

Üretken Çekişmeli Ağ Kullanarak Eski Görüntüleri Renklendirme ve İyileştirme

Arda Cem BİLECAN^{1*}, Simay HOŞMEYVE¹⁺ ve Bahadır KARASULU¹

¹Çanakkale Onsekiz Mart Üniversitesi, Bilgisayar Mühendisliği Bölümü, Çanakkale, Türkiye

*Sorumlu Yazar: ardacembilecan@gmail.com

+Sunucu: simayhosmeyve@gmail.com

Özet – Günümüzde derin öğrenme alanı için veri, çalışma sonuçlarının başarı oranı için büyük önem taşımaktadır. Bu ihtiyaca yönelik olarak görüntü işleme konusunda yapabileceğimiz iyileştirmeler ile daha kaliteli ve eksiksiz görüntüler elde etmek ve eski verilerde karşılaştığımız sorunlardan biri olabilen siyah beyaz görüntülerin renkli görüntüye çevrilmesi ve eksik piksellerin tamamlanması bu çalışmadaki amacı oluşturmaktadır. Çalışmada geliştirilen kullanıcı arayüzü sayesinde hem kullanıcıların kişisel görüntüleri düzenlemelerine hem diğer çalışmalarda destekleyici bir araç olarak kullanılmasına hizmet etmektedir. Yöntem olarak ön işleme, derin öğrenme modeli ile eğitim ve rötuş işlemleri sırasıyla kullanılmıştır. Mimari olarak kullandığımız Üretken Çekişmeli Ağ (GAN) çeşitlerinden bir tanesi olan Pix2pix modeli sayesinde kaynak görüntüden hedef görüntüye dönüşüm yapan bir yöntem geliştirilmiştir. Bu bakış açısıyla Pix2pix modeli görüntü iyileştirme ve dönüşümlerimizde ideal bir ağ olmuştur. Deneylerde elde edilen yüksek başarımlar sonucu %82 doğruluk oranıyla iç mekân ve dış mekân sınıflandırması için elde edilmiştir. Ayrıca iç mekân renklendirmesi, dış mekân renklendirmesi, insan yüz renklendirmesi, görüntü onarma ve görüntü kalitesini artırma için yapılan deneylerdeki ortalama değerler göz önüne alındığında en yüksek Yapısal Benzerlik Endeksi Ölçüsü (SSIM) değeri 0,9256 ve en düşük Tepe Sinyal Gürültü Oranı (PSNR) değeri ise 65,11 dB olmuştur. Çalışmamızda bilimsel bulgulara dayanan tartışma ve değerlendirmelere yer verilmektedir.

Anahtar Kelimeler – Derin Öğrenme, Üretken Çekişme Ağlar, Görüntü İyileştirme, Sınıflandırma, Kullanıcı Grafik Arayüzü

Colorization and Enhancement of Old Images Using Generative Adversarial Networks

Abstract– Nowadays, data for deep learning is crucial to the success rate of the study results. With the improvements we can make to image processing for this need, it is intended to achieve better quality and complete images and to translate black and white images into a color image, which may be one of the problems we face with older data, and to complete missing pixels. Thanks to the user interface developed in the study, it serves both as a user's ability to edit personal images and as a supporting tool for other studies. As a method, pre-processing, deep learning and training and retouching are used in order. One of the varieties of Generative Adversarial Network (GAN) architecture, the Pix2pix model has been developed as a way to transform from source image to target image. From this point of view, the Pix2pix model has been the ideal network for improvement and conversions of images. The highest performance result in experiments was achieved for indoor and outdoor classification with accuracy of 82%. In addition, the highest Structural Similarity Index Measure (SSIM) value was 0.9256 and the lowest Peak Signal to Noise Ratio (PSNR) value was 65.11 dB when the average values were taken into account with the experiments to improve indoor colorization, outdoor colorization, human face colorization, image repair and image quality. Our study includes discussions and assessments based on scientific findings.

Keywords – Deep Learning, Generative Adversarial Networks, Image Enhancement, Classification, Graphical User Interface

I. GİRİŞ

Günümüzde görüntü çözünürlüğünü iyileştirmek için eski görüntülerde piksel eksiklikleri ve siyah beyaz görüntülerin renkliye dönüştürülmesi oldukça popüler hale gelmiştir. Bu iyileştirmeler kişisel ihtiyaca yönelik kullanılabileceği gibi akademik bir çalışmada araç olarak da kullanılabilir. Görüntü üzerindeki çalışmalarda verinin kalitesi çalışmanın doğruluk oranına genellikle etki eden bir etkidir.

Görüntülerdeki piksel eksiklikleri makine öğrenmesi veya derin öğrenme gibi yöntemler sayesinde giderilerek yüksek doğrulukta sınıflandırma, görüntüden gürültü giderme ve görüntü iyileştirme sonuçları elde edilebilmektedir [1].

