

KONVOLÜSYONEL AĞLARLA YÜZ İFADE KARAKTERİSTİĞİ TANIMLAMA DERİN SİNİR AĞLARI VE UYGULAMALARI DERSİ

FİNAL RAPORU

22833301009 - Ceren Umay Özten

2. ÖZET

Yüz ifade karakteristiği tanımlama (FER) görevi proje uzanımında birtakım ifadesel piksel tabanlı görüntüden duygusal karşılık bağlamının etiket olarak (happy, sad, disgust, anger, contempt gibi) döndürülmesidir. Bağlantılı görüntü piksellerinin konvolüsyonel sinir ağlarında düzlemsel değerlendirilerek soyutlanması amaçlanmaktadır, paralelinde gerçek ifadesel duygusal eşleştirmelere ait doğruluk oranının artırılması düşünülmektedir. Örneğin bir profil fotoğrafı projenin konvolüsyonel sinir ağı mimarilerinden geçirimi sırada soyutlaştırmanın olasılık dağılımına yönlendirmesi incelenmektedir. Bu sebeple proje aksanı görsel sınıflandırmanın yaygın kullanımlarından konvolüsyonel derin sinir ağı katmanlarından oluşmaktadır, Dropout ara katmanı ile regularization hiper parametrelili kontroller (Earlystopping, weight decay, data augmentation gibi) takip etmektedir. Proje çalışması, aynı zamanda özgün konvolüsyon katman mimarilerinin Transfer Learning geliştirmesini (VGG16) daha ileri mimarilere taşınmasını (VGG19, InceptionV3, DenseNet121, EfficientNetB0) araştırmaktadır. Transfer Learning blok mimarilerinin kök katmanları dondurularak yüz hatlarının kenar, renk benzeri çizgisel kesişimleri “imagenet” ağırlık kalıbıyla korunmaktadır, özellikle özgün son fully-connected katmanları ise -Dropout regularization ara katmanlarıyla fazlalık aktivasyon düğümleri düşürülür- ağırlık parametre değişimli eğitime tabi tutulmaktadır. Farklı Transfer Learning mimarilerinin eğitimi ve paralelinde özgün fully-connected katman bloğuna hiper parametre kontrolleriyle uyarlanması denenmektedir. Yüz ifade görüntüleri konvolüsyonel derin sinir ağlarına tanıtılarak anger, disgust, fear, happy, neutral, sad, surprise 7 duygusal bağlamından birine atanmaktadır, yüz ifade karakteristikleri belirlenmektedir. Konvolüsyon katman hiyerarşisi, basit yüz hat ifadesel soyutlaştırmalardan birbirini tamamlayan karakteristik yüz bölgesel işaretlemelerine ve en son sınıflandırma etiketine kadar birleşimlerini sürdürmektedir. Sonuçta proje çözümü yaygın FER2013 yüz ifade karakteristiği tanımlama veri setinde eğitilmiştir ve FERPlus facial expression recognition veri setinde test edilmiştir. Aynı zamanda gerçek bir ham vesikalık fotoğrafında test gerçekleştirilmiştir.

3. PROBLEMİN TANIMI

Duygusal zeka analizinin önemli uzantılarından yüz ifade karakteristiği günlük insan karşılıklı iletişiminin ayrılmaz parçasıdır. Mental sağlık, eğlence sektörü, autism spectrum disorder araştırmaları, sosyal yazılım uygulamaları, devlet programları yüz ifade karakteristiği belirtiminin otonom programsal duygusal zeka analizine çözdürmeyi amaçlamıştır. Makineye ifadesel duygusal zekanın tanıtılması yaygın kullanım alanlarına uyarlanarak gerçek dünya sürecine ilerletilmiştir. Örneğin Disney şirketi, film seyreden insanların yüz ifadelerini, Factorized Variational Autoencoder (FVAE) adı verilen bir yöntemle izlemektedir, ardından alınan bilgiler çerçevesinde on dakika duygusal zeka analizi yapılarak insanların nasıl tepki verecekleri tahmin edilmektedir. FVAE yöntemi, izleyicilerin yüz ifadelerini belirli özelliklere göre sınıflandırmaktadır ve ortaya bir rakam çıkarmaktadır. Ve rakam yüzün gülmesi, göz ifadesel aralıkları gibi veriler bu rakamın oluşmasına katkıda bulunmaktadır. Disney ekibi, birçok farklı filmde 3000 izleyiciye FVAE uygulamıştır. Project Oxford (2015) tasarısında insanlar fotoğraflarını yollayarak duygu incelemesinde bulunmuştur. O zamandan bu zamana teknolojinin gelişimiyle artık yüz ifadeleri, gerçek zamanlı analizle sonuçlandırılabilmiştir. ABD ulaştırma güvenlik sistemi ise biyometrik yüz tanımlama sistemi programı için testler yürütmüştür. Bu tür bir sistemin devreye girmesiyle yapay zeka desteği tabanında yolcular arasında potansiyel terörist şüphelilerini tespit etmeyi denemek için kullanılabileceği tahmin edilmiştir. Aynı zamanda teknoloji firmaları, duygusal tanımlamadaki gelişimden yararlanmaktadır. Örneğin Apple telefon şirketine ait iPhoneX telefonuna entegreli Animoji uygulaması, telefon ekranına bakanın yüz ifade mimiklerini başarılı şekilde taklit edebilmektedir. Autism Spectrum Disorder tanılı insanların duygusal zeka analizi, yüz ifade tepkilerine dayanarak ortaya çıkarılmaya çalışılmaktadır. ASD tanımlı bireylerin nöral hafıza motor bozuklukları ya da mekanik tepkisel yönelimlerinin yavaşlığı nedenleri irdelenmiştir. Dolayısıyla duygusal gecikme bandının ASD bireylerinde daha geç zaman aralıklarında seyrettiği belirlenmiştir. Derin öğrenmenin bağlantılı bilgi kaynağından örüntü öğrenimini genelleyen birleşim yapısı, el-yapımı (hand-crafted) ile formal makine öğrenimini geçmektedir. Disney derin öğrenmenin autoencoders yapısını ve ASD bireylerin ifadesel tepki araştırmaları, konvolüsyonel derin sinir ağlarıyla gerçekleştirilmiştir. Konvolüsyonel derin sinir ağlarının hiyerarşisi, yüz ifade karakteristiği tanımlamada duygusal bağlantıların ifadesel yüz bölgeleriyle ilişkilendirmektedir.

