Dicle University Journal of Engineering (DUJE)

web: http://dergipark.gov.tr/dumf

Araştırma Makalesi / Research Article

Farklı Evrişimsel Sinir Ağı Mimarilerinin Yüz İfade Analizi Alanındaki Başarımlarının İncelenmesi

Investigation of the Performances of Various Convolutional Neural Networks Architectures on the Domain of Facial Expression Analysis

Ömer Faruk Söylemez[1*], Burhan Ergen[2]

[1] Dicle Üniversitesi, Bilgisayar Mühendisliği Bölümü, Diyarbakır, osoylemez@dicle.edu.tr

[2] Fırat Üniversitesi, Bilgisayar Mühendisliği Bölümü, Elazığ, bergen@firat.edu.tr

MAKALE BİLGİLERİ

Makale geçmişi:

Geliş: 25 Ocak 2020
Düzeltme: 12 Şubat 2020
Kabul: 10 Mart 2020

Anahtar kelimeler:

Evrişimler sinir ağları, yüz ifade analizi

ÖZET

Evrişimsel Sinir ağları (ESA), son yıllarda birçok çalışma tarafından öznitelik çıkarıcı olarak sıklıkla kullanılmaktadır. Geleneksel öznitelik çıkarım algoritmalarının aksine, etkileşim gerektirmeden otomatik olarak öznitelik çıkaran ESA'ların yardımıyla, birçok problem ve çalışma alanındaki başarımlar daha ileriye taşınmıştır. Bu çalışmada, farklı mimari özelliklere sahip olan ESA'ların yüz ifade analizi üzerindeki başarımları incelenmiştir. Öncelikle farklı ESA mimarileri tanıtılmış ve bu mimarilerin birbirlerinden farklılaştıkları kısımlar açıklanmıştır. FER2013 veri seti kullanılarak bütün ağ mimarileri üzerinde gerçekleştirilen eğitim ve doğrulama işlemleri sonucunda her bir mimariye ait başarım ve kayıp grafikleri sunulmuştur. Son olarak farklı mimarilerin yüz ifade analizi üzerindeki başarımlarının sebepleri tartışılmış ve gelecek çalışmalar için önerilerde bulunulmuştur.

Doi: 10.24012/dumf.679793

ARTICLE INFO

Article history:

Received: 25 January 2020
Revised: 12 February 2020
Accepted: 10 March 2020

Keywords:

Convolutional neural networks, facial expression analysis

ABSTRACT

Convolutional Neural Networks (CNN) have been widely used as a feature extractor by many studies in recent years. Contrary to the different feature extraction algorithms that extract features manually, with the help of CNNs that achieves feature extraction automatically without intervention, state of the art for many problem domains have been redefined. In this study, performances of various CNN architectures on the problem domain of facial expression analysis have been investigate. Initially, different CNN architectures have been introduced and then the implementations in which they differ from each other are explained. As a result of the training and verification processes performed on all network architectures using the FER2013 data set, performance and loss graphs of each architecture are presented. Finally, the reasons for the success of different architectures on facial expression analysis are discussed and suggestions are made for forthcoming studies.

* Sorumlu yazar / Correspondence
Ömer Faruk SÖYLEMEZ
✉ osoylemez@dicle.edu.tr

Please cite this article in press as O. F. Söylemez, B. Ergen, "Farklı Evrişimsel Sinir Ağı Mimarilerinin Yüz İfade Analizi Alanındaki Başarımlarının İncelenmesi", DUJE, vol. 11, no. 1, pp. 123-133, March 2020.

GİRİŞ

Yüz ifadeleri, geçmişten bu yana insanoğlunun duygularını aktarmak için kullandığı sözel olmayan iletişim türlerinin en önemlisidir. Duygu durumdaki değişiklikleri tasvir eden yüz ifadeleri, aynı zamanda bireyler arasındaki duygu paylaşımındaki en önemli rollerden birini üstlenmektedir. Birçok çalışma, yüz ifadelerinin, kültürlerden bağımsız olarak aynı hisleri temsil ettiğini göstermiştir [1], [2]. Yüz ifadeleri kişi veya toplumlara özgü değil evrenseldir ve bu, yüz ifadelerini duygu tasviri ve değişimi için ana elemanlardan birisi yapar.

Yüz ifadelerinin otomatik olarak analizi, geçtiğimiz son 20 yılın en güncel konularından birisidir [3], [4]. Başta insan bilgisayar etkileşimi olmak üzere, sürücü yorgunluk tespiti [5], kişisel güvenlik uygulamaları, sosyal pazarlama [6], interaktif oyunlar [7], sosyalleşebilir robotlar [8], yüz ifade sentezi [9] ve hasta duygu durum tespiti [10] gibi birçok uygulama alanına sahiptir.

Genel olarak öznitelik çıkarımı aşaması, bir örüntü tanıma sistemindeki en önemli adımdır. Veriyi ve verinin sınıfları arasındaki farkları doğru bir şekilde temsil edebilen öznitelik seçilmesiyle başarılı bir sınıflandırma işlemi gerçekleştirilebilir. Bu tarz bir senaryoda başarımı artırabilecek etmenler, verinin önişleme şekli ve seçilen sınıflandırma yöntemidir. Buna karşın, veriyi ve verinin sınıfları arasındaki farklılıkları temsil edemeyen öznitelikler yardımıyla gerçekleştirilen sınıflandırma işlemi sonucunda ise istenilen başarım yakalanamamaktadır.

Bir yüz ifade analizi yönteminin başarımı, bir örüntü tanıma problemi olduğu gerçeğini göz önüne alırsak, büyük bir şekilde seçilen özniteliklere bağlıdır. Bu öznitelikler geleneksel yöntemler yardımıyla manuel veya otomatik olarak çıkarılabilir [11]–[16]. Küçük ve kontrollü veri setlerinde geleneksel yöntemler yardımıyla seçilen öznitelikler iyi sonuçlar verebilmektedir. Fakat büyük veri setlerinde bu şekilde bir öznitelik çıkarımı yorucudur ve hatalara oldukça açıktır. Bu sebeple son 8 yıldır, büyük veri setlerinde öznitelik çıkarımı için ESA'lar sıklıkla kullanılmaktadır ve geleneksel öznitelik çıkarım yöntemlerine oranla daha başarılı olmuşlardır [3], [17], [18].

