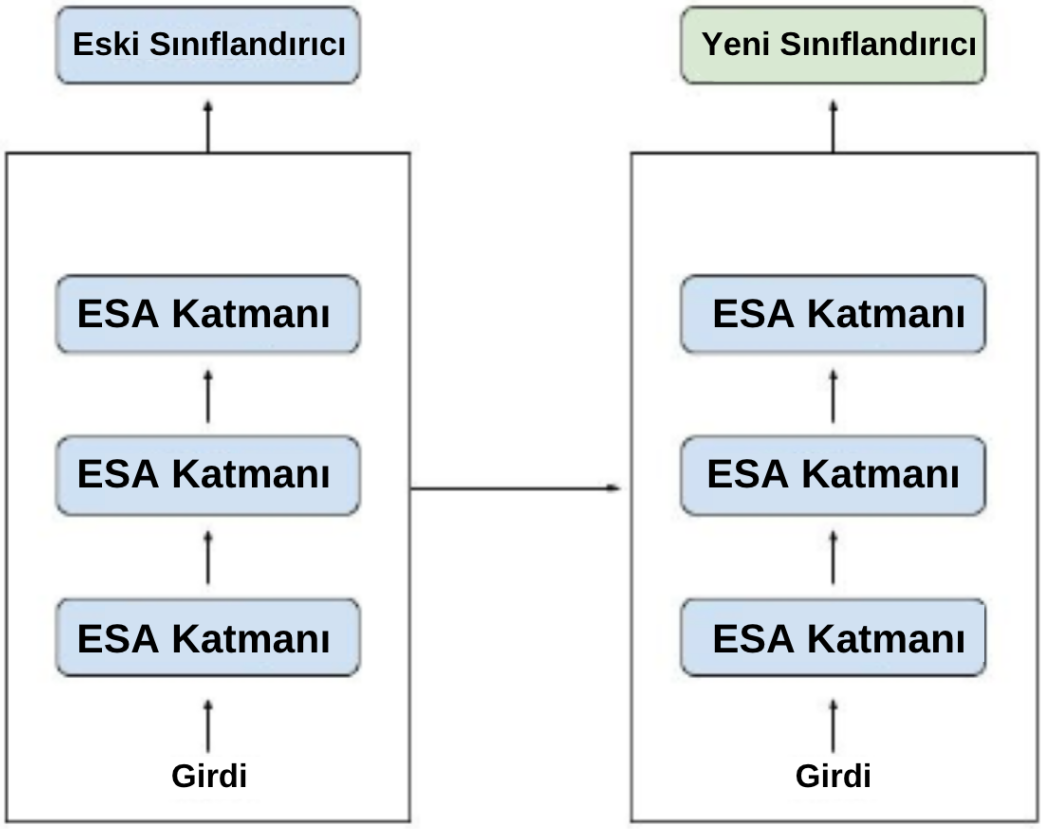
ConvNet özelliklerinin ilk katmanlarda daha genel özellikler ve sonraki katmanlarda daha orijinal veri kümesine özgü olduğunu göz önünde bulundurulumalıdır. (“CS231n Convolutional Neural Networks for Visual Recognition”)

Bu yapılmak istenen görev içerisindeki veri ile kullanılan mimarinin önceden eğitilmiş veri birebir aynı olmasa bile ince-ayar yönteminde çok etkili sonuçlar alınabileceği bilinmektedir.



Şekil 3.2.1 CONVNet Mimarilerin Katman bölgelerinde öğrenimlerinin temsili

Önerilen çalışmada aşağıdaki şekilde gösterildiği gibi önceden farklı bir veri setinde eğitilmiş evrişimli sinir ağı mimarisi ve parametreleri üzerine, belirlenmiş moda verisi ile ince-ayar yöntemi ile eğitildikten sonra sınıflandırma katmanı çıkarılarak gömülü vektörel temsil öğretime hazır hale getirilmiştir.



Şekil 3.2.2 İnce Ayar

3.3 - Benzerlik Modülü

Gömülü vektörel temsil öğrenim modülü, sistem içerisine giren her bir moda fotoğrafının gömülü vektörel temsilini çıkabilmektedir. Bu sayede her bir moda görselinin birbirleri arasındaki benzerliği vektörel yakınlık metrikleri üzerinden değerlendirme imkanı ortaya çıkmaktadır. Benzerlik modülünde amaç daha sonradan sisteme sorgu şeklinde gelecek bir gömülü vektörel temsilin, gömülü vektörel temsillerden oluşan bir veri kümesi içerisinden belirlenen metrik üzerinden yakınlık değeri en fazla olan 5 gömülü vektörel temsilin dönmesidir. Gömülü benzerlikler açısal benzerlikleri ile temsil edilir, bu nedenle kosinüs benzerliğini kullanarak benzer çiftler belirlenebilir. Ancak tüm gömülü vektörel temsiller içerisinde bu metriği hesaplamak ,gömülü vektörel temsillerinden oluşan veri kümesi içerisinde bulunan gömülü vektörel temsil adedi ile doğru orantılı olarak hem zaman üzerinden hem de hesaplama kaynağı üzerinden verimsiz olacaktır.

Bu nedenle vektörel temsiller içerisinde benzerlik arama üzerine oldukça optimize bir şekilde çalışan Annoy (“Annoy Library”) kütüphanesi kullanılacaktır. Annoy, verimli benzerlik arama ve yoğun

vektörlerin kümelenmesi için bir kitaplıktır. Muhtemelen RAM'e sığmayanlara kadar herhangi bir boyuttaki vektör kümelerinde arama yapan algoritmalar içerir. Ayrıca değerlendirme ve parametre ayarı için destekleyici kod içerir. Annoy, Python/numpy için eksiksiz sarmalayıcılarla birlikte C++ ile yazılmıştır ve Spotify AI Research'te geliştirilmiştir.

4 - UYGULAMA

Bu bölüm içerisinde öncelikle, metodoloji kısmında bahsedilen deney bölümleri içerisinde kullanılacak veri setleri tanıtılacaktır. Bununla birlikte her bir kısım için değerlendirme metrikleri tanıtılıp hesaplanma yöntemleri açıklanacaktır. Bununla birlikte gerçekleştirilen deneylerin eğitim verisi ve test verisi üzerindeki sonuçları incelenip değerlendirilecektir.

4.1 - Veri Setleri

4.1.1 - iMaterialist (Fashion) 2019 at FGVC6

iMaterialist (Fashion) 2019 veri Kaggle üzerinde gerçekleştirilen[*] aynı isimli yarışma için paylaşılmış bir veri setidir. Bu yarışma, Bilgisayarla Görme ve Örüntü Tanıma Konferansı CVPR 2019'daki İnce Taneli Görsel Kategorizasyon FGVC6 atölye çalışmasının bir parçasıdır (“iMaterialist (Fashion) 2019 at FGVC6”).

Giysilerin görsel analizi son yıllarda artan ilgi gören bir konudur. Giyim ürünlerini ve ilgili özellikleri resimlerden tanıyabilmek, tüketiciler için alışveriş deneyimini geliştirebilir ve moda profesyonelleri için iş verimliliğini artırabilir (“iMaterialist (Fashion) 2019 at FGVC6”).

Moda ve bilgisayarla görme toplulukları arasındaki güçleri birleştirerek yeni, ince taneli bir segmentasyon görevi sunma hedefiyle yeni bir giyim veri seti sunuyoruz. Önerilen görev, gerçek dünya uygulamalarına doğru önemli bir adım olan zengin ve eksiksiz giyim özelliklerinin hem kategorizasyonunu hem de segmentasyonunu birleştirir (“iMaterialist (Fashion) 2019 at FGVC6”).