4. LİTERATÜR TARAMASI

Proje çalışmasının geçmiş çalışmanın özgün konvolüsyon katman mimarileri ile VGG16 transfer learning geliştirmesine ait temel literatür taraması ile VGG19, InceptionV3, DenseNet121, EfficientNetB0 ileri mimari transfer learning geliştirmelerine ait esas literatür taraması yürütülmüştür. Özgün konvolüsyon katman mimarileriyle bağlantılı literatür taramalarında ara katman ile konvolüsyonel blok hiyerarşileri, aktivasyon fonksiyonları, hiper parametre tercihleri, optimizasyon fonksiyonları ve transfer learning blok mimarileri ele alınmıştır ve farklı kombinasyonlarda geçmiş proje çalışması dahilinde denenmiştir. İleri transfer learning geliştirmelerine ait literatür taramasının derin konvolüsyonel yapıları, eğitim ile test veri setleri ve test yöntemleri ele alınmıştır, yüz ifade karakteristiklerinden konvolüsyonel öznitelik çıkarımı proje tabanına değerlendirilmiştir.

Geçmiş Çalışmanın Özgün Katman Mimarileri

Geçmiş özgün katman mimarilerinin ilkinde max pooling katmanı her konvolüsyonel bloğun sonuna entegre edilmiştir, Dense SIFT CNN network modeli, katman organizasyonunda yararlanılmıştır. Dense SIFT mekanizması, görüntü verisini (12x12) piksel bölgelerine ayırarak piksel gradyanlarını çıkarmaktadır ve Dense SIFT networküne girmektedir. Dense SIFT networkü, altı konvolüsyonel katmana sahiptir, konvolüsyonel katmanlar, iki-aşamalı blok bağlamında her konvolüsyonel bloğa max pooling katman entegrasyonunu kurmaktadır. Üç max pooling katman ile altı konvolüsyonel katman total mimariyi oluşturmaktadır. Dense SIFT CNN network modeli, FER-2013 veri setinde 72.1% doğruluk oranındadır [1]. İlk özgün katman mimarisinde CNN network modeli, dört konvolüsyonel katman iki aşamalı blok bağlamında her konvolüsyonel bloğa max pooling katman entegrasyonunu kurmaktadır. Dense SIFT CNN network modelinin fully-connected katmanları arası Dropout entegrasyonunu desteklemektedir.

İkincil geçmiş özgün katman mimarisinde regularization tekniği ile Dropout-BatchNormalization-maxpooling üç-aşamalı uygulaması, Shimah Alizadeh ile Azar Fazel Derin CNN networkünden galvanize edilmiştir. Derin CNN networkü L2 regularization tekniği ile Dropout-BatchNormalization entegrasyonlarını overfitting önlemine katmaktadır. Derin CNN networkünün dört konvolüsyonel katmanı ile iki fully-connected katman mimarisinde ilk konvolüsyon katmanı (3x3) filtre matrisli 64 unit nöron, ikinci konvolüsyon katmanı, (5x5) filtre matrisli 128 unit nöron, üçüncü konvolüsyon katmanı (3x3) filtre matrisli 512 unit nöron ve dördüncü konvolüsyon katmanı (3x3) filtre matrisli 512 unit nöron

mevcuttur. Her konvolüsyon katmanı, BatchNormalization-Dropout-maxpooling üç-aşamalı entegrasyondadır. İlk 256 unit nöronlu, ikincisi 512 unit nöronlu fully-connected katmanları Dropout-BatchNormalization entegrasyonludur. Regularization değeri, $1e-7$. Derin CNN network doğrulama setindeki doğruluk oranı, 65%'tir [2]. İkincil geçmiş özgün CNN mimarisi, dört konvolüsyon katmanı ile iki fully-connected katmanlıdır. Her konvolüsyonel katman çifti, Dropout-BatchNormalization-maxpooling üç-aşamalı entegrasyonludur, konvolüsyon katmanları sırasıyla 64-128-256-512 unit nöronludur ve fully-connected katmanları 128-64 unit nörona sahiptir. Tüm konvolüsyonel mimari, (3x3) matrisli ve 2 filtre adımıdır.

Üçüncül geçmiş özgün katman mimarisinde Dropout-BatchNormalization-maxpooling üç-aşamalı entegrasyonu 1024 unit nöronlu ve Dropout-BatchNormalization entegrasyonlu fully-connected katmanlara sahiptir. Stanford araştırmasının beş katmanlı mimarisine ait üç konvolüsyon katmanı ile iki fully-connected katmanı (softmax) örneklenmiştir. Üçlü konvolüsyonel katmanlar, sırasıyla 32-32-64 unit nöronu ile (5x5), (4x4) ve (5x5) filtre matrislidir. Beş katmanlı CNN model, konvolüsyonel katmanlarda BatchNormalization-maxpooling entegrasyonludur. Fully-connected katmanlara BatchNormalization entegrasyonu ile 1024 unit nöron ataması, üçüncül katman mimarisinin önemli parçasıdır. Beş katmanlı CNN modeli, FER2013 test setinde 66.3% doğruluk oranına sahiptir.

