



Scalable image generation and super resolution using generative adversarial networks

Ceren Guzel Turhan^{1*} , Hasan Şakir Bilge² 

¹Computer Engineering Department, Gazi University, Ankara, 06570, Turkey

²Electric-Electronic Engineering Department, Gazi University, Ankara, 06570, Turkey

Highlights:

- A novel autoencoder and adversarial learning based hybrid model for image generation and reconstruction
- Compositional generative adversarial networks for scalable image generation and reconstruction
- More robust generative model using feature-wise objective rather than pixel-wise

Keywords:

- Deep learning
- variational autoencoders
- generative adversarial networks
- image generation
- scalable super resolution

Article Info:

Research Article
Received: 04.07.2019
Accepted: 15.11.019

DOI:

10.17341/gazimmfd.587010

Correspondence:

Author: Ceren Guzel Turhan
e-mail:
cerenguzel@gazi.edu.tr
phone: +90 312 582 31 30

Graphical/Tabular Abstract

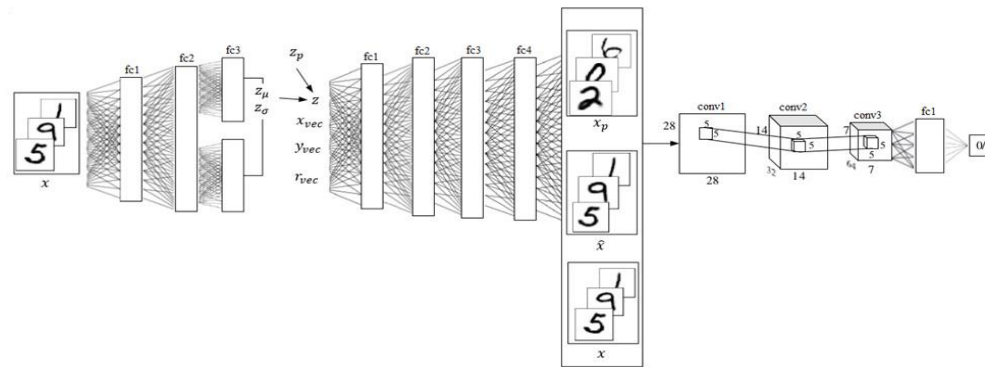


Figure A. The proposed VAE/CPGAN architecture

Purpose: In spite of their popularity, Generative Adversarial Network (GAN) based models seem to have limited generation capabilities for image generation. The lack of the inference mechanism of GAN models has led to the combination of GAN and autoencoder (AE) networks. Therefore, this study focuses on GAN and VAE combining with Compositional Pattern Producing Network (CPPN) like generator and feature-wise reconstruction idea to enhance the model performance while obtaining more robust model. Moreover, it is aimed to learn the image densities and sampling images at different scales that are super resolution images thanks to CPPN like generator.

Theory and Methods:

The introduction and development of a VAE/GAN hybrid model combining CPPN approach for scalability properties and feature-wise objective instead of pixel-wise. That differs from standard hybrid models

- in point of generator idea: CPPN model is provided as generator
- in addition to latent code, coordinate inputs are given for generation which allows us to generate any high dimensional version of images without any training with them as ground-truth.
- focusing blur problem on generated images. It improves image sharpness thanks to pixel-wise generation by CPPN.
- feature-wise objective instead of pixel-wise objective which cannot understand the underlying structure of image domains.

Results:

In order to evaluate the model performance, VAE and GAN based models have been compared with the proposed model based on MSE and IS score. According to given results, the proposed VAE/CPGAN approach is superior than compared models not only in point of image generation but also reconstruction performance.

Conclusion:

This study shows the effects of combining inference mechanism, adversarial learning and CPPN-like generator with feature-wise objective. The proposed network not only allows us to learn and sample the data distribution but also the image density, which enables the generation of high-resolution images from latent codes of the low-resolution images in addition to noiseless image reconstruction.



Çekişmeli üretici ağ ile ölçeklenebilir görüntü oluşturma ve süper çözünürlük

Ceren Güzel Turhan^{1*}, Hasan Şakir Bilge²

¹Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 06570, Maltepe, Ankara, Türkiye

²Gazi Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 06570, Maltepe, Ankara, Türkiye

Ö N E Ç İ K A N L A R

- Görüntü oluşturma ve yeniden yapılandırma için yeni bir oto kodlayıcı ve çekişmeli eğitime dayalı hibrit model
- Sunulan Çekişmeli Üretici Ağ modeli ile sentetik görüntü oluşturma ve görüntülerin yeniden yapılandırılması
- Piksel tabanlı amaç fonksiyonu yerine öznitelik tabanlı amaç fonksiyonu ile daha gürbüz bir üretici ağ modeli

Makale Bilgileri

Araştırma Makalesi

Geliş: 04.07.2019

Kabul: 15.11.019

DOI:

10.17341/gazimmfd.587010

Anahtar Kelimeler:

Derin öğrenme,
varyasyonel otokodlayıcı,
üretici çekişmeli ağ,
görüntü oluşturma,
ölçeklenebilir süper
çözünürlük

ÖZET

Çekişmeli eğitimin, derin öğrenme konusundaki ileri gelen araştırmacılar tarafından son 10 yılın en heyecan verici konularından biri olarak ilan edilmesi ile birlikte birçok araştırmacı, Üretici Çekişmeli Ağ (GAN) üzerine yoğunlaşmıştır. Bu gelişmeler ile birlikte Otokodlayıcı (AE) olarak daha önceden bilinen modelin bir versiyonu olan Varyasyonel Otokodlayıcı (VAE) ile birlikte Otokodlayıcı modeller literatürde yeniden önem kazanmıştır. GAN modellerinin bazı kısıtlamaları ile çıkarım mekanizması gibi bazı eksiklikleri, GAN ve VAE hibrit modellerini ortaya çıkarmıştır. Bu gelişmelerin etkisiyle, bu çalışmada da bir GAN ve VAE hibrit modeline odaklanılmıştır. Sentetik görüntüler ile eğitim yapılmaksızın istenen çözünürlükte görüntüleri elde edebilmek üzere, Kompozisyonel Örtüntü Üreten Ağlar (CPPN) ile GAN ve VAE modelleri bir arada kullanılarak VAE/CPGAN olarak adlandırılan model sunulmuştur. Ayrıca, VAE / GAN hibrit modelinde kullanılan objektif fonksiyonu önerilen model ile bir araya getirilmiştir. Model performanslarını değerlendirmek üzere, deney çalışmalarında VAE/CPGAN modeli, bu modele temel olan GAN, VAE ve VAE / GAN modelleri ile karşılaştırılmıştır. Önerilen modelin el yazısı görüntülerinden oluşan popüler veri kümesi üzerinde gerçekleştirilen deneyler ile gerçekçi ve süper çözünürlüklü, ölçeklenebilir, sentetik görüntüler oluşturmadaki başarısı bu çalışma ile ortaya koyulmuştur.

