BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

*Araştırma Makalesi - Research Article*

# Yeni Bir Veri Kümesi (RidNet) Kullanarak Kontrolsüz Ortamda Yüz İfadesi Tanımanın Derin Öğrenme Yöntemleri ile İyileştirilmesi

Rıdvan ÖZDEMİR[1], Mehmet KOÇ[2*]

***Geliş / Received: 11/11/2019*** ***Revize / Revised: 03/12/2019*** ***Kabul / Accepted: 03/12/2019***

**ÖZ**

Bu çalışmada, internetten genel erişime açık görüntüler kullanılarak oluşturulan veri kümesi (RidNet) ile yedi farklı yüz ifadesi için derin öğrenme yöntemleri kullanılarak duygu tanıma işlemi yapılmıştır. Daha sonra AlexNet, GoogLeNet ve ResNet101 gibi literatürdeki tanınmış evrişimli sinir ağları mimarileri ile RidNet üzerinden transfer öğrenmesi yapılmıştır. Compound Facial Expressions of Emotion (CE) ve Static Facial Expressions in the Wild (SFEW) veri kümeleri test veri kümeleri olarak belirlenmiştir. Yapılan ilk deneysel çalışmalar ile en iyi sınıflandırma performansını gösteren evrişimli sinir ağı mimarisi belirlenmiştir. Bu evrişimli sinir ağı AffectNet, The Karolinska Directed Emotional Faces (KDEF) ve RidNet ile eğitilmiştir. AffectNet, KDEF ve RidNet ile eğitilmiş ağlar kontrollü ortamda oluşturulan veri kümesi (CE) ile test edildiğinde benzer sınıflandırma başarımları elde edilmiştir. Kontrolsüz ortamdaki test veri kümesinde (SFEW) ise RidNet ile eğitilen ağ diğer ağlara belirgin bir üstünlük sağlamıştır.

**Anahtar Kelimeler- Yüz İfadesi Tanıma, Derin Öğrenme, Transfer Öğrenmesi, Evrişimli Sinir Ağları.**

[1]İletişim: ridvan.ozdemir@bilecik.edu.tr (https://orcid.org/0000-0002-8599-1709 )
*Endüstriye Dayalı Mesleki Eğitim Uygulama ve Araştırma Merkezi, Bilecik Şeyh Edebali Üniversitesi, Bilecik*
[2]*Sorumlu yazar iletişim: mehmet.koc@bilecik.edu.tr (https://orcid.org/0000-0003-2919-6011 )
*Elektrik Elektronik Mühendisliği, Bilecik Şeyh Edebali Üniversitesi, Bilecik*

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

# Enhancing Facial Expression Recognition in the Wild with Deep Learning Methods Using a New Dataset: RidNet

**ABSTRACT**

In this study, emotion recognition process is performed by using deep learning methods for seven different facial expressions from the dataset (RidNet) which is created by using images that are publicly accessible from internet. Afterwards, transfer learning over RidNet is done with well-known convolutional neural network architectures such as AlexNet, GoogLeNet and ResNet101. Compound Facial Expressions of Emotion (CE) and Static Facial Expressions in the Wild (SFEW) datasets are determined to be used as test datasets. In the first experimental studies, convolutional neural network architecture with the best classification performance is determined. This convolutional neural network is trained using AffectNet, The Karolinska Directed Emotional Faces and RidNet. Similar classification performances are achieved when the AffectNet, KDEF, and RidNet-trained networks are tested with the dataset (CE) generated in a controlled environment. In the test dataset (SFEW) in an uncontrolled environment, RidNet-trained network gives a significant advantage over the other networks.

**Keywords- Facial Expression Recognition, Deep Learning, Transfer Learning, Convolutional Neural Networks.**

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

## I. GİRİŞ

Yüz ifadeleri, insanların hislerini anlık olarak dışavurum biçimidir. Charles Darwin 1872'de yayımlanan kitabında, insanların ve hayvanların doğuştan sahip olduğu temel duygularının, tüm dünyada aynı anlama gelen, yüz ifadeleri ile dışa yansıtıldığını belirterek, yüz ifadelerinin evrenselliğinden bahsetmiştir. Görsel olarak nesne tanıma ve yüz ifadesine dayalı duygu tanıma problemlerinin çözümünde çok yüksek sayıda gizli katman içeren sinir ağları, bir başka deyişle, derin sinir ağları kullanımı son zamanlarda giderek artmıştır. Derin sinir ağları, sinir ağlarına göre birçok yönden daha gelişmiş olsa da sahip olduğu çok yüksek sayıdaki katman sebebiyle eğitim için büyük verilere ihtiyaç duyar.

Yüze dayalı duygu tanıma problemlerinin çözümünde derin öğrenme alanında gelişen son teknolojinin ancak makul yanıtlar verebileceği öngörülmektedir. Derin sinir ağları esnek öğrenme görevlerinde üstün çalışma kabiliyeti göstermektedir. Sinir ağlarının özellikle de derin sinir ağlarının eğitimiyle yüze dayalı duygu tanıma problemlerinin çözümü için öznitelik tanıma ve çıkarımında gerekli olan süre ciddi bir oranda düşürülebilmektedir. Daha sonra bu ağları yeni veriler üzerinde de test ettiğimizde yüksek başarım oranlarına erişilebilir.

Derin sinir ağları mimarisinin bir alt kümesi olan evrişimli sinir ağları (ESA) bilgisayarlı görü ve derin öğrenme araştırmacıları tarafından geleneksel bir yöntem olarak benimsenmiştir. 2014 yılında nesne tanıma için düzenlenen ImageNet yarışmasında da ilk üç finalist evrişimli sinir ağları yaklaşımını kullanmışlardır. Bunlar arasından sınıflandırmada %6,66'lık hata oranına inmeyi başaran GoogLeNet mimarisi dikkat çekici bir başarıya ulaşmıştır [1,2]. GoogLeNet mimarisi, geri yayılım için çoklu kaynaklar ile birleştirilen, çoklu sınıflandırma yapıları kullanan, yeni çok-ölçekli bir yaklaşım kullanır. Bu mimari sayesinde geri yayılım uygulamasının giriş katmanına ulaşmadan önce bozulması sonucu oluşan birçok problemin de önüne geçilmesi sağlanır. Lin ve arkadaşları, boyut azaltan ilave katmanlar sayesinde GoogLeNet'in genişlik ve derinlikte önemli bir kayba uğramadan karmaşık ağ içinde ağ mimarisi ile fark yaratan bir yenilik getirmişlerdir [3]. Ağ içindeki küçük mikro ağlar mimarinin daha karmaşık kararlar verebilmesine imkân sağlamaktadır. AlexNet [4] gibi daha eski evrişimli sinir ağlarının da dikkat çekici başarımları vardır. AlexNet mimarisi temel olarak maksimum havuzlama katmanları ve doğrultulmuş lineer birimlerin yığın olarak birbirini takip ettiği, bu yığınların üstünde birkaç adet tam bağlı katmanların olduğu yapılardan oluşan evrişimli mimarilerdir. AlexNet, ILSVRC-2012 yarışmasında %15,3'lük en iyi 5 hata oranıyla (top 5 error rate) birinci olmuştur.