Görüntü kalitesi ise nesne tanıma, sınıflandırma gibi konularda daha iyi bir öğrenim için faydalı olmaktadır. Diğer bir problem ise çok eski görüntüler üzerinde onarıma muhtaç çeşitli bozukluklar olabileceği ve renk uzayının yeterli doygunlukta olmamasıdır. Görüntüler genellikle uzman insan

tarafından el ile renklendirilebilmektedir. Mevcut çalışmalarda ise spesifik bir alana yönelik dönüşümler yapılmaktadır. Bu çalışmada ise genel kullanımına uygun bir sonuç hedeflenmiştir. Donanımsal kısıtlamalar ve veri kümesi boyutları göz önüne alınarak mümkün olan en iyi sonuçlara ulaşılmaya çalışılmıştır. Daha fazla veri ve daha iyi donanım kaynakları ile sonuçları iyileştirmek mümkündür. Çalışmada kullanılan Üretken Çekişmeli Ağ (Generative Adversarial Network-GAN) [1] yapısı günümüzde sentetik ama gerçekçi görünen görüntü oluşturan üretici ağ modellerinden en popüler olanıdır.

GAN'lar, sentetik bir veri oluşturmanın yanı sıra mevcut veriyi istediğimiz farklı bir hedef görüntüye benzetmek için de kullanılabilirler. Bu amaçla çalışmamızda kullanılan mimarinin temelini koşullu GAN (Conditional GAN-cGAN) [2] mimarisi türü olan Pix2pix modeli oluşturmaktadır. Çalışmada kullanılan veri kümeleri üç adet alınmış, bunlardan alınan verilerle toplamda beş adet alt küme oluşturulmuştur. Bu veri kümeleri; iç mekân, dış mekân, insan yüzü, görüntü onarımı için kullanılan görüntüler, düşük ve yüksek çözünürlüklü görüntüleri içermektedir.

Materyal ve yöntemler bölümünde veri kümesinin detaylarından ve örneklerinden, GAN mimarisinin teknik detaylarından, eklenen ön işlem (preprocessing) ve son işlem (post processing), kullanıcı arayüzü sayesinde bu yapının dinamik olarak kullanılabilmesinden bahsedilmektedir. Deneysel sonuçlar bölümünde çalışılan konular üzerine başarımlar ölçümü ve değerlendirmeleri, seçilen parametreler ve çalışmanın katkılarına değinilmektedir. Son olarak sonuçlar bölümünde çalışmanın başarısı sayısal veriler üzerinden nicel olarak yorumlanmakta ve bilimsel bulgulara dayanan tartışmaya yer verilmektedir.

II. MATERYAL VE YÖNTEMLER

Çalışmamızın bu bölümde çalışmamızda kullanılan üç adet veri kümesinden veriler alınarak beş ayrı alt küme oluşturulmuştur. Bu bölümde bu veri alt kümelerinden, çalışmaya temel oluşturan altyapıdan, bu altyapıda kullanılan cGAN mimarisindeki Pix2pix modeli, AlexNet ve YOLOv3 (You Only Look Once) modellerinin teknik detaylarından bahsedilmektedir.

A. Veri Kümeleri

Bu çalışmadaki tüm modeller için (256,256,3) piksel boyutunda üç renk kanalına sahip girdiler kullanılmıştır. İnsan yüz görüntülerini *Massachusetts Amherst Üniversitesi*'nin hazırladığı "Labeled Faces in the Wild" adlı görüntü veritabanından seçilerek kullanılmıştır [3], [4].

İlgili veri kümesinde 5749 farklı insan için toplamda 13233 adet insan yüzü görüntüsü bulunmaktadır. Eğitim için kullanılmak üzere bu görüntülerin 7180 tanesini seçilmiştir. İç mekân görüntüleri ise iki farklı veri kümesinden alınmıştır. Bunlardan ilki MIT Enstitüsünde hazırlanmış olan "Indoor Scene Recognition" adlı veri kümesidir. İçerisinde 67 farklı iç mekân kategorisi olup 15620 adet farklı görüntü bulunmaktadır [5].

İç mekân için ikinci olarak kullanılan veri kümesi ise 5 farklı mekândan 5250 adet görüntü içerir biçimde, makine öğrenmesi ve veri bilimi alanında çevrimiçi bir topluluk ve çeşitli yarışmalar içeren Kaggle [6] İnternet sitesinden elde edilen "House_Room_Dataset" [7] adlı veri kümesidir. Her iki veri kümesinden iç mekânlar için toplamda boyutsal

uyuşmazlıkları içeren görüntüler elenerek 15915 adet görüntü çalışmamızda kullanılmıştır.

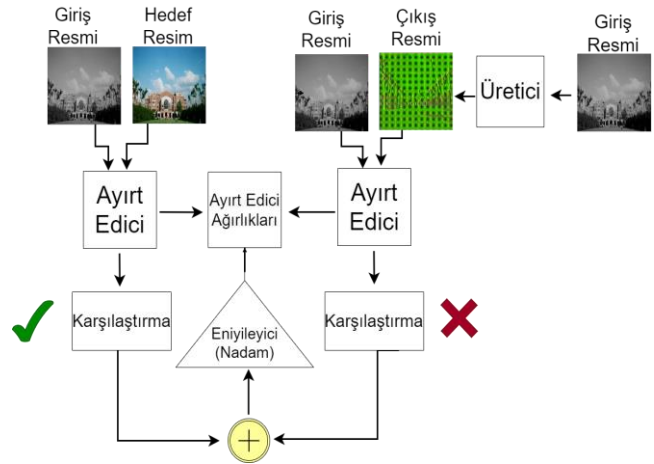
Dış mekân görüntüleri için ise yine aynı enstitünün "Places2" veritabanının "Data of Places365-Standard/Small Images"[8] alt başlığındaki "Test Images" alt veri kümesi kullanılmıştır. Kullanıcı arayüzündeki "Görüntü Onarma" işlemi için yukarıdaki veri kümelerinin her biri için 2000 adet görüntü içeren karma bir veri kümesi oluşturulmuştur. Kullanıcı arayüzündeki "Görüntü Onarma" işlemi için toplam 6000 adet görüntü kullanılmıştır. Tüm kullanılan görüntüler rastgele olarak seçilmiştir.