Dördüncül geçmiş özgün katman mimarisinde ASTEJ araştırmasının VGG-16 GAP CNN networküne ait Dropout-BatchNormalization-maxpooling üç-aşamalı entegrasyonun filtre matris boyut artırımı, Dropout-BatchNormalization-averagepooling ile SGD-early stopping tekniği uygulaması irdelenmiştir. ASTEJ araştırmasında VGG-16 mimarisinin softmax dışındaki fully-connected katmanları Global Average Pooling katmanıyla değiştirilmektedir. VGG-16 GAP CNN network, SGD optimizasyonu ile early stopping tekniğini taşımaktadır [3]. Dördüncül geçmiş özgün CNN modeli, tüm konvolüsyonel katmanları (5x5) filtre matris boyutuna ortaktır, fully-connected katmanların yanında averagepooling katmanına sahiptir. Early stopping 150 epoch sayısını 52 epoch sayısına indirmektedir.

Beşincil Transfer Learning Mimarisi ile Bu Raporun İleri Transfer Learning Geliştirmeleri

Beşincil transfer learning mimarisinde geçmiş özgün katman modelinin “base” iskeleti VGG16 mimariyle değiştirilmiştir ve özgün fully-connected katman mimarisi, VGG16 iskeletiyle birleştirilmiştir. Önceden eğitilmiş imageNet ağırlıkları, özgün fully-connected katman mimarisinden öznetelik çıkarımı için eğitim aktivasyonuna dahil edilmiştir. İleri

Transfer Learning geliřtirmeleri için yoğun baęlantılı konvolüsyonel sinir aęı ile hiyerarřik uzaysal attention mekanizmasını birleřtiren alıřmanın yoğun baęlantılı modülü, yüz ifade karakteristiklerinin öznitelik ıkarımında konvolüsyonel örnekleme tercih edilmiřtir. Yüz ifade karakteristiklerine ait lokal bilgi kaybının derin sinir aęında azaltılması adına özel baęlantı modu, yoğun baęlantılı modüle katılmıřtır. Baęlantılı öznitelik ıkarımında kayıp fonksiyonu cross-entropy, fully-connected softmax olasılık daęılımı düşünölmüřtür [4]. Bu raporun proje kapsamında densenet121 özgün fully-connected katman bloęu ile cross-entropy kayıp fonksiyonlu fully-connected softmax olasılık daęılımında denenmiřtir. Zorlu kořullarda yüz ifade karakteriřtięi tanımlama için görü ve geometri tabanlı transformer alıřmasında, self-attention mekanizmasının öznitelik örüntöleri, ok-boyutlu graf tabanlı transformer modölüne girilmektedir, ok-boyutlu graf tabanlı transformer modölünün BatchNormalization ile ReLU aktivasyonu, transfer learning mimarilerinde var olabilmektedir. Stochastic gradient descent (SGD), optimizasyonu transformer yapılarının eęitiminde edinilmiřtir [5]. Bu rapor arařtırmasında SGD optimizasyonu 0.01 öęrenme oranıyla edinilmektedir. Zorlu kořullarda yüz ifade karakteriřtięi tanımlama için az-veriyle öęrenme konvolüsyonel iliřki networkü alıřmasında konvolüsyonel iliřkisel öznitelik ıkarımı yeterli duygusal kategorilerin benzerlięi karřılařtırılmıřtır ve az-verili, önceden görölmemiř duygusal kategoriler tahmin edilmiřtir [6]. Rapor alıřması ise transfer learning mimarilerinin konvolüsyonel yapısına ait öznitelik ıkarımını iç-benzerliklerine göre olasılıksal daęılıma paralamaktadır.

5. YAPILAN ALIřMALAR

Transfer learning ile özgün fully-connected katman bloklarının entegresi, arařtırma raporunun ana izgisidir. Transfer learning, ImageNet gibi olduka yoğun veri ieren bilgi alanına ait önceden eęitilmiř aęırlık aktivasyonlarının daha spesifik ölekli probleme, konvolüsyonel derin sinir aęları uzanımında dahil edilmesidir -derin öęrenme-. Konvolüsyonel sinir aęı mimarilerinin yapısalıęı, özgün fully-connected katmanlarına uyarlanarak sil bařtan konvolüsyonel mimari kurulumunu gerektirmemektedir, önceden eęitilmiř aęırlıklara ait bilgi donanımı, baęlantılı sınıflandırma görevine geirmektedir. Yüz ifade karakteriřtięi tanımlama görevinde ifadesel veriler, düşük seviyeli katmanları dondurulmuř backbone transfer learning mimarisine girilmektedir ve softmax olasılık daęılımına ayrıřtırılmaktadır. İlgili yüz ifade verileri, en yüksek olasılıklı duygusal kategorileriyle eřleřtirilmektedir. Yedi temel duygu durumuna ait, angry, disgust, fear,

happy, neutral, sad, surprise, ifadesel öznitelik çıkarımının sinir nöron aktivasyonları duygusal sonuca dönüştürülmektedir.

Metodoloji

Backbone transfer learning mimarileri için sırasıyla InceptionV3, VGG19, DenseNet121, EfficientNetB0 sinir ağ yapıları tercih edilmiştir. FER2013 veri seti, bağlantılı transfer learning mimari eğitimleri için seçilmiştir, FERPlus veri seti, eğitim modellerinin testinde yararlanılmıştır. InceptionV3, VGG19 ile DenseNet121 mimarileri için Dropout uygulamalı (0.3) sırasıyla 64 unit nöron ile 7 unit nöron (softmax çıkış katmanı) özgün fully-connected katman bloğu kullanılmıştır. EfficientNetB0 mimari için 7 unit nöron softmax çıktı katmanı backbone mimarisine sonlanmıştır. Criterion kayıp fonksiyonu, DenseNet121 ile EfficientNetB0 sinir ağları için cross-entropy kayıp fonksiyonu olarak seçilmiştir. InceptionV3 ile VGG19 sinir ağları için categorical_crossentropy kayıp fonksiyonu atanmıştır. Eğitim sırasındaki optimizasyon fonksiyonu öğrenim oranı -lr- 0.01 ayarlı SGD algoritması InceptionV3, VGG19 ile DenseNet121, 0.001 ayarlı SGD algoritması EfficientNetB0 sinir ağı için tanımlanmıştır. Kaggle GPU destekli notebook ortamında InceptionV3 ile VGG19 sinir ağlarının eğitimi, NVIDIA CUDA 11.7 torchvision GPU destekli, Pytorch ortamında DenseNet121 ile EfficientNetB0 ağlarının eğitimi gerçekleştirilmiştir. InceptionV3 ile VGG19 eğitimleri 70 epoch döngüsüne, DenseNet121 10 eğitimi epoch adım döngüsüne, EfficientNetB0 eğitimi 25 epoch döngüsüne sahiptir.