Yüzden duygu tanıma problemlerine çözüm üretmek adına AU-Aware adı verilen yeni bir derin sinir ağı mimarisi geliştirilmiştir [5]. AU-Aware mimarisinde katman yığınının en altı yüzün bütün bir sunumunu oluşturmak için kullanılan evrişimli ağlar ve maksimum havuzlama katmanları içermektedir. Kahou ve arkadaşları [6]'daki çalışmada video analizindeki yüze dayalı duygu tanıma problemini çözmek için birden fazla derin sinir ağı mimarisini birleştirmişlerdir. Bu ağ mimarileri şunları içermektedir: Videonun her bir karesine ayrı ayrı odaklanan AlexNet mimarisine benzer bir evrişimli sinir ağı, ses verisi ile eğitilmiş derin inanç (deep belief) ağı, insan davranışlarının zaman-mekansal özelliklerini modellemek için otokodlayıcılar ve kişinin ağzına odaklanan yüzeysel bir ağdır. Evrişimli sinir ağı Toronto Face [7] adında özel bir veritabanı ile eğitilmiş ve ince ayarı ise AFEW veritabanı [8] üzerinden yapılarak, bağımsız verilerden oluşan AFEW veritabanı üzerinde verimliliği ölçüldüğünde %35,58'lik başarıma ulaştığı görülmüştür. 2013'deki EmotiW en yüksek başarım olan test setindeki %41,03 doğruluk oranı, beş mimarinin tekli bağımsız değişken ile birleştirilmesi sonucu elde edilmiştir. 2014 yılındaki yarışmada ise Riemann manifoldu üzerinde çoklu kernel yöntemi kullanarak test kümesi üzerinde %50,40 başarıma ulaşan çalışma [9] birinci olmuştur. Yapısal uzamsal kısıtlar dahilinde belirli yüz hareketlerini tespit edebilen bölüm tabanlı ayırt edici temsilini elde etmeye yarayan, şekil değiştirebilen hareket bölümlerine sahip ESA yöntemine ait kısıtlar ise [10]'da anlatılmıştır. İki adet poz verilerek oluşturulan CK+ ile MMI veri kümeleri ve spontane pozlar ile oluşturulmuş FERA veritabanı üzerindeki başarım sonuçları video tabanlı yüzden duygu tanıma alanında en iyi sonuçlardan biri olmuştur. Jung ve arkadaşları ise ESA'nın iki farklı tipini kullanarak, ilk adımda görüntü yığınlarından zamansal görünüş özelliklerini çıkartıp sonraki aşamada ise yüzdeki zamansal nirengi noktalarından geometrik öznitelik çıkartımı yapmışlardır [11]. Bu iki model birleştirilerek yeni bir metot geliştirilmiş ve yüzden duygu tanıma problemlerindeki başarım performansı arttırılmıştır.

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

Zhao ve arkadaşlarının sunduğu yöntem derin alan ve çok-etiketli öğrenme (deep region and multi-label learning) adında birleşik bir derin ağdır [12]. DRML önemli yüz alanlarını uyarmak için ileri beslemeli fonksiyon kullanan ve yüzün yapısal bilgilerini kaydetmek için eğitilmiş ağırlıkları zorlayan bir alan katmanıdır. Hasani ve Mahoor'un çalışmasında ise 3D Inception-ResNet mimarisi, video sekansları arasındaki farklı karelerden yüz görüntülerinin mekansal ilişkileri ve zamansal ilişkilerini beraber çıkartan bir LSTM birimi tarafından takip edilmiştir [13]. Ayrıca, yüzden ifade belirtmeye önemli bir katkısı olmayacak yüz bölgeleri yerine yüz bileşenlerinin önemini vurgulayacak, yüz nirengi noktaları da bu ağın girişleri olarak kullanılmıştır. Li ve Deng derin öğrenme yöntemlerinin duygu analizindeki başarımlarının kapsamlı bir şekilde incelemişlerdir [14]. Bunun dışında derin öğrenme yöntemleri kullanılarak duygu sınıflandırma problemine çeşitli çalışmalarda çözüm aranmıştır [15,16,17].

Sunulan çalışmada ise faklı yaş, ırk ve cinsiyetteki bireylerin, ağırlıklı olarak spontane pozlarından oluşan görüntüleri ile 7 farklı duygu durumu için bir veri kümesi meydana getirilmiştir. Bu veri kümesi, çeşitli anahtar kelimelerin internette aratılması ile bulunan genel erişime açık görsellerin özenle seçilmesi sonucunda oluşturulmuştur ve "RidNet" olarak adlandırılmıştır. Devamında ise bu yeni veri kümesi ile literatürün en çok kabul görmüş ESA mimarilerinden olan AlexNet, GoogLeNet ve ResNet101 ağları eğitilerek transfer öğrenmesi gerçekleştirilmiş. Elde edilen ağlar Compound Facial Expressions of Emotion (CE) ve Static Facial Expressions in the Wild (SFEW) veri kümeleri ile test edilerek başarımları ölçülmüştür. Bu sonuçlar incelenerek yüzden duygu tanıma uygulamaları için en başarılı ağ seçilmiştir. RidNet veri kümesinin başarıma katkısını ölçebilmek adına seçilen ağ AffecNet ve The Karolinska Directed Emotional Faces (KDEF) veri kümeleri ile de eğitilmiş sonra elde edilen ağlar CE ve SFEW veri kümelerinden oluşturulmuş test veri kümesi ile değerlendirilmiştir.

Çalışmanın katkıları şu şekilde özetlenebilir:

- İnternette erişime açık, farklı ortamlarda çekilmiş, yedi duyguyu içeren görüntülerden geniş bir veri kümesi oluşturulmuştur. Diğer veri kümelerinden farklı olarak, deneklerin yaş, cinsiyet ve ırkları bakımından oldukça zengindir.

- Bu veri kümesi ile eğitilen ESA'nın kontrolsüz ortamda elde edilen ve farklı duyguları içeren veri kümelerindeki sınıflandırma başarımları, literatürdeki eğitim için kullanılan veri kümelerine göre daha iyidir.