Düşük çözünürlükteki görüntülerin yüksek çözünürlüğe dönüştürülmesi adına yukarıda bahsi geçen iç mekân ve dış mekân veri kümelerinden her biri için 200 adet görüntü rastgele seçilerek yüksek çözünürlüğe dönüştürülerek ve orijinal düşük çözünürlüklü halleriyle birlikte deneylerde kullanılmıştır.

B. Üretken Çekişmeli Ağ Mimarisi

Kullanılan ağ ile Danışmansız Öğrenme ile genel kullanıma yönelik bir çalışma ortaya çıkarılması hedeflenmiştir. Danışmansız Öğrenme (Unsupervised Learning) [9], algoritmaya eğitim verileri için önceden atanmış herhangi bir etiket veya puan verilmediği bir tür makine öğrenimi olduğu için veri kümesi yardımıyla birçok görüntü etiketlenmeden işlenebilecek duruma gelmiştir. Danışmansız öğrenme algoritmaları, veri kümesinde doğal olarak oluşan örüntüleri kendi kendine keşfetmektedir.

GAN'ı temel aldığımız için burada da bir Üretici (Generator) ve Ayırt Edici (Discriminator) bulunur. Pix2Pix modelinde kullanılan Üretici, her iki yığımda da yansıtılmış katmanlar arasında atlama bağlantıları olan bir Kodlayıcı-Kod Çözücü içeren bir görüntü bölütleme ağı olan U-NET'tir [10]. Ayırt Edici ise hem girdinin istenen hedef görüntü ile karşılaştırır hem de Üretici'nin çıktısı ile karşılaştırır. Ayırt Edici ağırlıklarını yenileyerek farkı tespit etmede kendini geliştirmektedir. Çalışmada kullandığımız Pix2pix modelinin çalışma şeması aşağıdaki şekilde verilmektedir.



Şekil 1. Pix2pix modelinin çalışma şeması

Eğitim esnasında amaç fonksiyonu olarak verilen kayıp fonksiyonunun (Loss function) hesaplamasında İkili Çapraz Entropi (Binary Cross-Entropy) [11] kullanılmaktadır. Modelimizde kullanılan eniyileyici (optimizer) Karesel hatalar toplamı üzerinden geriye yayılım yapan RMSProp

[12] ilk olarak alınmıştır. Diğer bir eniyileyci olan Nadam [12], RMSProp'un benzeri bir şekilde ancak Nesterov momentumu [12] yöntemini uygulayarak çalışmaktadır. Deneylerin sonucunda ileriki bölümde anlatılan Tepe Sinyal Gürültü Oranı (Peak Signal To Noise Ratio - PSNR) ve Yapısal Benzerlik Endeks Ölçüsü (Structural Similarity Index Measure - SSIM) sonuçlarına bakılarak Nadam eniyileme yönteminin daha başarılı olmasından dolayı Üretici ve Ayırt Edici için eniyileyci olarak bu yöntem seçilmiştir.

Üstün parametre (hyperparameter) olarak öğrenme oranı (learning rate) değeri 0,001 olarak ayarlanmıştır. Deneyde eğitim için 500 eğitim yinleme (epoch) adımı boyunca eğitilmiştir. Görüntü renklendirme, iyileştirme ve onarma aşamalarında aynı mimari ve üstün parametreler kullanılmıştır. Görüntü renklendirme altyapısında ek olarak insan yüzü tespiti için YOLOv3 [13], [14] mimari modeli, iç ve dış mekân için AlexNet [15] mimari modeli kullanılmaktadır. Sınıflandırma için kullanılan bu modeller sonrasında çıktı görüntü Pix2pix modeline verilmektedir. Kullanılan mimarinin katmanları aşağıdaki Tablo 1'de gösterilmiştir:

