



Evrişimsel Sinir Ağlarında Duygu Tanıma Algoritmalarının Hiperparametreler ile Kapsamlı Analizi

Suha GÖKALP¹, İlhan AYDIN²,

¹ Bilgisayar Teknolojileri Bölümü, Pasinler Meslek Yüksek Okulu, Atatürk Üniversitesi, Erzurum, Türkiye

² Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Fırat Üniversitesi, Elâzığ, Türkiye

✉: suha.gokalp@atauni.edu.tr  ¹ 0000-0002-9705-1466.  ² 0000-0001-6880-4935

Geliş (Received): 07.05.2024

Düzeltilme (Revision): 20.06.2024

Kabul (Accepted): 10.07.2024

ÖZ

İnsanlar, davranışsal özelliklerini duygular aracılığıyla ifade ederler. Bir diğer deyişle ifadeler, insan davranış özelliklerinin duygusal dışavurumudur. Yüz ise bu ifadeleri insan-insan etkileşiminde en temel noktaya taşır. Bu etkileşimler sayesinde insanlardan geri tepkiler alınır, fakat makinelerin bu etkileşimi algılayabilmesi için üzerinde çalışılmalıdır. Karikatürize insan görünüşleriyle yalnızca duygulara odaklanılmasına olanak sağlayan FERG veri setinde daha yüksek performans elde edilmesi, gerçek görüntülerdeki başarı oranını da arttıracaktır. Bu amaçla, bu çalışmada, insan yüzündeki duyguların yapay zekâ yardımıyla daha hızlı ve isabetli tahmin edilmesi üzerine yoğunlaşmıştır. Bu bağlamda 7 duygu ifadesi olan Şaşkın, Üzgün, Normal, Mutlu, Korku, İğrenme, Kızgın isabetli bir şekilde tahmin edilmeye çalışılmıştır. FERG veri seti ile çalıştırılan ResNet 50, ResNet 50 32x4d, Vit_b_16, Vit_b32, EfficientNet B0, EfficientNet B1, EfficientNet B2 modellerinde %100 başarımla elde edilmiştir. Modellerin performans metrikleri karşılaştırmalı olarak sunulmuştur. FERG veri seti ile yapılan literatürdeki diğer çalışmalarla sonuçlar kıyas edilmiştir.

Anahtar Kelimeler: Derin Öğrenme Yöntemleri, Yüz Duygu Tanıma, Evrişimsel Sinir Ağları, FERG, ResNet, EfficientNet

Comprehensive Analysis of Emotion Recognition Algorithms in Convolutional Neural Networks with Hyperparameters

ABSTRACT

People express their behavioral characteristics through emotions. In other words, expressions are emotional expressions of human behavioral characteristics. The face carries these expressions to the most fundamental point in human-human interaction. Thanks to these interactions, feedback is received from people, but it must be studied so that machines can perceive this interaction. Achieving higher performance on the FERG dataset, which allows focusing only on emotions with cartoonish human appearances, will also increase the performance rate in real images. For this purpose, this study focused on faster and more accurate prediction of emotions on the human face with the help of artificial intelligence. In this context, 7 emotion expressions, Confused, Sad, Normal, Happy, Fear, Disgust and Angry, were tried to be predicted accurately. 100% performance was achieved in ResNet 50, ResNet 50 32x4d, Vit_b_16, Vit_b32, EfficientNet B0, EfficientNet B1, EfficientNet B2 models run with the FERG dataset. Performance metrics of the models are presented comparatively. The results were compared with other studies in the literature using the FERG dataset.

Keywords: Deep Learning Methods, Facial Emotion Recognition, Convolutional Neural Networks, FERG, ResNet, EfficientNet

GİRİŞ

Gerek hız, depolama ve hesaplama olanakları anlamında gelişen bilgisayar teknolojisi, gerek gelişen diğer imkânlar sayesinde insanlar yeni keşif alanlarına yönelmiştir. Duygu, davranış yoluyla ve birçok ifade yöntemiyle dışı doğru belirli bir uyaran, olay veya duruma verilen bir psikolojik ve fizyolojik tepkidir. Yüz duygu tanıma alanı da son birkaç yıldır araştırmacılar

arasında popüler olsa da keşfedilmeye muhtaç alt alanları barındıran bir konsepttir. Yüz duygu tanıma işleminin makinelerce hesaplanarak yapılması birçok güvenlik ve ticari faydalarını da beraberinde getirmektedir. İnsan yüzlerinden gelen yüz duygularını yorumlayabilen algoritmik çözümler tasarlamak ve geliştirmek, robotik, oyun, dijital pazarlama, akıllı öğretmen sistemleri ve diğerleri gibi insan-bilgisayar

etkileşimi bağlamında yeni olasılıklar penceresi açmaktadır [1].

Yapay zeka ile de otomatik yüz ifadesi tanıma (Facial Emotion Recognition – FER) sürekli gelişme göstererek duygusal zekanın önemli bir araştırma alanı olarak araştırmacıların artan ilgisini gördü [2]. Yüz görüntülerinden insan ifadesi tanıma için analitik modeller sağlamak amacıyla mühendisler, matematikçiler ve bilgisayar bilimcileri, verimli algoritmaları etkili bir şekilde uygulayabilen yaklaşımları yeniden üretmenin farklı yollarını araştırmaktadırlar. Bu konu üzerine çalışan araştırmacılar görüntü işleme, bilgisayarla görme, örüntü tanıma ve yapay zeka alanları arasında güçlü korelasyonlara ulaşmaktadır ve literatür üzerinden daha başka ilginç yaklaşımlara göz atılabilir [3]. Genel olarak, özellikle hesaplamalı işlem gücündeki son başarılar ve yüksek performanslı bilgi işlem için yeni mimariler nedeniyle, daha önce yukarıda belirtilen alanlarda hesaplama sınırlamaları ile kısıtlanmış olan araştırmacılar, yeni imkânlarla uygulamaları çözümlerine kavuşturmuşlardır [4].