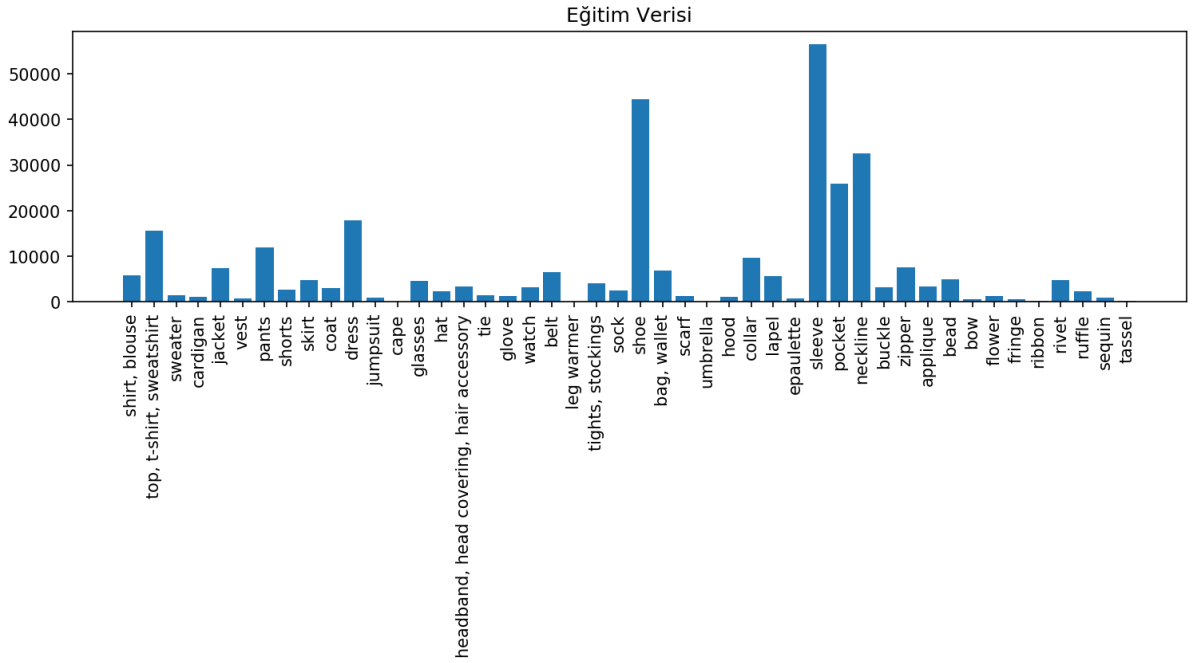
Giyim örneği segmentasyonları, 27 ana giyim nesnesini (ceketler, elbiseler, etekler vb.) ve 19 giyim parçasını (kollar, yakalar vb.) içerir. Ana giyim nesneleri için toplam 92 tane ince taneli öznitelik uzmanlar tarafından kataloglanmıştır. Veri seti toplamda 50.000 giyim görseli içermektedir (giyim örneği segmentasyonu ile 40,00 ve hem segmentasyon hem de ayrıntılı özniteliklerle 10.000). Nitelik içeren segmentler, verilerin yalnızca %3,46'sını oluşturuyor ve veri paylaşımcılarına göre görsellerin %80'inde hiçbir özellik paylaşılmamış. Bu nedenle, ilk adımda, görevin karmaşıklığını azaltmak için yalnızca kategorilerle üzerinden ilerlenmiştir.

Şekil 4.1.1 iMaterialist Veri Seti Girdi Ve Etiket Örnekleri

Toplam veri içerisinde eğitim,validasyon ve test kümelerinin ayrımı sırasıyla 43.387, 1.119, 689 şeklinde gerçekleşmiştir. Belirlenen kümeler sonrasında, eğitim ve validasyon kümesenin öznitelik dağılımları grafikleri aşağıdaki gibidir.