Çalışma FER2013 veri seti üzerinde gerçekleştirilmiştir [19]. Bu veri seti, 7 ifade sınıfına ait toplam 35887 adet 48x48 boyutunda gri-seviye imgeden oluşmaktadır. Veri setini hazırlayanlar tarafından rapor edilen insan başarımı %65±5'dir. Yarışma sonucunda Y. Tang [20] tarafından sunulan evrişimsel sinir ağı tabanlı yaklaşım doğrulama verisinde %69.7, test verisinde ise %71.1 başarım göstermiştir. Bu çalışma da kayıp fonksiyonu olarak L2-SVM kullanılmıştır. Yarışma sonrasında yayınlanan çalışmalar arasında en yüksek başarıma sahip olan çalışma Georgescu ve ark. [21] tarafından gerçekleştirilmiştir. Test verisi üzerinde %75.42 başarım gösteren bu çalışmada, öznitelik çıkarımı için 3 adet ESA ile birlikte BOVW kullanılmıştır. ESA'ların eğitimde sürecinde DSD [22] yönteminden faydalanılmıştır. Elde edilen öznitelik vektörü normalleştirilmiş ve ardından yerel karar destek makineleri ile sınıflandırılmıştır.

FER2013 veri seti üzerinde gerçekleştirilen çalışmalar, genel olarak birden fazla ESA ile birlikte geleneksel yöntemlerin de yardımıyla öznitelik çıkarma ve bu özniteliklerin farklı sınıflandırma yöntemleri ile sınıflandırılmasını içermektedir. Yukarıda bahsedilen çalışmalara ek olarak, aşağıdaki çalışmalar da farklı yaklaşımlarla benzer başarım oranları sergilemişlerdir: Connie ve ark. [23] %73.40, Kim ve ark. [24] %72.72, Yu ve Zang [25] % 72.

Bu çalışmanın asıl amacı, yüz ifade analizi problemi üzerine farklı ESA mimarilerinin başarımlarının incelenmesidir. Bu kapsamda farklı derinliklere ve parametre sayılarına sahip 8 adet özgün ESA, aynı parametrelerle eğitilmiş ve aralarındaki başarım farklarının kullanılan ESA'ya bağımlı bir değişken olması hedeflenmiştir. Bu kapsamda sabit tutulan parametreler şunlardır: Eğitim verisi, doğrulama verisi, yığın boyutu, imge boyutu (tasarım kısıtlamalarından ötürü InceptionV1 ağı hariç), epok, optimize edici, eğitim işleminin sıfırdan gerçekleştirilmesi, başlangıç ağırlık değerlerinin dağılımı ve yığınların eğitime gönderiliş sırası.

Bu çalışmanın devamı şu şekilde organize edilmiştir. "Materyal ve yöntem" bölümünde kullanılan FER2013 veri seti ile birlikte genel olarak ESA'lar hakkında önbilgi ve kullanılan ESA'ların mimarileri anlatılmıştır. Ek olarak ağların hazırlanması ve eğitim sürecinden bahsedilmiştir. "Sonuçlar ve Tartışma" bölümünde ise çalışma sonucunda elde edilen veriler paylaşılmış ve ilgili sonuçlar yorumlanarak ve ileriye dönük çalışmalar için önerilerde bulunulmuştur.

MATERYAL VE YÖNTEM

Evrişimsel Sinir Ağları, genellikle imgeler üzerinde görüntü işleme amacıyla kullanılan ve nesneye dayalı öznitelik çıkarma, sınıflandırma, segmentasyon ve sınıflandırma gibi farklı işlemleri de gerçekleştirebilen sinir ağlarına verilen genel bir isimdir. ESA'lar büyük boyutlu verileri çeşitli katmanlardan geçirerek küçük öznitelik vektörleri elde ederler. ESA'ları diğer sinir ağlarından ayıran ana özellik evrişimsel katmanları içermeleridir.

Temel olarak ESA'lar şu şekilde çalışır: Girdi olarak verilen görüntüye, evrişimsel katmanlarda evrişim çekirdeklerinin yardımıyla evrişim işlemi uygulanır. Bir filtre yardımıyla imge filtrelemeye benzeyen bu işlem neticesinde uzaysal düzlemdeki pikseller ve bu piksellerin komşuları ile olan ilişkilerinin oluşturduğu bir öznitelik haritası elde edilir. Kullanılan filtre adedi kadar elde edilen öznitelik haritalarının boyutları, havuzlama yardımıyla küçültülür. Elde edilen görüntülere tekrar evrişim işlemi veya havuzlama uygulanarak, imge boyutu elde edilmek istenilen öznitelik vektörü boyutuna ulaşıncaya kadar bu işlem tekrarlanır. ESA'lar oluşturulurken bu sıranın izlenmesi önemli değildir. Bir evrişimsel sinir ağında birbirini izleyen birden fazla evrişim katmanı olabilir.

ESA'lar ile ilgili gerçekleştirilen çalışmalar 1980'lere dayanmaktadır. Fakat günümüzde kullandığımız geri-yayılım ile gerçekleştirilen öğrenme işlemi sayesinde eğitilen ESA fikri ilk kez Y. Lecun tarafından 1998'de ortaya atılmıştır [26]. Son birkaç yıldaki ESA'lara olan yoğun ilginin asıl çıkış noktası ise AlexNet'in 2012 yılında ILSVRC [27] üzerinde gösterdiği başarımdır [28]. Grafik İşlem Birimlerinin (GİB), Merkezi İşlem Birimlerine (MİB) oranla veriyi paralel olarak çok daha seri bir şekilde işleyebilmeleri argümanından yola çıkılarak gerçekleştirilen AlexNET, aynı yöntem kümesini kullanan ve kendisinden sonra gerçekleştirilen birçok ESA mimarisine ilham kaynağı olmuştur.

