Transfer Öğrenme ve Topluluk Öğrenmesi Kullanarak Durağan Görüntüler Üzerinde Duygu Tanıma
Emotion Recognition on Static Images Using Deep Transfer Learning and Ensembling

Hüseyin Abanoz ve Zehra Çataltepe
Bilgisayar Mühendisliği Bölümü
İstanbul Teknik Üniversitesi
İstanbul, Türkiye
abanozh@itu.edu.tr, cataltepe@itu.edu.tr

***Özetçe* —Duygu tanıma bilgisayar ve insan etkileşiminin olduğu her alanda faydalı olabilmektedir. Evrişimsel Sinir Ağlarının (CNN) bilgisayarla görü görevlerinde iyi olduğu bilinmektedir. Fakat CNN'lerin eğitilmesi, özellikle eğitim verisinin ve hesaplama gücünün az olduğu durumlarda, oldukça zordur. Transfer öğrenme bu gibi durumlarda, ucuz ve etkili bir çözüm olarak göze çarpar. Transfer öğrenmede önceden eğitilmiş CNN sınıflandırıcıları kullanılır. Bu çalışma iki katkı sunar. Birinci katkı, farklı mimariler ve farklı veri kümeleri kullanılarak eğitilmiş CNN modelleri incelenerek duygu tanıma problemine uygun olanı bulunmaya çalışılmıştır. İkinci katkı olarak, her duygu için ayrı bir uzman sınıflandırıcı eğitilmiştir. Ana model, uzman modellerle topluluk öğrenmesi yöntemiyle birleştirilip daha iyi bir sınıflandırıcı elde edilmiştir. Deney sonuçları, transfer öğrenme ve topluluk öğrenmesi kullanılarak güçlü bir sınıflandırıcının elde edilmesinin mümkün olduğunu ortaya koymuştur. Eğilen sınıflandırıcı, FER13 doğrulama veri kümesi üzerinde %68.12 isabet oranı göstermiştir.**

Anahtar Kelimeler—Derin Öğrenme, Transfer Öğrenme, Kümeleme, Duygu Tanıma

***Abstract*—Emotion recognition may be useful in any area where human and computer interacts. CNNs are known to be good at computer vision tasks. However, CNNs are difficult to train, especially when the amount of data and computation power is limited. Transfer learning emerges as a cheap and efficient way of making use of pre-trained CNN classifiers. Our work has two contributions. Firstly, different CNN architectures and models trained using different datasets are investigated to find a suitable model to use in emotion recognition. Secondly, expert models for each emotion are trained. The Base model is ensembled with expert models to create a better classifier. Experiments show that our use of ensembling together with transfer learning helps to create a good classifier. Final classifier shows 68.12% accuracy on FER13 validation set.**

Keywords—Deep Learning, Transfer Learning, Ensembling, Emotion Recognition

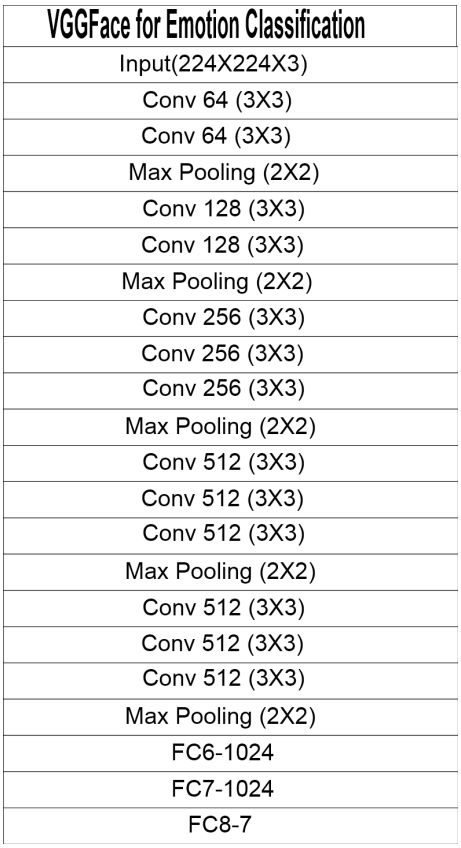
978-1-5386-1501-0/18/$31.00 ©

I. GİRİŞ

Duygu tanıma makine öğrenmesi alanındaki popüler konular arasındadır. Yüz ifadeleri duygu tanıma için önemli bir kaynaktır. Derin Sinir Ağları, özellikle de Evrişimsel Sinir Ağları (CNN), bilgisayarla görü alanında yaygın olarak kullanılmaktadır. Duygu tanıma, CNN'lerin başarılı olduğu problemler arasındadır [1]–[5].