Yüz Tanıma ve Duygusal Yüz Tanıma

Derin öğrenme tekniklerinden CNN, yüz tanıma uygulamalarında oldukça popülerdir. Bu teknik, birçok farklı yüz verisini öğrenerek, yeni yüzlerin benzersiz özelliklerini tanımlayabilir. Bu sayede, yüz tanıma doğruluğu artırılmış olur [3].

Yüz tanıma uygulamaları genellikle 2D görüntüler üzerinden gerçekleştirilirken, 3D yüz tanıma çalışmaları

da yapılır. Bu yaklaşım, yüzün daha fazla detayının yakalanmasına olanak sağlar ve yüz tanıma doğruluğunu artırır [5].

COVID-19 pandemisi döneminde yüz maskesi takmak yaygın hale geldi. Bu nedenle, yüz maskesi takılıken bile yüz tanıma uygulamalarının çalışabilmesi için özel teknikler geliştirildi. Bu teknikler, yüz maskesi takılıken bile yüz özelliklerinin tanınabilmesini sağlar [6].

Yüz tanıma uygulamalarında, yüz özellikleri çıkarımı için farklı teknikler kullanılır. Bu teknikler arasında özellikle PCA (Principal Component Analysis) ve LDA (Linear Discriminant Analysis) sık kullanılır. Bu teknikler, yüz özelliklerini daha doğru bir şekilde çıkarmayı ve tanımlamayı sağlar [7].

Yüz ifadeleri, insanların duygularını ifade etmek için önemli bir araçtır. Duygusal yüz tanıma, yüz ifadelerini analiz ederek, insanların duygusal durumlarını belirleme amacıyla kullanılır. Bu teknik, özellikle psikoloji ve tıp alanlarında kullanılır [3].

Yukarıda belirtilen örneklerin yanı sıra, yüz tanıma alanında birçok farklı çalışma ve geliştirme yapılmaktadır [3]. Özellikle, derin öğrenme tekniklerinin kullanımı ile yüz tanıma doğruluğu artırılmış ve bu teknolojinin birçok farklı alanda kullanımı yaygınlaşmıştır.

Duygusal yüz tanıma alanında kullanılmak üzere üretilmiş FERG veri seti üzerinden en iyi isabeti yapabilen modellerin eğitim ve tahmin zamanı olarak en performanslısını seçmek üzere yapılmış olan bu çalışmada, literatürde bu alanda yapılmış çalışmalarla kıyas yapılmış, sonuçlar analiz edilmiştir.

Tablo 1. Kullanım amacı bakımından derin öğrenme ile yapılan duygusal yüz tanıma çalışmaları.

Makale	Veri Seti	Özellik Çıkarımı	Yöntem	Başarım
[8]	KDEF, GENK, CK+, SFEW	Ağaç Yapısı Parça Modeli, CNN	Deep-CNN	%82.63 (KDEF), %96.75 (GENKI), %96.81(CK+), %36.79 (SFEW)
[9]	10000 resim (154 kişi)	İfade Vektörü (EV)	İki Parçalı CNN	%96
[10]	FER13, JAFFE, CK+	HOG	CNN	%87.05 (FER13), %99 (JAFFE), %98.79% (CK+)
[11]	JAFFE, CK+	CNN	CNN	%95.65 (JAFFE), %99.36 (CK+)
[12]	JAFFE, CK+	Viola-Jones Algoritması	Geliştirilmiş Özellik Haritası ile CNN	%92.66 (karışık veri seti), %94.94 (çapraz veri seti)
[13]	AffectNet (420.299 resim)	HOG	CNN	%69.3 (Veri setindeki resimlerde insanların yüzünde maske vardır)
[14]	RAVDESS	SSL	Hadamard, IMA, Transofmer temelli füzyon	%86.4

Tablo 2. FERG Veri Seti ile farklı parametrelerle daha önceden yapılan araştırmalar.

Makale	Yıl	Veri Seti	Özellik Çıkarımı	Yöntem	Başarım
[15]	2017	Geliştirilmiş CK+, DISFA, KDEF, MMI, FERG	CNN ve Transfer Öğrenme	CNN	%85.27 (İnsan Yüzü), %89.02 (Stilize Karakter)
[16]	2022	FER2013, CK+, FERG	CNN, VGG16, ResNet50	Logistic Regression, Decision Tree, Bayes	%76.2 FER, %99.4 CK+, %99.6 FERG
[17]	2023	CK+, FER2013, FERG	Ayrık Derin Öğrenme – SDL	CNN Temelli STN	%74.92 FER, %99.47 CK+, %98.5 FERG
[18]	2022	SFEW, CK+, JAFFE, FERG	CNN temelli OFELBW metodu	OFELBW metodu İkili Balina Alg.	%64.98, %98.35, %96.6, %99.42
[19]	2023	FERG, CK+, KDEF, RaF-DB, FER-2013H, FER-2013C	DCNN	DCNN	99.98%, 99.25%, 88.17%, 84.09%, 69.87% ve 69.16%
[20]	2021	JAFFE, CK, FEI ve FERG	LGC-HD operator	SVM	%84.2, %92.3, %98.2 ve 97.9%
[21]	2023	Aff-Wild2, RAF-DB, JAFFE, FER+, FERG	window-based cross-attention ViT	MLP	%72.48, %92.47, %96.67, %93.09, %98.18
[22]	2022	FER2013, FERG, JAFFE	RCNN, DenseNet121	IFER-DTFL Derin Transfer Öğrenme, WKELM	92.6% FER2013, 99.86% FERG, 99.18% JAFFE

Literatür Taraması

Tablo 1’de yer alan araştırmalar, farklı yüz ifadeleri tanıma yöntemlerini ele almaktadır ve sonuçları, derin öğrenme algoritmalarının kullanımının artması ile daha yüksek doğruluk oranları elde edildiğini göstermektedir. Tablo 2’den çıkan bir sonuç da son yıllarda özellikle FERG veri seti ile başarımların artmış olmasıdır. Fakat FERG veri setinin animasyon karakterleri ile yapıldığı unutulmamalı, animasyon karakterlerinden gerçek insan yüzlerine nasıl geçileceğine dair araştırmalar yapılmalıdır.