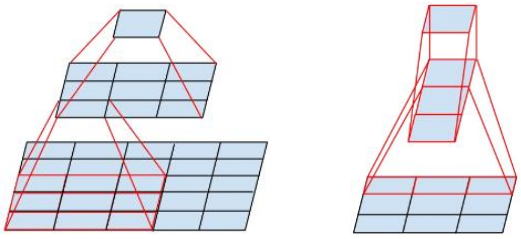
Bu çalışmamızda farklı ESA mimarilerinin yüz ifadesi tanıma işlemi üzerindeki başarımları incelenmiştir. Çalışmamızda kullanılan ESA'lar şunlardır: VGG16 [29], VGG16bnorm [29], InceptionV1 [30], InceptionV3 [31], Xception [32], ResNet50V1 [33], ResNet50V2 [34], MobileNetV1 [35] ve MobileNetV2 [36].

VGG16: Çalışmamızda ilk olarak VGG-16 ağı kullanılmıştır. VGG-16, AlexNet ile popülerleşen ESA'ların daha fazla katman ile daha yüksek başarım elde edebilecekleri fikri üzerine kurulmuştur. 13 evrişim ve 3 tam bağlı katman'dan (TBK) oluşan ağ, 138 milyon parametre içermektedir. VGG-16, AlexNet'e oranla daha küçük boyutlu filtreler (2x2 ve 3x3) kullanırken aktivasyon fonksiyonu olarak aynı şekilde ReLu kullanmaktadır.

Inceptionv1, Inceptionv3 ve Xception: Inceptionv1 ya da diğer adıyla GoogLeNET, "Network-In-Network" [37] çalışmasından esinlenerek Google mühendisleri tarafından geliştirilmiştir. Incepitonv1 ağını kendisinden önceki ağlardan ayıran ana özellik, ağ derinliği sağlamak için evrişim katmanlarını üst üste yığmaktan farklı olarak tasarlanan Inception modüllerini kullanmasıdır. Şekil 1'de gösterilen her bir Inception modülü, paralel olarak birbirlerine bağlı olan bir dizi evrişim katmanı içerir. Bu evrişim katmanlarında, görüntüden farklı öznitelikler çıkarmak için 1 x 1, 3 x 3 ve 5 x 5 boyutlarında filtreler kullanılmaktadır. Her bir katman dizisi, modülün sonunda birleştirilerek modülün çıktısını oluşturmaktadır. 1 x 1 filtreler, boyut indirgenmesi için kullanılmakta ve bu sayede işlem yükünü azaltmaktadırlar. Buna ek olarak 1 x 1 filtrelerde kullanılan aktivasyon fonksiyonları lineerliği bozmakta, bu da ağın genellemesine yardımcı olmaktadır. Inceptionv1 ağı, 22 katmandan oluşmakta ve 5 milyon parametre içermektedir.

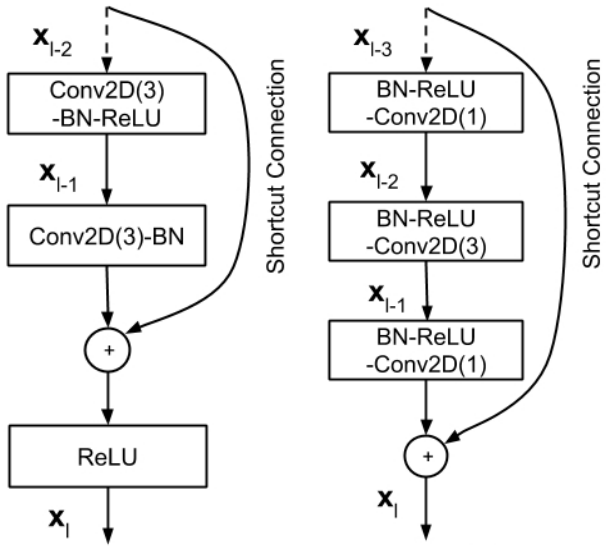
Inceptionv1 mimarisinin halefi olarak geliştirilen Inceptionv3 mimarisinde temsil darboğazı

giderilmeye çalışılmış, büyük filtreler daha küçük filtrelere ayrıştırılmış ve ağın genişliği ile derinliği dengelenmeye çalışılmıştır. Temsil darboğazının giderilmesi, evrişim katmanlarından sonra havuzlama yapılması ile sağlanmıştır. İşlem gücü bakımından daha maliyetli olsa da bu sayede ağın temsil yeteneği bir alt katmana daha iyi bir şekilde aktarılmıştır. n x n boyutlarındaki filtreler, 1 x n ve n x 1 olarak faktörize edilmiştir (Şekil 2.a). Bu asimetrik filtrelerle gerçekleştirilen evrişim işlemi, n x n boyutundaki filtre ile gerçekleştirilen sürümüne oranla işlem miktarını %33 azaltmaktadır. İşlem gücü ihtiyacını azaltmak için önerilen bir diğer yöntem ise büyük boyutlu filtrelerin gerçekleştirdiği işlemin, aynı işlemi yapabilecek daha fazla sayıda küçük boyutlu filtrelerle gerçekleştirilmesidir (Şekil 2.b).



(a) (b)

Şekil 2. a) 5x5 evrişimlerin yerine geçen mini ağ. b) 3x3 evrişimlerin yerine geçen mini ağ.



(a) (b)

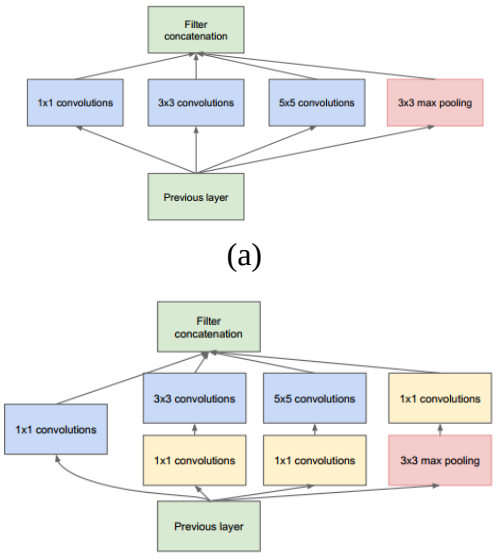
Şekil 3. a) Residual Modül v1
b) Residual Modül v2

Bu bağlamda 5 x 5 filtreler için iki adet 3 x 3 filtre, 7 x 7 filtreler için ise bir dizi 3 x 3 filtre kullanılmıştır. Bu uygulama neticesinde ise 5 x 5 filtreler için %28, 7 x 7 filtreler için ise %26 oranında işlem miktarı azaltılarak performans artışı sağlanmıştır.