Toplam Eğitim Görsel Adedi: 43387

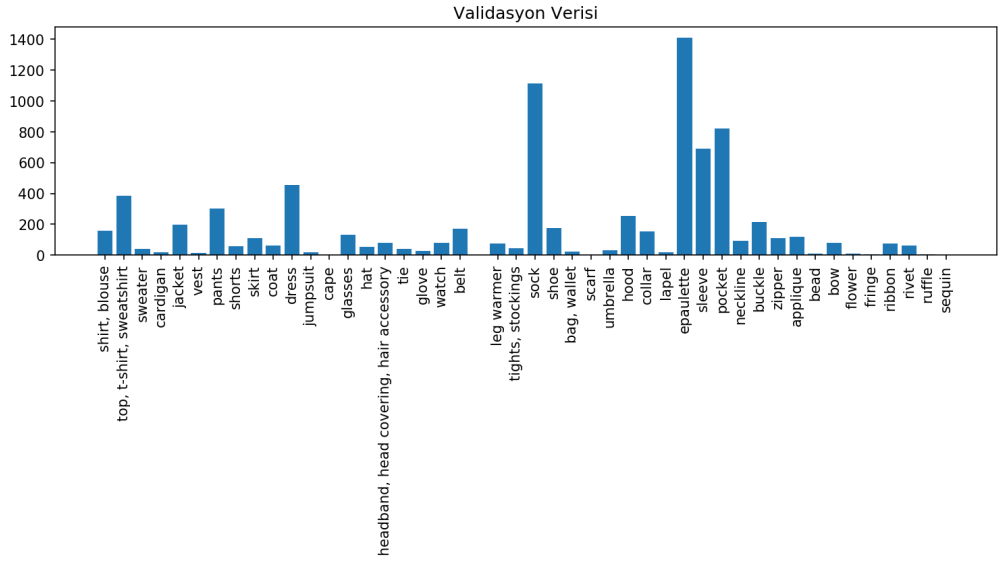
Toplam Eğitim Segment Adedi: 318049



Tablo 4.1.1.1 iMaterialist Eğitim Verisi Segment Dağılımı

Toplam Validasyon Görsel Adedi: 1119

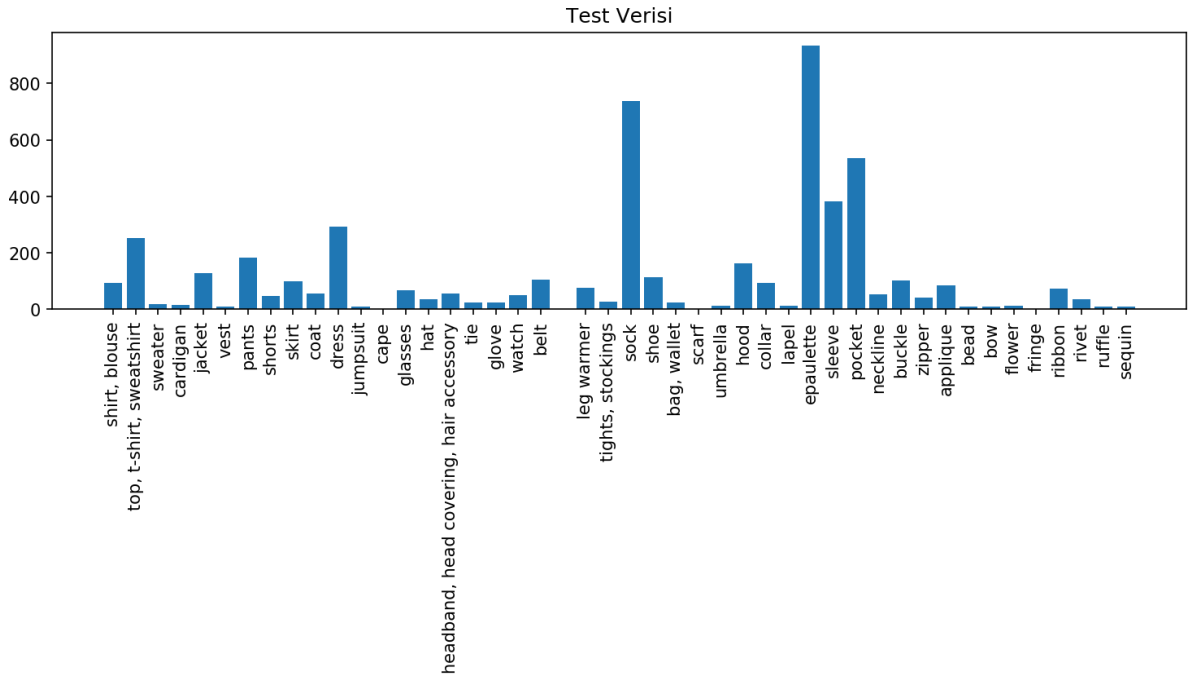
Toplam Validasyon Segment Adedi: 8022



Tablo 4.1.1.2 iMaterialist Validasyon Verisi Segment Dağılımı

Toplam Test Görsel Adedi: 689

Toplam Test Validasyon Adedi: 5142

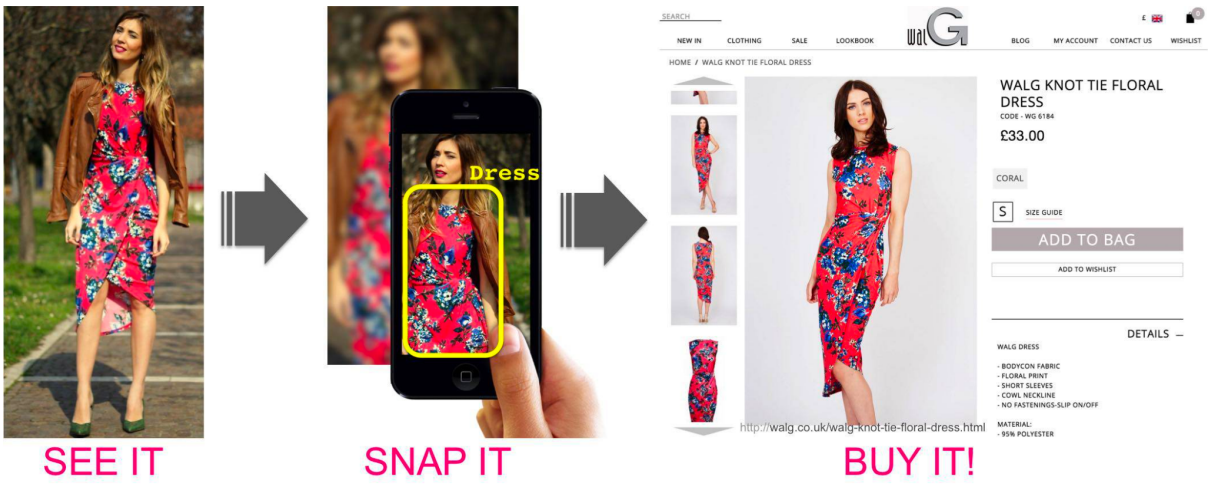


Tablo 4.1.1.3 iMaterialist Test Verisi Segment Dağılımı

Tüm dağılımlar birbirleri ile karşılaştırıldığında, öznitelik dağılımlarının birbirlerine yakınsadığı görülmektedir.

4.1.2 - Sokaktan Dükkana Veri Seti (Exact Street2 Shop)

Exact Street2Shop veri kümesi, gerçek dünya sokak fotoğrafları ve profesyonel olarak çekilmiş mağaza fotoğraflarından oluşan bir giysi ürünleri koleksiyonudur. Veri kümesi, sokakta giyilen giysilerle ilgili online satın alınabilecek öğeleri eşleştirebilen bilgi kazanım algoritmalarının geliştirilmesini sağlamayı amaçlamaktadır. Veri kümesi, sokak fotoğrafları ve mağaza fotoğrafları olmak üzere iki tür resim içermektedir. Sokak fotoğrafları, günlük ortamlarda giyilen giysileri gösteren insanların fotoğraflarıdır, mağaza fotoğrafları ise insanlar, mankenler veya izolasyon olarak giyilen giysilerin daha kontrollü ortamlarda çekilen fotoğraflarıdır. Veri kümesi, sokak ve mağaza fotoğraflarındaki belirli giysi öğelerini işaretleyerek, bilgi kazanım algoritmalarını değerlendirmek için tam sokak-mağaza çiftleri oluşturularak oluşturulmuştur.

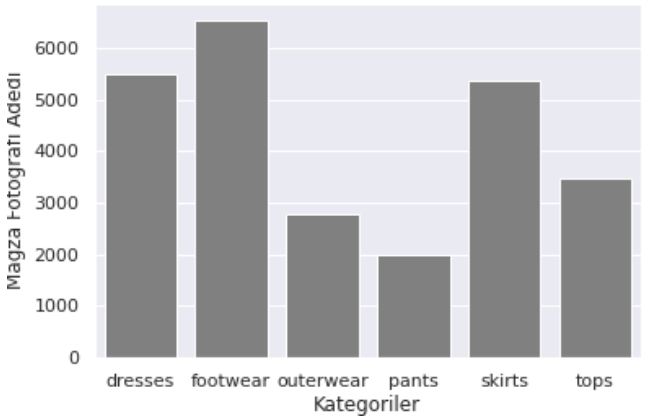


Şekil 4.1.2.1 Sokaktan Dükkana (Exact Street2 Shop) Gösterimi

Veri setinde, 25 farklı çevrimiçi perakendeciden toplanan 404.683 mağaza fotoğrafı ve 20.357 sokak fotoğrafı içeren, sokak ve mağaza fotoğrafları arasında toplam 39.479 giyim eşyası eşleşmesi bulunmaktadır. Ancak 2015 te yayımlanan veri setinin tamamına günümüzde ulaşmak, geçen süre içerisinde silinen ve ulaşıma kapanan fotoğraflar nedeniyle oldukça güçtür. Bununla birlikte ilgili veri setinin kullanıldığı deney içerisinde araştırılan konu, modelin belirli bir alt küme veri içerisinde nasıl bir sonuç vereceğini gözlemlemek olması nedeniyle küme içerisinden, kümeyi temsil edebilecek bir alt küme eğitim ve test için seçilmiştir. Bu kapsamda;

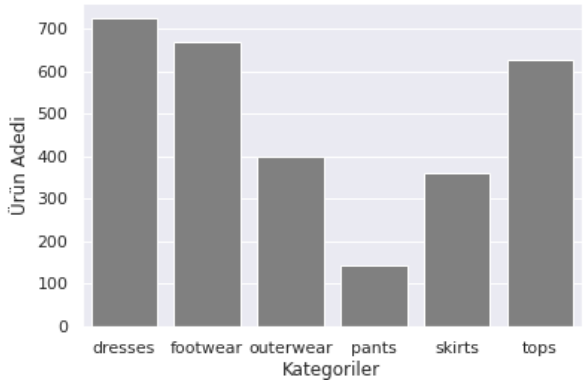
Eğitim verisi için, 3.897 adet sokak fotoğrafı ve karşılığı 12.215 mağaza fotoğrafı olmak üzere toplamda 16.112 farklı fotoğraf kullanılmıştır. Kullanılan fotoğraflar sırasıyla {'dresses', 'footwear', 'outerwear', 'pants', 'skirts', 'tops'} kategorilerinden seçilmiştir.

Eğitim verisi içerisinde kullanılan mağaza görsellerinin kategori bazında dağılımı



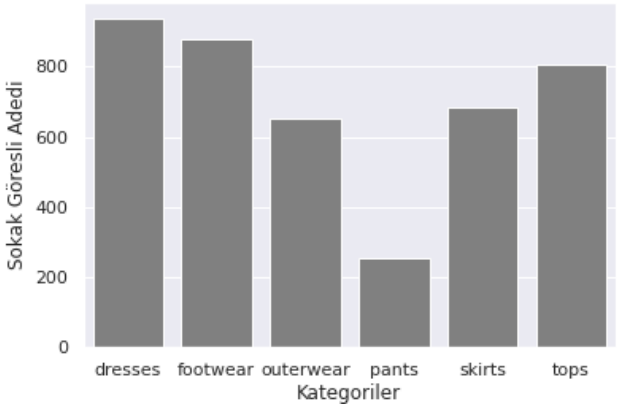
Tablo.4.1.2.1 Eğitim Verisi İçerisinde Kullanılan Sokak Görsellerinin Kategori Bazında Dağılımı

Eğitim verisi içerisinde kullanılan ürünlerin kategori bazında dağılımı



Tablo 4.1.2.2 Eğitim Verisi İçerisinde Kullanılan Ürünlerin Kategori Bazında Dağılımı

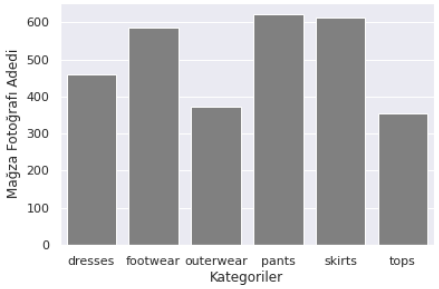
Eğitim verisi içerisinde kullanılan sokak fotoğraflarının kategori bazında dağılımı



Tablo 4.1.2.3 Eğitim Verisi İçerisinde Kullanılan Sokak Fotoğraflarının Kategori Bazında Dağılım

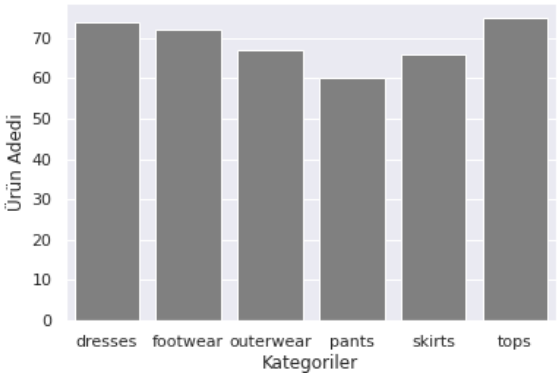
Test datası için, 445 adet sokak fotoğrafı ve karşılığı 1783 mağaza fotoğrafı olmak üzere toplamda 2228 farklı fotoğraf kullanılmıştır. Kullanılan fotoğraflar sırasıyla {'dresses', 'footwear', 'outerwear', 'pants', 'skirts', 'tops'} kategorilerinden seçilmiştir.