Scalable image generation and super resolution using generative adversarial networks

H I G H L I G H T S

- A novel autoencoder and adversarial learning based hybrid model for image generation and reconstruction
- A compositional generative adversarial network for synthetic image generation and reconstruction
- More robust generative model using feature-wise objective rather than pixel-wise objective of recent studies

Article Info

Research Article

Received: 04.07.2019

Accepted: 15.11.019

DOI:

10.17341/gazimmfd.587010

Keywords:

Deep learning,
variational autoencoder,
generative adversarial
network,
image generation,
scalable super resolution

ABSTRACT

Generative adversarial training has been one of the most active research topics and many researchers have conducted their studies on Generative Adversarial Network (GAN) shortly after it is claimed to be one of the most promising research area of the last decade by pioneers of the deep learning community. On the other hand, the idea behind generators has also reemerged autoencoder models such as Variational Autoencoder (VAE). Therefore, autoencoder models have gained their popularity back. Some restrictions of GAN models such as lack of inference mechanism, GAN and VAE based hybrid models have proposed addressing image generation. With the effect of these notions and studies, we have also considered VAE and GAN hybrid models. For obtaining synthetic but at the same time high-resolution handwritten-looking images without any training, Compositional Pattern Producing Network (CPPN) is adapted from previous studies for combining with VAE and adversarial training. For improving generation capabilities, an objective from a previous VAE/GAN model is also adapted for our VAE/CPGAN hybrid model. For analyzing the proposed model performance, baseline models such as GAN, VAE and VAE/GAN are also evaluated for comparisons. In this paper, it is clearly seen the proposed model is capable of the generating realistic and scalable super resolution synthetic images on a common dataset.

*Sorumlu Yazar/Corresponding Author: cerenguzel@gazi.edu.tr, bilge@gazi.edu.tr / Tel: +90 312 582 31 30

1. GİRİŞ (INTRODUCTION)

Derin öğrenme evrimsel ağ (CNN) modellerinin ardından son zamanlarda ortaya çıkan derin üretici modellerin içerik üretme konusundaki potansiyel kullanımları nedeniyle literatürde oldukça ses getirmiştir. Üretici modeller esas olarak yeni örneklerin üretilmesine odaklanırken, CNN modelleri gibi ayrımcı modeller örnekleri sınıflandırarak ayırt edebilmek üzere bir karar sınırı bulmaya çalışmaktadır. İlk derin çekişmeli üretici ağ modeli olan GAN [1], üretici ağ ile ayrımcı ağları çekişmeli eğitim yaklaşımı ile bir araya getiren öncül bir model olarak birçok çalışmaya yön vermiştir. Yeni bir kavram olmamakla birlikte GAN modeli ile gündeme gelen çekişmeli eğitim kavramı, çekişmeli örnekler kullanılarak modellerin eğitilerek daha gülbüz hale getirilmesine dayanır. GAN modelinde, çekişmeli eğitim ile daha güçlü bir model elde edilirken gerçekçi sentetik örnekler de elde edilerek bu örnekler ile gerçek örnekleri ayırt etmek üzere eğitilen ayrımcı modellerin performansını artırılır [2]. GAN modeli, yüksek seviyeli öznelikler elde ederek veri yapısı ve dağılımını modelleyebilmesi ile çekişmeli eğitim aracılığıyla öğrenilen veri dağılımından örnekleme yapılmasına olanak tanınması nedeniyle araştırmacıların dikkatini çekerek çok sayıda GAN tabanlı modellerin ortaya çıkmasına neden olmuştur.

Üretici modeller ile sentetik fakat gerçekçi örnekler oluşturmaya ek olarak verilerin tanımlı niteliklere göre modifiye edilmesi [3, 4] belirlenmiş cümle ve kelimeleri tanımlayan görüntülerin oluşturulması [5, 6], eksik görüntülerin tamamlanması [7, 8], görüntülerdeki sahnelerin değiştirilmesi [9, 10] gibi çok sayıda uygulama gerçekleştirilebilmiştir.

Önceki çalışmalarda, GAN tabanlı modellerin popülerliklerinin aksine sınırlı üretim yeteneklerine sahip oldukları görülmüştür. GAN modellerinin çıkarım mekanizması eksikliğini çözmek üzere GAN ile kodlayıcı ağların bir arada kullanılması ile hibrit modellerin ortaya çıkmasına neden olunmuştur. Modellerin eksikliklerini gidererek performansların arttığının değerlendirilmesi nedeniyle bu çalışmada da bir GAN ve VAE [11] hibrit modeline odaklanılmıştır.

Süper çözünürlük (SÇ), daha yüksek çözünürlüklü görüntülere gereksinim duyulması, hareket ve görüntüleme kaynaklarına uzaklık gibi gerçek hayatta karşılaşılabilen problemler nedeniyle ele alınan bir bilgisayarlı görü problemidir. Düşük çözünürlüklü (DÇ) görüntülerden yüksek çözünürlüklü görüntülerin elde edilmesi işlemi olarak ele alınmaktadır. GAN tabanlı çok sayıda çalışma [25-38] ile bu alanda önemli gelişmeler kaydedilmiştir.

Bu çalışmada CPPN [12] olarak bilinen bir modelden yararlanılarak literatürde yer alan süper çözünürlük modellerinden farklı olarak gerçekçi fakat sentetik yüksek çözünürlüklü görüntülerin herhangi bir eğitim yapılmaksızın oluşturulabilmesi hedeflenmiştir. CPPN modeli çalışmada

önerilen hibrit modelin üretici kısmını oluşturmaktadır. Bu çalışmada önerilen model VAE/CPGAN olarak adlandırılmıştır. Bu üretici model, daha önceki hibrit VAE-GAN modellerinde farklı olarak, orijinal görüntülerden elde edilen veya örneklenen gizli koda ek olarak koordinat girdilerine de gereksinim duymaktadır. Bu koordinat verilerine göre gizli koddan istenen boyutta görüntülerin oluşturulabilmesi önerilen model ile mümkün olmuştur. Önerilen modelin en önemli özelliği, literatürdeki diğer üretici modellerden farklı olarak ölçeklenebilir bir üretici ağ modeli olması ile süper çözünürlüklü görüntü oluşturmada yüksek çözünürlüklü görüntüler ile eğitime gerek duyulmamasıdır. Önerilen model, el yazısı karakter veri kümesi üzerinde deneyerek bu modele temel olan ağ modelleri ve diğer hibrit modeller ile karşılaştırmalara deneysel çalışmalarda yer verilmiştir. Bu çalışmalarda, üretilen gerçekçi ama sentetik görüntüler ile birlikte bir eğitim yapılmamasına rağmen farklı ölçeklerde sentetik görüntülerin oluşturulabildiği görülmüştür.

2. İLGİLİ ÇALIŞMALAR (RELATED WORKS)

Üretici çekişmeli ağlarının ortaya çıkması ile birlikte birçok GAN tabanlı model ortaya çıkmıştır. Koşullu Üretici Çekişmeli Ağ (ConGAN) [13] bir koşul tabanında sentetik yüz görüntüleri oluşturmak üzere önerilmiştir. Laplas Piramit Üretici Çekişmeli Ağ (LAPGAN) [14] modelinde, birden çok üretici ve ayrımcı ağ ile düşük çözünürlüklü, kaba bir görüntü ile eğitime başlanarak Laplas piramidi şeklindeki bir mimari ile her bir adımda daha gerçekçi, yüksek çözünürlüklü ve iyi görüntüler elde edilebilmiştir. Derin Evrimsel Üretici Çekişmeli Ağ (DCGAN) [15] GAN modelinin üretici ve ayrımcı ağında de kullanılan çok katmanlı perseptron ağları yerine bu ağların evrimsel ağlar olarak tanımlanması ile elde edilmiş bir GAN modeli olarak sunulmuştur. Geliştirilmiş GAN (Improved GAN) [42], GAN modelinin yığın normalleştirme, etiket yumuşatma ve geri yayılım gibi farklı yaklaşımlar ile performansını geliştirmek üzere sunulan bir model olmuştur.