Xception mimarisi, Incepiton modüllerinin derinlemesine ayrıştırılabilir evrişim modülleri ile değiştirildiği bir Inception mimarisi varyantıdır. InceptionV1 ile hemen hemen aynı sayıda parametreye sahiptir.

ResNetV1 ve ResNetV2: Derin ağlara daha fazla katman eklenerek ağların daha da derinleştirilmesi bir yere kadar başarıma katkıda bulunmaktadır. Fakat bu işlem her zaman başarımı arttıramayabilir. Belirli bir derinliğin üzerindeki ağlarda temsil yeteneği kaybolmakta, eğitim ve doğrulama başarım sabitleşmekte ve devamında ise azalmaya başlamaktadır. Microsoft Research tarafından geliştirilen ResNetV1, derin ağlara özgü olan bu durumun çözümüne "Residual" kısayol bağlantılar sunarak katkıda bulunmuştur. Bu şekilde tasarlanan ağlar, kaybolan ve patlayan gradyan problemlerine karşı daha gürbüz çalışmaktadırlar. ResnetV2'nin ResnetV1'den farklılaştığı nokta, kısayol bağlantılarının daha fazla blok sonrasında gerçekleşmesi ve blokların içerisindeki işlem sıralarının değişmesidir (Şekil 3).

MobileNetV1 ve MobileNetV2: Mobil ve gömülü görü uygulamaları için sunulan



(a)

(b)

Şekil 1. a) Saf Inception modülü
b)Boyut azaltmalı Inception modülü

MobileNet mimarisi Google ekibi tarafından geliştirilmiştir. Derinlemesine ayrıştırılabilir evrişim katmanları ile oluşturulan bu hafif mimaride, iki adet global hiperparemetre sayesinde gecikme ve başarım arasında seçim yapılabilmektedir. Bu parametreler sayesinde geliştirici, kendi problem kıstaslarına göre istediği ağı elde edebilmektedir. MobileNetV2 ağının MobileNetV1 ağından en büyük farkı, kullanılan ters "residual" yapıdır.

FER2013 Veri Seti:

Bu çalışmada FER2013 [38] veri seti kullanılmıştır. Kaggle üzerinde gerçekleştirilen bir yüz ifade analizi yarışması için oluşturulan veri seti, 28709'u eğitim, 3589'u doğrulama ve 3589'u test verisi olmak üzere toplam 35887 48x48 boyutunda gri-seviye yüz ifadesinden oluşmaktadır. Duygu içeren 184 anahtar kelimenin (öfkeli, keyifli, şaşkın v.b.), cinsiyet, yaş veya etnik köken gibi kelimelerle birleştirilmesiye yaklaşık 600 kelime dizisi oluşturulmuş ve bu diziler Google Image Search API'sinde yüz ifadesi sorguları olarak kullanılmıştır. Her bir sorgudan elde edilen ilk 1000 görüntü bir sonraki işlem aşamasında kullanılmıştır. Bu aşamada OpenCV yüz tanıma kütüphanesi yardımıyla yüz imgelerini sınırlayan çizgiler belirlenmiştir. Devamında ise bir ekip tarafından yanlış etiketlenmiş imgeler ayıklanmış, gerektiği takdirde yüz imgeleri sınır çizgileri tekrar belirlenmiş ve aynı resme ait diğer kopyalar veri setinden çıkarılmıştır. Son olarak elde edilen imgeler, Toronto Yüz Veri Setinde [39] belirtilen yüz ifadesi sınıfları ile eşleştirilmiştir. Ortaya çıkan nihai veri seti 4953 "Kızgın", 547 "İğrenme", 5121 "Korku", 8989 "Mutlu", 6077 "Üzgün", 4002 "Şaşkın" ve 6198 "Nötr" olmak üzere 35887 görüntüden oluşmaktadır. Veri seti üzerindeki insan duygu tespit başarımı %65±5 olarak belirlenmiştir. Veri setine ait örnek imgeler ve dağılımları Şekil 4'te gösterilmiştir.

Eğitim için Ağların Hazırlanması ve Eğitim Süreci:

Veri setindeki 48x48 boyutundaki imgelerin ağlara girdi olarak verilmek üzere yeniden boyutlandırılması için "en yakın komşu interpolasyonu" kullanılmıştır. İşlem yükü oldukça düşük olan bu interpolasyon, yeni oluşturulan pikselin değerinin, en yakın komşusuna eşitlenmesiyle gerçekleştirilir. InceptionV1 ağı hariç olmak üzere bütün ağlar için girdi imgesi boyutu 128 x 128 olarak belirlenmiştir. InceptionV1 mimarisinin kısıtlamalarından dolayı sadece bu ESA için girdi imgesi boyutu 192 x 192 olarak atanmıştır.

Şekil 4. FER2013 veri setindeki sınıflara ait örnekler, örnek sayısı ve dağılımı

Yeniden boyutlandırılma haricinde önişlem olarak imge yoğunluk değerleri "255" ile bölünmüştür. Bu şekilde girdi olarak verilen imgelerin [0, 1] aralığında yoğunluk değerlerine sahip olmaları hedeflenmiştir. Bir diğer önişlem ise yığınlar üzerinde gerçekleştirilmiştir. Yığındaki piksel değerlerinden yığının piksel yoğunluk ortalaması çıkartılmış ve ardından yeni piksel değerleri yığının standart sapmasıyla bölünmüştür. Bütün bu işlemler sonucunda piksel yoğunluk değerleri [-1, 1] aralığında dağılıma sahip olmuşlardır. Bunlara ek olarak veri artırmak amacıyla herhangi bir veri artırma yöntemi kullanılmamıştır.