Test verisi içerisinde kullanılan mağaza görsellerinin kategori bazında dağılımı



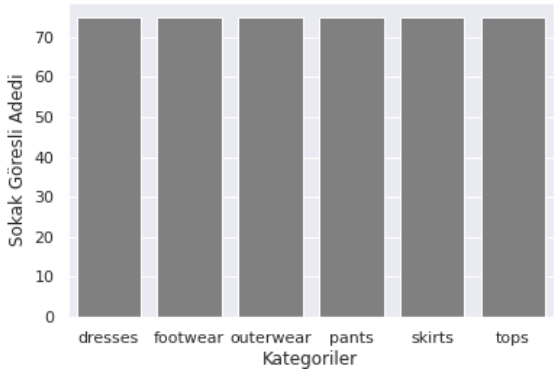
Tablo.4.1.2.4 Eğitim Verisi İçerisinde Kullanılan Sokak Görsellerinin Kategori Bazında Dağılımı

Test verisi içerisinde kullanılan ürünlerin kategori bazında dağılımı



Tablo 4.1.2.5 Test Verisi İçerisinde Kullanılan Ürünlerin Kategori Bazında Dağılımı

Test verisi içerisinde kullanılan sokak fotoğraflarının kategori bazında dağılımı

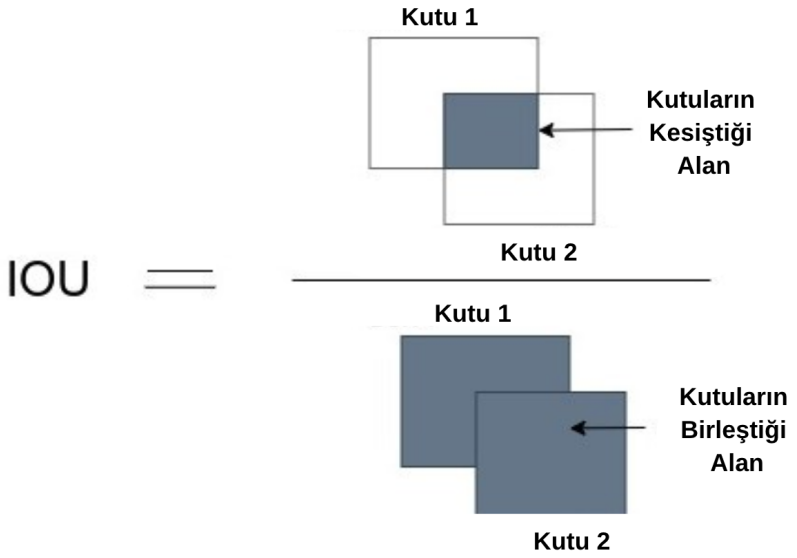


Tablo 4.1.2.6 Test Verisi İçerisinde Kullanılan Sokak Fotoğraflarının Kategori Bazında Dağılım

4.2 - Değerlendirme Metrikleri

4.2.1 - Nesne Tespiti

İşlem oranı (IoU), nesne tespitinde sıklıkla kullanılan bir değerlendirme metriğidir. Bu metrik, tahmin edilen sınırlama kutusu ve gerçek sınırlama kutusunun kesişim oranına göre hesaplanır. IoU metriği, bir nesnenin görüntüde ne kadar doğru bir şekilde tahmin edilebildiğini anlamamıza yardımcı olması nedeniyle nesne tespiti modellerinin performansını değerlendirmede sıklıkla kullanılır.



Şekil 4.2.1.1 İşlem Oranı Metriği Görsel Anlatımı

AP Ortalama hassasiyet, nesne tespiti modellerinin performansını değerlendirmede kullanılan bir metriktir. Bu metrik, ilk olarak farklı güvenilirlik eşiklerinde hassasiyet ve recall değerlerini hesaplamak suretiyle hesaplanır ve daha sonra bu değerlerin eşikler aralığındaki ortalaması alınır. Ortalama hassasiyet metriği, farklı nesne tespiti modellerini karşılaştırmada yararlı bir araçtır ve iyileştirme alanlarını belirleyebilir. Bu metrik, işlem oranı (IoU) metriği ile birlikte kullanıldığında, bir modelin performansı hakkında daha tam bir resim sunabilir.

$$AP = \int_0^1 p(r)dr$$

mAP@IoU, ya da intersection over union değerine göre ortalama hassasiyet, nesne tespiti modellerinin performansını değerlendirmede kullanılan bir metriktir. Bu metrik, modelin farklı IoU eşiklerindeki genel hassasiyetini ölçer. mAP metriği, her sınıf için ortalama hassasiyeti hesaplamak suretiyle hesaplanır ve sonra bu değerlerin tüm sınıflar için ortalaması alınır. Bu metrik, bir modelin genel performansı hakkında anlamaya yardımcı olur ve iyileştirme alanlarını belirleyebilir.

4.2.2 - Öneri Sistemi

Precision@K, öneri sistemlerinin performansını değerlendirmede kullanılan bir metriktir. Bu metrik, ilk olarak sistem tarafından önerilen en iyi K öğeyi seçmek suretiyle hesaplanır ve daha sonra bu öğelerin kullanıcıya uygun olan bölümünü ölçer. Örneğin, sistem 10 öğeyi önerir ve bunlardan 5'i kullanıcıya uygundur, precision@10 değeri 0.5 olur. Bu metrik, bir öneri sisteminin kullanıcılara

uygun öneriler sunabilme yeteneğini anlamamıza yardımcı olur. Sıklıkla recall@K metriği ile birlikte kullanılır ve bir öneri sisteminin performansı hakkında daha tam bir resim sunar.

Recall@K, öneri sistemlerinin performansını değerlendirmede kullanılan bir metriktir. Bu metrik, ilk olarak sistem tarafından önerilen en iyi K öğeyi seçmek suretiyle hesaplanır ve daha sonra öneriler içinde bulunan tüm uygun öğelerin bölümünü ölçer. Örneğin, sistem 10 öğeyi önerir ve toplamda 20 uygun öğe vardır, ancak öneriler içinde sadece 5 tanesi bulunur, recall@10 değeri 0.25 olur. Bu metrik, bir öneri sisteminin kullanıcı için tüm uygun öğeleri yakalama yeteneğini anlamamıza yardımcı olur. Sıklıkla precision@K metriği ile birlikte kullanılır ve bir öneri sisteminin performansı hakkında daha tam bir resim sunar.

$$P = \frac{\text{Önerilen ürünler içerisinde ilgili olan ürün adedi}}{\text{Önerilen toplam ürün adedi}}$$

$$r = \frac{\text{Önerilen ürünler içerisinde ilgili olan ürün adedi}}{\text{Toplam ilgili ürün adedi}}$$

4.3 Deney Sonuçları

Bu bölümde gerçekleştirilen deneylerin sonuçları paylaşılıp değerlendirilecektir ve iki alt başlıktan oluşacaktır.

1. Gömülü Temsil Öğrenim Performans Deneyi

   Gömülü Temsil Öğrenimi Performans deneyi Gömülü Temsil Öğrenim Modülü ile Benzerlik Modülünün birleşimi ile tasarlanmıştır.Deney öğrenilen vektörel temsillerin Benzerlik Modülü sonrası performansının analizi ile gerçekleştirilir.

2. Nesne Tespiti Performans Deneyi

   Çıkarım Modülü içerisinde gerçekleşen işlemin performans analizi ile gerçekleşir.

4.3.1 - Gömülü Temsil Öğrenim Performans Deneyi

Bu kısımda, farklı mimariler kullanılarak triplet kayıp fonksiyonu üzerinde eğitilen ürün vektörel temsillerinin benzerliğini öğrenebilmek için uygulanan metod açıklanacaktır. Tripletler, bir kişinin rastgele bir ortamda giydiği giysiyi içeren street2shop veri kümesinden bir görüntü ile aynı giysiyi içeren bir katalog görüntüsünü eşleştirerek oluşturulur. Bu şekilde elde edilen çiftler, semantik olarak benzer ancak aralarında ciddi değişiklikler bulunan "sıkı çekirdek-pozitif çiftler" oluşturmada yardımcı olur. Tripletleri tamamlamak için aynı makale türünde farklı bir giysi içeren bir negatif görüntü rasgele örneklenebilir. Bu, triplet ağının görsel görünümlerinde ciddi değişikliklere rağmen semantik benzerliği tanımasını sağlar.