MATERYAL ve METOT

Bu bölümde kullanılan veri seti, temel derin öğrenme işlemi ve kullanılan birçok algoritmanın başarımları yer almaktadır.

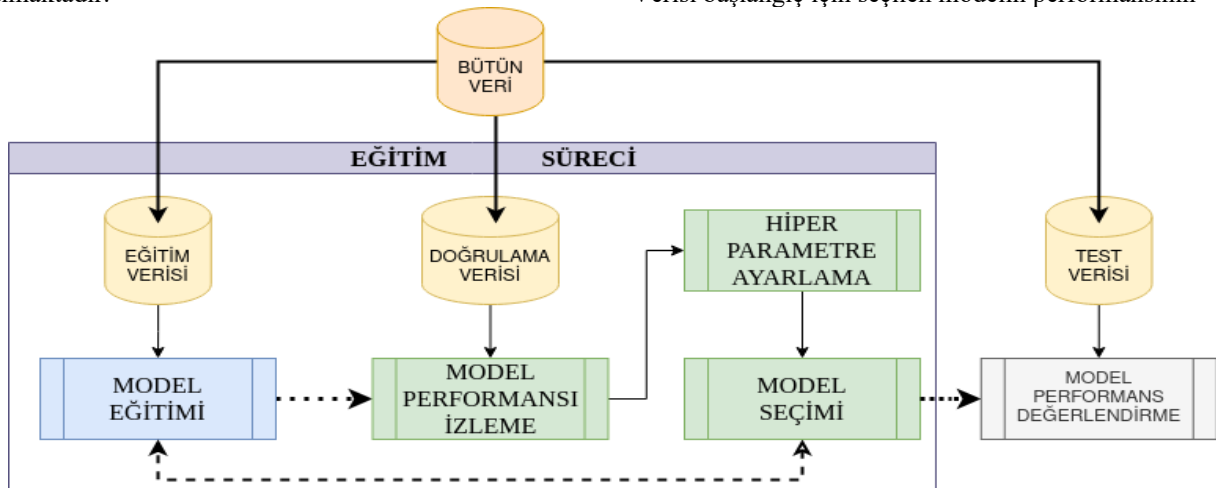
Veri Seti

Çalışmada FERG Veri Seti [15] kullanılmıştır. FERG veri setinde toplam 55768 resim dosyası (Kızgın 9170, İğrenme 8571, Korku 7419, Eğlenme 7330, Nötr 6939, Üzgün 7627, Şaşkın 8712 dosya ile temsil edilmek üzere) bulunmaktadır.

Şekil 2’de veri setinden bir kesit verilmiştir. Veri seti kurgu karakterler üzerinden modelleri eğitmeyi ve uygun modeli bulmayı amaçlamaktadır.

Metodoloji

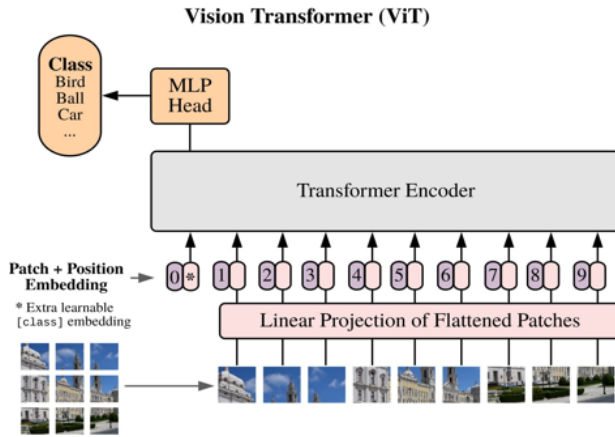
FERG Veri Seti kaynağından edinildikten sonra Eğitim Verisi, Doğrulama Verisi, Test Verisi başlıklarına bölünerek farklı yerlere alınmıştır. Eğitim Verisi başlangıç için seçilen modelin eğitiminde, Doğrulama Verisi başlangıç için seçilen modelin performansını



Şekil 1. Veri setinin bölünmesi ve modelle kullanım tercihleri

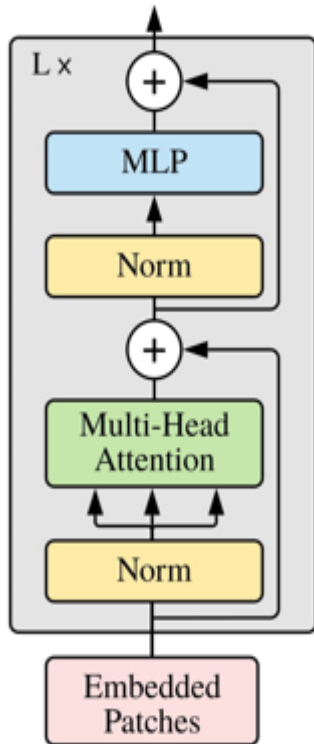
Vision Transformer (ViT)

Hugging Face topluluğunun bünyesindeki ViT, Google'ın araştırma ve beyin takımı tarafından 2021'de [24], 2017'de Vaswani ve diğerleri [26] tarafından geliştirilen Transformers mimarisi üzerine temellenmiş, yeni bir görsel işleme modeli olarak ortaya konmuştur. Aşağıda Şekil 4'te ViT mimarisi yer almaktadır.