Çalışmada kullanılan bütün ağların en baştan eğitildiğini daha öncesinde belirtmiştik. Bu kapsamda evrişim ve tam bağlı katmanlardaki ağların başlangıç ağırlıkları, "Glorot sürekli düzgün dağılım" [40] ile belirlenmiştir. Buna ek olarak bias terimi bütün ağırlıklar için "0" olarak atanmıştır.

Ağların eğitiminde optimize edici olarak "Adam" [41] kullanılmıştır. Adam, stokastik hedef fonksiyonlarının birinci dereceden gradyan tabanlı optimizasyonu için sunulan bir algoritmadır. Bu algoritma, uygulaması kolay, hesaplama açısından verimli olmakla birlikte düşük bellek gereksinimine sahiptir ve veri veya parametre bakımından büyük olarak sınıflandırılabilecek problemlere karşı başarılıdır.

Çalışmamızda veri seti geliştiricileri tarafından oluşturulan eğitim ve doğrulama verileri kullanılmıştır. Veri setine dışarıdan veri eklenmemiş veya veri seti içerisindeki olumsuz örnekler ayıklanmamıştır. Her bir ESA, 200 epok boyunca eğitim verisi ile eğitilmiş ve her epok sonrasında doğrulama verisi ile doğrulama başarımı ölçülmüştür. ESA'ların gösterdiği en yüksek doğrulama başarımı ilgili ESA'nın başarımı olarak kaydedilmiştir.

Bunlara ek olarak, ağların aynı kıstaslar altında eğitilebilmesi için eğitim ve test süreçleri boyunca imgeler aynı sıra ile gönderilmiştir.

SONUÇLAR ve TARTIŞMA

Bu bölümde ESA'ların eğitim ve test süreçlerinden bahsedilmiş ve çalışma sonucunda elde edilen veriler paylaşılmıştır. Ayrıca ilgili sonuçlar yorumlanmış ve ileriye dönük çalışmalar için önerilerde bulunulmuştur.

Çalışmamıza söz konusu olan her bir ESA mimarisi, FER2013 veri seti ile eğitilmiş ve ardından doğrulamaya tabii tutulmuştur. Çalışmamızda kullanmış olduğumuz bütün ESA'lar, farklı yaklaşımları tasarımlarının temellerine oturtarak farklı mimariler sunmuşlardır. Sunulan bu mimarilerin temel amacı, ILSVRC2012 [27] veri seti üzerindeki başarımı artırmak veya aynı başarımı daha düşük işlem karmaşıklığı ile elde etmektir. ESA'ların en önemli kullanım alanlarından birisi olan imge sınıflandırma üzerine sunulan ve toplam 1000 sınıf ve her sınıfta 1000 adet imge içeren ILSVRC2012 veri seti, ESA başarımı için temel endüstri standardı olmuştur.

Genel olarak ESA'ların eğitimi için "en baştan eğitme", "öğrenme transferi" ve "ince ayar" olmak üzere üç farklı yöntem izlenebilir. "En baştan eğitme", daha öncesinde herhangi bir eğitimden geçmemiş ağların eğitilme işlemi için kullanılan bir tabirdir. En baştan eğitme esnasında ağırlık değerleri, farklı dağılım yöntemleri yardımıyla hesaplanır ve atanır. Devamında ise belirlenen optimizasyon algoritması ve öğrenme oranı yardımıyla, ağ yakınsayana kadar öğrenme işlemi gerçekleştirilir. "Öğrenme transferi", daha öncesinde eğitilmiş olan bir ağın, aynı sınıflara fakat farklı sınıf dağılımına sahip bir veri seti ile veya eğitim kümesindeki sınıflardan tamamen farklı bir veri seti ile eğitilmesi vasıtasıyla gerçekleştirilir. Öğrenme transferindeki asıl amaç, daha öncesinde eğitilmiş ve belirli bir başarıma ulaşmış bir ağın öznitelik çıkarma yeteneğinin, farklı bir problem üzerinde kullanılmasıdır. ESA eğitimi için kullanılabilecek olan son yöntem ise "ince ayar"dır. Önceden eğitilmiş ağın sınıfları ile aynı sınıflara ait daha fazla verinin eğitilmesi ile gerçekleştirilen ince ayarda, öğrenme oranı düşük tutularak ağın başarımının yükseltilmesi hedeflenmektedir. Çalışmamızda yer alan ESA'lar, eğitildikleri problem uzayının (imge sınıflandırma) eğitilecek problem uzayından (yüz ifade analizi) farklı olması sebebiyle en baştan eğitilerek kullanılmışlardır.

Çalışmamızda ilk olarak VGG-16 ağı kullanılmıştır. Gerçekleştirdiğimiz eğitim işlemi sonucunda ağ eğitim verisine aşırı uymuş, %25 civarında bir başarım elde edilmiş ve eğitim başarısız olmuştur. VGG-16 ağı yaklaşık olarak 138 milyon parametre içermektedir. Bu denli

yüksek parametre içeren ağların sınırlı büyüklükte eğitim verileriyle eğitilmesi için regularizasyon yöntemlerinin kullanılması gerekmektedir. Ağ performansının iyileştirilmesi için, "Yığın normalizasyonu" ile regüle edilen VGG-16 ağı ile gerçekleştirdiğimiz eğitim işlemi sonucunda %65.0 doğrulama başarımı elde edilmiştir.

Çalışmamızda yer verdiğimiz ve Inception mimarisini temel alarak oluşturulan ağlar InceptionV1, InceptionV3 ve Xception'dır. Eğitimlerini gerçekleştirdiğimiz ağlardan 7 milyon parametre içeren InceptionV1 ağı %65.8 oranında bir başarım sergilemiştir. InceptionV1 ağına, getirdiği imge boyutu kısıtlamalarından ötürü diğer bütün ağlardan farklı olarak imge boyutu 192x192 piksel olarak verilmiştir. Aynı mimari temelli 24 milyon parametre içeren InceptionV3 ve 23 milyon parametre içeren Xception ağlarının sırasıyla %63.2 ve %61.1 başarım sağladığı gözlemlenmiştir.