VGG19, VGG16 mimarisinin derinleştirilmiş versiyonudur ve (3x3) filtre matris boyutlu, 1 filtre adımlı üçlü konvolüsyon blokları ile ardı sıra (2x2) filtre matris boyutlu, 2 filtre adımlı maxpooling katmanları yoğunluktadır. Maxpooling katmanlarını ReLU aktivasyon fonksiyonu, takip etmektedir, lineer-olmayan veri örüntüleri bu sayede öğrenilmektedir. Derin konvolüsyon katmanları, başarılı doğruluk oranları yakalayabilmektedir. (5x5) ile (7x7) filtre matrisleri yerine (3x3) filtre matrisli üçlü konvolüsyon bloklarının kullanımı hesaplama maliyetini düşürmektedir.

InceptionV3 mimarisi, derinleşmeyle overfitting probleminin önlenmesi için farklı filtre matris boyutlu katman yapıları, aynı yatay seviyede kullanılmıştır. Derin katman bloklarının yerine paralel genişleyen bir katman yapısıllığı mevcuttur. Konvolüsyon ile maxpooling katmanları, nx1, 1xn, nx1 matris boyut parçalamalarıyla 42 katman boyunca paralel genişlemeye dahil edilmiştir. (3x3) matris boyut parçalamaları parametre yoğunluğunu

düşürmüştür ve 5 milyon parametrelili InceptionV3 mimarisini daha hafif ve yüksek başarılı hale getirmiştir.

DenseNet121 mimarisi, konvolüsyonel derinleşmenin ardı sıra bağlantılarının gradyan kaybolmasını önlemek için formal CNN bağlantıları basitleştirerek her katmana ait bağlantının direkt diğer katmanlara bağlanması sağlamıştır. Bu sayede her L adet katman için $L(L + 1)/2$ direkt bağlantı mevcuttur. Öznitelik çıktılarının CNN mimarisindeki boyutsal düşürümünün gerçekleştirilmesi için DenseNet yapıları Dense bloklara parçalanmaktadır. Ve Dense blok yapılarının ardı sıra gelen (3x3) konvolüsyon yapılarının parametre yoğunluğu (1x1) bottleneck konvolüsyon katmanlarıyla bastırılmaktadır.

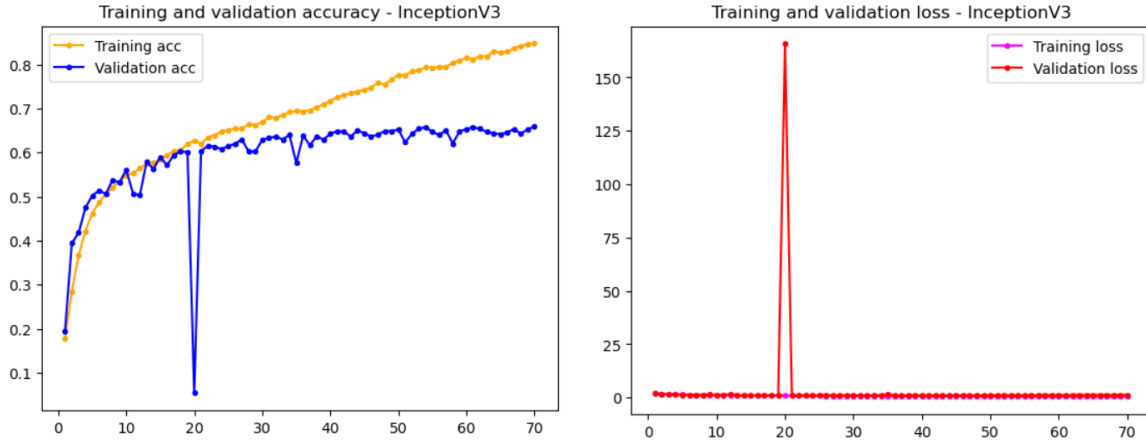
EfficientNetB0 mimarisi, parametre yoğunluğunu model genişleme (model scaling) özelliğine ait yatay genişleme, dikey derin genişleme ve veri çözünürlük genişlemesi seçeneklerini kombine ederek etkin şekilde ayarlayabilmektedir. Harici konvolüsyonel modellere göre oldukça düşük parametre sayısına rağmen daha başarılı sonuçlar döndürebilmektedir. EfficientNetB0, totalde 237 katmandan oluşmaktadır, yalnız katman yapıları basit modüler yapılarıyla derin konvolüsyon, BatchNormalization, global average pooling, Dropout, Rescaling birleşmelerini tamamlamaktadır.