Bu çalışmanın 2. kısmında, kullanılan veritabanları ve derin öğrenme mimarileri hakkında kısa bilgiler verilmiş, ayrıca oluşturulan RidNet veritabanı anlatılmıştır. 3. kısmında yapılan deneyler hakkında ayrıntılı bilgiler verilmiştir. Son kısımda ise sonuçlar yorumlanmıştır.

## II. MATERYAL VE METOTLAR

Yapılan çalışma için sırasıyla şu adımlar izlendi: İnternet üzerindeki, herkesin kullanımına açık görüntülerden yedi farklı yüz ifadesi için bir veri kümesi oluşturuldu. Elde edilen bu veri kümesi "RidNet" olarak isimlendirildi. Veri kümesindeki görüntülerdeki yüzleri tespit etmek için MATLAB ortamında Viola-Jones algoritması [18] uygulandı. Bu uygulamadan sonra veri kümesindeki görüntüler tekrar gözden geçirilerek, yüzün doğru tespit edilemediği görüntüler ayıklandı ve veri kümesinden çıkartıldı. Bu sayede görüntülerdeki yüz kısımları belirlenerek, 224×224 boyutlarında kare şeklinde kırpıldı. Sadece yüz kısmından oluşan 224×224 boyutlarındaki bu görüntülere bazı görüntü işleme yöntemleri uygulanarak orijinal görüntülerin daha parlak ve keskin, daha karanlık ve keskin kopyaları elde edildi. Böylece Viola-Jones algoritması uygulanıp hatalı tespit edilen yüzler çıkartıldıktan sonra veri kümesinde kalan görüntü sayısı da 3 katına çıkartılmış oldu.

Veri kümesi, eğitim ve doğrulama veri kümeleri olmak üzere %80'e %20 oranında ikiye bölündü. Ayrıca eğitim ve doğrulama veri kümesindeki görüntülerden tamamen bağımsız olması adına CE veri kümesinden her farklı yüz ifadesi için 100'er adet görüntü seçilerek 700 görüntüden oluşan bir test veri kümesi oluşturuldu. Daha sonra elde edilen veri kümesi ile sırasıyla "AlexNet", "GoogLeNet" ve "ResNet-101" ağları üzerinden transfer öğrenmesi yapıldı. Yapılan deneysel çalışmalar sonucu en yüksek başarım oranına ulaşan ağın "ResNet-101" olduğu belirlendi. Sonraki adımda ise RidNet veri kümesi yüzden duygu tanıma uygulamalarında sıkça başvurulan, literatürde kabul görmüş AffectNet ve KDEF veri kümeleri ile kıyaslandı. Bunun için ResNet-101 ağı RidNet, AffectNet ve KDEF veri kümeleri ile ayrı ayrı eğitime tabi tutulup CE ve SFEW veri kümeleri kullanılarak başarımları test edildi.

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

### *A. Compound Facial Expressions of Emotion ve Static Facial Expressions in the Wild Veri Kümeleri*

Compound Facial Expressions of Emotion (CE) veri kümesi 21 farklı duygu sınıfından oluşmaktadır. Bu 21 farklı duygu sınıfı için 230 farklı katılımcıdan örnek görüntüler alınmıştır. 7 temel duygu sınıfı yanısıra aynı anda iki duygunun da bulunduğu (örneğin mutlu-şaşkın duygularını göstermeye yarayan kasların birlikte kasılması sonucu oluşan ifade) örnek görüntüler de oluşturulmuştur [19]. CE veri kümesi ile oluşturulan test veri kümesine ait örnek görüntüler ise Şekil 1'de verilmiştir.

**Şekil 1.** CE test veri kümesinden örnek imgeler

Static Facial Expressions in the Wild (SFEW) veri kümesi sıradan bir veri kümesi değildir. Buradaki "in the Wild" koşulların kontrollü olmaması manasına gelmektedir. Duygu tanıma üzerine yapılan çalışmalarda çoğunlukla kontrollü bir ortamda (laboratuar veya benzeri gibi) elde edilmiş veri kümeleri kullanılmıştır. Kontrolsüz koşullarda iyi performans elde edebilmek için, kontrolsüz ortamlardan elde edilmiş etiketli verilere ihtiyaç vardır [20]. SFEW veri kümesi filmlerdeki video karelerinden seçilerek oluşturulduğu için çok farklı koşullara sahip örneklerden oluşmaktadır. Şekil 2'de SFEW test veri kümesine ait örnek görüntülerden bazıları verilmiştir.

**Şekil 2.** SFEW test veri kümesine ait bazı imgeler

### *B. RidNet Veri Kümesi*

RidNet veri kümesini literatürdeki çoğu veri kümesinin aksine kontrollü ortamda poz verilerek değil çok değişken ortamdaki, farklı yaş, renk ve cinsiyetteki bireylerin çoğunlukla spontane pozlarından oluşmaktadır. Hatta görüntülerin bazılarında modelin kamera ile doğrudan göz teması yoktur. Bu durum eğitilen ağın kontrollü ortamda oluşturulmuş veri kümesi ile eğitilen ağlara göre gerçek hayattan alınan örneklerle yapılan testlerde daha başarılı olmasını sağlamaktadır. RidNet veri kümesindeki görseller internette herkesin erişimine açık görüntülerden belirli anahtar kelimeler ile yapılan aramalar sonucuna göre seçilmiştir [21]. Şekil 3'te eğitim veri kümesindeki farklı yüz ifadelerine ait örnek görüntüler verilmiştir. Bu görüntülerin mutlu, üzgün, sinirli, şarkın, korkmuş, iğrenmiş ve yalın yüz ifadelerine göre alt gruplara ayrılması ile sınıflar oluşturulmuştur. Veri kümesinde her bir sınıf için küçük kenarı en az 240 piksel olacak şekilde görüntüler seçilmiştir. Yüz ifadelerine ait görüntü sayıları Tablo 1'de gösterilmiştir.