Şekil 4. Orijinal makaleden [24] alınan ViT mimarisi.

ViT model parametreleri üzerinde çalışılan platformun kapasiteleri ve veri setinin yapısı dikkate alınarak belirlenmiştir.



Şekil 5. ViT mimarisinde bulunan Transformer Encoder yapısı [24]

Şekil 5'te ViT mimarisinde bulunan Transformer Encoder yapısı yer almaktadır. Resim boyutları eğitim, doğrulama ve test verileri için PyTorch modülünün transforms metodu ile ayarlanmıştır.

Modelin inşasında ViTForImageClassification sınıfı (class) kullanılmış, büyük ölçüde ilk makale yazarlarının metoduna bağlı kalmıştır.

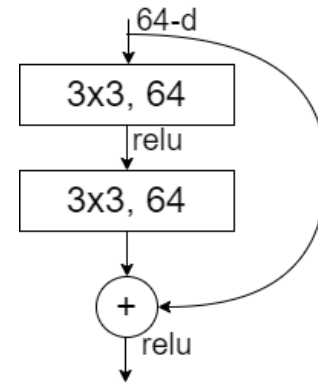
ResNet 50 Modeli

ResNet 50 [27], 34 katmanlı ağdaki her 2 katmanlı bloğu 3 katmanlı darboğaz bloğuyla değiştirilerek elde edilir. Her artık fonksiyon F için 2 yerine 3 katmandan oluşan bir yığın kullanılır. Bu üç katman, 1×1 , 3×3 ve 1×1 'lik boyutlara sahiptirler; burada 1×1 katmanı boyutları azaltmaktan ve ardından artırmaktan (geri yüklemekten) sorumludur. 3×3 katmanı ise daha küçük girdi/çıkış boyutlarına sahip bir darboğaz bırakır.

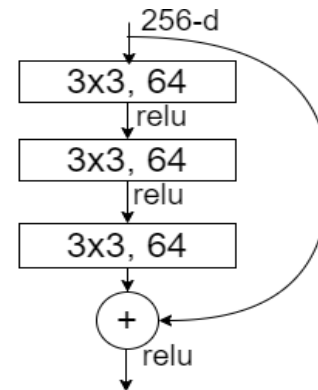
ResNet 50 için belirlenen model parametrelerinin $([64,128,256,512],[3,4,6,3],4, True)$ açılımı şöyledir;

- Her ara blok için kanal adedini $[64,128,256,512]$ kısmı göstermektedir.
- Her blok için darboğazları $[3,4,6,3]$ kısmı göstermektedir.
- Genişleme faktörü 4 olarak verilmiştir. Yani 64, 256'ya, 128, 512'ye dönmektedir. Bütün ResNet katmanları aynı faktörü kullanmaktadır.
- "True" ifadesi ise darboğaz katmanının durumunu deklare eder.

ResNet 50'deki dar boğaz yapı taşını verebilmek amacıyla Şekil 6 ve Şekil 7 eklenmiştir.



Şekil 6. ResNet-34 için bir yapı taşı [27]



Şekil 7. ResNet-50/101/152 için bir "darboğaz" (bottleneck) yapı taşı [27]

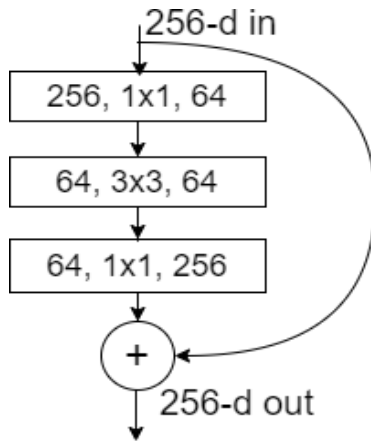
Tablo 3'te ResNet mimarileri yer almaktadır.

Tablo 3. ResNet dokümanından [27] ResNet mimarileri

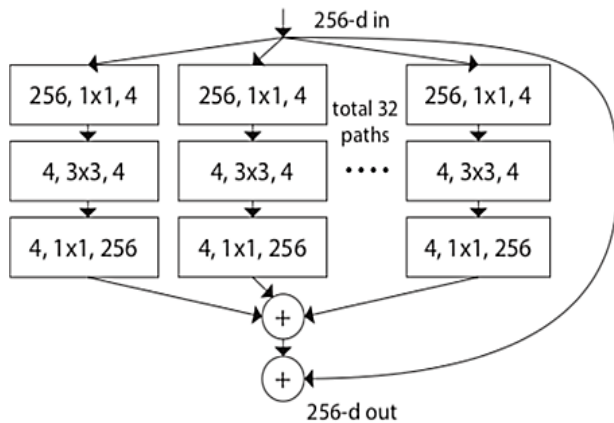
Katman Adı	Çıktı Ebatı	18-katman	34-katman	50-katman	101-katman	152-katman
Conv1	112x112			7x7, 64, stride 2 3x3 max pool, stride 2		
Conv2.x	56x56	$\begin{bmatrix} 3x3, 64 \\ 3x3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3x3, 64 \\ 3x3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 3$
Conv3.x	28x28	$\begin{bmatrix} 3x3, 128 \\ 3x3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3x3, 128 \\ 3x3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 4$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 4$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 8$
Conv4.x	14x14	$\begin{bmatrix} 3x3, 256 \\ 3x3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3x3, 256 \\ 3x3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 23$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 36$
Conv5.x	7x7	$\begin{bmatrix} 3x3, 512 \\ 3x3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3x3, 512 \\ 3x3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{bmatrix} \times 3$
	1x1			Average pool, 1000-d fc, softmax		
FLOPs		$1,8 \times 10^9$	$3,6 \times 10^9$	$3,8 \times 10^9$	$7,6 \times 10^9$	$11,3 \times 10^9$

Resnext 32x4d Modeli

Darboğaz bloğu (Şekil 9) içindeki 3x3 gruplu evrişimlerin yerine 3x3 evrişimleri koyan normal ResNet modelini temel alır [28].