Otokodlayıcı [16] kodlayıcı ve kod çözücü yapısı ile verileri sıkıştırarak verilerin sıkıştırılmış versiyonlarından geri elde edilebilmesi için eğitilen gözetimsiz bir ağ modelidir. Üretici ağ modellerinin ortaya çıkması ile birlikte otokodlayıcı modelleri yeniden gündeme gelmiştir. Otokodlayıcı modellerinde karşılaşılan gradyan düşüşü ve aşırı öğrenme problemini çözmek üzere bir varyasyon bileşeni kullanılarak KL (Kullback Leiber) -ıraksama metriği tabanında model eğitilmesine dayanan VAE modeli ortaya çıkmıştır. Çekişmeli eğitim yaklaşımındaki gelişmeler ile birlikte Çekişmeli Otokodlayıcı (AAE) [17] modeli geliştirilmiştir. Modelde VAE modelindeki KL-ıraksama terimi yerine GAN objektifi kullanılarak çekişmeli eğitime dayalı bir otokodlayıcı modeli önerilmiştir.

GAN modellerinde olduğu gibi rassal bir vektör şeklindeki veri yerine sıkıştırılmış bir veriden görüntü oluşturabilmek üzere VAE ve GAN modellerinin bir arada kullanıldığı

birçok hibrit model ortaya çıkmıştır. İlk çalışmalardan biri olarak, VAE ve GAN modelleri VAE/GAN [18] olarak adlandırılan bir modelde birleştirilmiştir. Bu modelin ayrımcı modeli ile çıkarılan özelliklere göre elde edilen benzerlik, otokodlayıcı yeniden yapılandırma objektif fonksiyonu yerine kullanılmıştır. VAE/GAN'ın kayıp fonksiyonu, VAE önceki kayıp fonksiyonu, öznelik benzerliği ve üretici ayrımcı kayıp fonksiyonunun bir kombinasyonu şekilde tanımlanarak model eğitilmiştir. Başka bir VAE-GAN hibrit çalışmada da [19], önceden eğitilmiş yardımcı bir ağ ile elde edilen kayıp fonksiyonu yeniden yapılandırma hatası yerine kullanılmıştır. Benzer şekilde, elde edilen görüntülerdeki bulanıklık problemini çözmek üzere [20] numaralı çalışmada tanımlanmış DeePSiM kayıp fonksiyonunu tabanında GAN modeli önerilmiştir. Çekişmeli Öğrenilmiş Çıkarım (ALI) [21] modelinde, AAE modelinden farklı olarak bir çekişmeli otokodlayıcı modeliyle GAN bir arada kullanılmıştır.

GAN modellerinin eğitiminin tutarsız olması sebebi ile bu modellerin eğitilebilmesi üzerine farklı yaklaşımlara dayalı modeller üzerine çalışmalar dikkat çekmiştir. Wasserstein Çekişmeli Üretici Ağ (WGAN) [22], GAN modelinin eğitiminin daha kararlı hale getirilebilmesi için Earth-Mover (EM) uzaklığı ya da Wasserstein metriği olarak ifade edilen metriğe dayalı olarak modelin eğitilmesi ile elde edilen ağ modelidir. DRAGAN modeli [23], WGAN modelinde olduğu gibi GAN modellerinin eğitimi kararlı hale getirerek yakınsama hızını artırmak üzere önerilen bir eğitim yaklaşımıdır.

Üretici modellerin düşük boyutlarda görüntü oluşturma ve yeniden yapılandırmadaki etkinliklerine ek olarak bu modellere ölçeklenebilirlik sağlayabilmek üzere farklı yaklaşımlar incelenmiştir. CPPN [12] yeni bir kodlayıcı yapısı ile verideki yapısal ilişkileri evrimsel olarak fonksiyonların bir kompozisyonu şeklinde tanımlayan ağ modelidir. NEAT [24] olarak adlandırılan yöntem ile evrimsel olarak daha karmaşık bir ağ yapısı oluşturulur. Basit motifler şeklindeki görüntüler her bir iterasyonda daha kompleks motiflere dönüştürülerek model ile daha karmaşık soyut görüntüler oluşturulabilmiştir. Bu yapının ölçeklenebilirliğinden ilham alınarak, CPPN modelinin, GAN'ın üretken ağı olarak kullanılabilceği bu çalışmada değerlendirilmiştir. CPPN ağının nokta odaklı olarak görüntü oluşturma kabiliyeti sayesinde, yüksek boyutlu görüntüler elde edilebileceğinden CPPN-GAN-VAE olarak bir araya getirilen modelin performansını daha da arttırmaya odaklanılmıştır. Süper çözünürlüklü görüntü oluşturma problemi için literatür incelendiğinde, son yıllarda CNN ve üretici modellere dayalı çalışmaların ağırlık kazandığı görülmektedir. Bu alandaki ilk çalışmalardan biri olan SRCNN [25] modelinde evrişim işleminden oluşan az sayıda katmanla düşük çözünürlüklü görüntülerden süper çözünürlüklü görüntülere dönüşüm gerçekleştirilmiştir. Modelin hız problemini çözebilmek üzere FSRCNN [26] varyasyonu sunulmuştur. Benzer şekilde 16 tekrarlı evrimsel katmandan oluşan DRCN [27] modeli; artık ağ, atlanan bağlantı ve ağ içinde ağ yapılarının kullanıldığı

DCSCN [28]; LAPGAN modeline benzer olarak piramit yapısındaki LAPSNN [29] ve artık öğrenmeye dayanan EDSR [30] gibi yaklaşımlarda literatürde oldukça ses getirmiştir. Son zamanlarda, üretici ağ yaklaşımlarının süper çözünürlük problemi için kullanılmasına yönelik çalışmaların ağırlık kazandığı görülmüştür. Johnson ve arkadaşları [31], VGG [32] modelinden elde edilen yüksek seviyeli öznelikler ve çekişmeli eğitime dayalı bir model ile SRCNN gibi modellerdeki piksel seviyesindeki Öklid metriği tabanında maliyet fonksiyonuna odaklanmışlardır. VDSR [33], VGG benzeri 20 katmanlı bir ağ modeli olarak sunulmuştur. Modelin son katmanında elde edilen çıktıya eklenen girdi olarak alınan düşük çözünürlüklü görüntüler ile SRCNN modelinden daha iyi bir performans elde edilebilmiştir. SRGAN [34] modeli, önceden eğitilen ResNet [35] olarak literatürde bilinen model ağırlıkları ile elde edilen süper çözünürlük görüntülerinin VGG yüksek seviyeli öznelikleri ve çekişmeli ağ maliyet fonksiyonuna göre iyileştirilmesine dayalı bir süper çözünürlük modelidir. Benzer şekilde ENet-PAT [36] modelinde, çekişmeli eğitime ek olarak algısal ve doku eşleme maliyet fonksiyonları kullanılmıştır. Daha önce sunduğumuz FSRCNN [37] modelinde, FSRCNN modeli ile çekişmeli eğitim yaklaşımını bir araya getirerek görüntülerin çözünürlükleri iyileştirilmeye çalışılmıştır. ProGANSR [38], asimetrik piramit mimarisi olarak nitelendirilen evrişim, yoğunluk sıkılaştırma birimi ve ayrı-piksel evrişim katmanından oluşan progresif bir GAN modeli olarak sunulmuştur. ESRGAN [39] modelinde, artık blok için artık bağlantıya dayalı bir yaklaşım ve gerçekçi ayrımcı ağ modeli ile SRGAN modelinin performansını iyileştirmek hedeflenmiştir.