Duygu tanıma temel olarak bir sınıflandırma problemidir. Bu çalışmada 7 duygu ele alınmıştır: Kızgınlık, İğrenme, Korku, Mutluluk, Üzüntü, Şaşkınlık ve herhangi bir duygunun ifade edilmediği Yalın.

Duygu tanıma problemi yayın biçimde çalışılmaktadır. Çalışmaları teşvik etmek için yarışmalar düzenlenmektedir. EmotiW bu yarışmalardan birisidir. EmotiW yarışmasında, CNN temelli yöntemlerin ağırlıklı olarak kullanıldığı görülmektedir. Hong-Wei et al. önceden eğitilmiş CNN modelleri kullanmıştır [1]. AlexNet ve VGG-CNN-M-2048 modellerini transfer öğrenme için kullanılmıştır. Eğitim kümesinin boyutunu artırmayı böylece daha iyi genelleme elde etmeyi ve aşırı öğrenmeden kaçınmayı hedeflemiştir. Bu amaçla, yarışma tarafından sağlanan veri kümesine ek olarak FER2013 veri kümesini de kullanmıştır. Zhiding et al. aynı mimariye sahip olan fakat farklı ilk değerlerle eğitilmiş CNN sınıflandırıcıları kullanmıştır [2]. Eğitim için FER2013 veri kümesini kullanmıştır. Sonra da yarışma tarafından sağlanan SFEW veri kümesi kullanılarak ince ayar yapılmıştır. Son olarak, farklı yitim fonksiyonları kullanılarak, eldeki CNN sınıflandırıcıları için en iyi model topluluğu elde edilmiştir. Bo-Kyeong et al. kural tabanlı hiyerarşik model topluluğu kullanmıştır [3], [4]. Topluluk öğrenmesi için, hem farklı ilk değerlerle eğitilmiş aynı mimariye sahip CNN modelleri hem de farklı mimariye sahip CNN modelleri kullanmıştır. Farklı sayıda hiyerarşi katmanı kullanarak deneyler gerçekleştirmiştir. [14] transfer öğrenme ve topluluk öğrenmesi yöntemlerini kullanmıştır. Transfer öğrenme, FER13 veri kümesi üzerinde eğitilen modeller üzerinde yapılmış, topluluk öğrenmesi aşamasında aynı modeller öznitelikleri ayrıştırmak için kullanılmıştır. Sonuçta, her sınıf için ayrı bir doğrusal sınıflandırıcı eğitilirken, aynı



Şekil 1: Duygu Tanıma için Kullanılan VGGFace Modeli

CNN sınıflndırıcısının ürettiği derin öznitelikler kullanılmıştır.

Bu çalışma [14] ile benzerlikler içermekle birlikte metodların uygulanması farklı icra edilmiştir. Öncelikle, duygu tanımada transfer öğrenmeye uygun modeller araştırılmıştır. Transfer öğrenme için Imagenet gibi çok büyük veri kümeleri üzerinde eğitilmiş modeller kullanılmıştır. Topluluk öğrenmesi için, yeniden yapılandırılan veri kümeleri kullanılarak yine transfer öğrenme yöntemiyle her sınıfa özel CNN sınıflandırıcıları eğitilmiştir. Topluluk öğrenmesi uygulanırken bütün sınıfların çıktılarını kullanılmamış, hata dizeyine göre seçim yapılmıştır.

II. KULLANILAN YÖNTEM

CNN'ler çok sayıda katmanlardan oluşur. İlk katmanlar görüntülerden düşük seviyeli özniteliklerin ayrıştırılması için kullanılır. En üstteki katmanlar, nesnelerin sınıflandırılması için uygun olan öznitelikleri ayrıştırır. Bu yüzden, farklı sınıflandırma problemleri için, önceden eğitilmiş bir CNN modeli, ilk katmanlarına dokunmadan, son katmanları yeniden eğitilerek kullanılabilir. Bu yönteme Transfer Öğrenme denilmektedir. Transfer Öğrenme, az sayıda veri ve düşük hesaplama kabiliyetleriyle güçlü sınıflandırıcılar eğitilmesini mümkün kılar.