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

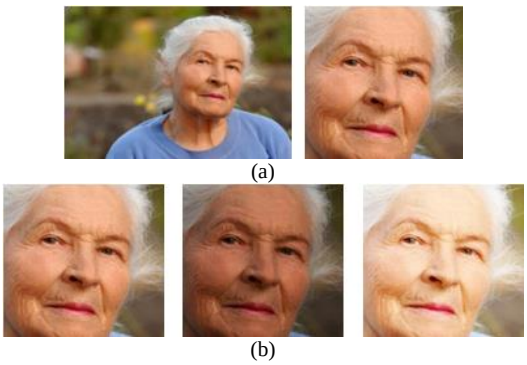
*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

**Şekil 3.** RidNet veri kümesine ait örnek imgeler

**Tablo 1.** Yüz ifadelerine ait görüntü sayıları

| **İfade** | Sinirli | İğrenme | Korkmuş | Mutlu | Yalın | Üzgün | Şaşkın | Toplam |
|---|---|---|---|---|---|---|---|---|
| **Sayısı** | 591 | 712 | 591 | 586 | 575 | 561 | 600 | 4216 |

Veri kümesi RGB formatındaki toplam 4216 görüntüden meydana gelmektedir. İmgelere Viola-Jones algoritması uygulandıktan sonra hatalı bulunun yüzlerin bulunduğu veya hiç yüzün bulunamadığı görüntüler kontrol edilip elenmiştir. Görüntü ön-işleme işlemleri görüntünün tamamına uygulanmak yerine, Viola-Jones algoritması [18] ile yüz kısmı belirlendikten sonra sadece bu bölgeye uygulanmıştır. Viola-Jones algoritması görüntüdeki yüzü belirlemek için dikdörtgen özellikleri kullanmakta, bu sayede oldukça hızlı çalışmasının yanı sıra yüksek oranda doğru sonuçlar da vermektedir [18]. RidNet veri kümesindeki görüntü sayısını arttırmak için görüntü ön-işleme uygulamaları ile daha parlak ve keskin kopyaları oluşturularak veri kümesinin genişletilmesi sağlanmıştır. Böylece toplamda 10989 görüntüye sahip sadece kırpılmış yüzlerden oluşan bir veri kümesi elde edilmiştir. Daha sonra bu veri kümesi ağın eğitimi sırasında %80 - %20 oranında, eğitim ve doğrulama veri kümesi olarak ayrılmıştır. Yani, görüntülerin 8791 tanesi eğitim, 2198 tanesi ise doğrulama kümelerinde kullanılmıştır. Bu işlemin ardından her bir görüntü 224×224 olacak şekilde yeniden boyutlandırılmıştır. Şekil 4-(a)'da örnek bir görüntü ve Viola-Jones algoritması ile bu görüntüde bulunan yüz bölgesinin kırpılmış hali gösterilmiştir. Şekil 4-(b)'de ise bulunan yüz bölgesinin orijinal hali ve filtrelerin uygulanması ile elde edilen varyasyonları gösterilmektedir.



(a)

(b)

**Şekil 4.** Yüz görüntüsünden (a) yüz bölgesinin tespiti, (b) yüz bölgesinin çeşitli filtreler kullanılarak çoğaltılması

### *C. Evrişimli Sinir Ağları*

ESA bilgisayarlı görü uygulamalarında kullanılmak üzere geliştirişmiş çok katmanlı yapay sinir ağlarının özel bir modelidir. Yapısında evrişim (convolution), ortaklama (pooling) ve tam bağlı (fully connected) gibi, kendine özgü görevleri olan ayrı katmanları barındırır. Bunlar birbirini takip edecek şekilde dizilerek ESA oluşturulur. Bu yapının ilk kısımlarında öznitelik çıkartım işlemleri gerçekleştirilirken sınıflandırma işlemi ise son katmanlarda gerçekleşir [22].

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

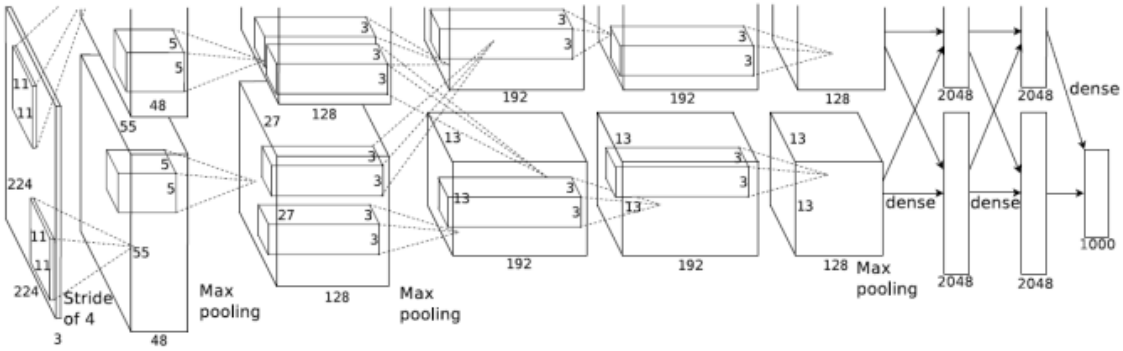
BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

1) *AlexNet*: 2012 yılında düzenlenen ImageNet yarışmasını, o zaman kadarki en düşük hata oranı olan %26,2'yi %15,4'e indiren AlexNet kazanmıştır. AlexNet üzerinde, 60 milyon parametre ve 650 bin nöron bulunmaktadır. Ağ yapısı, bazılarının sonunda maksimum ortaklama (max-pooling) bulunan beş evrişimli katman ve üç tam bağlı katman ile bunların sonunda bulunan 1000 yollu bir eşiksiz en büyük işlev (softmax) fonksiyonundan oluşmaktadır [4]. AlexNet mimarisi Şekil 5'te verilmiştir.

2) *GoogLeNet:* 2014 yılında düzenlenen "The ImageNet Large Scale Visual Recognition Challenge" (ILSVRC) yarışmasında %6,66 ile en düşük hata oranını sağlayarak birinci olmuştur. Ağın çıktılarının tekrar daha önceki ağa giriş olarak uygulanması ile ağ içinde ağ yapısı elde edilir. Bu yapı sebebi ile daha kompleks bir mimari oluşmuştur. 1×1'lik filtreler sayesinde parametre sayısı 10 kat azalmıştır. GoogLeNet 22 adet katmandan oluşmaktadır [23].

*3) ResNet-101:* ResNet 2015 yılında düzenlenen ILSVRC yarışmasında %3,37 ile en düşük hata oranını sağlayarak birinci olmuştur. ResNet, 101 adet katmandan oluşmaktadır. ResNet'te katmanlar arasında yapılan atlama işlemine ResBlock ismi verilir. ResBlock sayesinde bir önceki katmanda bir şey öğrenilmese bile eski katmandaki bilgiyi yeni katmana uygulanarak model daha güçlü hale getirilir. Böylece gradyan silinmesi problemi de ResBlock ile çözülmüş olur. Optimizasyon algoritması olarak eğim düşümü kullanılır [24].




**Şekil 5.** AlexNet CCN mimarisinin temsili bir gösterimi [4].

4) *Transfer Öğrenmesi:* Büyük bir veri kümesi modelinden öğrenilmiş ağırlıkların alındığı ve bunların çeşitli sabit katmanlara uygulandığı, kalan katmanların tekrar eğitildiği veya ağın ince ayarının (fine tuning) yapıldığı, yaygın olarak kullanılan bir derin öğrenme tekniğidir [25].