GAN modelleri ile birlikte son zamanlarda artık bloklara dayalı çalışmaların ağırlık kazandığı görülmüştür. RCAN [40] modeli, artık bloklara dayalı CNN modeli olarak sunulmuştur. SRFBN [41] modeli ise ileri-beslemeli blok tabanlı bir CNN modeli olarak süper çözünürlük problemi için önerilen son yıllardaki çalışmalardan biri olmuştur. SAN [42], LSRAG olarak adlandırılan modül yapısı ile artık bağlantıya dayalı bir CNN modeli olarak sunulmuş son çalışmalardan biridir.

Sunulan bu çalışmada, bu bölümde incelenen çalışmalardan farklı olarak, ölçeklenebilir bir üretici ağ modeli geliştirilerek istenen boyutta görüntülerin elde edilebilmesine odaklanılmıştır. Böylece modelimiz ile süper çözünürlüklü görüntülerin eğitimde kullanılmaksızın elde edilebilmesi hedeflenmektedir.

3. METOTLAR (METHODS)

3.1. Temel Modeller (Background Models)

3.1.1. Üretici çekişmeli ağ (Generative adversarial network)

GAN, üretici ağ (G) ve ayrımcı ağ (D) olmak üzere iki ağdan oluşan çekişmeli eğitime dayalı ağ modelidir. Üretici ve ayrımcı ağlar çok katmanlı perseptron ağından oluşmaktadır.

Üretici ağ, z vektörü olarak adlandırılan p_z olarak bilinen normal dağılımdan örneklenen gürültü vektörünü, p_g dağılımından gelen \hat{x} görüntüsüne dönüştürmektedir. Ayrımcı ağ ise, p_{data} dağılımından gelen x görüntüsü ile üretici ağdan elde edilen \hat{x} görüntülerini girdi olarak 0 ya da 1 olarak etiketlemeye çalışmaktadır. Ayrımcı ağ, sentetik örneklerin 0, gerçekçi örneklerin ise 1 olarak etiketlenerek ayırt edilebilmesi üzerine eğitilmektedir. Model ile hedeflenen, üretici ağı ile elde edilen görüntülerin ayrımcı ağ tarafından gerçek olarak etiketlenmesi iken ayrımcı modelin sentetik görüntüleri de ayırt edebilecek kadar hassas olmasıdır. Ayrımcı ağ objektif fonksiyonu, x verileri için $D(\hat{x}; \theta_d)$ değerinin maksimize etmek ile birlikte \hat{x} için $D(\hat{x}; \theta_d) (= D(G(z; \theta_g); \theta_d))$ 'yi minimize etmektir. Bu durum Eş. 1'de ifade edilmiştir.

$$L_d^{GAN} = \max_{\theta_g, \theta_d} (\mathbb{E}_{x \sim p_{data}} [\log (D(x; \theta_d))] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z; \theta_g); \theta_d))] \quad (1)$$

Üretici ağ için hedeflenen $D(G(z; \theta_g); \theta_d)$ değerini maksimize etmek, Eş. 2'de de gösterildiği gibi bu durum, $1 - D(G(z; \theta_g); \theta_d)$ değerini minimize etmektir.

$$L_g^{GAN} = \min_{\theta_g, \theta_d} (\mathbb{E}_{z \sim p_z} [\log (1 - D(G(z; \theta_g); \theta_d))] \quad (2)$$

GAN objektif fonksiyonu Eş. 3'te bir araya getirilmiştir.

$$L_{GAN} = \min_{\theta_g, \theta_d} \max (\mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z; \theta_g); \theta_d))] + \mathbb{E}_{x \sim p_{data}} [\log (D(x; \theta_d))] \quad (3)$$

GAN ile hedeflenen optimum durum $p_g = p_{data}$ durumudur.

3.1.2. Varyasyonel Otokodlayıcı (Variational Autoencoder)

Gözetimsiz bir ağ modeli olan VAE, klasik otokodlayıcı modelinde olduğu gibi kodlayıcı (E) ve kod çözücü (G) ağlarından oluşan bir ağ modelidir. Klasik otokodlayıcı modelinde karşılaşılan gradyan düşüşü ve aşırı öğrenme problemlerini çözmek üzere otokodlayıcıya bir varyasyon terimi eklenmesi ile elde edilmiştir. Klasik AE modelinden farklı olarak kodlayıcı ağı, x görüntülerini girdi olarak alarak sıkıştırılmış z_μ ortalama ve z_σ standart sapma vektörlerine daha sonra ise bu vektörlerden z gizli kodu örneklenmektedir.

$$z_\mu, z_\sigma = E(x; \theta_e), z = z_\mu + \epsilon z_\sigma = q(z|x) \quad (4)$$

Eş. 4 'de verildiği gibi, gizli kod z , verilerinin gerçek dağılımından, $q(z|x)$, elde edilir. Kod çözücü, klasik otokodlayıcıda olduğu gibi, gizli kodu girdi olarak verilen veriye dönüştürmek üzere kullanılmaktadır. Eş. 5'te olduğu gibi \hat{x} ile ifade edilen yeniden yapılandırma görüntüleri $p(x|z)$ ile ifade edilen dağılımdan elde edilmektedir.

$$\hat{x} = G(z; \theta_d) = p(x|z) \quad (5)$$

L_{pixel} piksel tabanlı yeniden yapılanma maliyet fonksiyonu Eş. 6'da verildiği şekilde hesaplanmaktadır.

$$L_{pixel} = -\mathbb{E}_{q(z|x)} [\log p(x|z)] \quad (6)$$

Ayrıca, L_{prior} olarak Eş. 7'de ifade edilen KL-ıtraksama maliyeti model dağılımı ile veri dağılımını örtüştürmek üzere kullanılmaktadır.

$$L_{prior} = \mathbb{D}_{KL}(q(z|x)||p(z)) = \int_{-\infty}^{\infty} q(z|x) \log(q(z))|p(z) dz \quad (7)$$

VAE objektif fonksiyonu, Eş. 8'de verildiği gibi L_{pixel} ve L_{prior} maliyet fonksiyonlarının toplamı şeklinde ifade edilebilmektedir.

$$L_{VAE} = L_{pixel} + L_{prior} \quad (8)$$

3.1.3. Kompozisyonel örüntü üreten ağlar (Compositional pattern producing network)

CPPN modeli, girdi olarak alınan verileri hedeflenen çıkış değerlerine eşlemek için kullanılan fonksiyonlar bileşimi şeklindeki bir ağ modelidir. Geleneksel yapay sinir ağlarından farklı olarak, sigmoid ve Gauss gibi aktivasyon fonksiyonlarından farklı fonksiyonlarının bir araya gelmesiyle oluşmaktadır. Daha önceki çalışmalarda, (x, y) koordinat sisteminden iki boyutlu ikili görüntülerin NEAT gibi bir evrişim algoritması ile evrişimsel olarak öğrenilen $f(w, x, y, d)$ fonksiyonu ile üretilen görüntü [12, 43]. Eş. 9'da bu model ile öğrenilebilecek bir kompozisyonel ağ örneği temsili olarak sunulmuştur.