Tablo 1. Deneylerde kullanılan Üretici Çekişmeli Ağ mimarisini

Model: "model"	1	2	3
Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 512, 512, 3)]	0	[]
sequential_1 (Sequential)	(None, 256, 256, 64)	3072	['input_1[0][0]']
sequential_2 (Sequential)	(None, 128, 128, 64)	65792	['sequential_1[0][0]']
sequential_3 (Sequential)	(None, 64, 64, 128)	131584	['sequential_2[0][0]']
sequential_4 (Sequential)	(None, 32, 32, 256)	525312	['sequential_3[0][0]']
sequential_5 (Sequential)	(None, 16, 16, 512)	2099200	['sequential_4[0][0]']
sequential_6 (Sequential)	(None, 8, 8, 512)	4196352	['sequential_5[0][0]']
sequential_7 (Sequential)	(None, 4, 4, 512)	4196352	['sequential_6[0][0]']
sequential_8 (Sequential)	(None, 2, 2, 512)	4196352	['sequential_7[0][0]']
sequential_9 (Sequential)	(None, 1, 1, 512)	4196352	['sequential_8[0][0]']
concatenate (Concatenate)	(None, 2, 2, 1024)	0	['sequential_9[0][0]', 'sequential_7[0][0]']
sequential_10 (Sequential)	(None, 4, 4, 512)	8390656	['concatenate[0][0]']
concatenate_1 (Concatenate)	(None, 4, 4, 1024)	0	['sequential_10[0][0]', 'sequential_6[0][0]']
sequential_11 (Sequential)	(None, 8, 8, 512)	8390656	['concatenate_1[0][0]']
concatenate_2 (Concatenate)	(None, 8, 8, 1024)	0	['sequential_11[0][0]', 'sequential_5[0][0]']
sequential_12 (Sequential)	(None, 16, 16, 512)	8390656	['concatenate_2[0][0]']
concatenate_3 (Concatenate)	(None, 16, 16, 1024)	0	['sequential_12[0][0]', 'sequential_4[0][0]']
sequential_13 (Sequential)	(None, 32, 32, 256)	4195328	['concatenate_3[0][0]']
concatenate_4 (Concatenate)	(None, 32, 32, 512)	0	['sequential_13[0][0]', 'sequential_3[0][0]']
sequential_14 (Sequential)	(None, 64, 64, 128)	1049088	['concatenate_4[0][0]']
concatenate_5 (Concatenate)	(None, 64, 64, 256)	0	['sequential_14[0][0]', 'sequential_2[0][0]']
sequential_15 (Sequential)	(None, 128, 128, 64)	262400	['concatenate_5[0][0]']
concatenate_6 (Concatenate)	(None, 128, 128, 12)	0	['sequential_15[0][0]', 'sequential_1[0][0]']
sequential_16 (Sequential)	(None, 256, 256, 32)	65664	['concatenate_6[0][0]']
concatenate_7 (Concatenate)	(None, 256, 256, 96)	0	['sequential_16[0][0]', 'sequential[0][0]']
conv2d_transpose_8	(None, 512, 512, 3)	4611	['concatenate_7[0][0]']

Eğitim öncesi ön işlemler ve sonrasında son işlemler yapılarak modelin sonuçlarının desteklenmesi ve GAN'ın uzun zamanda yapacağı iyileştirmeler ve eğitim süresinin kısaltılması amacıyla eklenmiştir. Ön işleme için yapılan adımların ilki histogram eşitlemesidir. Histogram eşitleme, renk dağılımının eşit bir şekilde yayılmasına olanak sağlamaktadır. Bazı görüntülerde karanlık tonların veya silik kalmış bölgelerin görüntülerdeki detaylarının

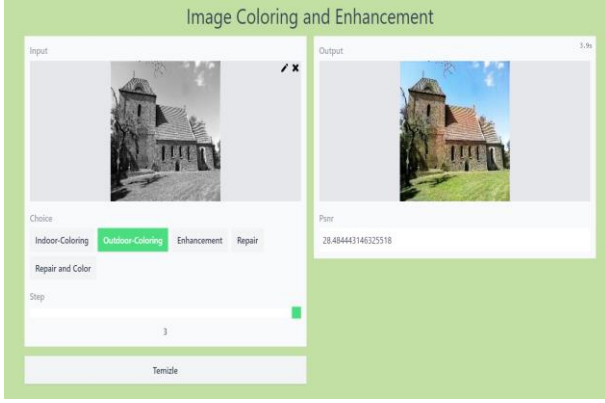
artırılabilmesini sağlamaktadır. Bu adımda, histogram eşitleme ve uyarlanabilir histogram eşitleme olmak üzere iki yöntem denenmiştir. Deneysel sonuçlarda uyarlanabilir histogram eşitleme daha iyi sonuçlar verdiği ve fazla parlak alanlarda histogram eşitlemede olduğu gibi görüntüde bozukluklar oluşturmadığı için uyarlanabilir histogram eşitlemesi tercih edilmiştir. Daha sonraki adım ise Gamma (Işıklendirme) düzeltmesidir. Bu yöntem görüntüdeki parlaklık ayarını yapmayı sağlamaktadır. Her bir görüntü için özel olarak Gamma değeri ayarlanabilir ve parlaklık değiştirilebilir ancak girdilerde gelen Gamma değerleri farklı olacağı ve her görüntüde aynı parlaklık artırma veya azaltma işleminin uygulanması doğru olmayacağı için bu değer hesaplandıktan sonra gelen değere göre yöntemin uygulandığı koşullu durumlar eklenmiştir.