ESA yapısı oluştururken eğitim için bir veri kümesine ihtiyaç duyulur ve bu veri kümesi yeteri kadar büyük değilse oluşturulan ağ teste tabi tutulduğunda düşük bir doğruluk oranı verir. Bu durumda eğitim veri kümesinin büyütülmesi gerekir. Fakat bazı durumlarda bu mümkün değildir ve sınırlı sayıdaki etiketli eğitim verisi ile çalışmak gerekebilir. Bu koşullarda en iyi doğruluk derecesine sahip sonuçların transfer öğrenmesi ile elde edildiği görülmüştür. Yani transfer öğrenmesi, küçük veri kümeleri ile yüksek doğruluk derecesine sahip sonuçlar alınmasını sağlayabilir.

## III. DENEYSEL ÇALIŞMALAR

Yapılan çalışmaların ilk kısmında, RidNet veri kümesi ile AlexNet, GoogLeNet ve ResNet-101 mimarileri üzerinden transfer öğrenmesi gerçekleştirilmiştir. Daha sonra elde edilen ağların performansı CE veri kümesi ile ölçülmüş nihayetinde de en başarılı olan ağ bulunmuştur. İkinci kısımda ise yapılan deneysel çalışmalar sonucu saptanan en başarılı ağ RidNet, AffectNet, KDEF veri kümeleri ile eğitilmiş ve elde edilen ağ yine bağımsız test veri kümelerinde sınanarak ortaya çıkan sonuçlara dayalı bazı analizlerde bulunulmuştur.

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

**Tablo 2.** AlexNet, GoogLeNet, ResNet-101 ağlarının eğitim sonuçları

| ESA | Küçük Yığın | Döngü | Öğrenme Oranı | Doğrulama Frekansı | Eğitim Zamanı(dk) | Doğrulama Başarımı(%) | Test Başarımı(%) |
|---|---|---|---|---|---|---|---|
| AlexNet | 10 | 6 | 0,001 | 30 | 106 | 82,43 | 57,14 |
| AlexNet | 100 | 10 | 0,001 | 3 | 176 | 90,44 | 61,29 |
| AlexNet | 10 | 6 | 0,001 | 3 | 606 | 88,76 | 49,29 |
| GoogLeNet | 100 | 10 | 0,001 | 3 | 480 | 73,83 | 44,14 |
| GoogLeNet | 10 | 6 | 0,003 | 30 | 162 | 71,37 | 43,43 |
| GoogLeNet | 64 | 6 | 0,001 | 3 | 273 | 70,14 | 42,71 |
| ResNet101 | 16 | 6 | 0,001 | 30 | 993 | 92,49 | 60,0 |
| ResNet101 | 10 | 6 | 0,003 | 3 | 8342 | 94,9 | 63,57 |

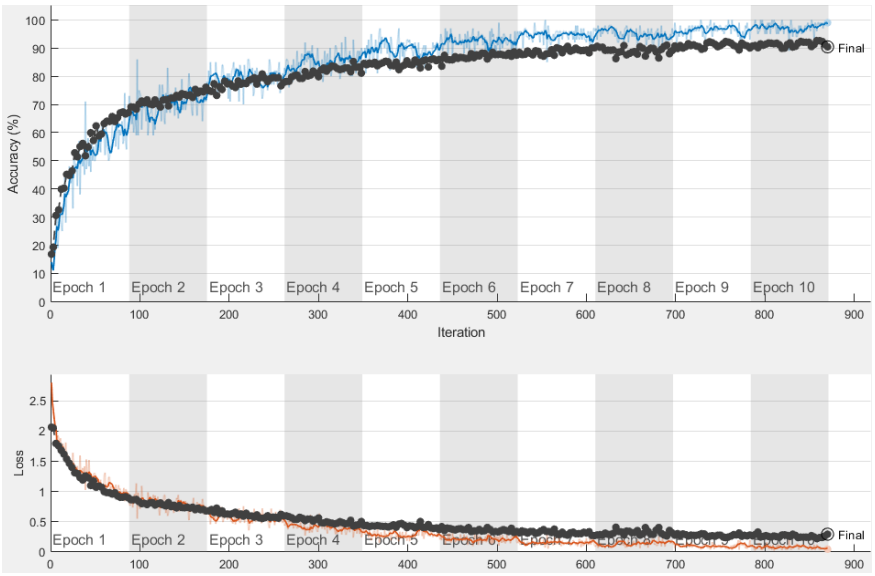
### *A. RidNet Veri Kümesi ile Eğitilen Evrişimli Sinir Ağlarının Kıyaslanması*

RidNet veri kümesi %80 eğitim ve %20 doğrulama veri kümesi olarak ikiye bölünmüş ve AlexNet, GoogLeNet, ResNet-101 mimarileri üzerinden transfer öğrenmesi gerçekleştirilmiştir. Her ağın eğitimi için kullanılan parametre değerleri ve elde edilen başarım sonuçları Tablo 2'de sunulmuştur. Deneysel çalışmaların bu aşamasında toplamda 8 farklı eğitim yapılmıştır. Yapılan ilk 3 eğitimde AlexNet mimarisi kullanılırken ilk eğitimde 10 olan küçük yığın (mini batch) ölçüsü ikinci eğitimde 100'e, 6 olan döngü (epoch) sayısı 10'a çıkartılırken doğrulama sıklığı (validation frequency) da 30'dan 3'e çıkarılmıştır. Daha sonraki eğitimde ise ilk eğitimdeki parametrelerin hepsi aynı kalırken sadece doğrulama frekansı arttırılmıştır. AlexNet mimarisi üzerinden yapılan transfer öğrenmesinde en yüksek test doğruluğuna %61,29 ile 2. eğitimdeki parametreler kullanılarak ulaşıldığı görülmüştür. Sonraki adımda ise yapılan eğitimlerde GoogLeNet mimarisi kullanılmış küçük yığın ölçüsü, döngü sayısı ve öğrenme oranı gibi parametreler değiştirilmesine rağmen test doğruluk yüzdelerinde belirgin farklar ortaya çıkmamıştır. Son iki eğitimde ise ResNet-101 mimarisi kullanılmış eğitim parametrelerinde yapılan değişiklikler ile test doğruluk oranı %60'tan %63,57'ye çıkartılmış. Yapılan bütün eğitimlerde fonksiyon olarak moment ile rastgele gradyan inişi (Stochastic Gradient Descent with Momentum, SGDM) kullanılmıştır. Gerçekleştirilen eğitimler toplamda 11138 dakika yani yaklaşık olarak 185 saat sürmüştür. Küçük yığın ölçüsüne bağlı olarak değişen iterasyon sayısı ve doğrulama sıklığı eğitim süresine etki eden en büyük etkenler iken ağın sahip olduğu katman sayısı da bu süreye etki etmektedir. Deneyler Intel Core i7-4510U CPU @ 2.00 GHz işlemci, 8 GB RAM ve Nvidia Geforce 840M ekran kartına sahip bir dizüstü bilgisayarda gerçekleştirilmiştir. Yapılan deneysel çalışmalar sonucunda en yüksek başarıma sahip ağın %94,9 eğitim ve %63,57 test ile ResNet-101 olduğu görülmüştür. Diğer ağlar ile yapılan çalışmaları incelendiğinde, AlexNet, %90,44 doğruluk oranına ulaşırken test veri kümesindeki başarımının %61, GoogLeNet ise, %73,83 doğruluk oranına ulaşırken test veri kümesindeki başarımının ise %44 olduğu görülmüştür.