Precision at K (P@K) ve Recall at K (R@K), bir yöntemin performansını nicel olarak değerlendirmek için kullanılan değerlendirme metrikleridir. Bu metrikler, değerlendirmede dahil edilen öğelerin sayısını temsil eden farklı K değerlerinde hesaplanır. P@K, yöntem tarafından önerilen ilk K öğenin içindeki ilgili öğelerin yüzdesini ölçerken, R@K, yöntem tarafından önerilen ilk K öğelerin içindeki ilgili öğelerin yüzdesini ölçer. Bu metrikler, yöntemin doğruluğunu ve tamamını değerlendirmek için bir yol sağlar.

Eğitim sürecinde, NVIDIA Tesla K80 GPU üzerinde 0.0001 öğrenme hızı ve 256'lık bir batch boyutu ile ADAM optimizatörü kullanılmıştır. ADAM, derin öğrenme modellerinin eğitiminde yaygın olarak kullanılan bir stokastik gradient inişi optimizasyon algoritmasıdır. Öğrenme hızı, eğitim sırasında gradient yönünde atılacak adımın boyutunu belirlerken, batch boyutu her eğitim iterasyonunda kullanılan eğitim örneklerinin sayısını belirler. NVIDIA Tesla K80 GPU'nun kullanımı, derin öğrenme modelinin verimli ve hızlı bir şekilde eğitimini sağlar. Bununla birlikte her görsel için belirlenen gömülü temsil vektörü boyutu 128'dir.

Oluşturulan Annoy dizini için tercih edilen n_trees(dizin ormanı için kullanılacak ağaç sayısı) 500 olarak belirlenmiştir.

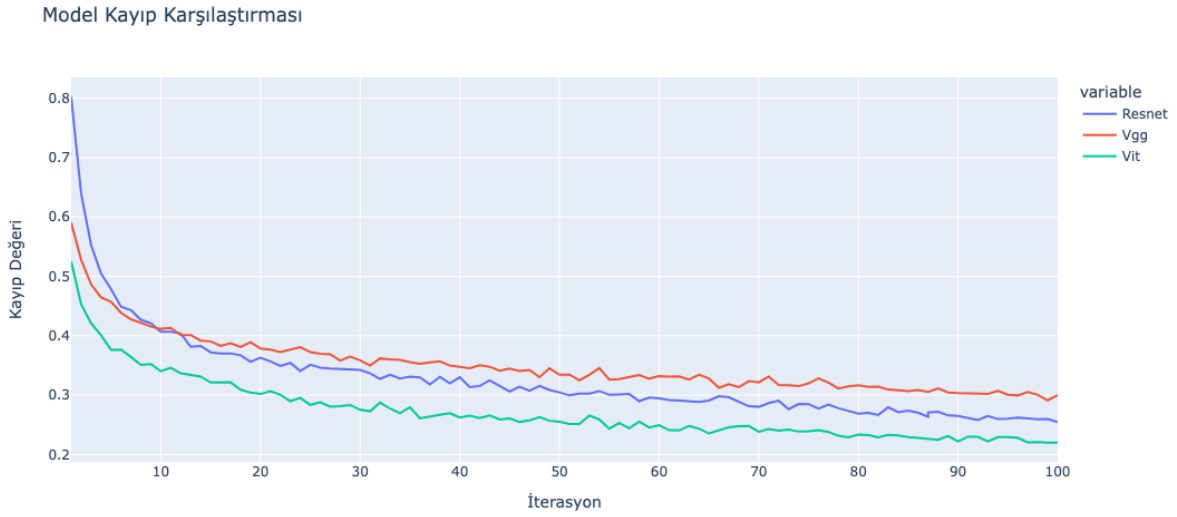
Deney içerisinde kullanılan Resnet18, Vgg16 ve Vision Transformer modeli yukarıda belirtilen şartlarda ve belirtilen yöntemlerle 100'er epoch eğitilmiştir.Eğitim sürecinde alınan kayıp fonksiyon değerleri tabloda belirtilmiştir.

İterasyon	Resnet	Vgg	Vit
1	0,8027	0,5905	0,5254
25	0,3511	0,3723	0,2836
50	0,3047	0,3343	0,2550
75	0,2848	0,3198	0,2389
100	0,2544	0,3000	0,2200

Tablo 4.3.1.1 Model ve İterasyonlara Göre Kayıp Değerleri

Kayıp eğrileri, eğitim sırasında farklı modellerin performansını karşılaştırmak için yararlı bir araçtır. Her modelin kayıp eğrisini çizerek, zamanla kaybın nasıl değiştiğini görebiliriz, bu da en iyi modeli belirlemeye yardımcı olur. Farklı modellerin kayıp eğrilerini karşılaştırarak, görev için en iyi performansı sergileyecek olan modeli belirleyebilir ve seçilen modelin performansını iyileştirmek için uygun adımlar atabiliriz. Bu, daha doğru ve etkili bir makine öğrenimi modeli oluşturmamıza yardımcı olur.

Grafik 6, yöntemimizin gömme öğrenme bileşenindeki temel mimarinin değişiminin, her epochta kayıp fonksiyonu bazında değerlendirilmesini göstermektedir



Grafik 4.3.1.2 Model Kayıp Karşılaştırması Grafiği

Kayıp eğrileri farklı modellerin performansını karşılaştırmak için kullanılan tek yöntem değildir. Diğer metrikler, bu amaçla kullanılabilenler arasında, k değerine göre doğruluk ve geri çağırma sayıları bulunur. Bu metrikler, sınıflandırma modellerinin performansını değerlendirmek için kullanılır. Farklı modellerin k değerindeki doğruluk ve geri çağırma sayılarını karşılaştırarak, görev için en iyi performansı gösterecek modeli belirleyebiliriz.

Öncelikle, hem eğitim hem de test kümesi için precision@k ve recall@k hesaplanabilir. Eğitim kümesinde, bu ölçümler modelin eğitim sırasında gördüğü veriler üzerindeki performansını değerlendirmek için kullanılabilir, test kümesinde ise modelin görmediği verilere yönelik performansını değerlendirmek için kullanılabilir.

Önemli bir nokta, precision@k ve recall@k'nın k değerinin seçimi tarafından etkilenebildiğidir. k değeri daha yüksek olursa, genellikle recall@k daha yüksek olur, ancak precision@k daha düşük olur. Diğer taraftan, k değeri daha düşük olursa, precision@k daha yüksek olur, ancak recall@k daha düşük olur. Precision ve recall arasındaki dengeyi belirlemek burada önemli bir tercihtir.

Metrikleri değerlendirmenin önemli bir koşulu ise, dizayn edilecek öneri sisteminin kullanımm hedefidir, eğer hedef kullanıcılara yapılan ilgili önerilerin sayısını maksimize etmekse, yüksek bir

recall değeri önemli olacaktır. Diğer taraftan, eğer hedef kullanıcılara yapılan ilgisiz önerilerin sayısını minimize etmekse, yüksek bir precision önemli olacaktır.

4.3.1.1 Eğitim Veri Kümesi

Model	AP@3	AP@5	AP@7	AP@10
Resnet18	0,3826	0,2981	0,2353	0,1746
Vgg16	0,3344	0,2015	0,1446	0,1018
Vit	0,3973	0,3294	0,2711	0,2116

Tablo 4.3.1.1.1 Eğitim Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Precision Değerleri

Model	AR@3	AR@5	AR@7	AR@10
Resnet18	0,170904	0,208125	0,222812	0,232671
Vgg16	0,164904	0,165392	0,165878	0,166532
Vit	0,180931	0,235451	0,263096	0,28776

Tablo 4.3.1.1.2 Eğitim Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Recall Değerleri

Hem precision hem de recall değerleri farklı K tresholdları göz önüne alındığında Vision Transformer modeli Evrişimli Yapay Sinir Ağı tabanlı mimarilerden daha iyi sonuçlar elde etmiştir

Model	Kategori	AP@3	AP@5	AP@7	AP@10
Resnet18	dresses	0,401834	0,304136	0,237748	0,174221
Resnet18	footwear	0,471827	0,363653	0,280601	0,203702
Resnet18	outerwear	0,3044	0,214105	0,15908	0,112948
Resnet18	pants	0,340319	0,299002	0,249572	0,19516
Resnet18	skirts	0,33575	0,293363	0,24909	0,195594
Resnet18	tops	0,343295	0,239253	0,177217	0,125259
Vgg16	dresses	0,334305	0,201421	0,144367	0,101457
Vgg16	footwear	0,334812	0,201989	0,145392	0,102845
Vgg16	outerwear	0,333695	0,200796	0,143684	0,100796
Vgg16	pants	0,333999	0,201397	0,144568	0,101697
Vgg16	skirts	0,334697	0,201896	0,144929	0,102082
Vgg16	tops	0,334195	0,200862	0,143555	0,100833
Vit	dresses	0,432198	0,363454	0,306353	0,24591