Son adımda bu işlemleri birleştirerek görüntülerimize uygulayabileceğimiz genel bir ön işleme fonksiyonu oluşturularak kullanılmıştır. Bu fonksiyonda işleme tutulan görüntüler Pix2pix modelimizde girdi olarak kullanılmıştır. Pix2pix modelimizden çıkan iyileştirilmiş görüntüler daha sonra son işlemleri içeren bir fonksiyonda işlem görmüştür. Gauss Bulanıklığı [16] eklenerek öncelikle renklendirilmiş görüntünün piksel değerlerinin ortalaması alınarak uygun bir değışkene atanmıştır. Böylelikle görüntüdeki gürültüler azaltılmıştır. Görüntü işleme teknikleri ile görüntülerin keskinlik (sharpen) değerleri artırılmıştır ve kenarlar belirginleştirilmiştir. Böylece bina, sütun gibi kenarlı nesnelerin kenarları daha belirgin ve daha ayırt edilebilir hale getirilmiştir. Bu son işlemler ile deneysel sonuçlar elde edilmiştir.

C. Kullanıcı Grafik Arayüzü

Son kullanıcının verdiği girdi görüntüsünü istenilen görüntü iyileştirme işlemlerini uygulayabileceği ve bu yolla çıktı görüntüsünü oluşturulabilecek bir kullanıcı grafik arayüzü meydana getirilmiştir. Aynı zamanda seçilen işlem ile beraber basamaklı olarak belirli eğitim yinleme adımı sayısı için (birinci, ikinci ve üçüncü basamaklar) arayüz üzerinden seçilerek kullanıcının farklı adım sayılarındaki çıktıları ve değışikliği görmesi sağlanmıştır. Kullanıcının seçeceği görüntüler dışında örnek olarak işlenebilecek görüntüler de arayüze eklenmiştir. Arayüzde ayrıca kullanımın açıklandığı bir bilgilendirme dosyası bulunmaktadır. Bu arayüz Gradio [17] ve Huggingface [18] servislerinin ve alt yapısının kullanımı sayesinde oluşturulmuştur. Çalışmada geliştirilen arayüze (Color, Enhancement, Restoration for Old Images-ImgCERO) erişim ilgili Internet sitesi üzerinden sağlanabilmektedir [19].

Aşağıdaki Şekil 2'de kullanıcı grafik arayüzünü kullanarak deney sonuçları üzerinden tahminleme, sınıflandırma, renklendirme ve görüntü iyileştirme yapılan bir örnek görülmektedir.



Şekil 2. Kullanıcı Grafik Arayüzü

Gradio'nun kendi özelliklerinden biri olarak görüntünün düzenlenebilmesi (kırpma, çeşitli efektler ve renk ayarları gibi) imkanı da bulunmaktadır.

III. DENEYSEL SONUÇLAR

Başarım ölçümü için deneylerdeki görüntü çıktıları literatürde sıkça kullanılan SSIM [20], PSNR [21] ölçütleri ile nesnel olarak değerlendirilmiştir. SSIM sayısal görüntülerin ve video görüntülerinin algılanan kalitesini tahmin etmek için kullanılan bir yöntemdir. İki görüntü arasındaki benzerliği ölçmek için de kullanılır. SSIM kontrast, parlaklık ve görüntü yapısının karşılaştırma ölçümüne dayanır.

PSNR, görüntü kalitesini ölçmek için literatürde yaygın olarak kullanılan bir ölçüttür. Bit derinliğinin 8 bit olması halinde, PSNR olarak ifade edilen gürültü değerleri 30 ile 50 desibel (dB) arasındadır. PSNR değeri 12 bit görüntülerin işlenmesinde 60 dB veya daha yüksek olabilir. 16 bitlik veriler için ise PSNR 60 ile 80 dB değerleri arasındadır [22]. Eski görüntülerin onarılması sayesinde eski görüntüler ile yeni onarılmış görüntüler kıyaslandığında SSIM değerinin oldukça yüksek olarak ortalama 0,8113 değeriyle elde edildiği görülmüştür. Deneyler sonucunda elde edilen görüntü çıktıları için veri kümesine dair ortalama SSIM, PSNR değerleri Tablo 2'de görülmektedir.

Tablo 2. Görüntü renklendirilmesine dair deney sonuçları

Ölçüt	İç Mekân	Dış Mekân	İnsan Yüzü	Kalite İyileştirme
SSIM	0,7354	0,8270	0,7917	0,9256
PSNR	67,86	66,89	65,11	70,22

Modellerin eğitimi sonucunda yapılan testlerde deneysel sonuçlar olarak çeşitli başarım değerleri içeren sonuçlar elde edilmiştir. Şekil 3'te görüldüğü gibi "Indoor Scene Recognition" veri kümesinden alınan (a) şıkkında örnek bir iç mekân görüntüsünün kendisi, (b) şıkkında ise iç mekân renklendirme sonucunda elde edilen çıktı görüntüsü verilmiştir.



(a)

(b)

Şekil 3. Indoor Scene Recognition [5] veri kümesinden alınan örnek için, (a) iç mekân görüntüsü, (b) renklendirme sonucu sonuç görüntüsü

Şekil 4'te görüldüğü gibi "Data of Places365-Standard/Small Images" alt başlığındaki "Test Images" alt veri kümesinden alınan (a) şıkkında örnek bir dış mekân görüntüsünün kendisi, (b) şıkkında ise dış mekân renklendirme sonucunda elde edilen çıktı görüntüsü verilmiştir.