$$f(w, x, y, d) = (f_1(x, y), f_5(f_2(x, y), f_4(f_2(x, y), f_3(x, y)))) \quad (9)$$

3.1.4. Önerilen model: VAE/CPGAN (Proposed model: VAE/CPGAN)

Bu çalışmada, son yıllardaki çalışmalarda da tercih edildiği gibi VAE ve GAN modelleri bir arada kullanılmıştır. Çalışma kapsamında, daha önceki hibrit modellerden farklı olarak, GAN modelindeki üretici ağ modeli yerine CPPN ağına benzer bir yaklaşımla piksel tabanlı yeniden yapılandırma hedeflenmiştir. Böylece, istenen çözünürlükte görüntülerin o çözünürlükte herhangi bir eğitim yapılmaksızın yeniden yapılandırma ile elde edilmesi mümkün olmuştur. VAE modelinin çıkarsama mekanizmasından yararlanmak üzere görüntülerin sıkıştırılabilmesi için kod çözücü olarak VAE kod çözücü modelinden yararlanılarak, VAE modelinde kullanılan yakınsama objektifi model objektiflerine eklenmiştir. Bu çalışma ile CPPN yaklaşımı çekişmeli eğitim yaklaşımı ile bir araya getirilerek CPGAN olarak adlandırılan yeni bir ağ modeli elde edilmiştir. Çekişmeli eğitim yaklaşımı ile daha gerçekçi görüntüler elde edilerek daha gülbüz bir model elde edilmesi planlanmıştır. Otokodlayıcı ile bir araya getirilen bu yeni ağ modeli VAE/CPGAN modeli olarak adlandırılmıştır.

Kod çözücü olarak, VAE kod çözücü modeli yerine CPPN benzeri bir yaklaşım kullanılmıştır. Böylece VAE çıkarsama mekanizması ile CPPN ölçeklenebilirliğinden yararlanılmıştır.

VAE/CPGAN model mimarisine Şekil 1'de yer verilmiştir. Şekilde görüldüğü gibi model, kodlayıcı (E), üretici ağ (G) ve ayırıcı ağdan (D) oluşmaktadır. Kodlayıcı ağ (E), VAE modelinde olduğu gibi x görüntüsünü alarak z_μ ve z_σ ortalama ve standart sapma vektörlerine, bu vektörlerin doğrusal kombinasyonu aracılığıyla ise z gizli kod vektörüne dönüştürmektedir. Modelin ikinci kısmı olan üretici ağ modelini (G) CPPN benzeri bir üretici ağ oluşturmaktadır. Klasik üretici/kod çözücü ağlardan farklı olarak z veya $z_p \in \mathbb{N}(0, 1)$ vektörü ile koordinat sistemi vektörleri girdi olarak alınmaktadır. Koordinat matrisleri, s ölçek değeri, m ise x kare matrisinin olarak genişlik ve yükseklik değeri olmak üzere, $s * (m \times m)$ boyutludur. Model ile x eksenini için x_{coord} , y eksenini için y_{coord} ve yarıçap yani merkezden uzaklık vektörü olarak r girdi olarak alınarak gizli kod vektörü olan z veya z_p ile füzyon işlemi yapabilmek üzere x_{vec} , y_{vec} ve r_{vec} vektörleri haline getirilir. Üretici ağ z veya z_p , x_{vec} , y_{vec} ve r_{vec} vektörlerini \hat{x} veya x_p görüntülerine dönüştürmektedir. Modelin son bölümünü, GAN modellerinde olduğu gibi, ayırıcı ağ modeli oluşturmaktadır. Ayırıcı ağ (D), gerçek görüntüler (x) ile üretici ağdan elde edilen x_p ve \hat{x} görüntülerini girdi olarak 0 ya da 1 şeklinde sınıflandırmaya çalışarak, gerçek görüntüleri sentetik görüntülerden ayırt edebilmek üzere kullanılmaktadır. Ayrıca, ağın üst katmanlarından (l) elde edilen yüksek seviyeli gösterimler, $f_l(x)$, kullanılarak gerçek verilerden elde edilen öznitelikler ile sentetik verilerden elde edilen öznitelikler tabanında objektif fonksiyonu güncellenmiştir. Elde edilen öznitelikler tabanında karşılaştırma yapabilmek üzere kullanılan L_{fea} objektifi Eş. 10'da verilmiştir.

$$L_{fea} = -\mathbb{E}_{q(z|x)} [\log p(f_l(x)|z)] \quad (10)$$

VAE modellerinde olduğu gibi elde edilen dağılımı yakınsayabilmek üzere L_{prior} , KL-ıraksama maliyet fonksiyonundan (Eş. 7) yararlanılmıştır. Çekişmeli eğitim objektif fonksiyonu ise x_p ve x için girdileri model

eğitildiğinden Eş. 11'de gösterildiği şekilde genişletilmiştir.

$$L_{CPGAN} = \min_{\theta_g, \theta_d} \max_{z \sim p_z} (\mathbb{E}_{z \sim p_z} [\log (1 - D(G(z_p, x_{vec}, y_{vec}, r_{vec}; \theta_g); \theta_d))] - [\log (1 - D(G(z_p, x_{vec}, y_{vec}, r_{vec}; \theta_g); \theta_d))] + \mathbb{E}_{x \sim p_{data}} [\log (D(x; \theta_d))] \quad (11)$$