Vit	footwear	0,473714	0,372434	0,290784	0,213936
Vit	outerwear	0,324653	0,243617	0,191062	0,144846
Vit	pants	0,34348	0,313373	0,270388	0,220908
Vit	skirts	0,338105	0,31794	0,280907	0,230191
Vit	tops	0,378927	0,289885	0,227586	0,172328

Tablo 4.3.1.1.3 Eğitim Veri Kümesi İçin Kategori Bazında Modellerin Precision Değerleri

Kategori özelinde Resnet18 , Vgg18 ve VİT modellerinin precision değerlerini karşılaştırmak için, ilgili kategori özelinde model sütunu "Resnet18" ,"Vgg18",VİT olan satırlardaki precision değerlerine bakılır. Elde edilen sonuçlara göre, Tüm kategorilerde VİT modelinin Resnet18 ve Vgg16 modellerine göre tüm precision seviyelerinde daha yüksek precision değerlerine sahip olduğu görülüyor.

category	**AP@3**	**AP@5**	**AP@7**	**AP@10**
dresses	0,389446	0,28967	0,229489	0,173863
footwear	0,426784	0,312692	0,238926	0,173494
outerwear	0,320916	0,219506	0,164609	0,11953
pants	0,339266	0,271257	0,221509	0,172588
skirts	0,336184	0,271066	0,224975	0,175956
tops	0,352139	0,243333	0,182786	0,132807

Tablo 4.3.1.1.4 Eğitim Veri Kümesi İçin Kategorilerin Precision Değerleri

Kategori özelinde incelendiğinde precision değerlerinin ortalaması değerlendirildiğinde en “footwear” kategorisinin precision@3,precision@5,precision@7 metriklerinde en başarılı kategori olarak görülürken ,”outerwear” kategorisinin tüm precision threshold metriklerinde en az başarılı kategori olduğu görülmektedir. Bu sonucun nedenlerinden bir tanesi outerwear kategorisi içerisinde olan ürünlerin nesne tespiti sürecinden geçtikten sonra dahi farklı kategoriden ürünleri kapsayabilir olma potansiyeli olabilir.

4.3.1.2 - Test Veri Kümesi

Model	AP@3	AP@5	AP@7	AP@10
Resnet18	0,4935	0,3617	0,2760	0,1999
Vgg16	0,3382	0,2081	0,1512	0,1092
Vit	**0,5282**	**0,4493**	**0,3755**	**0,2969**

Tablo 4.3.1.2.1 Test Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Precision Değerleri

Model	AR@3	AR@5	AR@7	AR@10
Resnet18	0,2023	0,2274	0,2363	0,2419
Vgg16	0,1511	0,1538	0,1557	0,1595
Vit	**0,2256**	**0,2966**	**0,3376**	**0,3743**

Tablo 4.3.1.2.2 Test Veri Kümesi İçin Farklı Öneri Adedi İçerisinde Modellerin Recall Değerleri

Test veri kümesi incelendiğinde hem precision hem de recall değerleri farklı K tresholdları göz önüne alındığında Vision Transformer modeli Evrişimli Yapay Sinir Ağı tabanlı mimarilerden daha iyi sonuçlar elde etmiştir

Model	Kategori	AR@3	AR@5	AR@7	AR@10
Resnet18	dresses	0,5000	0,3435	0,2469	0,1746
Resnet18	footwear	0,5852	0,4188	0,3114	0,2198
Resnet18	outerwear	0,3729	0,2701	0,1937	0,1358
Resnet18	pants	0,4700	0,3871	0,3236	0,2466
Resnet18	skirts	0,5487	0,4095	0,3167	0,2299
Resnet18	tops	0,4058	0,2599	0,1872	0,1333
Vgg16	dresses	0,3370	0,2070	0,1484	0,1061
Vgg16	footwear	0,3396	0,2116	0,1543	0,1123
Vgg16	outerwear	0,3351	0,2022	0,1456	0,1062
Vgg16	pants	0,3403	0,2122	0,1543	0,1113
Vgg16	skirts	0,3399	0,2075	0,1529	0,1108
Vgg16	tops	0,3343	0,2040	0,1473	0,1048

Vit	dresses	0,5739	0,5048	0,4314	0,3485
Vit	footwear	0,5920	0,4591	0,3504	0,2602
Vit	outerwear	0,4394	0,3499	0,2926	0,2372
Vit	pants	0,4791	0,4286	0,3723	0,2979
Vit	skirts	0,5590	0,5240	0,4586	0,3664
Vit	tops	0,4896	0,3723	0,2930	0,2311

Tablo 5.3.1.2.3 Test Veri Kümesi İçin Kategori Bazında Modellerin Precision Değerleri

Test veri seti içerisinde kategori özelinde Resnet18 , Vgg18 ve VİT modellerinin precision değerlerini karşılaştırmak için, ilgili kategori özelinde model sütunu "Resnet18" ,"Vgg18",VİT olan satırlardaki precision değerlerine bakılır.

Elde edilen sonuçlara göre, test veri seti üzerinde, tüm kategorilerde VİT modelinin Resnet18 ve Vgg16 modellerine göre tüm precision seviyelerinde daha yüksek precision değerlerine sahip olduğu görülmektedir.

Kategori	**AR@3**	**AR@5**	**AR@7**	**AR@10**
dresses	0,4703	0,3517	0,2756	0,2097
footwear	0,5056	0,3632	0,2720	0,1974
outerwear	0,3825	0,2740	0,2106	0,1597
pants	0,4298	0,3427	0,2834	0,2186
skirts	0,4825	0,3803	0,3094	0,2357
tops	0,4099	0,2787	0,2092	0,1564

Tablo 4.3.1.2.4 Test Veri Kümesi İçin Kategorilerin Precision Değerleri

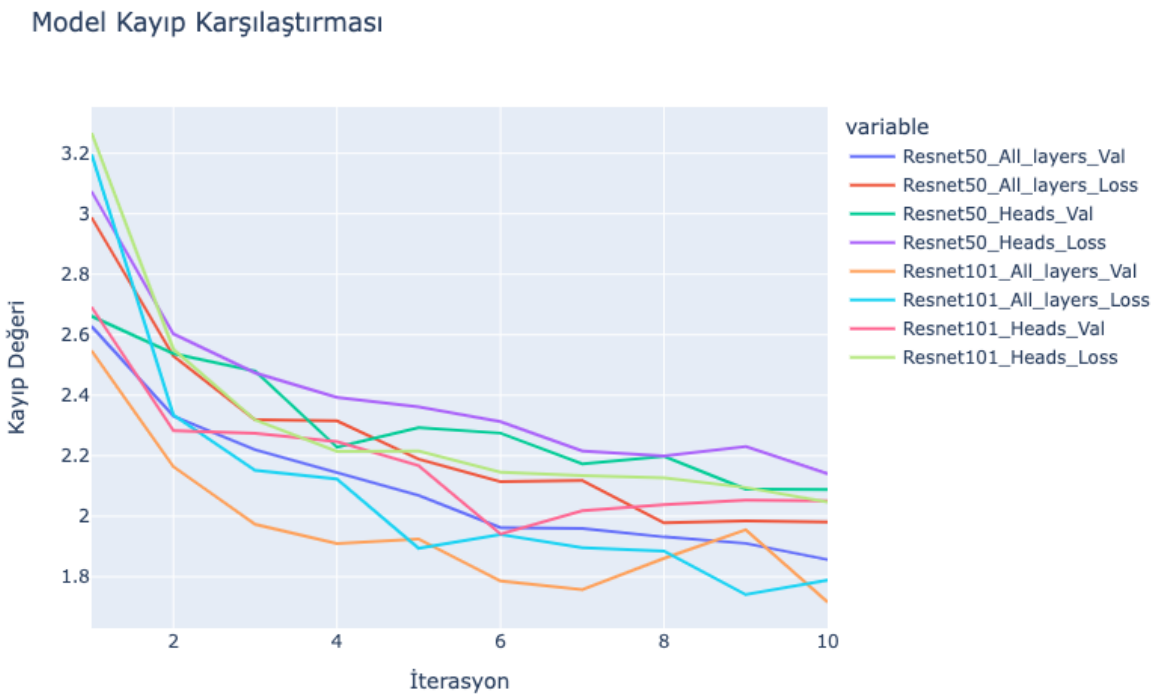
Her kategoride, ilgili doğruluk değeri, o kategoride yapılan tüm pozitif tahminlerin yüzdesine karşılık gelen doğru pozitif tahminlerin yüzdesini gösterir. Örneğin, "elbiseler" kategorisi için, en az 3 pozitif tahmin yapması gereken modelin doğruluk oranı 0.47028985507246374 idi. Bu, modelin "elbiseler" kategorisi için yaptığı tüm pozitif tahminlerin yüzdesinin %47'sinin doğru pozitif tahminler olduğu anlamına gelir.