Residual bağlantıların sunulduğu ResNet50V1 ve ResNet50V2 ağları ile gerçekleştirilen eğitim işlemleri sonucunda sırasıyla %59.5 ve %59.8 oranında başarım elde edilmiştir.

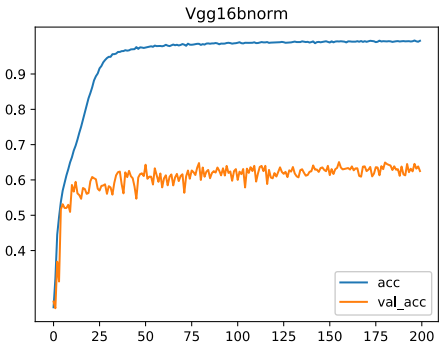
Son olarak, mobil ve gömülü görü uygulamaları için geliştirilen MobileNetV1 ve MobileNet V2 ağları çalışmamız kapsamında eğitilmiş ve iki ağın da %58.5 başarım sergilediği gözlemlenmiştir.

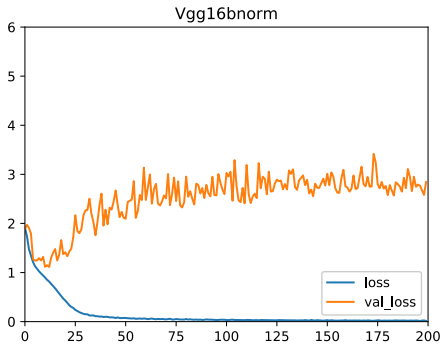
Tablo 1'de, çalışmamızda kullanılan ağlar ve bu ağlara ait parametre sayısı bilgisi ile birlikte eğitim esnasında kullanılan yığın boyutu, imge boyutu, eğitim süresi ve başarıma ilişkin bilgiler verilmiştir. Söz konusu ağlara ilişkin eğitim başarımı, eğitim kaybı, doğrulama başarımı ve doğrulama kaybı değerlerine ilişkin veriler Şekil 5'de gösterilmiştir.

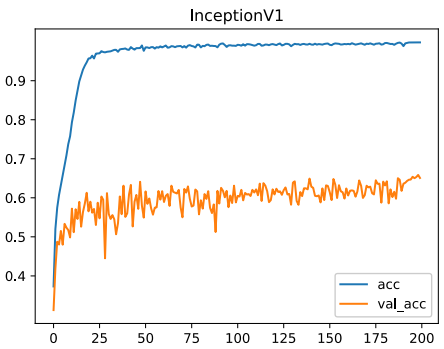
InceptionV3 ve MobileNetV2 ağlarının kayıp grafikleri incelendiğine, kayıpların bir değere yakınsamadığı ve büyük bir aralıkta dalgalandığı görülmektedir. Buna birden fazla etken sebep olabilir. Fazla parametre ve düşük regularizasyona sahip ESA'lar bu şekilde bir davranış sergileyebilmektedirler. Bu tarz ağlar, eğitim verilerini sınıflandırmak için birçok farklı yola sahiptirler. Fakat bu seçimlerden bazıları, eğitim verilerine aşırı uyum gösterebilmekte ve doğrulama verisinde yüksek hataya sebep olabilmektedir. Kayıp dalgalanmasına sebep olabilecek bir diğer etken ise örneklerin "güven" değerlerinin düşmesidir. İki sınıfla gerçekleştirilen ve doğru sınıfın ilk sınıf olduğu iki adet sınıflandırılma işlemi sonucunda [0.9 0.1] ve [0.6 .04] olasılık dağılımları elde edilsin. İki fonksiyon da örneği doğru bir şekilde sınıflandırmış olmasına rağmen ikinci fonksiyonun güven aralığı daha düşük ve kayıp fonksiyonunun çıktısı daha yüksektir. Çok sayıda örnekle gerçekleştirilen tahmin işlemi sonucunda her ne kadar başarım farklılık göstermese de kayıp bu yolla artmaktadır.

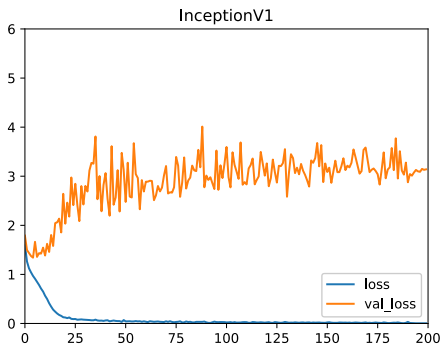
Tablo 1. Çalışmada kullanılan ağlar ve bu ağlara ait parametre sayısı ile birlikte eğitim esnasında kullanılan yığın boyutu, imge boyutu, epok ve başarıma ilişkin bilgiler.

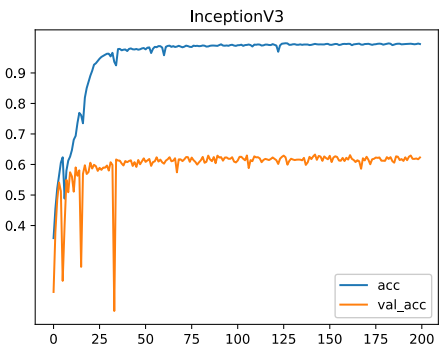
Ağ	Parametre	Yığın Boyutu	İmge Boyutu	Epok	Başarım (%)
VGG16	138 m	128	128x128	200	25.4
VGG16-Bnorm	138 m	128	128x128	200	65.0
InceptionV1	7 m	128	192x192	200	65.8
InceptionV3	24 m	128	128x128	200	63.2
Xception	23 m	128	128x128	200	61.1
Resnet50V1	25.6 m	128	128x128	200	59.5
Resnet50V2	25.6 m	128	128x128	200	59.8
MobileNetV1	4.2 m	128	128x128	200	58.5
MobileNetV2	3.5 m	128	128x128	200	58.5