Büyük veri kümeleri kullanılarak eğitilmiş popüler CNN modelleri bulunmaktaktadır. VGGNet [6], VGGFace [5], ResNet50 [8] and Inception [10] ağları Imagenet [7] veri kümesi kullanılarak eğitilmiştir. Imagenet veri kümesi 1000 sınıfa ait olan on milyondan fazla görüntü içermektedir. Imagenet Large Scale Visual Recognition Competition (ILSVRC) yarışması her sene düzenlenmektedir.

Tablo I: FER13 YENIDEN EĞITME GEÇERLEME SONUÇLARI

Önceden Eğitilmiş Model	FER13 Geçerleme İsabet
Inception	0.567
ResNet50	0.562
VGG16	0.6718
VGGFace	0.6779

VGGNet , ILSVRC 2014 sınıflandırma yarışmasında ikinci sırayı elde etmiştir. Küçük evrişimsel filtrelerden(3x3) oluşan 16-19 katman içerir. Inception ağı ILSVRC 2014 yarışmasından birinci sırada yer almıştır. Inception ağı, CNN mimarilerinin sıralı katmanlar yığını şekilde olmak zorunda olmadığını ortaya koymuştur. Modüler CNN mimarisi fikrini ortaya atmıştır. ResNet50 ağı ILSVRC 2015 yarışmasının kazananıdır. Girdinin aktivasyon fonksiyonuna ilave edildiği artık bloklar fikrini ortaya atmıştır. 152 katmandan oluşur. VGGFace modeli temel olarak 16 katmandan oluşan bir VGGNet ağıdır. 2.6 milyon yüz görüntüsünden oluşan VGG Face veri kümesi kullanılarak eğitilmiştir [5].

Transfer öğrenme iki aşamada uygulanır. Asıl yeniden eğitme aşamasından önce, buna uzun yeniden eğitme diyeceğiz, yeni eklenen katmanlara ilk değerleri atamak için kısa bir yeniden eğitme, kısa yeniden eğitme diyeceğiz, uygulanır. Kısa yeniden eğitme esnasında sadece yeni eklenen tam bağlı katmanlar ve softmaks katmanı eğitilir. Uzun yeniden eğitme aşamasında, tam bağlı katmanlar ve softmaks katmanı sonda bulunan evrişimsel katmalarla birlikte yeniden eğitilir.

III. VERİ KÜMESİ

Fer2013 [9] veri kümesi, 32K gri renkli 48X48 piksel boyutunda, yüz görüntülerinden oluşur. Küçük boyutlu görüntüler içermesine rağmen, veri kümesi duygu tanıma çalışmalarında yaygın biçimde kullanılır [1]–[5]. Bunun nedeni veri kümesinin içerdiği örnek sayısının çok olmasıdır. Deneylerimiz esnasında yüz içermeyen görüntüler, manuel olarak veri kümesinden çıkarılmıştır. Toplamda 72 tane görüntü silinmiştir. Bu sayı veri kümesinin büyüklüğü göz önünde tutulunca göz ardı edilebilir. Tablo II'de veri kümesinde bulunan örnek görüntüler görülebilir. Tablo III'de her sınıfta bulunan örnek sayısı gösterilmiştir.

IV. DENEY ORTAMI

Öğrenme katsayısının seçimi oldukça önemlidir. Transfer öğrenmede, küçük öğrenme katsayısı tercih edilmelidir. [1] de olduğu gibi öğrenme katsayısı 0.001 olarak kullanılmıştır. Yapılan deneyler, bu değerin optimum değer olduğunu göstermiştir. Eğitimi hızlandırmak için moment değeri olarak 0.9 kullanılmıştır. Daha küçük ağırlıklar öğrenmek için sönüm değeri olarak 0.0005 kullanılmıştır. Eniyileyici olarak Rastgele Eğim Düşüşü (Stochastic Gradient Descent) kullanılmıştır.

Aşırı öğrenmeden kaçınmak için Dropuot [11] 0.5 oranı ile 0.5 kullanılmıştır ve sadece softmaks katmanından önce uygulanmıştır.