Şekil 8. ResNet-34 için bir yapı taşı [28]



Şekil 9. Önem Derecesi (Cardinality) 32 olarak belirtilmiş ResNext bloğu [28]

ResNext 32x4d ifadesi ResNet üzerine Şekil 9'daki yapı gelmekte ve Önem Derecesi 32, darboğaz genişliği de 4 olarak belirlenmektedir.

MODELLERİN UYGULANMASI ve BULGULAR

EfficientNet B0 – B2, Vision Transformer (ViT), ResNet 50, ResNext50_32x4d modellerinin uygulaması anlatılmaktadır.

FERG için kullanılan veri adetlerinin ayrıntıları Tablo 4'te verilmiştir.

EfficientNet B0 – B2

B0 için veri dizini, girdi boyutu ayarlanmış, yığın ebatı 32 olarak belirlenmiştir. Eğer daha önceden B0 modeli çalıştırılmışsa ağırlıkları hazır olarak tekrar eğitim yapılmıştır.

Model 50 çevrim için uygulanmıştır. Optimizer modeli Stokastik Dağılım Azaltımı (Stochastic Gradient Descent – SGD) olarak belirlenmiş, öğrenme oranı 10^{-3} ve momentum değeri $9 \cdot 10^{-1}$ olarak alınmıştır. Metrik kriteri ise Cross Entropy Loss seçilmiştir.

Veri seti eğitim, doğrulama ve test olarak alt gruplara ayrılmış ve eğitim ve doğrulama verileriyle modelin eğitilme işlemleri yapılmıştır. Modelde kullanılan “device” parametresi ile var ise GPU kullanımı direktifi verilmiştir.

B0 – B2 arası modeller birbirlerine eğitilmiş parametre vererek çalıştırılmıştır. Çalıştırılmadan önce gc.collect() metodu çalıştırılarak sonraki modellerde kullanılmayacak olan parametreler temizlenmiştir (Garbage Collector – GC). PyTorch kütüphanesindeki GPU çalıştırmak için kullanılan CUDA kütüphanesinden de empty_cache metodu çalıştırılarak kullanılmayacak CUDA parametreleri temizlenmiştir. En son test fazına geçilmiştir.

Tablo 4. FERG Veri Seti Eğitim, Doğrulama ve Test verileri adetleri.

Verilerin kullanım yerleri	Kızgın	İğrenme	Korku	Eğlenme	Nötr	Üzgün	Şaşırmiş
TEST	1867	1710	1462	1486	1350	1542	1737
EĞİTİM	5826	5553	4797	4608	4470	4849	5586
DOĞRULAMA	1476	1308	1160	1236	1119	1236	1388

Test fazında eğitim aşamasında kaydedilen modeller B0'dan başlanarak tekrar yüklenerek işlem yapılmıştır. 32'li yığınlar tahminler ve etiketler kullanılarak İsbet, Hassasiyet, Özgüllük ve F1-Skoru metrikleri hesaplanmıştır.

Vision Transform (ViT)

ViT önceden 16-224 ebatlarıyla ve önceden eğitilmiş olarak doğrudan alınmıştır. Ayrıca sınıflandırıcı olarak Linear ve Dropout katmanı 10-1 parametresiyle seçilmiştir. Optimizasyon fonksiyonu ADAM'dır. Kayıp fonksiyonu ise Cross Entropy Loss olarak seçilmiştir. Model uygulanırken varsa CUDA yoksa CPU kullanacak şekilde sistem kontrol edilmiştir.

Torch ile yapılan bu çalışma bir parçalı öğrenme metodudur. Eğitim, Doğrulama ve Test verileri 32 olarak belirlenen yığın parçalarına bölünüp bir döngüye sokularak öğrenme işlemi gerçekleştirilmiş, eğitim kaybı, doğrulama ve test isabet oranları her yığınla birlikte takip edilmiştir.

Parametrelerin belirlenmesinde platform kısıtları ve veri seti gereklilikleri gözetildiği belirtilmiştir. Buna göre;