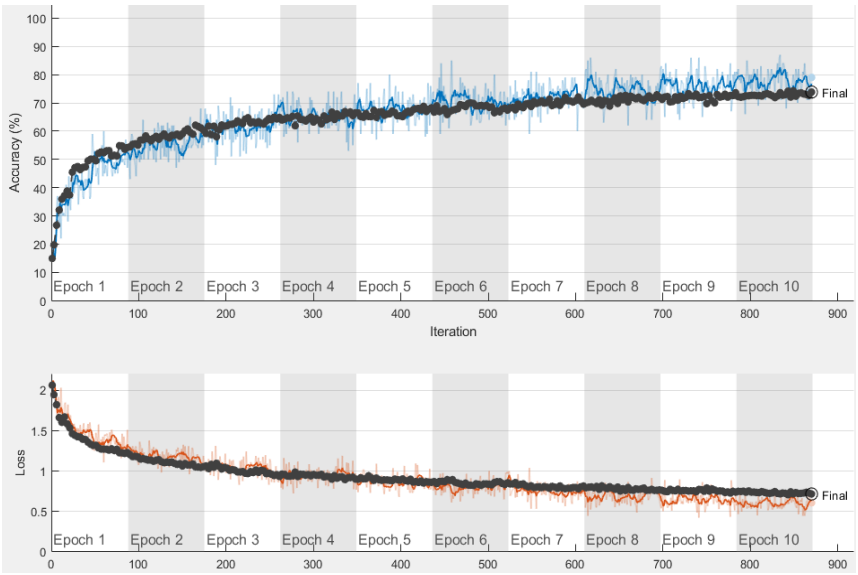
BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*




**Şekil 6.** AlexNet ağının RidNet ile eğitimi

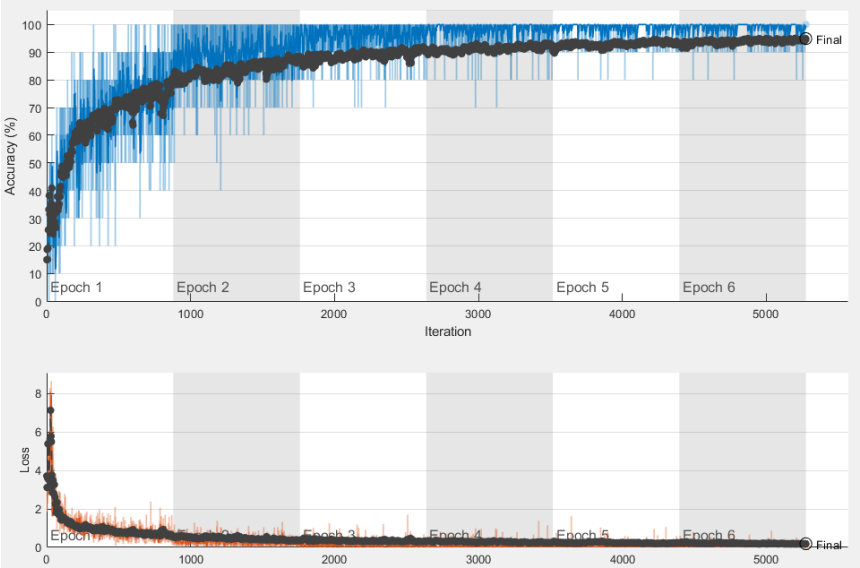



**Şekil 7.** GoogLeNet ağının RidNet ile eğitimi

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

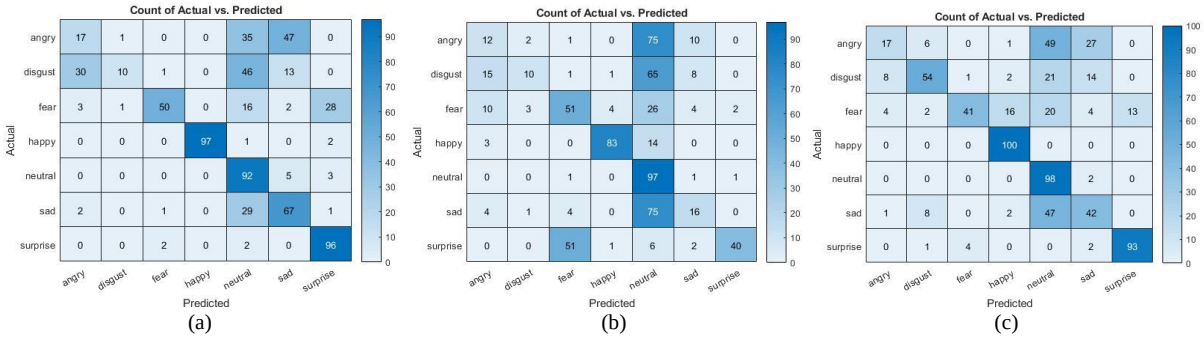
BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*




**Şekil 8.** ResNet-101 ağının RidNet ile eğitimi

Şekil 6 incelendiğinde AlexNet ağının eğitimi sırasında her iterasyonda elde edilen başarım ve kayıp oranlarının grafiği görülmektedir. Grafiğe detaylı olarak baktığımızda ilk döngüde ağın hızlı bir şekilde öğrenme işleminin gerçekleştiği daha sonra ise 6. döngüye kadar yavaşta olsa öğrenme işleminin devam ettiği, kalan diğer döngüler boyunca ise belirgin bir artışın olmadığı görülmektedir. Doğrulama veri kümesinin belirli frekanslar ile karıştırılması sonucu grafikte inişler ve çıkışlar olduğu görülmektedir. Bu da ağın öğrenmeye devam ettiğini göstermektedir. Şekil 7'de GoogLeNet ağının eğitimi sırasında her iterasyonda elde edilen başarım ve kayıp oranlarının grafiği görülmektedir. Grafikten AlexNet'e oranla daha yavaş bir öğrenme sürecinin gerçekleştiği 5. döngüden sonra da belirgin bir artışın olmadığı görülmektedir. Şekil 8'de ResNet-101 ağının RidNet ile eğitimine ait başarım ve kayıp oranlarının grafikleri verilmiştir. Grafikten en başarılı eğitimin ResNet-101 ağı ile gerçekleştirildiği görülmektedir. İlk döngü sonunda %80 oranında bir başarım yakalandığı 3. döngüden sonra ise de başarım oranında belirgin bir artış olmadığı görülmektedir.