Görüntü Artırma ile veri kümesinde bulunan görüntüler üzerinde doğrusal dönüşümler uygulanarak yeni eğitim örnekleri

Tablo II: VERI KÜMESINDEN ÖRNEKLER

Kızgınlık	Korku	İğrenme	Mutluluk	Üzüntü	Şaşkınlık
					

Tablo III: SINIF BAŞINA DÜŞEN ÖRNEK SAYILARI

	Kızgınlık	Korku	İğrenme	Mutluluk	Üzüntü	Şaşkınlık	Yalın
Eğitim	3979	435	4090	7198	4951	4827	3163
Geçerleme	467	56	496	895	607	653	415

oluşturuldu. Uygulanan doğrusal dönüşümler, rastgele belirlenen çevirme, döndürme ve yaklaştırma kombinasyonlarıdır. Bu şekilde modelin aşırı öğrenmeden korunması hedeflenir. Ön işleme esnasında, bütün görüntüler 224x224 olacak şekilde yeniden boyutlandırılmıştır.Piksel değerleri [0,1] aralığında olacak şekilde yeniden boyutlandırılmıştır. Bellek alanı sınırlı olduğu için, rastele seçilen 64 görüntülük küçük öbekler diskten okunmuş, doğrusal dönüşümler uygulanmış ve eğitim için ağı beslemiştir.

V. SONUÇLAR

VGG16, Inception, ResNet50 ve VGGFace modelleri üzerinde deneyler yapılmıştır. Tablo I'te FER13 geçerleme kümesi üzerindeki isabetlilik oranı gösterilmiştir.

En yüksek isabet değeri VGG16 ve VGGFace modelleri kullanılarak elde edilmiştir. VGG16 ve VGGFace aynı mimariye sahip olduğu için bu sonuç şaşırtıcı değildir. VGGFace yüz görüntüleri ile eğitildiği için, duygu tanıma probleminde kullanılmak için daha elverişlidir. İlerideki deneyler VGGFace modeli kullanılarak yapılmıştır. VGGFace modelinin yapısı Şekil 1'de görülebilir.

Eğitim ve geçerleme isabet değerlerinin grafiği Şekil 2'de gösterilmiştir. Kırmızı çizginin sağ tarafı, kısa yeniden eğitme aşamasını, sol tarafı ise uzun yeniden eğitme aşamasını temsil eder. Kırmızı çizgi üzerindeki ani düşüş, iki ayrı eğitim aşamasına ait grafiklerinin birleştirilmesinden kaynaklanmaktadır. Uzun yeniden eğitme aşaması başladığı zaman, evrişimsel katmanların ağırlık değerleri güncellenmeye başlanmış, bunun sonucunda isabet değerinde düşüş gözlemlenmiştir.

Hata dizeyi (Confusion Matrix) elde edilen sınıflandırıcının güçlü ve zayıf yönlerini özetlemektedir. İğrenme sınıfındaki düşük başarım, bu sınıfa ait olan örneklerin sayısının düşük olmasıyla açıklanabilir. Korku ve Üzüntü sınıfları çok sayıda örnek içermelerine rağmen, başarım düşük kalmıştır. Bunun nedeni yine hata dizeyine bakılarak anlaşılabilir. Bu iki sınıf birbirleriyle karıştırılmıştır. Bu duygu tanıma probleminde sık gözlemlenen bir sorundur. Farklı duygulara ait mikro ifadeler, büyük oranda benzerlikler gösterebilmektedir. Bu gerçek, durağan yüz görüntüleri üzerinde duygu tanıma problemini zorlaştırır. Başarımı etkileyen diğer etmen, veri kümesinde bulunan gürültü miktarıdır [12], [13]. Bir çalışmaya göre FER13 veri kümesindeki etiketlerin doğruluk oranı %65 artı eksi%5 civarındadır [13]. Tablo II bu konuda fikir verebilir. Korku örneği Kızgın örneğini andırmaktadır. Üzgün örneği ise Yalın örneğine benzemektedir. İğrenme örneği, Yalın örneği gibi hatta Şaşırmış örneği gibi durmaktadır.

Yeninde eğitim esnasında eğitilecek evrişimsel katmanların sayısının belirlenmesi önemlidir ve veri kümesi büyüklüne bağlı olarak deneysel olarak tespit edilmelidir. Eğer veri kümesi büyük ise, daha fazla evrişimsel katman eğitilebilir. Eğer veri kümesi küçük ise, az sayıda evrişimsel katman eğitime dâhil edilmelidir. Tablo IV deneylerin sonuçlarını göstermektedir. En iyi sonuç 5 tane katmanın eğitilmesiyle elde edilmiştir.