Bu çalışmada önerilen VAE/CPGAN modelinin eğitim algoritması Şekil 2'de verilmektedir.

```

1:  $\theta_E, \theta_G, \theta_D \leftarrow$  parametrelerin ilk değer atamaları
2:  $k \leftarrow$  yığın boyutu
3:  $m \leftarrow$  örnek sayısı
4: while  $\theta_E, \theta_G, \theta_D$  yakınsanmamış do
5:    $x \leftarrow x^{(i)}$ ,  $i = 1, \dots, m$ 
6:    $z_\mu, z_\sigma \leftarrow E(x; \theta_E)$ 
7:    $z \leftarrow z_\mu + \epsilon z_\sigma$ ,  $\epsilon \in \mathcal{N}(0, 1)$ 
8:    $L_{prior} \leftarrow \mathbb{D}_{KL}(q(z|x)||p(z))$ 
9:    $x_{vec}, y_{vec}, r_{vec} \leftarrow m \times m$  grid
10:   $\hat{x} \leftarrow G(z, x_{vec}, y_{vec}, r_{vec}; \theta_G)$ 
11:   $z_p \leftarrow \mathbb{N}(0, 1)$ 
12:   $\hat{x}_p \leftarrow G(z_p, x_{vec}, y_{vec}, r_{vec}; \theta_G)$ 
13:   $L_{fea} \leftarrow -\mathbb{E}_{q(z|x)} [\log p(f_l(x)|z)]$ 
14:   $L_{CPGAN} \leftarrow \mathbb{E}_{x \sim p_x} [\log D(x; \theta_D)]$ 
     $+ \mathbb{E}_{x \sim p_{\hat{x}}} [\log (1 - D(\hat{x}; \theta_D))]$ 
     $+ \mathbb{E}_{x_p \sim p_{x_p}} [\log (1 - D(x_p; \theta_D))]$ 
15:   $\theta_E \leftarrow \theta_E - \nabla_{\theta_E} (L_{prior} + L_{fea})$ 
16:   $\theta_G \leftarrow \theta_G - \nabla_{\theta_G} (L_{fea} - L_{CPGAN})$ 
17:   $\theta_D \leftarrow \theta_D - \nabla_{\theta_D} L_{CPGAN}$ 
18: end while

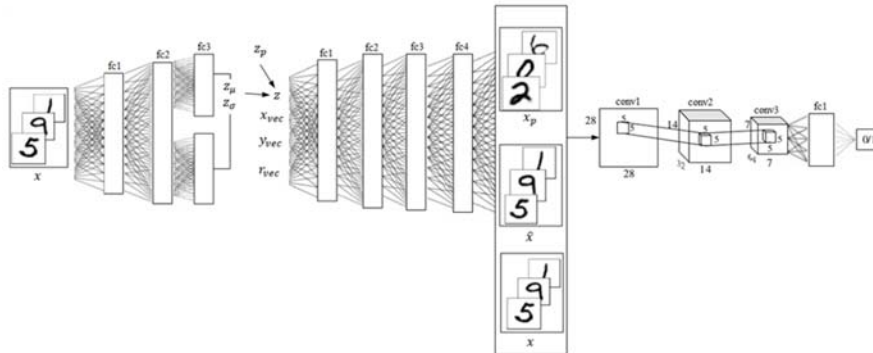
```

Şekil 2. VAE/CPGAN model eğitim algoritması (The VAE/CPGAN training algorithm)

4. DENEYSEL ÇALIŞMALAR (EXPERIMENTAL STUDIES)

4.1. Veri Kümesi: MNIST (Dataset: MNIST)

Görüntü oluşturma ve yeniden yapılandırma için literatürde yer alan diğer otokodlayıcı ve GAN modellerinde de kullanılan MNIST veri kümesi [44], optik karakter tanıma gibi problemler ile diğer makine öğrenmesi/bilgisayarlı görü problemleri için kullanılan bir veri kümesidir. 0 ile 9 arasındaki rakamların el yazısı şeklinde yazılması ile elde



Şekil 1. VAE/CPGAN ağ mimarisi (The VAE/CPGAN model architecture)

edilmiştir. Veri kümesi, el yazısı şeklindeki rakamların 28×28 boyutlarındaki 80000 adet ikili görüntülerinden oluşmaktadır. MNIST veri kümesinde, 60000 görüntü eğitim verisini, 10000 görüntü geçerleme verisini kalan 10000 görüntü ise test verisini oluşturmak üzere veri kümesi eğitim, geçerleme ve test olarak ayrılmıştır.

4.2. Model Mimarisi ve Ağ Parametreleri (Model Architecture and Network Parameters)

Bu çalışmada sunulan VAE/CPGAN modeli mimarisi 3 ağ modelinin bir araya gelmesinden oluşmaktadır. Modelin kodlayıcı ile üretici ağ kısmı sırasıyla 3 katmandan ve 5 katmandan oluşan çok katmanlı perseptron ağı olarak tanımlanmıştır. Kodlayıcı ağ modelinde son katman dışında yığın normalleştirme işlemi yapılarak ReLU aktivasyon fonksiyonu kullanılmıştır. Üretici ağ modelinde, benzer olarak, son katman dışında yığın normalleştirme ve ReLU aktivasyon fonksiyonu, son katmanda ise yığın normalleştirme yapılmaksızın sigmoid aktivasyon fonksiyonundan yararlanılmıştır. Ayrımcı ağ modeli ise 3 evrimsel katman (conv), 1 tam-bağlı katmandan (fc) oluşan bir evrimsel sinir ağıdır. İlk katman ve son katman dışında, evrimsel katman ve tam-bağlı katmandan sonra yığın normalleştirme işlemi yapılmıştır. Aktivasyon fonksiyonu olarak ilk 4 katmanda Leaky ReLU [45] son katmanda ise Sigmoid fonksiyonu tercih edilmiştir.

Deney çalışmalarında yığın boyutu 100, gizli kod vektör boyutu da 100 olarak kullanılmıştır. Modeli eğitmek üzere Adam en iyileme [46], $\beta_1 = 0.65$, $\beta_2 = 0.999$, $\epsilon = 1e - 08$, öğrenme oranı (α) = $1e - 03$ parametreleri ile kullanılmıştır. Deney çalışmalarındaki modellerin kısa bir eğitim süresindeki performanslarını karşılaştırabilmek üzere tüm modeller 10 epok (6000 iterasyon) boyunca eğitilmiştir. Model parametreleri ise Tablo 1'de detaylı olarak verilmiştir.

4.3. Deney Sonuçları (Experiment Results)

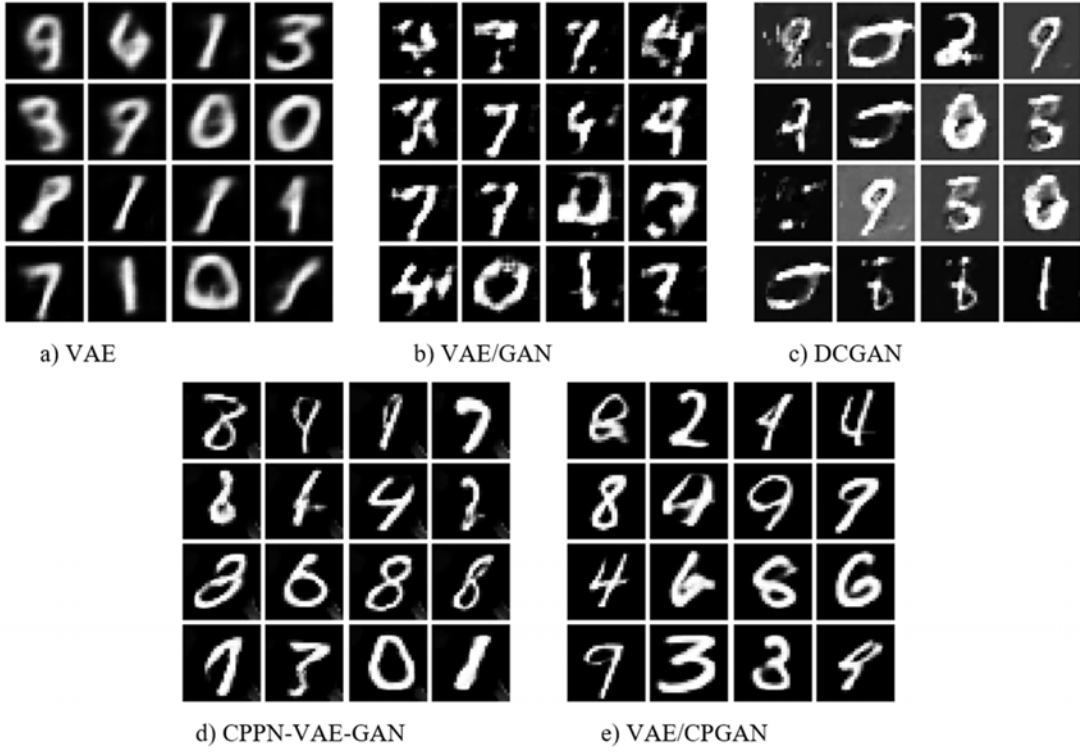
Önerilen model performansını değerlendirmek üzere modele temel olan VAE, VAE/GAN, DCGAN, CPPN-GAN-VAE modelleri ve GAN modellerini iyileştirmek üzere sunulan popüler yöntemler olan WGAN, DRAGAN ile de karşılaştırmalar yapılmıştır. Bu modeller için aynı derinlikteki ağlar aynı parametreler ile tanımlanmıştır. Modeller, hem niceliksel hem de niteliksel olarak karşılaştırılmıştır. VAE, VAE/GAN, DCGAN, CPPN-GAN-VAE ve VAE/CPGAN ile elde edilen örnek görüntüler Şekil 3'te gösterilmiştir.