- Çevrim adedi 50,

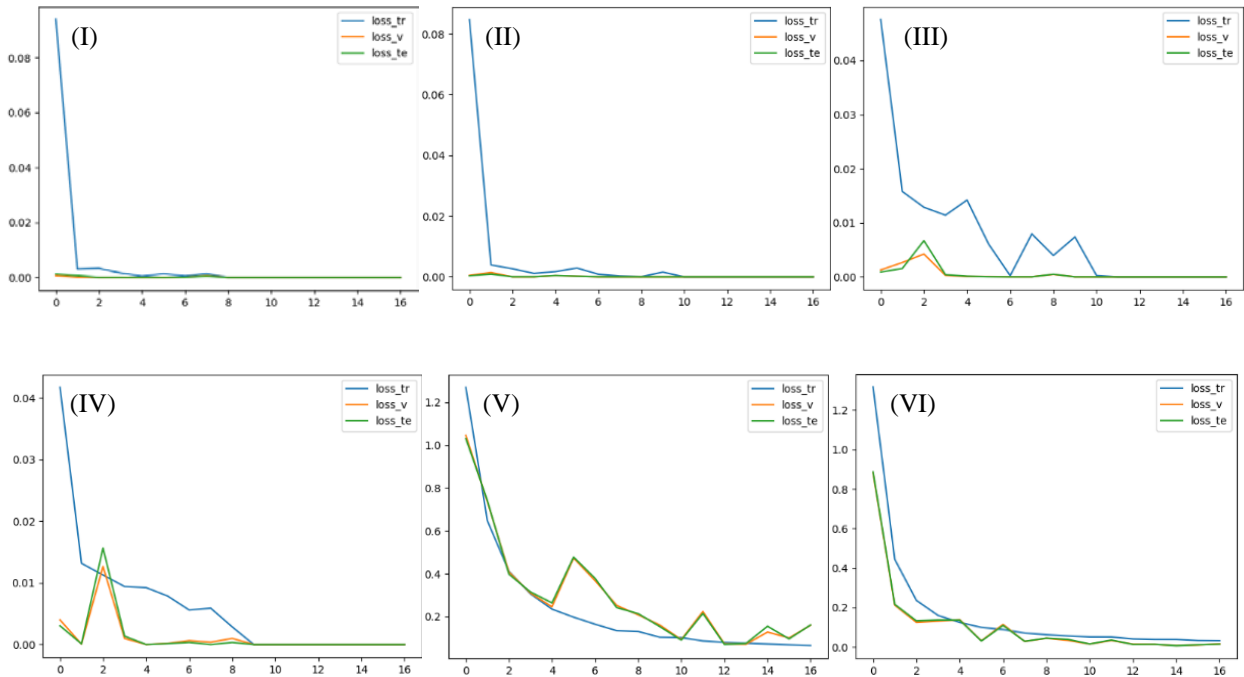
- Yığın boyutu 21,
 - Öğrenme oranı (LR) $2.e^{-5}$,
 - Metriklerin bozulmaya başladığında öğrenmenin durdurulabileceği ilk çevrim 10,
 - Resim boyutu 48x48,
- parametreleriyle çalışılmıştır.

ResNet 50 ve ResNext50_32x4d Modelleri

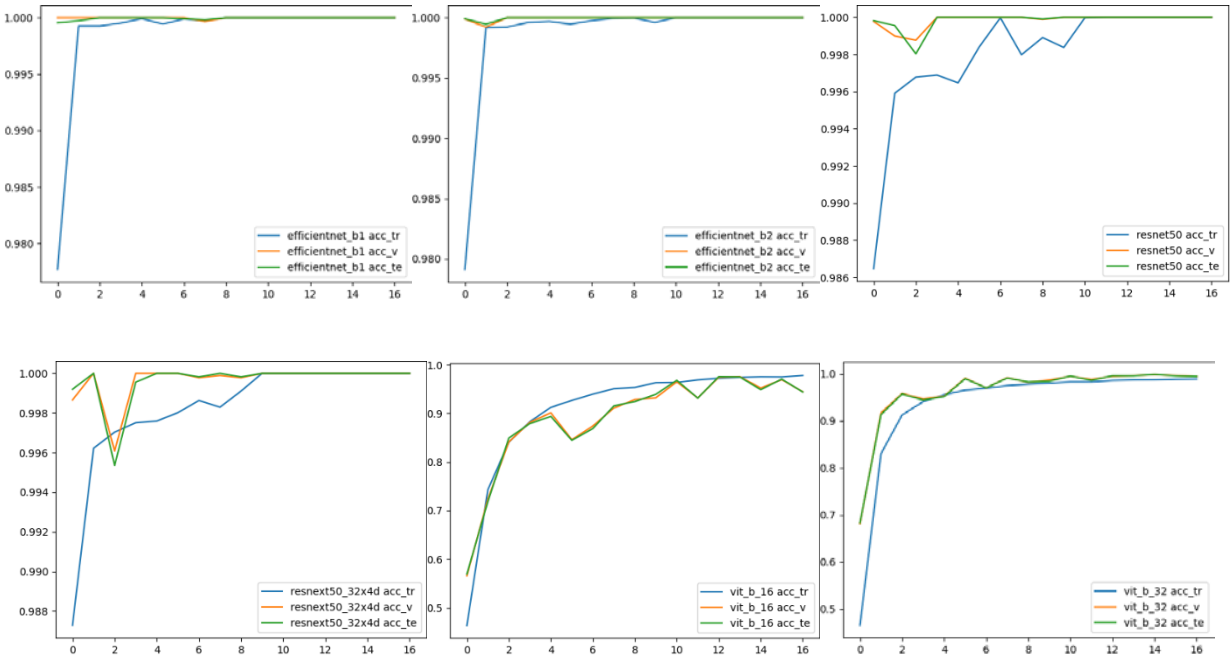
Çevrim adedi 50, yığın boyutu 70 olarak seçilmiştir. Platform performans kısıtları da göz önünde bulundurulmuştur. Metriklerin bozulmaya başladığında öğrenmenin durdurulabileceği ilk çevrim 5 olarak belirlenmiştir.

Döngüsel öğrenme metoduyla her yığının eşit sayıda eğitim, doğrulama, test verisi aynı anda modelde denenmiş, bütün metrikler anlık olarak alınmış ve çevrim zamanı not edilmiştir.

Pretrain_utils kütüphanesinden loaderdef, modeldef, showimages, evaluate, megatrain, printaccm, get_predictions, plot_confusion_matrix metodları kullanılmıştır. Eğitimler Megatrain metodu ile, metrik hesaplamaları evaluate, printaccm ve plot_confusion_matrix ile yapılmıştır.