Sonuçlar

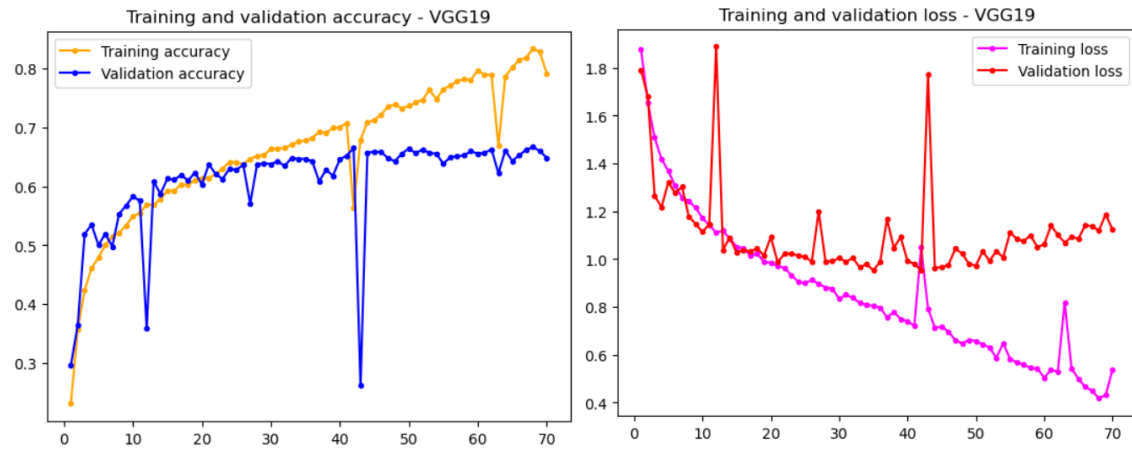
VGG19, InceptionV3, DenseNet121 ile EfficientNetB0 mimarilerinin FER2013 eğitim ile doğrulama doğruluk oranları ve FERPlus test veri setindeki doğruluk oranları dahildir.

| Backbone Mimari | FER2013 eğitim doğruluk oranı | FER2013 doğrulama doğruluk oranı | FERPlus test doğruluk oranı |
|---------------------------------|--------------------------------------|---|------------------------------------|
| InceptionV3 | 0.84 | 0.65 | 0.58 |
| VGG19 | 0.79 | 0.64 | 0.59 |
| DenseNet121 | 0.61 | 0.66 | 0.49 |
| EfficientNetB0 | 0.61 | 0.66 | 0.15 |
| InceptionV3 (early_stopping) | 0.6839 | 0.6174 | 0.58 |
| VGG19 (early_stopping) | 0.60 | 0.5914 | 0.63 |

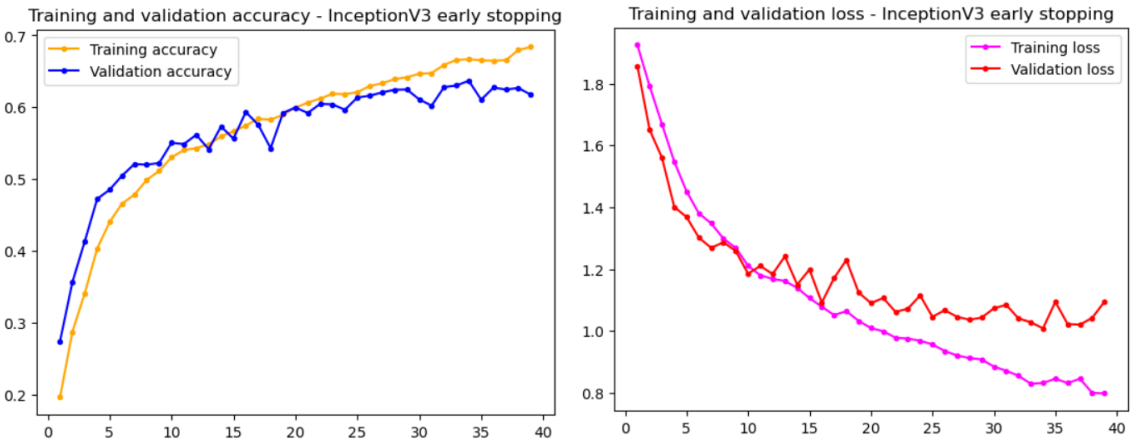
*pytorch eğitim doğruluk oranını açıkça belli etmemektedir, kayıp oranını belirtmektedir.



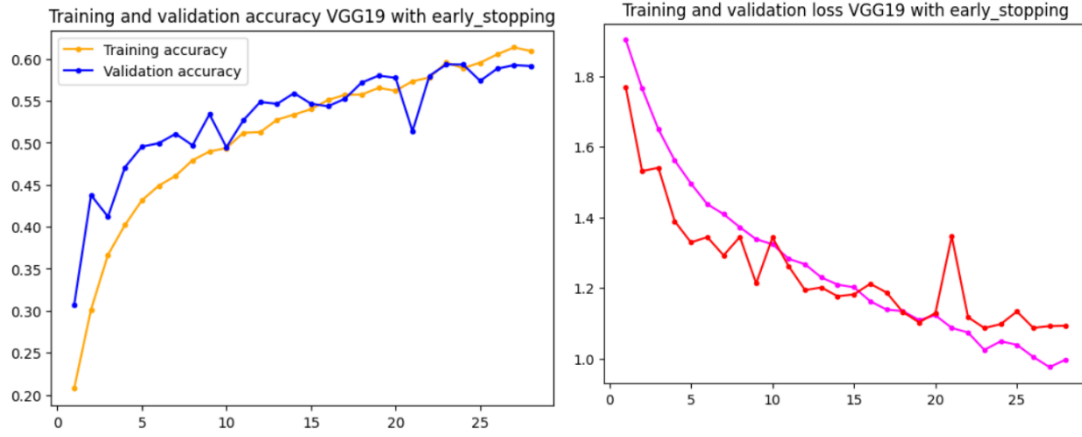
InceptionV3 backbone mimarisinin eğitim ile doğrulama doğruluk oranı ile kayıp oranları gösterilmiştir. Eğitim doğruluk oranı ile doğrulama doğruluk oranı arasında kaymalar vardır.



VGG19 backbone mimarisinin eğitim ile doğrulama doğruluk oranı ile kayıp oranları gösterilmiştir. Eğitim doğruluk oranı ile doğrulama doğruluk oranı zikzaklıdır.