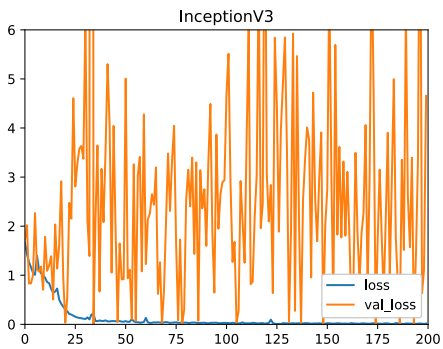

Vgg16bnorm
acc
val_acc

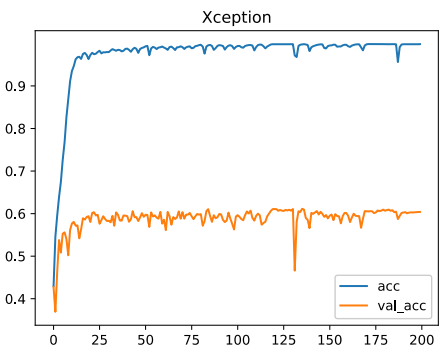

Vgg16bnorm
loss
val_loss

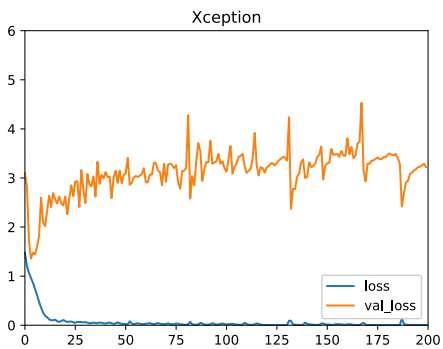

InceptionV1
acc
val_acc

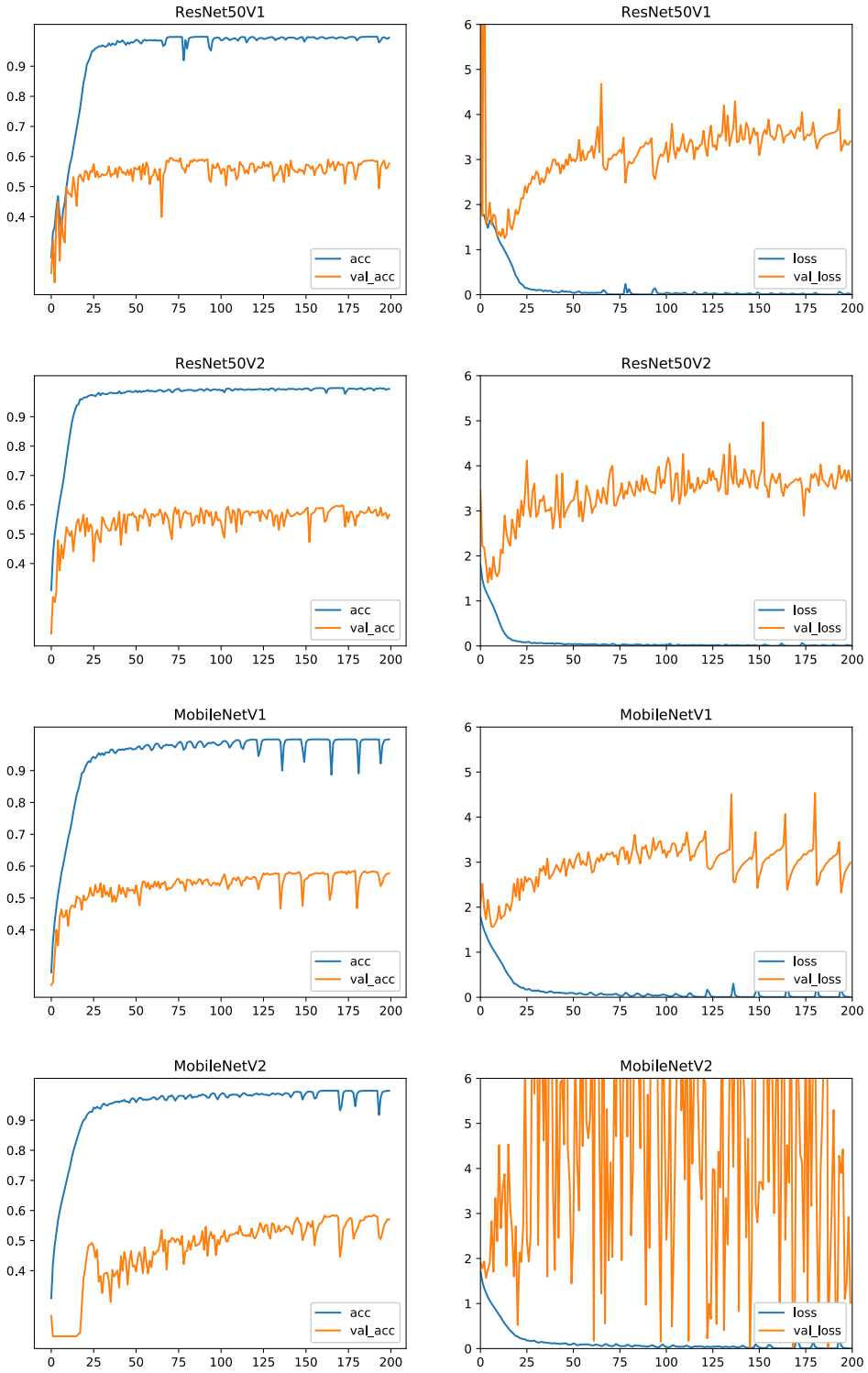

InceptionV1
loss
val_loss


InceptionV3
acc
val_acc


InceptionV3
loss
val_loss


Xception
acc
val_acc


Xception
loss
val_loss



Şekil 5. Söz konusu ağlara ilişkin eğitim ve doğrulama verileri. Lejant: (acc:test başarımı) (val_acc:doğrulama başarımı) (loss: test kaybı) (val_loss:doğrulama kaybı)

Derin yüz ifade analizi için FER2013 veri seti kullanılarak gerçekleştirilen çalışmamızın sonucunda en yüksek başarım %65.8 ile InceptionV1 mimarisi tarafından sergilenmiştir. Her ne kadar kendisine en yakın başarıma sahip olan VGG16bnorm (%65) ile arasında küçük bir fark gözlemlense de, parametre sayıları göz önüne alındığında InceptionV1 mimarisinin bunu yaklaşık 20 kat daha az parametre ile gerçekleştirmesi, ağın yüz ifade analizi