Şekilde verilen görüntülere göre VAE/CPGAN modeli ile CPPN benzeri üretici ağ modeli kullanılmayan modellere göre daha gerçekçi ve net görüntüler oluşturulabildiği görülmektedir. Çalışmada, karşılaştırmak üzere kullanılan öznelik tabanlı objektif yerine, klasik yeniden yapılandırmaya dayalı diğer CPPN benzeri modelin örneklerine göre ise daha çeşitli ve farklı yazım stillerinde örneklerin, gürültü olmaksızın, oluşturulabildiği görülmektedir.

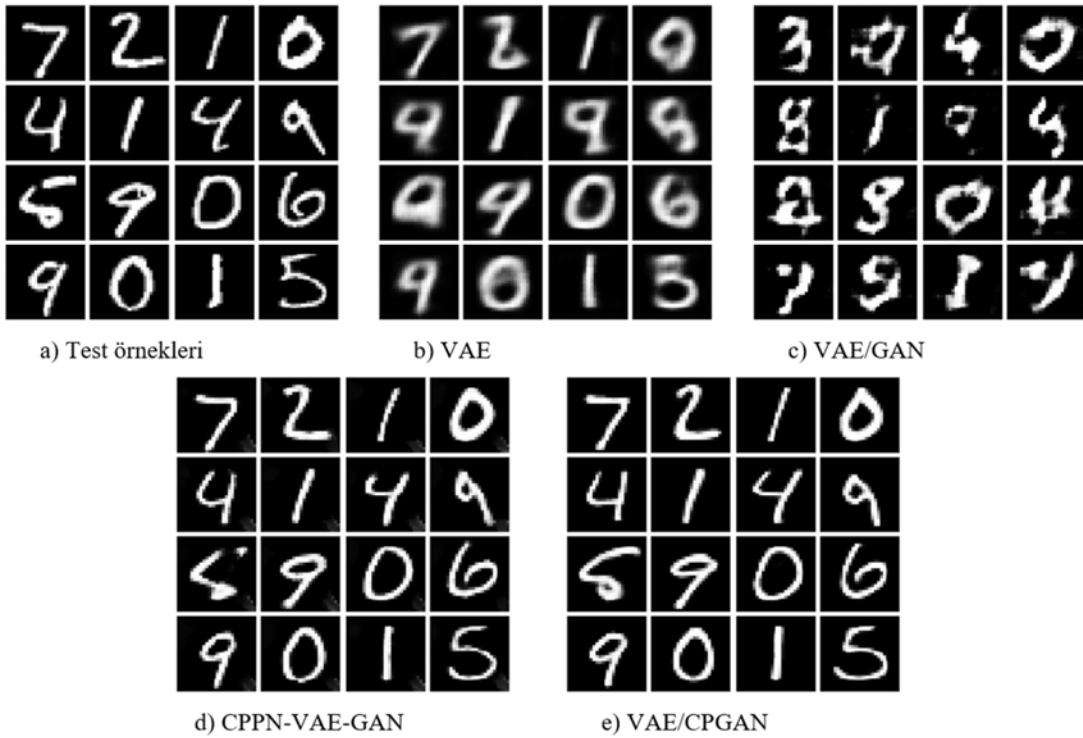
Otokodlayıcı tabanlı modeller kullanılarak yeniden yapılandırma ile elde edilen görüntülere Şekil 4'te yer verilmiştir. Şekildeki görüntüler, VAE/CPGAN modelinin diğer modellerden üstünlüğünü göstermektedir. Karşılaştırmak üzere farklı objektife dayalı sunulan CPPN benzeri hibrit modelin diğer modellere göre performanslı gözükmekte, fakat VAE/CPGAN modeli ile karşılaştırıldığında elde edilen görüntülerde bazı gürültülerin olduğu görülmektedir. VAE modelinin, beklendiği şekilde, diğer modellerden daha bulanık görüntüler oluşturduğu görülmektedir. VAE/GAN modeli ise bu eğitim koşullarına ve parametrelere göre optimize edilemediğinden beklenen performansı gösterememiş, eksik ve anlamsız görüntüler elde edilmesine neden olmuştur.

Tablo 1. VAE/CPGAN model parametreleri (VAE/CPGAN model parameters)

Kodlayıcı ağ (E)	fc1	fc2	fc3	fc4	fc5
Filtre sayısı	512	512	100	-	-
Veri normalleştirme	Yığın normalleştirme	Yığın normalleştirme	-	-	-
Aktivasyon fonksiyonu	ReLU	ReLU	-	-	-
Üretici Ağ (D)	fc1	fc2	fc3	fc4	fc5
Filtre sayısı	128	128	128	128	128
Veri normalleştirme	Yığın normalleştirme	Yığın normalleştirme	Yığın normalleştirme	Yığın normalleştirme	-
Aktivasyon fonksiyonu	ReLU	ReLU	ReLU	ReLU	Sigmoid
Ayrımcı ağ (D)	conv1	conv2	conv3	fc1	fc2
Filtre sayısı	32	64	128	128	1
Filtre boyutu	5×5	5×5	5×5	-	-
Stride	2	2	2	-	-
Padding	0	0	0	-	-
Veri normalleştirme	-	Yığın normalleştirme	Yığın normalleştirme	Yığın normalleştirme	-
Aktivasyon fonksiyonu	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU	Sigmoid



Şekil 3. Modeller ile elde edilen sentetik görüntüler (Generated synthetic images)



Şekil 4. Otokodlayıcı tabanlı modeller ile elde edilen yeniden yapılandırma görüntüleri (Reconstructed images by autoencoder based models)

Niceliksel olarak modelleri karşılaştırmak üzere literatürde tercih edilen inception skoru (IS) [47] kullanılmıştır. Ayrıca otokodlayıcı modellerinde yeniden yapılandırma performansı MSE tabanında değerlendirilmiştir. Bu skorlar dikkate alınarak yapılan karşılaştırmalar göre, Tablo 2’de gösterildiği gibi, en yüksek sonuçlar yine VAE/CPGAN modeli ile elde edilmiştir. CPPN-VAE-GAN modeli, inception skoru ve MSE tabanında performans olarak önerilen VAE/CPGAN modelini takip eden model olmuştur. Bu durum, CPPN tabanlı modelin üretici ağ olarak etkisini gözler önüne sermektedir. Inception skoruna göre GAN tabanlı modellere bakıldığında ise önerilen modele yakın sonuçlar elde edildiği gözlemlenmiştir.