Test veri seti üzerinde , Kategori özelinde precision değerlerinin ortalaması değerlendirildiğinde precision@3 te en başarılı kategori “footwear” precision@5,precision@7 ve precision@10 da skirts kategorisi dir

4.3.2 Nesne Tespiti Performans Deneyi

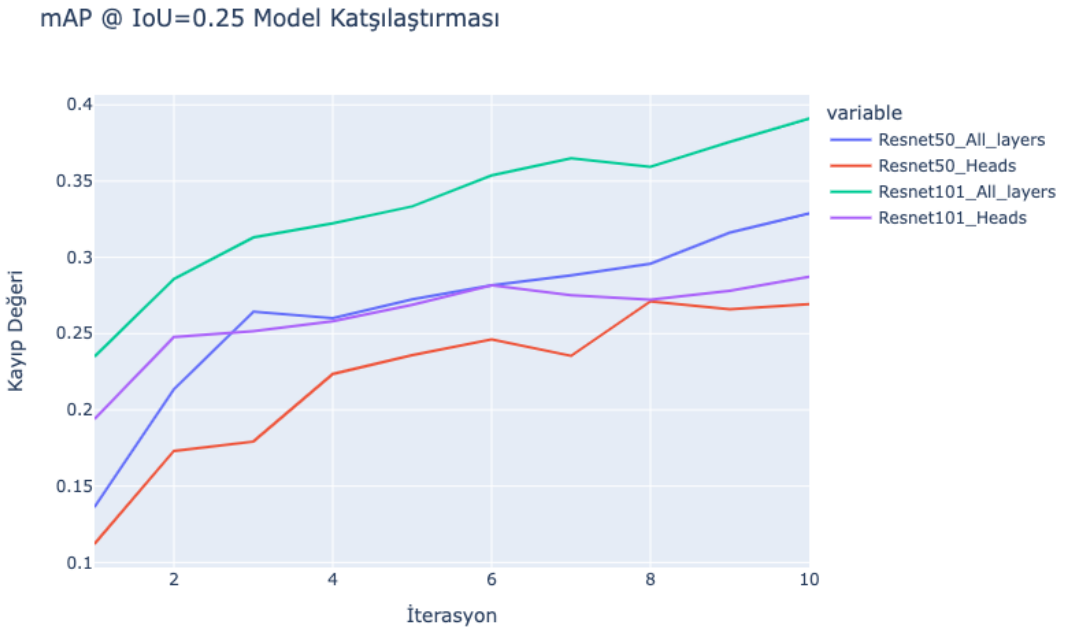
Bir nesne algılama modelinin bir veri kümesindeki performansı, algılanan nesnelerin karmaşıklığı, eğitim verisinin kalitesi, model mimarisi ve kullanılan optimize ve eğitim prosedürü gibi çeşitli faktörlere bağlıdır.

Mask R-CNN modeli resnet50,resnet 101 olmak üzere iki farklı backbone üzerinden iki farklı fine tune yönteminin oluşturduğu tüm kombinasyonlarda eğitimli ve iMaterialist veri kümesindeki performansını değerlendirmek için ortalama hassasiyet (mAP) değerinin farklı IOU tresholdları altında değerlendirilmiştir.

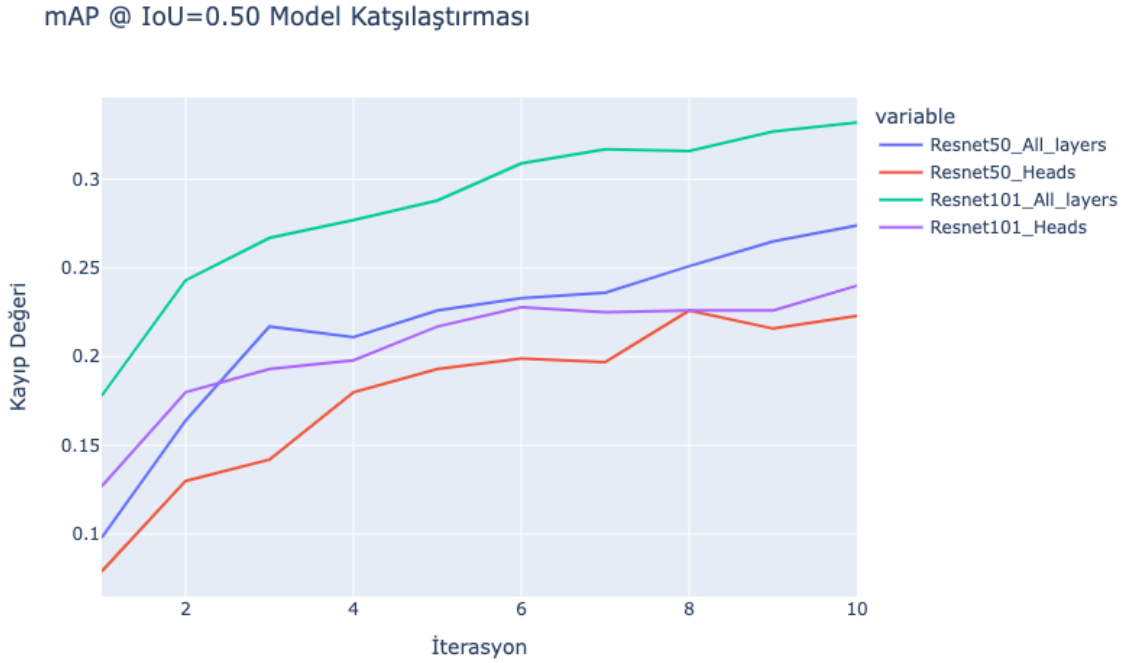


Grafik 4.3.2.1 Model Kayıp Değeri Karşılaştırması

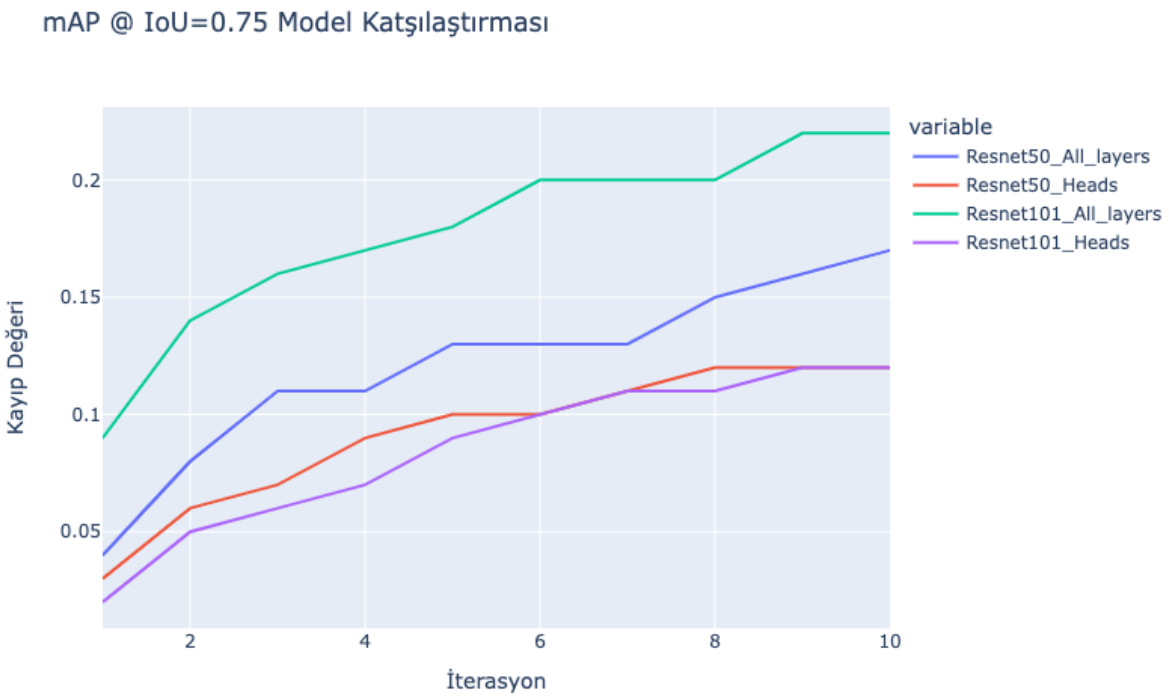
Tüm kombinasyonları eğitim ve test seti üzerinde aldığı sonuçlar görselde belirtilmiştir. Alınan sonuçlara göre tüm layerların öğrenilmiş parametreler üzerinden tekrardan öğrenime devam ettiği ce resnet101 backbone'una sahip model kombinasyonu en iyi sonucu vermiştir.