**Şekil 9.** Test veri kümesinin hata matrisleri; a) AlexNet, b) GoogLeNet, c) ResNet-101

Şekil 9-(a)'da AlexNet ile test veri kümesinde yapılan sınıflandırma işlemi sonucunda elde edilen hata matrisi görülmektedir. Burada en başarılı tahminlerin mutlu ifadesi için yapıldığı görülürken en başarısız durum ise sinirli ifadesinde elde edilmiştir. En çok tahminde bulunulan ifadenin ise yalın olduğu ve iğrenme ve sinirli ifadelerinin yalın olarak yanlış sınıflandırıldığı görülmektedir. Eğitim veri kümesindeki görüntülere bakıldığında bu durumun sebebinin, yalın ifadesi ile sinirli ifadesinin kaşların çatılması dışında benzer olması olduğu söylenebilir. Benzer şekilde iğrenme durumunda da sadece dudakların aldığı şekil değişmektedir. Şekil 9-(b)'de ise GoogLeNet ile test veri kümesinde yapılan sınıflandırma işlemi sonucunda elde edilen hata matrisi görülmektedir. Burada en başarılı tahminlerin yalın ifadesi için yapıldığı görülürken en başarısız durum ise

BŞEÜ Fen Bilimleri Dergisi
6 (2), 384-396, 2019

BSEU Journal of Science
DOI: 10.35193/bseufbd.645138

*e-ISSN: 2458-7575 (http://dergipark.gov.tr/bseufbd)*

iğrenme ifadesinde elde edilmiştir. Hemen devamında Şekil 9-(c)'de ise ResNet-101 ile test veri kümesinde yapılan sınıflandırma işlemi sonucunda elde edilen hata matrisi görülmektedir. Burada en başarılı tahminlerin %100 başarım ile mutlu ifadesi için yapıldığı görülürken ne başarısız durum ise %17 başarım ile sinirli ifadesinde elde edilmiştir.

### *B. RidNet ile Diğer Veri Kümelerinin Kıyaslanması*

Yapılan deneysel çalışmalarda en başarılı ağ olarak tespit edilen ResNet-101'e RidNet'ten sonra, AffectNet ve KDEF veri kümeleri ile de ayrı ayrı transferi öğrenmesi yapılmıştır. Daha sonra eğitilen yeni ağ CE ve SFEW veri kümeleri ile test edilmiştir. Elde edilen sonuçlar Tablo 3'de görülebilir.

**Tablo 3.** Veri kümelerinin kıyaslanması

| Veri Kümesi | Boyut | Küçük Yığın | Döngü | Öğr. Oranı | Doğ. Oranı | Eğitim Süresi | Doğruluk(%) | CE Test(%) | SFEW Test(%) |
|---|---|---|---|---|---|---|---|---|---|
| RidNet | 10089 | 16 | 6 | 0,001 | 30 | 993 | 92,49 | 60 | 34,13 |
| KDEF_FER | 7341 | 16 | 6 | 0,001 | 30 | 584 | 96,46 | 79,86 | 11,98 |
| AffectNet | 17500 | 16 | 6 | 0,001 | 30 | 2072 | 86,86 | 66,86 | 22,16 |

KDEF veri kümesi ile ResNet-101 ağı eğitildiğinde doğrulama veri kümesinde %96,46 doğruluk oranına CE veri kümesi ile oluşturulan test veri kümesi üzerinde yapılan tahminlerde ise %79,86, SFEW veri kümesi ile yapılan tahminlerde ise %11,98 başarıma ulaşılmıştır. AffecNet veri kümesi ile ResNet-101 ağı eğitildiğinde doğrulama veri kümesinde %86,86 doğruluk oranına CE veri kümesi ile oluşturulan ters veri kümesi üzerinde yapılan tahminlerde ise %66,86, SFEW veri kümeleri ile yapılan tahminlerde ise %22,16 başarıma ulaşılmıştır. RidNet veri kümesi ile eğitilen ResNet-101 ağı ise doğrulama veri kümesinde %92,49 ile doğru tahminde bulunurken, CE test veri kümesinde %60, SFEW test veri kümesinde ise %34,13 doğru tahminde bulunmuştur. Yapılan kıyaslamalar sonucu SFEW test veri kümesinde en başarılı ağın RidNet veri kümesi ile eğitilen ağ olduğu görülmektedir.

## IV. SONUÇLAR VE TARTIŞMA

Literatür incelendiğinde yüzden duygu tanıma çalışmalarında kullanılan veri kümelerinin büyük çoğunluğunun kontrollü ortamlarda oluşturulduğu görülmektedir. Kontrollü ortamda oluşturulan veri kümesi ile eğitilen ESA'ların gerçek hayattaki uygulamalarda test veri kümelerindeki kadar başarılı olamadığı görülmektedir. Bu durumda alışılagelmiş veri kümelerinden farklı olarak daha gerçekçi görüntü örneklerine sahip veri kümesi ihtiyacı doğmaktadır. RidNet adı verilen, internette belirli anahtar kelimelerin aratılması ile elde edilen görüntülerden oluşan veri kümesi ile bu ihtiyaç giderilmeye çalışılmıştır. Daha sonra RidNet ile AlexNet, GoogLeNet ve ResNet-101 mimarileri üzerinde transfer öğrenmesi gerçekleştirilmiştir. Eğitimlerin tamamlanmasının ardından bağımsız test veri kümesi üzerinde en iyi sonuçlara ResNet-101 ile ulaşılmıştır. Fakat sahip olduğu 101 katman sebebiyle ResNet için uygulanan eğitim sürelerinin oldukça yüksek olduğu görülmüştür. (Aynı parametreler ile AlexNet'ten 13 kat daha fazla.)