Şekil 10. Loss fonksiyonu grafikleri: I) Efficient Net B1, II) Efficient Net B2, III) ResNet 50, IV) ResNext 50 32x4d, V) Vit_b_16, VI) Vit_b_32



Şekil 11. İsalet oranları eğitim, doğrulama ve test grafikleri: Efficient Net B1, Efficient Net B2, ResNet 50, ResNext 50 32x4d, Vit_b_16, Vit_b_32

Tablo 5. Çalıştırılan modellerin sonuçları

Model	Çevrim Süresi (Saniye)	Çevrim Adedi	Performans		
			Eğitim	Doğrulama	Test
ResNet50	318,27680	27	99,20%	100,00%	100,00%
ResNext50_32x4d	480,9140	31	99,71%	100,00%	100,00%
ViT_b_16	1522,0624	49	98,75%	99,74%	99,67%
ViT_b_32	500,11860	37	91,59%	96,57%	96,84%
EfficientNet B0	287,02220	27	99,85%	100,00%	100,00%
EfficientNet B1	369,86960	26	97,62%	100,00%	100,00%
EfficientNet B2	360,95280	27	99,85%	100,00%	100,00%

Tablo 6. Önerilen modellerin literatür ile karşılaştırılması

Çalışma	Yılı	Modeller	Ferg Performansı
Soman vd. [16]	2022	CNN, VGG16 + ResNet 50	%99.6
Khan vd. [17]	2023	DCNN, SVM	%98.5
Muthamilselvan vd. [18]	2022	CNN Temelli OFELBW Metodu	%99.42
Kalsum ve Mehmood [19]	2023	DCNN	%99.98
Kola ve Samayantula [20]	2021	LGC-HD	%97.9
Wasi vd. [21]	2023	MLP	%98.18
Albraikan vd. [22]	2022	RCNN, DenseNet121, IFER-DTFL	%99.86
Önerilen Model - 1	2024	ViT_b_16	%99.67
Önerilen Model - 2	2024	ResNet ve EfficientNet	%100

TARTIŞMA ve SONUÇ

Geniş kullanım alanı olan görüntüden, sestem veya yazıdan duygu tanıma konsepti çok canlıdır ve sürekli gelişmektedir. Mevcut çalışmada bilinen bir veri seti üzerinde başka alanlarda başarısını ispatlamış pek çok sinir ağı modeli hiperparametrelere dikkat edilerek imkânlar ölçüsünde uygulanmıştır. Elde edilen sonuçların ileride yapılacak olan çalışmalara motivasyon olması ve desteklemesi beklenmektedir. Yukarıda Tablo 5'te çalıştırılan modellerin sonuçları, Tablo 6'da ise önerilen modellerin literatür karşılaştırılması verilmiştir.

Çalıştırılan ResNet50 ve ResNext50_32x4d modelinde %100 başarımlar, ViT_b_16 modelinde %99.67 ve ViT_b_32 modelinde ise %96,84 başarımlar elde edilmiştir. Somon ve ark. 2022 yılında yaptıkları çalışmada %99,6 sonucuna ulaşmışlardır [16]. Literatürde duygu tanıma ile ilgili farklı modellerle yapılan diğer çalışmalarda ise başarımlar %97,9 ile %99,98 arasında değişmektedir [17-22].

Şekil 10'da çalıştırılan modellerin her çevriminde kayıp fonksiyonundan elde edilen çıktıları verilmiştir. Lejantta verilen *loss_tr* çizgisi eğitim verisinden gelen kaybı, *loss_v* çizgisi doğrulama verisinden gelen kaybı, *loss_te* çizgisi test verisinden gelen kaybı göstermektedir.

Şekil 11'de çalıştırılan modellerin her çevriminde elde edilen isabet oranları verilmiştir. Lejantta model isminden sonra verilen *acc_tr* çizgisi eğitim verisinde elde edilen isabeti, *acc_v* çizgisi doğrulama verisinde elde edilen isabeti, *acc_te* çizgisi test verisinde elde edilen isabeti göstermektedir.

Çalıştırılan modellerin başarımlar oranları yönüyle literatürdeki diğer çalışmalarla kıyası yapılmış, modellerin çevrim adedi, hiper parametreleri ve geçen süreler kıyas yapılabilmesi amacıyla tablolaştırılarak verilmiştir. Başarımlar oranından kayıp vermeden modellerin nasıl optimize edildiğine dair ayrıntılar çalışmanın ortaya koyduğu sonuçlardan en önemlisidir. FERG Veri Seti, EfficientNet (B0 – B2), ViT Hugging, ResNet 50 ve ResNext_32x4d modelde derin öğrenme algoritmalarına tabi tutulmuş ve sonuçları not edilmiştir. Elde edilen bulgulara göre veri setindeki verilerin arttırılması veya görsellerin sadeleştirilmesi gerektiği yorumu yapılmıştır. Çalışma buradan itibaren aynı modellerle fakat veri setinde veri zenginleştirilmesi veya veri setinin değiştirilerek tekrar denemesi yolu ile devam ettirilmiştir.

Makine öğrenimi ve yapay zekâ alanının gelişmesi ile yeni yüz tanıma teknolojileri ortaya çıkmıştır [4,13,29]. Derin öğrenmenin ilerlemesi şüphesiz insanların yaşamlarını iyileştirecek ve işleri kolaylaştıracaktır. İnsan-bilgisayar etkileşiminde, en önemli durumlardan biri yüz duygu tanımadır. Günümüzde yüz duygu tanıma, Google ve Facebook gibi büyük şirketlerin de dikkatini çekiyor. Bu yüzden yüz duygu tanıma gibi alanlar parlak bir geleceğe sahiptir. Derin öğrenme alanının daha da çok ilerlemesi birçok alanda (akıl hastası bir kişinin ruh halini belirleme vb. araştırma-inceleme alanları) üretkenliği arttırmaya ve ürün kalitesini

geliştirmeye de yardımcı olacağı görülmektedir. Bu bol örnekli karşılaştırmalı çalışmanın gelecekte yapılacak diğer çalışmalar için ışık tutacağı ön görülmektedir. Aynı modellerin veri zenginleştirme ve veri setinin değiştirilmesiyle başarımlar oranlarının değişmesi bu problemin ön işlem fazının iyileştirilmesi ve daha dengeli bir veri setiyle daha yüksek başarımlar oranıyla çözülebileceğine işaret etmektedir.

Etik Hususlar

Etik kurallara uyum

Yazarlar, çalışmalarının etik kurallara uygun olduğunu beyan ederler.