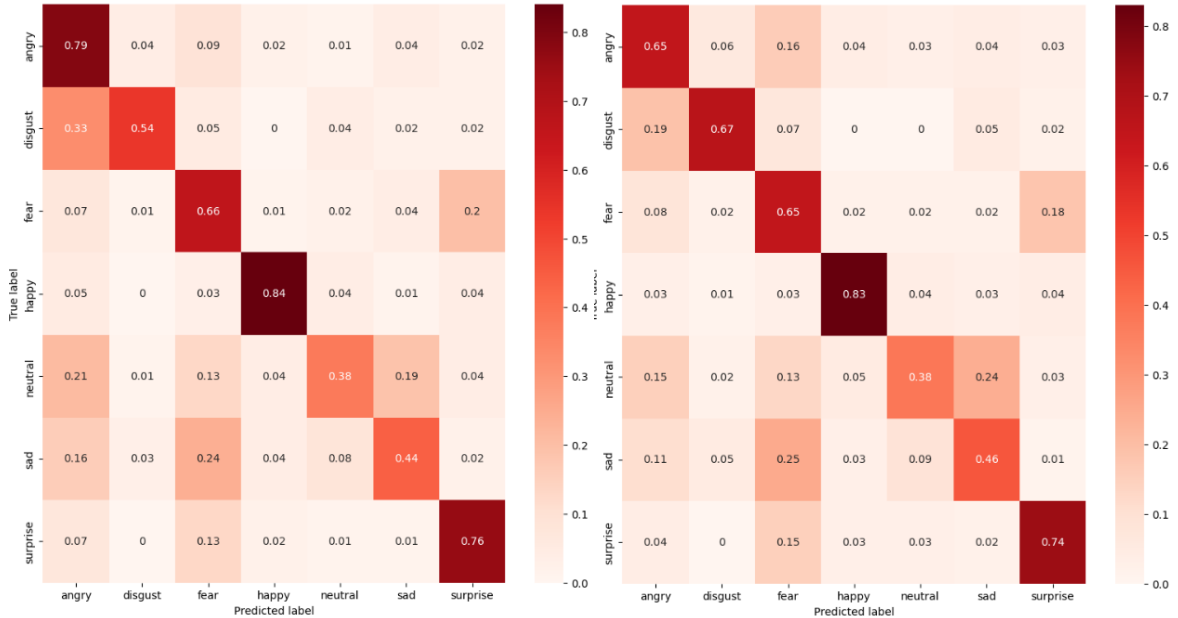
InceptionV3 backbone mimarisinin early_stopping regularize versiyonunda eğitim ile doğruluk oranları birbirine daha yakın ve pürüzsüz seyretmektedir.



VGG19 backbone mimarisinin early_stopping regularize versiyonunda eğitim ile doğruluk oranları birbirine daha yakın seyretmektedir.

| Backbone Mimari | Facial Expression test doğruluk oranı |
|------------------------------|---------------------------------------|
| InceptionV3 (early stopping) | 0.70 |
| VGG19 (early stopping) | 0.64 |
| DenseNet121 | 0.61 |
| EfficientNetB0 | 0.16 |

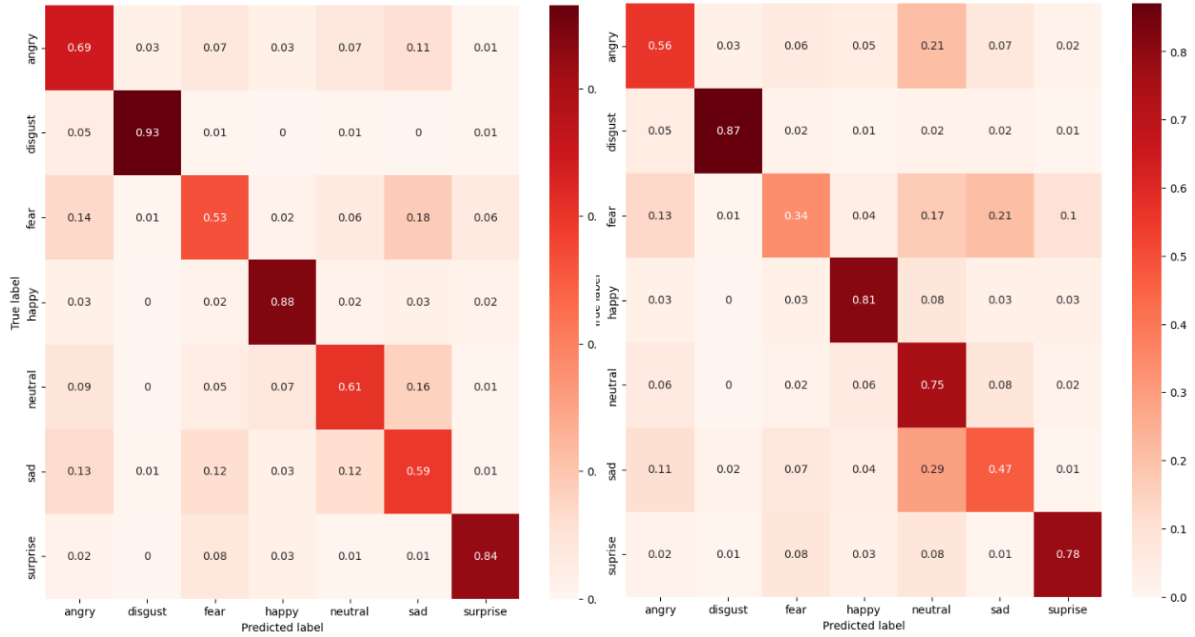
InceptionV3 mimarisinin early_stopping regularize versiyonu, Facial Expression veri setinde karşılıklarına kıyasla daha başarılı doğruluk oranı göstermektedir.