İkinci aşamada ise en başarılı ağ olarak belirlenen ResNet-101 ile eğitim veri setlerinin başarıma etkisi incelenmiştir. Stüdyo ortamında durağan görüntülerden oluşan CE veri kümesinin test için kullanıldığı ilk aşamada, KDEF veri kümesi ile eğitilen Resnet-101 ağının en başarılı sonuçları aldığı görünürken, en iyi ikinci sonuç ise AffecNet ile eğitilen ağda elde edilmiştir. CE test veri kümesi için sonuçlara baktığımızda KDEF, RidNet'ten %33 daha başarılı sonuçlar verirken, AffectNet de %11 daha başarılı sonuçlar vermiştir. Bunun sebebi KDEF eğitim veri kümesinde bulunan görüntülerin de stüdyo koşullarında çekilen durağan görüntüler olması şeklinde yorumlanabilir. AffectNet eğitim veri kümesinde ise RidNet'te olduğundan %73 daha fazla görüntü vardır. Eğitim veri kümesinde ne kadar fazla görüntü olursa ağın daha başarılı sonuçlar verme ihtimali de o kadar artar. AffectNet'in daha başarılı sonuçlar vermesi de böyle yorumlanabilir.

SFEW veri kümesi ile yapılan test sonuçlarına bakıldığında ise RidNet ile eğitilen ağın en başarılı sonuçları verdiği görülmektedir. Bu sonuçlar KDEF'e göre %184, AffectNet'e göre ise %54 daha iyidir. SFEW

veri kümesi görüntüleri filmlerden alınan akan görüntüler olduğu için sınıflandırılması daha zordur. Stüdyo ortamındaki görüntülerden oluşan veri kümeleri ile eğitilen ağların performansının düşük olmasının sebebi budur. Ayrıca kendisinden %73 daha büyük bir veri kümesi olan AffectNet'e göre RidNet'in %54 daha iyi doğruluk oranına ulaşması veri kümesindeki örneklerin nitelikli, çeşitliliği zengin olan görüntülerden oluştuğunu göstermektedir. Literatüre baktığımızda SFEW veri kümesi ile yapılan yarışmalarda birincilerin %35,9 civarında doğruluk oranına ulaştığı düşünülürse RidNet'in yapılacak düzenleme ve genişletme işlemleri ile %34 başarım oranını bu değerlerin üstüne çıkartma ihtimali oldukça yüksektir.

## TEŞEKKÜR

Bu çalışma Bilecik Şeyh Edebali Üniversitesi, 2019-01.BŞEÜ.03-05 nolu bilimsel araştırma projesi tarafından kısmi olarak desteklenmektedir. Yazarlar, Prof. Dr. Atalay Barkana'ya yapıcı yorum ve önerilerinden dolayı teşekkür etmektedir.

## KAYNAKLAR

[1] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). *Going deeper with convolutions*. arXiv preprint arXiv:1409.4842.

[2] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., & Li, F.-F. (2014). *Imagenet large scale visual recognition challenge*. arXiv preprint arXiv:1409.0575.

[3] Lin, M., Chen, Q., & Yan, S. (2013). *Network in network*. arXiv preprint arXiv:1312.4400.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). *Imagenet classification with deep convolutional neural networks*. 25th International Conference on Neural Information Processing Systems - Volume 1, 1097–1105.

[5] Liu, M., Li, S., Shan, S., & Chen, X. (2013) *Au-aware deep networks for facial expression recognition*. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 1–6.

[6] Kahou, S.E., Pal, C., Bouthillier, X., Froumenty, P., Gulcehre, C., Memisevic, R., Vincent, P., Courville, A., Bengio, Y., Ferrari, R.C., et al. (2013). *Combining modality specific deep neural networks for emotion recognition in video*. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction, 543–550.

[7] Susskind, J.M., Anderson, A.K., & Hinton, G.E. (2010). *The toronto face database*. Technical report, UTML TR 2010-001, University of Toronto.

[8] Dhall, A., Goecke, R., Joshi, J., Wagner, M., & Gedeon, T. (2013). *Emotion recognition in the wild challenge 2013*. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction, 509–516.

[9] Liu, M., Wang, R., Li, S., Shan, S., Huang, Z., & Chen, X. (2014) Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild. In Proceedings of the 16th International Conference on Multimodal Interaction, 494–501.

[10] Liu, M., Li, S., Shan, S., Wang, R., & Chen, X. (2014). *Deeply learning deformable facial action parts model for dynamic expression analysis*. In Computer Vision–ACCV 2014, 143–157.

[11] Jung, H., Lee, S., Yim, J., Park, S., Kim, J. (2015). *Joint fine-tuning in deep neural networks for facial expression recognition*. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2983–2991.

[12] Zhao, K., Chu, W.S. & Zhang, H. (2016). *Deep region and multi-label learning for facial action unit detection*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 3391–3399.

[13] Hasani, B., Mahoor, M.H. (2017). *Facial expression recognition using enhanced deep 3D convolutional neural networks*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Hawaii, HI, USA, 1–11.

[14] Li, S. & Deng, W. (2018) Deep facial expression recognition: A survey. arXiv preprint arXiv:1804.08348.

[15] Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y. & Dobaie, A.M. (2018) *Facial expression recognition via learning deep sparse autoencoders*. Neurocomputing, 273, 643–649.

[16] Otberdout, N., Kacem, A., Daoudi, M., Ballihi, L. & Berretti, S. (2018) *Deep covariance descriptors for facial expression recognition*. 29th British Machine Vision Conferernce.

[17] Li, D., Li, Z., Luo, R., Deng, J. & Sun, S. (2019) *Multi-Pose Facial Expression Recognition Based on Generative Adversarial Network*. IEEE Access, 7, 143980-143989.

[18] Viola, P., & Jones, M.J. (2004). *Robust Real-Time Face Detection*. International Journal of Computer Vision, 57(2), 137-154.

[19] Du, S., Tao, Y., & Martinez, A.M. (2014). *Compound facial expressions of emotion*. Proceedings of the National Academy of Sciences, 111(15), E1454–E1462.

[20] Dhall, A., Murthy, O.R., Goecke, R., Joshi, J., & Gedeon, T. (2015). *Video and image based emotion recognition challenges in the wild: Emotiw 2015*. International Conference on Multimodal Interaction, 423–426.

[21] Adobe Stock, https://stock.adobe.com/, (Erişim tarihi: 08.05.2019).

[22] Aydilek, İ.B. (2017). *Approximate estimation of the nutritions of consumed food by deep learning*. International Conference on Computer Science and Engineering (UBMK), 160-164.

[23] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). *Going deeper with convolutions*. IEEE Conference on Computer Vision and Pattern Recognition, 1–9.

[24] He, H., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

[25] Savoiu, A., & Wong, J. (2017). *Recognizing Facial Expressions Using Deep Learning*.