Finansman

Yazarlar, fon kuruluşlarından özel bir hibe almamıştır.

Çıkar çatışması

Makalemiz ile ilgili herhangi bir kurum, kuruluş, kişi ile mali çıkar çatışması yoktur ve yazarlar arasında çıkar çatışması bulunmamaktadır.

KAYNAKÇA

- [1] Liu, M., Raj, A. N. J., Rajangam, V., Ma, K., Zhuang, Z., Zhuang, S. 2024. Multiscale-multichannel feature extraction and classification through one-dimensional convolutional neural network for Speech emotion recognition. *Speech Communication*, 156, 103010.
- [2] Gong, W., Qian, Y., Zhou, W., Leng, H. 2024. Enhanced spatial-temporal learning network for dynamic facial expression recognition. *Biomedical Signal Processing and Control*, 88, 105316.
- [3] Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., Sobieranski, A. C. 2022. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, 582, 593-617.
- [4] Cheng, J., Deng, Y., Meng, H., Wang, Z. 2013. A facial expression based continuous emotional state monitoring system with gpu acceleration, in. 10th IEEE International Conference And Workshops On Automatic Face And Gesture Recognition (FG), IEEE, 22-26 Nisan, Shanghai, China, 1-6.
- [5] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I. 2010. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in. IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, IEEE, 13-18 Haziran, San Francisco, USA, 94-101.

- [6] Akgül, İ., Kaya, V., Baran, A. 2021. Examination of facial mask detection using deep learning methods against coronavirus. 4. Uluslararası İpek Yolu Akademik Çalışmalar Sempozyumu, 17-18 Aralık, Nevşehir, Türkiye, 149-154.
- [7] Hasimah, A., Hariharan, M., Yaacob, S., Adom, A.H. 2015. Facial emotion recognition using empirical mode decomposition, *Expert Systems with Applications*, 42, 1261–1277.
- [8] Hossain, S., Umer, S., Asari, V., Rout, R. K. 2021. A unified framework of deep learning-based facial expression recognition system for diversified applications. *Applied Sciences*, 11(19), 9174.
- [9] Mehendale, N. 2020. Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, 2(3), 446.
- [10] Fakhar, S., Baber, J., Bazai, S. U., Marjan, S., Jasinski, M., Jasinska, E., ... Hussain, S. (2022). Smart classroom monitoring using novel real-time facial expression recognition system. *Applied Sciences*, 12(23), 12134.
- [11] Khattak, A., Asghar, M. Z., Ali, M., Batool, U. 2022. An efficient deep learning technique for facial emotion recognition. *Multimedia Tools and Applications*, 81(2), 1649-1683.
- [12] Qazi, A. S., Farooq, M. S., Rustam, F., Villar, M. G., Rodríguez, C. L., Ashraf, I. 2022. Emotion Detection Using Facial Expression Involving Occlusions and Tilt. *Applied Sciences*, 12(22), 11797.
- [13] Mukhiddinov, M., Djuraev, O., Akhmedov, F., Mukhamadiyev, A., Cho, J. 2023. Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people. *Sensors*, 23(3), 1080.
- [14] Chaudhari, A., Bhatt, C., Krishna, A., Travieso-González, C. M. 2023. Facial emotion recognition with inter-modality-attention-transformer-based self-supervised learning. *Electronics*, 12(2), 288.
- [15] Aneja, D., Colburn, A., Faigin, G., Shapiro, L., Mones, B. 2016. Modeling stylized character expressions via deep learning. In *Computer Vision—ACCV 2016: 13th Asian Conference on Computer Vision*, 20-24 Kasım, Taipei, Taiwan, 136-153.
- [16] Soman, G., Vivek, M. V., Judy, M. V., Papageorgiou, E., Gerogiannis, V. C. 2022. Precision-based weighted blending distributed ensemble model for emotion classification. *Algorithms*, 15(2), 55.
- [17] Khan, N., Singh, A. V., Agrawal, R. 2023. Enhancing feature extraction technique through spatial deep learning model for facial emotion detection. *Annals of Emerging Technologies in Computing (AETiC)*, 7(2), 9-22.
- [18] Muthamilselvan, T., Brindha, K., Senthilkumar, S., Saransh, Chatterjee, J. M., Hu, Y. C. 2023. Optimized face-emotion learning using convolutional neural network and binary whale optimization. *Multimedia Tools and Applications*, 82(13), 19945-19968.
- [19] Kalsum, T., Mehmood, Z. 2023. A novel lightweight deep convolutional neural network model for human emotions recognition in diverse environments. *Journal of Sensors*.
- [20] Kola, D. G. R., Samayamantula, S. K. 2021. Facial expression recognition using singular values and wavelet-based LGC-HD operator. *IET Biometrics*, 10(2), 207-218.
- [21] Wasi, A. T., Šerbetar, K., Islam, R., Rafi, T. H., Chae, D. K. 2023. ARBEx: Attentive feature extraction with reliability balancing for robust facial expression learning. *Arxiv Preprint Arxiv*, 2305.01486.
- [22] Albraikan, A. A., Alzahrani, J. S., Alshahrani, R., Yafoz, A., Alsini, R., Hilal, A. M., ... Gupta, D. 2022. Intelligent facial expression recognition and classification using optimal deep transfer learning model. *Image and Vision Computing*, 128, 104583.
- [23] Tan, M., Le, Q. 2019, May. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- [24] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [25] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... Rush, A. M. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.
- [26] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- [27] He, K., Zhang, X., Ren, S., Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [28] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K. 2017. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).
- [29] Utku, A., Akcayol, M. A. 2024. Neural network based a comparative analysis for customer churn prediction. *Muş Alparslan Üniversitesi Fen Bilimleri Dergisi*, 12(1), 137-148.