(a)

(b)

Şekil 4. Places365-Standard/Small Images [8] veri kümesinden alınan örnek için, (a) dış mekân görüntüsü, (b) renklendirme sonucu sonuç görüntüsü

Şekil 5'te görüldüğü gibi Massachusetts Amherst Üniversitesi tarafından hazırlanmış olan "Labeled Faces in the Wild" veri kümesinden alınan (a) şıkkında örnek bir insan yüzü görüntüsünün kendisi, (b) şıkkında ise insan yüzü renklendirme sonucunda elde edilen çıktı görüntüsü verilmiştir. Şekil 5'ten görüldüğü üzere yüz bölgesini etkili bir şekilde renklendirilmesine karşın arka plan renklendirilmesinde etkili bir sonuç çıkmamıştır. Burada önemli olan yüz bölgesinin renklendirilmesi olduğu için ön plan kalitesi önemsenmemektedir.



(a)

(b)

Şekil 5. Massachusetts Amherst Üniversitesi'nin Labeled Faces in the Wild [4] veri kümesinden alınan örnek için, (a) insan yüzü görüntüsü, (b) renklendirme sonucu görüntüsü

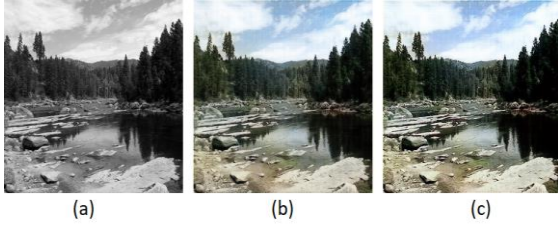
Şekil 6'da görüldüğü gibi yukarıda da bahsedildiği üzere her üç veri kümesinden karma olarak hazırlanmış olduğu veri kümesinden alınan (a) şıkkında örnek bir görüntünün kendisi, (b) şıkkında ise görüntünün onarılması sonucunda elde edilen çıktı görüntüsü verilmiştir. Ayrıca, (c) şıkkında ise görüntünün bozuk hali verilmiştir. Görüntüde bozulma işlemleri yapılırken, siyah beyaz görüntü üstüne rastgele noktalara ve

rastgele kalınlıklardan beyaz çizgiler oluşturularak görüntünün bütünlüğü bozulmuştur.



Şekil 6. Bahsi geçen veri kümelerinden karma olarak oluşturulan veri kümesinden alınan örnek için, (a) örnek görüntü, (b) görüntü onarılması, sonucu daha yüksek kalitedeki sonuç görüntüsü, (c) görüntünün bozuk hali.

Şekil 7’de görüldüğü gibi dış mekân veri kümesinden alınan örnek görüntünün siyah beyaz görüntü haline çevrildikten sonraki hali (a) şıkkında, siyah beyaz görüntünün çalışmamızdaki yaklaşımla renklendirilmesi sonucunda ortaya çıkan görüntü (b) şıkkında, renklendirilmiş görüntünün kalitesi iyileştirilmiş hali (c) şıkkında gösterilmiştir.



Şekil 7. Veri kümesinden [8] alınan örnek görüntü için, (a) siyah beyaz görüntü, (b) görüntünün renklendirilmesi sonucu, (c) görüntünün kalitesi iyileştirilmiş hali.

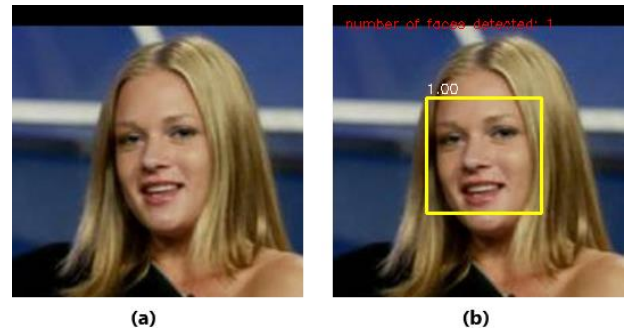
Ön işlem adımlarından biri olarak kullanıcıdan alınan görüntünün iç veya dış mekân olduğunun kararının verilmesinde çalışmamızdaki oluşturulan sınıflandırma altyapısı kullanılmıştır. Bu altyapıda derin öğrenme mimari modellerinden AlexNet kullanılarak yüksek başarımla öncelikli olarak mekânın türü belirlenmiş buna uygun olarak renklendirme aşamalarına geçilmiştir. Çapraz tahmin (confusion matrix) tablosu [23] literatürde sınıflar ve tahmin sonuçlarını yorumlamayı sağlayan istatistiksel temelli bir olgudur. Buna göre; Doğru negatif (True negative-TN) ölçütü gerçekte var olmayan bir şeye eğitilen model de yok demişse bunun değerini gösterir, Yanlış pozitif (False positive-FP) ölçütü gerçekte olmayan bir şeye eğitilen model de var demişse bunun değerini gösterir, Yanlış negatif (False Negative-FN) ölçütü gerçekte var olan bir şeye eğitilen model de yok demişse bunun değerini gösterir ve Doğru pozitif (True positive-TP) ölçütü ise gerçekte var olan bir şeye eğitilen model de var demişse bunun değerini gösterir. Aşağıdaki çapraz tahmin tablolarındaki diyagonal kesişimlerinde sol üst köşeden başlayarak sırasıyla TN ve TP değerleri bulunmaktadır. Eğer tahmin edilen etiket ve gerçekte var olan etiket aynı şekilde negatif etiketliyse buna TN denilmekte, aksine pozitif etikete sahip ise buna TP denilmektedir.