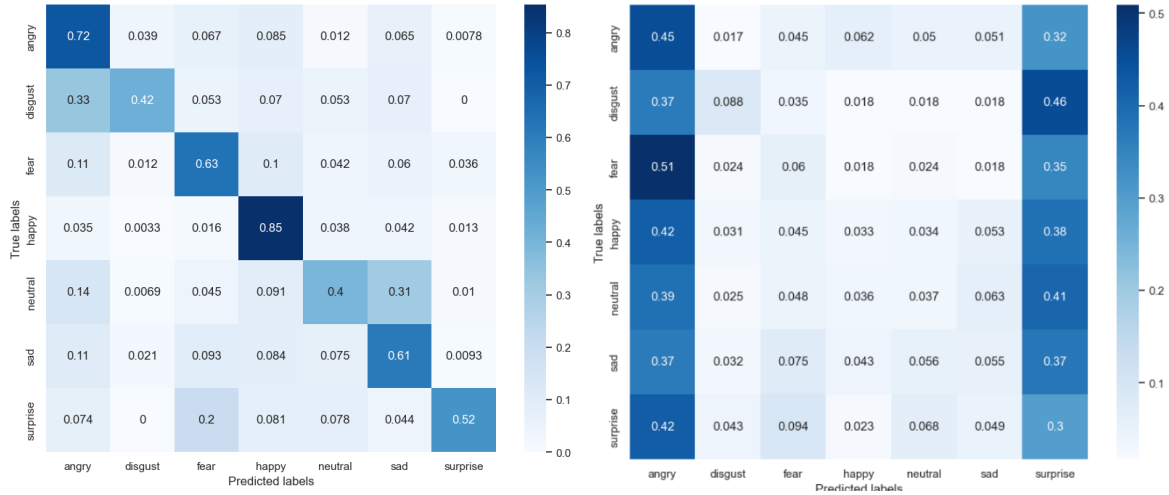
VGG19 FERPlus duygusal konfüzyon matrisi InceptionV3 FERPlus duygusal konfüzyon matrisi

Mimarilerin konfüzyon matrisleri incelendiğinde en yüksek doğruluk oranının happy duygu kategorisinde, en düşük doğruluk oranlarının neutral ile sad duygu kategorilerinde olduğu görülmektedir. InceptionV3 mimarisinin disgust duygu kategorisindeki doğruluk oranı

(0.67), VGG19 mimarisinin disgust duygu kategorisindeki doğruluk oranından (0.54) gözle görülür derecede yüksektir. Yalnız VGG19 mimarisi, angry duygu kategorisinde InceptionV3 mimarisine kıyasla daha yüksek başarı yakalamıştır. Disgust ile angry duygu kategorileri haricinde VGG19 ile InceptionV3 mimarileri benzer karakteristiğe sahiptir. Duygu kategorilerinin ifadesel yüz kas ayrışmaları birbirinden farklıdır.



Soldan sağa InceptionV3 ile VGG19 early_stopping Facial Expression konfüzyon matrisleri



DenseNet121 FERPlus konfüzyon matrisi

EfficientNetB0 FERPlus konfüzyon matrisi

6. TARTIŞMA VE YORUM

VGG19 ile InceptionV3 mimarilerinin FER2013 eğitimleri incelendiğinde VGG19 mimarisinin yakınsamasında tutarsızlıklar görülmektedir, eğitim-doğrulama kayıp yönelimleri zıtlılaşmaktadır. InceptionV3 mimarisinin yakınsaması, VGG19 mimarisine kıyasla daha pürüzsüzdür, eğitim-doğrulama kayıp yönelimleri aynı çizgidedir. VGG19 ile

InceptionV3 mimarilerinin duygusal konfüzyon matrislerinde neutral ile sad duygu kategorilerinin zayıflığı, ifadesel yüz kas ayrışmalarının birbirleriyle benzeşmesinde ve veri yetersizliğinden kaynaklanmaktadır. Görü ve geometri tabanlı transformer modelinin (AGT) FERPlus duygusal konfüzyon matrisinde “fear” duygu durumu en zor seçilendir, “fear” ile “surprise” duygu durumlarının ayrışmaları çoğunlukla birbirini takip etmektedir ve “fear” gibi negatif duygu durumlarının doğruluk oranı, “happy” gibi pozitif duygu durumlarına nazaran daha düşüktür, negatif duygu durumları geçişli tabiattadır. Projede “sad” duygu durumunun düşük doğruluk oranı, negatif duygu durumlarının tabiatı sebeplidir. Görü ve Geometri tabanlı transformer modelinin 3589 görüntülü FERPlus test setindeki doğruluk oranı, projenin VGG19, InceptionV3, DenseNet121 ile EfficientNetB0 mimarilerinin 7048 görüntü sayılı FERPlus test setindeki doğruluk oranından daha yüksek seyretmektedir. Bu durum sebebi, bağlantılı transformer modeline ait self-attention mekanizmalı derin sinir ağının ifadesel ayrışmaları odaklayan işaretlemesidir. Proje kapsamındaki transfer learning mimarilerinin güçlü self-attention mekanizma eksikliği, her ne kadar performanslı doğrusal veri eğitimi var olsa da duygu kategorilerinin bağlantılı veri temsilleriyle eşleşmesini bastırmaktadır, doğruluk oranını düşürmektedir. Zorlu koşullarda yüz ifade karakteristiği tanımlaması için konvolüsyonel ilişki network çalışmasının Facial Expression veri setindeki doğruluk oranı, projenin VGG19, InceptionV3 ile EfficientNetB0 mimarilerinin test setindeki doğruluk oranından düşük seyretmektedir, FER2013 test setindeki doğruluk oranı ise projenin VGG19, InceptionV3, DenseNet121, EfficientNetB0 mimarilerinin test setindeki doğruluk oranını geçmektedir. Test setleri arasında farklı doğruluk oranlarıyla karşılaşılmaktadır. Proje kapsamında özellikle DenseNet121 ile EfficientNetB0 mimarilerinin eğitim epoch döngü sayısı, harici çalışmalarla karşılaştırıldığında yetersiz kalmaktadır. Ek hazır modellerin, AGT modelinin “teacher” özütleme eklentisi gibi, proje kapsamına alınmaması bir etkindir. Yalnız proje kapsamı, konvolüsyonel transfer learning modellerine kuruludur, yüz ifade karakteristiği tanımlama görevinin konvolüsyon katman temelinde öznitelik çıkarımıyla bahsedilen çalışmalarla birleşmektedir.