Tablo IV: EĞITILEN EVRIŞIMSEL KATMAN SAYISI DENEYI

# of Convolutional Layers Trained	Validation Accuracy
2	0.6712
5	0.6779
8	0.6676

VI. TOPLULUK ÖĞRENMESİ

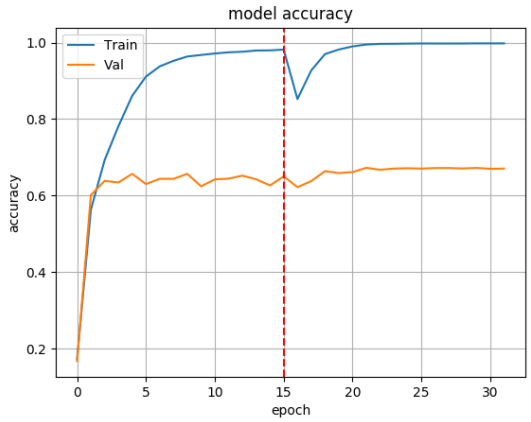
Farklı kabiliyetlere sahip modellerin Topluluk Öğrenmesi yöntemiyle birleştirilmesi daha iyi bir sınıflandırıcı ile sonuçlanabilmektedir. Deneyler esnasında farklı modeller yeniden eğitilmiştir: VGGFace, ResNet50 ve Inception. Fakat hata dizeyleri benzer sonuçlar göstermiştir. Bu yüzden topluluk öğrenmesi için uygun modeller elde edilememiştir.

Farklı kabiliyette sınıflandırıcılar elde etmek için, FER13 veri kümesi kullanılarak, yeni veri kümeleri elde edilmiştir. İki sınıflı veri kümeleri oluşturulmuştur İki sınıflı bir veri kümesi oluşturmak için, 7 sınıftan birisi ana sınıf olarak seçilmiştir. Yeni veri kümesinde, ana sınıfa ait örnekler ilk sınıf olarak etiketlenmiştir. Diğer sınıflara ait örnekler ise ikinci sınıf olarak etiketlenmiştir. Bu yöntemle, her birinde farklı bir ana sınıf seçilerek, 7 ayrı ikili veri kümesi elde edilmiştir.

Yeni veri kümeleri kullanılarak VGGFace modeli yeniden eğitilmiştir. Elde edilen yeni sınıflandırıcılar eğitildikleri veri kümesinin ana sınıfını sınıflandırma konusunda uzmandırlar. Örneğin, eğer Korku sınıfı ana sınıf seçildiyse, yeni veri kümesinde bütün korku örnekleri ilk sınıfa konulmuştur. Diğer sınıflara ait örnekler ise ikinci sınıfa konulmuştur. Sonuçta, bu veri kümesi ile eğitilen model, Korku sınıfında uzmandır. Bu şekilde eğitilen model, Korku örnekleriyle diğer örnekler arasındaki farka neden olan desenleri bulabilir.

Önemli bir husus şudur: Eğer diğer bütün örnekler ikinci sınıfa eklenirse, oluşacak veri kümesi dengesiz olacaktır. Bu yüzden, ikinci sınıf oluşturulurken diğer sınıflardan rastgele örnekler seçilir. Seçilen örneklerin sayısı ilk sınıftaki örneklerin sayısı ile sınırlandırılır. Uzman sınıflandırıcılar topluluk öğrenmesi kullanılarak birleştirilebilir. Bu şekilde başlangıçta eğitilen 7 sınıflı VGGFace modelinin başarısı artırılabilir. Bu amaçla hata dizeyi incelenirse şu gözlem yapılır: Korku ve Üzüntü sınıflarının başarısı düşüktür ve bu sınıflar bir birleriyle karıştırılmıştır (Şekil 3). 7 sınıflı VGGFace modelini Korku-Uzmanı VGGFace ve Üzüntü-Uzmanı VGGFace modelleriyle birleştiren başarılı bir topluluk öğrenmesi sonucunda performansın artması beklenir.

Bu deneyi gerçekleştirmek için ana model ve uzman modellerden elde edilen öznitelikler diske yazılmıştır. Özniteliklerin hangi katmandan alınması gerektiğine dair FC6, FC7 ve FC8 katmanları kullanılarak deneyler yapılmıştır. FC6 katmanından alınan özniteliklerin en iyi sonuç verdiği gözlemlenmiştir. Elde edilen öznitelikler kullanılarak tek gizli katmanlı, softmaks aktivasyonu kullanan yapay sinir ağı eğitilmiştir. Sonuçta elde edilen sınıflandırıcının geçerleme kümesindeki isabet değeri 0.6812 olarak gözlenmiştir. Bu değer topluluk öğrenmesi kullanılmadan elde edilmiş 0.6779 değerinden daha iyidir ve topluluk öğrenmesinin başarısını göstermektedir.