üzerindeki başarımını perçinlemektedir. Sırasıyla %63.2, %61.1, %59.5 ve %59.8 başarım sergileyen InceptionV3, Xception, Resnet50V1 ve ResNet50V2 ağları, birbirleri ile hemen hemen aynı parametre sayılarına sahip olan ağlardır. Diğer ağlara nazaran en az parametre sayısına sahip olan MobileNetV1 ve MobileNetV2 ağları %58.5 başarım göstermişlerdir. Eğitim ve test süreçleri için bu iki ağ arasında yapılacak bir seçimde daha az parametre içeren MobileNetV2 ağının seçilmesi daha uygundur. Son olarak VGG16 ağı, içerdiği fazla parametre ve regularizasyon eksikliğinden dolayı aşırı uyum göstermiştir ve başarımı %25.4 ile kısıtlı kalmıştır.

Bu çalışmada, yüz ifadesi analizi için farklı ESA mimarilerinin performansları karşılaştırılmıştır. Gelecek çalışmalarda, eğitimden önceki önişlemlerin, ilk parametre dağılımının, filtre boyutlarının ve ağ derinliğinin başarım üzerindeki etkisinin incelenmesi planlanmaktadır.

Kaynaklar

[1] C. Darwin, *The expression of the emotions in man and animals.* London: John Murray, 1872.

[2] P. Ekman *et al.*, "Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion," *J. Pers. Soc. Psychol.*, vol. 53, no. 4, pp. 712–717, 1987.

[3] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *arXiv Prepr. arXiv1804.08348*, pp. 1–22, 2018.

[4] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image Vis. Comput.*, vol. 30, no. 10, pp. 683–697, 2012.

[5] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A Survey on State-of-the-Art Drowsiness Detection Techniques," *IEEE Access*, vol. 7, pp. 61904–61919, 2019.

[6] A. M. Barreto, "Application of facial expression studies on the field of marketing," *Emot. Expr. brain face*, no. June, pp. 163–189, 2017.

[7] P. M. Blom *et al.*, "Towards personalised gaming via facial expression recognition," *Proc. 10th AAAI Conf. Artif. Intell. Interact. Digit. Entertain. AIIDE 2014*, pp. 30–36, 2014.

[8] L. Zhang, M. Jiang, D. Farid, and M. A. Hossain, "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot," *Expert Syst. Appl.*, vol. 40, no. 13, pp. 5160–5168, 2013.

[9] Y. Zhou and B. E. Shi, "Photorealistic facial expression synthesis by the conditional difference adversarial autoencoder," *2017 7th Int. Conf. Affect. Comput. Intell. Interact. ACII 2017*, vol. 2018-Janua, pp. 370–376, 2017.

[10] G. Muhammad, M. Alsulaiman, S. U. Amin, A. Ghoneim, and M. F. Alhamid, "A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities," *IEEE Access*, vol. 5, pp. 10871–10881, 2017.

[11] L. Wang, R. F. Li, K. Wang, and J. Chen, "Feature representation for facial expression recognition based on FACS and LBP," *Int. J. Autom. Comput.*, vol. 11, no. 5, pp. 459–468, 2014.

[12] Y. Chang, C. Hu, R. Feris, and M. Turk, "Manifold based analysis of facial expression," *Image Vis. Comput.*, vol. 24, no. 6, pp. 605–614, 2006.

[13] R. Shbib and S. Zhou, "Facial Expression Analysis using Active Shape Model," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 8, no. 1, pp. 9–22, 2015.

[14] U. Tekguc, H. Soyel, and H. Demirel, "Feature selection for person-independent 3D facial expression recognition using NSGA-II," *Comput. Inf. …*, pp. 35–38, 2009.

[15] H. Soyel and H. Demirel, "Facial expression recognition based on discriminative scale invariant feature transform," *Electron. Lett.*, vol. 46, no. 5, pp. 343–345, 2010.

[16] D. Al Chanti and A. Caplier, "Improving bag-of-Visual-Words towards effective facial expressive image classification," *VISIGRAPP 2018 - Proc. 13th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl.*, vol. 5, pp. 145–152, 2018.

[17] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," *Pattern Recognit.*, vol. 61, pp. 610–628, 2017.

[18] V. Tümen, Ö. F. Söylemez, and B. Ergen, "Facial emotion recognition on a dataset using Convolutional Neural Network," *IDAP 2017 - Int. Artif. Intell. Data Process. Symp.*, 2017.

[19] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, 2015.

[20] Y. Tang, "Deep Learning using Linear Support Vector Machines," 2013.

[21] M. I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted

features for facial expression recognition," *IEEE Access*, vol. 7, pp. 64827–64836, 2019.

[22] S. Han *et al.*, "DSD: Dense-sparse-dense training for deep neural networks," *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, 2019.

[23] T. Connie, M. Al-Shabi, W. P. Cheah, and M. Goh, "Facial expression recognition using a hybrid CNN-SIFT aggregator," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10607 LNAI, pp. 139–149, 2017.

[24] B. K. Kim, J. Roh, S. Y. Dong, and S. Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," *J. Multimodal User Interfaces*, vol. 10, no. 2, pp. 173–189, 2016.

[25] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," *ICMI 2015 - Proc. 2015 ACM Int. Conf. Multimodal Interact.*, pp. 435–442, 2015.

[26] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[27] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[28] A. Krizhevsky and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1–9.

[29] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, vol. abs/1409.1, 2015.

[30] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016.

[32] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1800–1807, 2017.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. abs/1512.0, pp. 770–778, 2016.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9908 LNCS, pp. 630–645, 2016.

[35] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[36] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018.

[37] M. Lin, Q. Chen, and S. Yan, "Network in network," *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, 2014.

[38] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, 2015.

[39] S. Rifai, Y. Bengio, A. Courville, P. Vincent, and M. Mirza, "Disentangling factors of variation for facial expression recognition," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7577 LNCS, no. PART 6, pp. 808–822, 2012.

[40] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, 2010.

[41] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015.