InceptionV3 mimarisine uygulanan early_stopping fonksiyonu ile takiben Dropout artırımı eğitim sırasında overfitting probleminin bastırılmasında yararlanılmıştır. InceptionV3 mimarisi, early_stopping-Dropout bloğu ile eğitim ile doğruluk oranı arasındaki eğitim açıklığını azaltmaktadır, paralelinde VGG19 mimarisi early_stopping-Dropout bloğunun etkisiyle FERPlus test setindeki doğruluk oranını yükseltmektedir. Geçişli (Compound) ifadesel Facial Expressions test setinde en yüksek doğruluk oranı, early_stopping-Dropout

bloklı InceptionV3 mimarisine aitken en düşük doęruluk oranı EfficientNetB0 mimarisine aittir. InceptionV3 early_stopping-Dropout ile VGG19 early_stopping mimarilerinin Facial Expression konfüzyon matrisleri incelendięinde disgust ile neutral duygu durumlarına ait doęruluk oranlarının geliřtięi gözlenmektedir, fear duygu durumuna ait doęruluk oranının ise zayıfladıęı görölmektedir. Özellikle VGG19 early_stopping-Dropout bloklı mimarisine ait genellemenin eęitim süresince iyileřtirilmesi, FERPlus test setindeki doęruluk oranını artırmaktadır. Early_stopping ile Dropout arttırım aksı InceptionV3 ile VGG19 temel mimarilerinin aęırlık aktivasyonlarını overfitting ezberinden doęrusal öęrenime yönlendirmektedir.

7. SONUÇ VE ÖNERİLER

Konvolüsyonel transfer learning modelleri, sil bařtan kurulan mimarilere kıyasla daha etkili eęitim aęırlık parametrelerine sahiptir. Proje alıřması, yüz ifade karakteristięi tanımlamasını transfer learning backbone mimarisine baęlı özgün fully-connected bloęunda gerekleřtirmektedir, ifadesel yüz verileri, farklı konvolüsyonel derin sinir aęı mimarilerinden geirilerek öznitelik ıkarımları olasılık daęılımına dönüřtürölmektedir. InceptionV3 mimarisi, VGG19 mimarisine kıyasla parametre hafiflięi, projenin FER2013 eęitiminde deneyimlenmiřtir. DenseNet121 mimarisine ait Dense blok ile InceptionV3 mimarisine ait yatay geniřlemeli konvolüsyonel blok gibi modöler yapıların yüz ifade karakteristikleriyle etkileřimi FER2013 eęitimi ile FERPlus test setinde kaydedilmiřtir. Özellikle Google tarafından geliřtirilen yeni, ok yönlü model ayarı imkanı sunan (model scaling) EfficientNetB0 mimarisinin FER2013 veri setinde eęitilerek denenmesi, projenin transfer learning gidiřatına yönelimsel bakıř aısı getirmiřtir. VGG19 transfer learning mimarisi ile gemiř alıřmaya ait VGG16 transfer learning mimarisinin FER2013 veri setindeki doęrulama doęruluk oranları, karřılařtırıldıęında -sırasıyla VGG19 iin 0.64 ile VGG16 iin 0.63- VGG16 transfer learning mimarisinin az bir farkla daha düşük performans sergilemektedir. Projenin ileriki ařamalarında VGG, DenseNet ile EfficientNet mimari iskeletleri harici konvolüsyonel mimariler yüz ifade karakteristięi görevinin özgün fully-connected katman bloęuna dahil edilebilmektedir, hiper parametreler baęlantılı mimarilere odaklanabilmektedir. Sonuç olarak proje kapsamında konvolüsyonel yapısallıęın sinir aęı aktivasyonları, duygusal analizi mümkün kılan yüz ifade karakteristięi tanımlama görevi iin deęerlendirilmektedir, anlamlandırılmaktadır.

KAYNAKLAR

1. Al-Shabi, M., Cheah, W.P., Goh, M., (2017), "Connie, T. Facial Expression Recognition Using a Hybrid CNN-SIFT Aggregator", International Workshop on Multi-disciplinary Trends in Artificial Intelligence Lecture Notes in Computer Science, vol. 10607, pp. 139-149 [10.1007/978-3-319-69456-6_12](https://doi.org/10.1007/978-3-319-69456-6_12)
2. Alizadeh, S., Fazel, A., (2017), "Convolutional Neural Networks for Facial Expression Recognition", Stanford University Deep Learning <https://doi.org/10.48550/arXiv.1704.06756>
3. Khanzada, A., Bai, C., Celepcikay, F.T., (2004), "Facial Expression Recognition with Deep Learning", Stanford University - CS230 Deep Learning <https://doi.org/10.48550/arXiv.2004.11823>
4. Negara, I.G.P.K., Lim, A.P., Jonathan, J., (2020), "Emotion Recognition on FER-2013 Face Images Using Fine Tuned VGG-16", Advances in Science Technology and Engineering Systems Journal, vol. 5, pp. 315-322 [10.25046/aj050638](https://doi.org/10.25046/aj050638)
5. Sun, N., Song, Y., Liu, J., Chai, L., Sun, H., (2023), "Appearance and geometry transformer for facial expression recognition in the wild", Computers and Electrical Engineering, vol. 107, 108583, ISSN 0045-7906, <https://doi.org/10.1016/j.compeleceng.2023.108583>
6. Zhu Q., Mao, Q., Jia, H., Noi, Q.E.N., Tu, J., (2022), "Convolutional relation network for facial expression recognition in the wild with few-shot learning", Expert Systems with Applications, vol. 189, 116046, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2021.116046>