Tablo 3’te AlexNet modeli ile elde edilmiş çapraz tahmin tablosu gösterilmektedir. Tablo 3’te iç ve dış mekân sınıflandırma sonuçlarında gerçek etiketlerle tahmini etiketler karmaşıklık matrisi üzerinde kıyaslandığında yüksek bir başarımın elde edildiği görülmektedir.

Tahmin Etiketi			
		Negatif	Pozitif
Gerçek Etiket	Negatif	TN: 720	FP: 31
	Pozitif	FN: 1200	TP: 5200

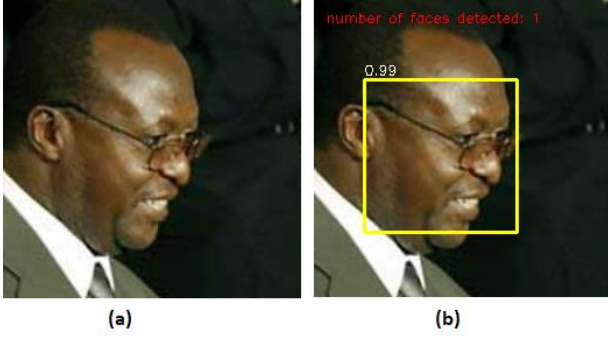
Toplamda 7151 adet görüntü üzerinde değerlendirme yapılmıştır. Tablo 3’teki sonuçlardan görüldüğü üzere doğru negatif ve doğru pozitif tahminlerin toplamının tüm tahmin değerlerine oranından iç ve dış mekân sınıflandırma modelinde %82 doğruluk oranında başarı elde edilmiştir. İç ve dış mekân sınıflandırması görüntü renklendirmede bir araç olarak kullanılmıştır.

YOLOv3 modeli kullanılarak insan yüzü tespiti yapılmıştır. Yüz tespiti için önceden eğitilmiş modelin ağırlıklarından faydalanılmıştır [24]. İnsan yüzü tespiti görüntü renklendirmede bir araç olarak kullanılmıştır. Görüntü içerisindeki işlenecek nesne çeşidini en aza indirmek için görüntüde insan yüzünün kapsadığı yüzdelik oran (%60 ve daha fazlası olan aralıktaki oran) ve buna uygun yüz tespiti yapılmıştır. Aralık oranından daha az alan kapsayan görüntüler için arayüzde girdinin uygun olmadığına dair uyarı verilmekte ve renklendirme işlemi buna göre yapılmamaktadır. Şekil 8 ve Şekil 9’da “Labeled Faces in the Wild” veri kümesinden alınan örnekler kullanılarak insan yüzü tespiti sonuçları görülmektedir. Şekil 8’de ön cepheden çekilmiş bir görüntüde tespit edilen insan yüzünün sinir ağına verilen görüntünün %85’ini, Şekil 9’da yan profilden çekilmiş bir görüntüde tespit edilen insan yüzünün sinir ağına verilen görüntünün %66’sını kapsadığı görüntü boyutu görülmektedir. Bunlar için insan yüzü alanı oranına göre hesaplama programatik olarak yapılmıştır. Şekil 8’de bir adet insan yüzü %100 ihtimalle tespit edilirken, Şekil 9’da bir adet insan yüzü %99 ihtimalle tespit edilmiştir.



Şekil 8. Massachusetts Amherst Üniversitesi’nin Labeled Faces in the Wild [4] veri kümesinden alınan ön cepheden çekilmiş örnek görüntü için, (a) renkli insan yüzü görüntüsü, (b) yüz tespitinin sonuç görüntüsü.

Tablo 3. İç ve dış mekân sınıflandırma sonuçları



Şekil 9. Massachussets Amherst Üniversitesi'nin Labeled Faces in the Wild [4] veri kümesinden alınan yan profilden çekilmiş örnek görüntü için, (a) renkli insan yüzü görüntüsü, (b) yüz tespiti sonuc görüntüsü.

IV. TARTIŞMA VE SONUÇ

Bu çalışmada eski ve yıpranmış, çoğunlukla siyah beyaz veya gri tonlamalı doğal görüntüleri sırası ile görüntü renklendirme, görüntü iyileştirme, onarım ve restorasyon işlemlerine dair derin öğrenme yöntemleri kullanılarak son kullanıcı dostu bir arayüz ile işlem yapma denemiştir, sonuçları olumlu bulunmuştur. Özellikle insan yüz bölgesinin tespit edilmesi sonucu renklendirme yapılması çalışmamızın öne çıkan katkılarından birisidir. Son kullanıcının bütünleşik olarak görüntüyle ilgili birçok aşamayı bir arada zahmetsizce ve hızlıca birkaç arayüz işlemiyle gerçekleştirebilmesi bu çalışmanın ana katkılarından birisidir.