Grafik 4.3.2.2 mAP @ IoU=0.25 İçin Model Karşılaştırması



Grafik 4.3.2.3 mAP @ IoU=0.50 İçin Model Karşılaştırması



Grafik 4.3.2.4 mAP @ IoU=0.75 İçin Model Karşılaştırması

Grafik 5.3.2.2, grafik 5.3.2.3 ve grafik 5.3.2.4'te gösterilen sonuçlar üzerine, tüm IoU tresholdlarında map en yüksek map değerine sahip olan kombinasyon resnet101 backbone ile birlikte tüm layerların ilave eğitimle tekrardan eğitildiği eğitim biçimidir.

grafikte, modelin "map" değerleri ("mean average precision") farklı "iou threshold" ("intersection over union") değerlerinde değiştirilirken gösterilmiştir. Ayrıca, bu grafiğe göre, modelin "backbone" olarak "resnet101" kullanıldığı durumda ve tüm "layerlar"ın ilave olarak eğitildiği durumda, map ("mean average precision") değeri en yüksektir.

Bunun anlamı, modelin "backbone" olarak "resnet101" kullanıldığı ve tüm "layerlar"ın ilave olarak eğitildiği durumda, modelin nesne tanıma performansı diğer kombinasyonlara göre daha iyidir. "Iou threshold" değerlerinin değiştirilmesi, nesne tanıma modelinin nasıl "doğru" olduğunu ölçmek için kullanılan bir metriktir ve daha yüksek iou threshold değerleri, modelin daha yüksek bir doğruluk seviyesi gerektirdiği anlamına gelir. Bu nedenle, modelin map değerlerinin daha yüksek olduğu durumlarda, modelin daha yüksek iou threshold değerlerinde daha doğru çalıştığı anlamına gelir.

5 - SONUÇ

Tez çalışmasının amacı,ürün sayfası içerisinde bulunan görselin kapsadığı tüm ürün gruplarına benzer ürün önerebilecek bir yapay öğrenme sistemi tasarlamaktır. Bu problemi çözmek için 3 aşamadan oluşan bir yapay öğrenme sistemi tasarlanmıştır. Her aşamada ilgili problem üzerine odaklanan yapay öğrenme modelleri geliştirilmiştir. Tez çalışması, literatürdeki araştırmalardan farklı olarak az sayıda veri ve az sayıda iterasyon üzerindeki performansı ortaya çıkarmakla birlikte,transformer mimarisine sahip bir modeli moda öneri sistemi probleminde değerlendirmektedir. Bu şekilde sadece evrişimli yapay sinir ağları değil, transformer mimarisinin de moda kategorisine ait bir veride performansı değerlendirilmiştir.

Yapılan deneysel çalışmalarda; Her aşamada en iyi sonuç veren modeller ve model hiperparametreleri belirlenmiş ve oluşturulmak istenen sisteme entegre edilmiştir.

Önerilen sistemin ilk aşamasında MaskRCNN modeli temel model olarak ele alınıp,farklı backbone ve ince ayar teknikleri ile test edilmiştir. İlgili deneyler sonucunda resnet101 backbone ile tüm katmanların önceden eğitilmiş ağırlıkların yeni veri üzerinden ince ayar yapılması tekniğiyle ilgili metriklerde en yüksek sonucu aldığı ortaya çıkmıştır.

Önerilen sistemin ikinci ve üçüncü aşamalarında, Resnet19, Vit, Vgg16 modelleri Cifar veri setinde eğitilmiş parametreleri üzerinden sadece sınıflandırma katmanının yeni veri ile eğitilmesi tekniği ile eğitilip karşılaştırılmıştır. Deney sonucunda Transformer tabanlı mimari olan Vision Transformer modelinin, evrişimli yapay sinir ağları tabanlı modellere göre, ilgili metriklerin tamamında daha iyi sonuçlar verdiği gözlemlenmiştir.

Literatür benzerlerine göre daha az sayıda veri kümesi ve daha az sayıda iterasyon ile gerçekleştirilen deney, özellikle teknik donanım yetersizliğinin karşılaşıldığı durumlarda,deney sonucunda önerilen yöntemin yüksek sayıda veri ve iterasyon ile aynı başarıyı sağlayamadığı ancak, farklı modellerin test edilebilmesi noktasında önemli bir araç olarak kullanılabileceğini göstermiştir.

Bu araştırma belirli mimariler üzerinden tasarlanmış olup, güncel mimariler her geçen gün yenilenmektedir.Tüm modüller için daha güncel mimarilerin ilgili görev düzleminde denenmesi, optimum kombinasyonu bulmak açısından önem arz edecektir.Bununla birlikte, çalışmada vision transformer modelinin veri alt kümesinde diğer mimarilere nazaran daha iyi performans gösterdiği belirlenmiştir ancak tüm veri içerisindeki performansının değerlendirilmesi gelecek iş olarak çalışmanın önünde durmaktadır.

6 - KAYNAKLAR

"Annoy Library." *url{https://github.com/spotify/annoy}*, https://github.com/spotify/annoy. Accessed 23 December 2017.

Black, Samuel, et al. "Visualizing Paired Image Similarity in Transformer Networks." *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 1534-1543.

"CS231n Convolutional Neural Networks for Visual Recognition." *CS231n Convolutional Neural Networks for Visual Recognition*, https://cs231n.github.io/transfer-learning/. Accessed 23 December 2022.

Dosovitskiy, Alexey, et al. "AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE." *ICLR 2021*, 2021.

Goodfellow, Ian, et al. *Deep Learning*. MIT Press, 2016.

He, Kaiming, et al. "Deep residual learning for image recognition." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

He, Kaiming, et al. "Mask R-CNN." *2017 IEEE International Conference on Computer Vision (ICCV),*, 2017, pp. 2980-2988.

"iMaterialist (Fashion) 2019 at FGVC6." *Kaggle*, 11 July 2020, https://www.kaggle.com/competitions/imaterialist-fashion-2019-FGVC6/overview. Accessed 23 December 2022.

Kalantidis, Yannis, et al. "Getting the Look: Clothing Recognition and Segmentation for Automatic Product Suggestions in Everyday Photos." *ICMR '13: Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, 2013.

Kiapour, M. Hadi, et al. "Where to Buy It: Matching Street Clothing Photos in Online Shops." *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3343-3351.

Magelssen, Brittany. “Recommended for you: Role, impact of tools behind automated product picks explored: Pros, cons of recommender systems.” *ScienceDaily*, 4 March 2021, https://www.sciencedaily.com/releases/2021/03/210304145157.htm. Accessed 21 December 2022.

O’Shea, Keiron, and Ryan Nash. “An Introduction to Convolutional Neural Networks.” *arXiv: Neural and Evolutionary Computing*.

Ravi, Abhinav, et al. “Buy me that look: An approach for recommending similar fashion products.” *2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2021.

Resnick, Paul, et al. “GroupLens: An Open Architecture for Collaborative Filtering of Netnews.” *ACM 1994 Conference on Computer Supported Cooperative Work, Chapel Hill*, 1994, pp. 175-186.

Russakovsky, Olga, et al. “ImageNet Large Scale Visual Recognition Challenge.” *International Journal of Computer Vision*, 2015.

Schrage, Michael. “Great Digital Companies Build Great Recommendation Engines.” *Harvard Business Review*, 1 August 2017, https://hbr.org/2017/08/great-digital-companies-build-great-recommendation-engines. Accessed 21 December 2022.

Simonyan, Karen, and Andrew Zisserman. “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION.” *t ICLR 2015*.

Vasilev, Ivan. *Advanced Deep Learning with Python: Design and Implement Advanced Next-Generation AI Solutions Using TensorFlow and Pytorch*. Packt Publishing, 2019. Accessed 22 December 2022.

Vaswani, Ashish, et al. “Attention Is All You Need.” *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, no. 30, 2017, pp. 6000–6010.

Veit, Andreas, et al. “Conditional Similarity Networks.” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Yamaguchi, Kota, et al. “Parsing clothing in fashion photographs.” *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, 3570-3577,.

Yao, Guanwen, and Lifeng Cai. “User-Based and Item-Based Collaborative Filtering Recommendation Algorithms Design.” 2015.

Yosinski, Jason, et al. “How transferable are features in deep neural networks?” *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, no. 2, 2014.