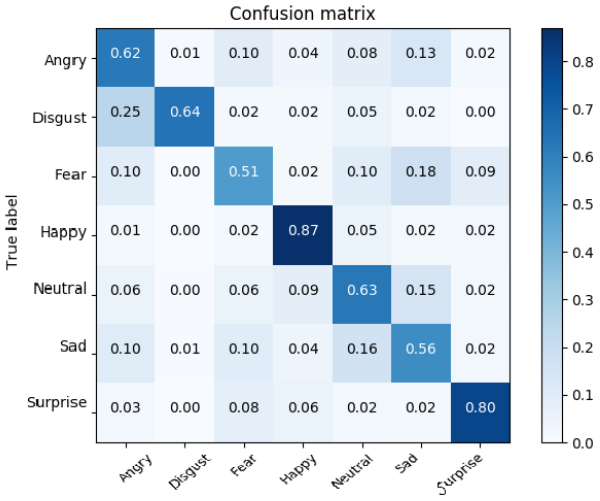
Şekil 2: VGGFace FER13 Eğitim/Geçerleme İsabet Grafiği

VII. SONUÇ

Evrişimsel Sinir Ağlarının en baştan eğitilmesi zordur. Çok miktarda veri ve işlem gücü gerektirir. Transfer Öğrenme önceden eğitilmiş CNN sınıflandırıcılarını kullanan kolay ve etkili bir öğrenme yöntemidir.

Transfer Öğrenme için probleme uygun olan modelin seçilmesi önemlidir. Örneğin, duygu tanıma probleminde, yüz görüntüleri kullanılarak eğitilmiş VGGFace benzeri bir model seçimi yerinde olacaktır. Eğitilecek evrişimsel katmanların sayısı eldeki veri miktarına göre deneysel olarak belirlenmelidir.

Farklı kabiliyetlere sahip CNN sınıflandırıcıları Topluluk Öğrenmesi yöntemiyle daha güçlü sınıflandırıcıların elde edilmesinde kullanılabilir.



Şekil 3: VGGFace FER13 Geçerleme Hata Dizeyi

KAYNAKÇA

[1] H.Ng, V.D. Nguyen, V. Vonikakis, S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning", Proceedings of the 2015 ACM on international conference on multimodal interaction 443–449, 2015.

[2] Z. Yu, C. Zhang, "Image based static facial expression recognition with multiple deep network learning", Proceedings of the 2015 ACM on International Conference on Multimodal Interaction 435–442, 2015.

[3] B. Kim, J. Roh, S. Dong, S. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition", Journal on Multimodal User Interfaces Vol 10 pages 173–189, 2016.

[4] B. Kim, H. Lee, J. Roh, S. Lee, "Hierarchical committee of deep cnns with exponentially-weighted decision fusion for static facial expression recognition", Proceedings of the 2015 ACM on International Conference on Multimodal Interaction 427–434, 2015.

[5] O.M. Parkhi, A. Vedaldi, Andrea, A. Zisserman, "Deep Face Recognition", PBMVC Vol 1 page 6, 2015.

[6] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.

[7] J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database", Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on pages 248–255, 2009.

[8] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE conference on computer vision and pattern recognition pages 770–778, 2016.

[9] P.L. Carrier, A. Courville, "FER-2013 face database", Technical report, 1365, Université de Montréal, 2013.

[10] S. Christian , L. Wei, J. Yangqing, S. Pierre, R. Scott, A. Dragomir, E. Dumitru, V. Vincent, R. Andrew, "Going deeper with convolutions", CoRR Vol. abs/1409.4842, 2015.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", The Journal of Machine Learning Research Vol. 15 pages 1929–1958, 2014.

[12] E. Barsoum, C. Zhang, C.C. Ferrer, Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution", Proceedings of the 18th ACM International Conference on Multimodal Interaction pages 279–283, 2016.

[13] I. J. Goodfellow, D. Erhan, P.L. Carrier, A. Courville, "Challenges in representation learning: A report on three machine learning contests", International Conference on Neural Information Processing pages 117–124, 2013.

[14] A. Savoiu, J. Wong, "Recognizing Facial Expressions Using Deep Learning",, 2017.