Çalışma kapsamında çeşitli sinir ağı modelleri denenerek bunlar arasında en iyi sonuç verenler için çeşitli üstün parametreler belirlenerek en uygun deney sonuçları elde edilmiştir. Bu deney sonuçları son kullanıcı arayüzü kullanarak işlemleri otomatik ve kıyaslamalı şekilde gerçekleştirilebileceğini göstermektedir. Ayrıca deneysel sonuçlar çalışmanın yüksek doğrulukta bir başarıya sahip olduğunu göstermektedir. Gelecek çalışmalarımızda daha farklı derin öğrenme mimari modelleri ve veri kümeleriyle deney ortamının ve deney kapsamının genişletilmesi hedeflenmektedir.

KAYNAKLAR

- [1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Nets", 2014, arXiv Preprint: arXiv:1406.2661 [stat.ML] <https://doi.org/10.48550/arXiv.1406.2661>
- [2] M. Mirza, S. Osindero, "Conditional Generative Adversarial Nets", 2014, arXiv Preprint: arXiv:1411.1784 [cs.LG] <https://doi.org/10.48550/arXiv.1411.1784>
- [3] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled Faces in the Wild: A Survey". In *Advances in Face Detection and Facial Image Analysis*, Editörler: M. Kawulok, M. E. Celebi ve B. Smolka, Springer, pp. 189-248, 2016, ISBN: 978-3-319-25958-1
- [4] Labeled Faces in the Wild [Online]: <http://vis-www.cs.umass.edu/lfw/>
- [5] A. Quattoni, A. Torralba, "Recognizing indoor scenes," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 20-25 June 2009, pp. 413-420, doi: 10.1109/CVPR.2009.5206537.
- [6] Kaggle İnternet sitesi [Online]: <https://www.kaggle.com>
- [7] House Rooms Dataset İnternet sitesi [Online]: <https://www.kaggle.com/datasets/robinreni/house-rooms-image-dataset>
- [8] Places2 Dataset İnternet sitesi [Online]: <http://places2.csail.mit.edu/download.html>
- [9] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, A. Joulin, "Unsupervised Learning of Visual Features by Contrasting Cluster Assignments", 2021: arXiv Preprint arXiv: 2006.09882[cs.CV] <https://doi.org/10.48550/arXiv.2006.09882>
- [10] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 2015,

- arXiv Preprint: arXiv:1505.04597 [cs.CV], <https://doi.org/10.48550/arXiv.1505.04597>
- [11] Y. Ho, S. Wookey, "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling", arXiv Preprint: arXiv:2001.00570 [cs.LG], <https://doi.org/10.48550/arXiv.2001.00570>
- [12] S. Ruder, "An overview of gradient descent optimization algorithms", 2017, arXiv Preprint : arXiv:1609.04747 [cs.LG] <https://doi.org/10.48550/arXiv.1609.04747>
- [13] A. Farhadi, J. Redmon, "YOLOv3: An Incremental Improvement", 2018, arXiv Preprint: arXiv:1804.02767 [cs.CV], <https://doi.org/10.48550/arXiv.1804.02767>
- [14] YOLOv3 in PyTorch, Github İnternet Sitesi: <https://github.com/ultralytics/yolov3>
- [15] G. E. Hinton, A. Krizhevsky, I. Sutskever, "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems*, Editörler: F. Pereira and C.J. Burges and L. Bottou and K.Q. Weinberger, Curran Associates, Inc., Vol. 25, 2012. [Online] : <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- [16] E. S. Gedraite, M. Hadad, "Investigation on the effect of a Gaussian Blur in image filtering and segmentation," *Proceedings ELMAR-2011*, 14-16 Sept. 2011, pp. 393-396. ISBN: 978-1-61284-949-2
- [17] Gradio İnternet sitesi [Online]: <https://gradio.app>
- [18] Hugging Face İnternet sitesi [Online]: <https://huggingface.co>
- [19] ImgCERO Project İnternet Sitesi: https://huggingface.co/spaces/simayhosmeyve/Image_Enhancement
- [20] T. A. Möller, J. Nilsson, "Understanding SSIM", 2020: arXiv Preprint: arXiv:2006.13846 [eess.IV] <https://doi.org/10.48550/arXiv.2006.13846>
- [21] O. Keleş, M. A. Yılmaz, A. M. Tekalp, C. Korkmaz, Z. Doğan, "On the Computation of PSNR for a Set of Images or Video", 2021: arXiv Preprint : arXiv:2104.14868 [eess.IV] <https://doi.org/10.48550/arXiv.2104.14868>
- [22] A. Horé and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," 2010 20th International Conference on Pattern Recognition, 2010, pp. 2366-2369, doi: 10.1109/ICPR.2010.579.
- [23] E. van der Knaap, A. Ralescu, B. Ramsay, S. Visa, "Confusion Matrix-based Feature Selection", MAICS , volume 710 of CEUR Workshop Proceedings, page 120-127.: <http://ceur-ws.org/Vol-710/paper37.pdf>
- [24] T. Nguyen, "Deep learning-based Face detection using the YOLOv3 algorithm" İnternet Sitesi [Online]: <https://github.com/ssthanhg/yoloface>