Tablo 2. MSE ve IS tabanında modellerin niceliksel analiz karşılaştırmaları

(Quantitative analysis of models based on MSE and IS scores)

Metot	Türü	MSE	IS	
			Ortalama	Standart sapma
VAE	Tekil	3,6940	1,5830	0,0153
DCGAN	Tekil	*	1,6141	0,0335
WGAN	Tekil	*	1.0060	0.0004
DRAGAN	Tekil	*	1.0500	0.0016
VAE/GAN	Hibrit	11,1919	1,3587	0,0249
CPPN-VAE-GAN	Hibrit	2,0118	1,6266	0,0518
VAE/CPGAN	Hibrit	1,7512	1,8538	0,0337

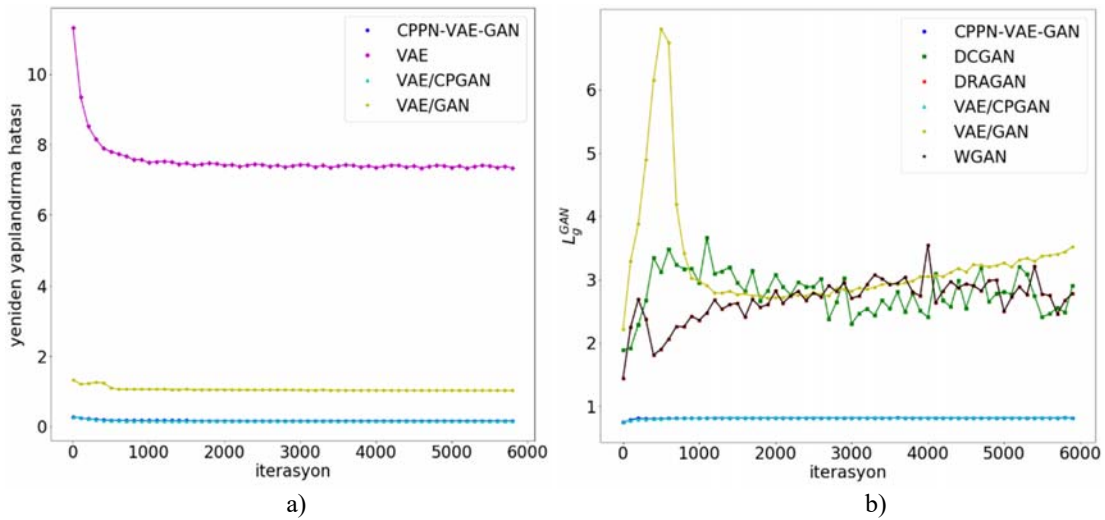
* ile gösterilen modeller ile yeniden yapılandırma görüntüleri elde edilemediğinden MSE skorları hesaplanamamıştır.

Şekil 5’te 6000 iterasyon süresince kaydedilen yeniden yapılandırma ve üretici ağ hata eğrileri verilmiştir. Şekil 5a’da verilen eğrilere bakıldığında VAE ve VAE/GAN modellerinin CPPN tabanlı modeller gibi yakınsanamadığı görülmektedir. GAN tabanlı modellerde sık karşılaşılan bir

problem olarak, literatürde ele alınan, eğitimde kararlı olamama durumunu analiz edebilmek üzere ise Şekil 5b’de GAN tabanlı modellerin üretici ağ hata grafiği ele alınmıştır. GAN modellerinin eğitiminin kararlı olamamasına çözüm getirmek üzere daha önce sunulan DRAGAN ve WGAN modelleri ile de karşılaştırmalara yer verilmiştir. Şekil 5b’ye göre VAE/GAN modelinin yakınsanamadığı ayrıca DCGAN, DRAGAN ve WGAN gibi modellerin eğitimde kararlı performans sergileyemediği görülmektedir. Bu nedenle, modelimizin diğer modellere göre daha kararlı olmakla birlikte test sonuçlarına bakıldığında da daha performanslı olduğu görülmektedir. Genel olarak, çekişmeli eğitime dayalı modellerde görüldüğü üzere, çekişmeli eğitim ile daha gülbüz ve performanslı modeller elde edilmesinin sağlandığı, buna rağmen karşılaştırılan diğer GAN modellerinin eğitiminin önerdiğimiz VAE/CPGAN modeline göre daha az kararlı performans sergilediği değerlendirilmiştir.

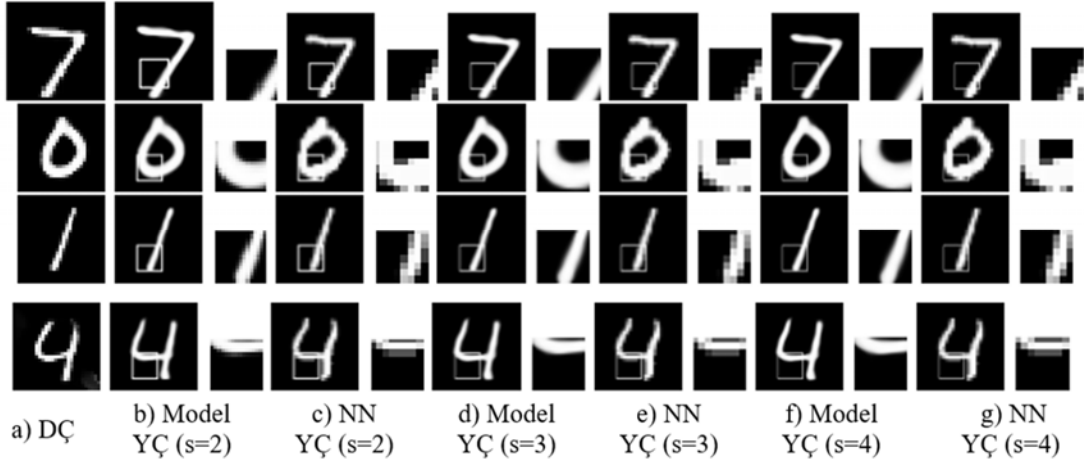
CPPN tabanlı modellerin ölçeklenebilirlik özelliği sayesinde, VAE/CPGAN modeli süper çözünürlüklü görüntüler oluşturma yeteneğine de sahip olmuştur. Süper çözünürlük görüntüleri, verilen ölçekler için bir eğitim yapılmadan elde edildiğinden literatürde bulunan süper çözünürlük modelleri yerine NN interpolasyon ile Şekil 6’da karşılaştırılmıştır. Elde edilen örnekler incelendiğinde, önerilen modelin farklı ölçekler için herhangi bir eğitim olmadan süper çözünürlük görüntüleri üretebildiği görülmektedir.

Önerilen model ve karşılaştırmak üzere oluşturulan CPPN tabanlı model ile ölçek 2, 3, 4, 5 ve 6 için elde edilen YÇ görüntüleri Şekil 7’de karşılaştırılmıştır. Bu görüntülere göre, ölçek = 6 gibi büyük ölçeklerde dahi başarılı sonuçlar elde edildiği görülmektedir. Şekilde verilen görüntülere bakıldığında, diğer model ile oluşturulan görüntülerin bazı gürültüler içerdiği görülmektedir.

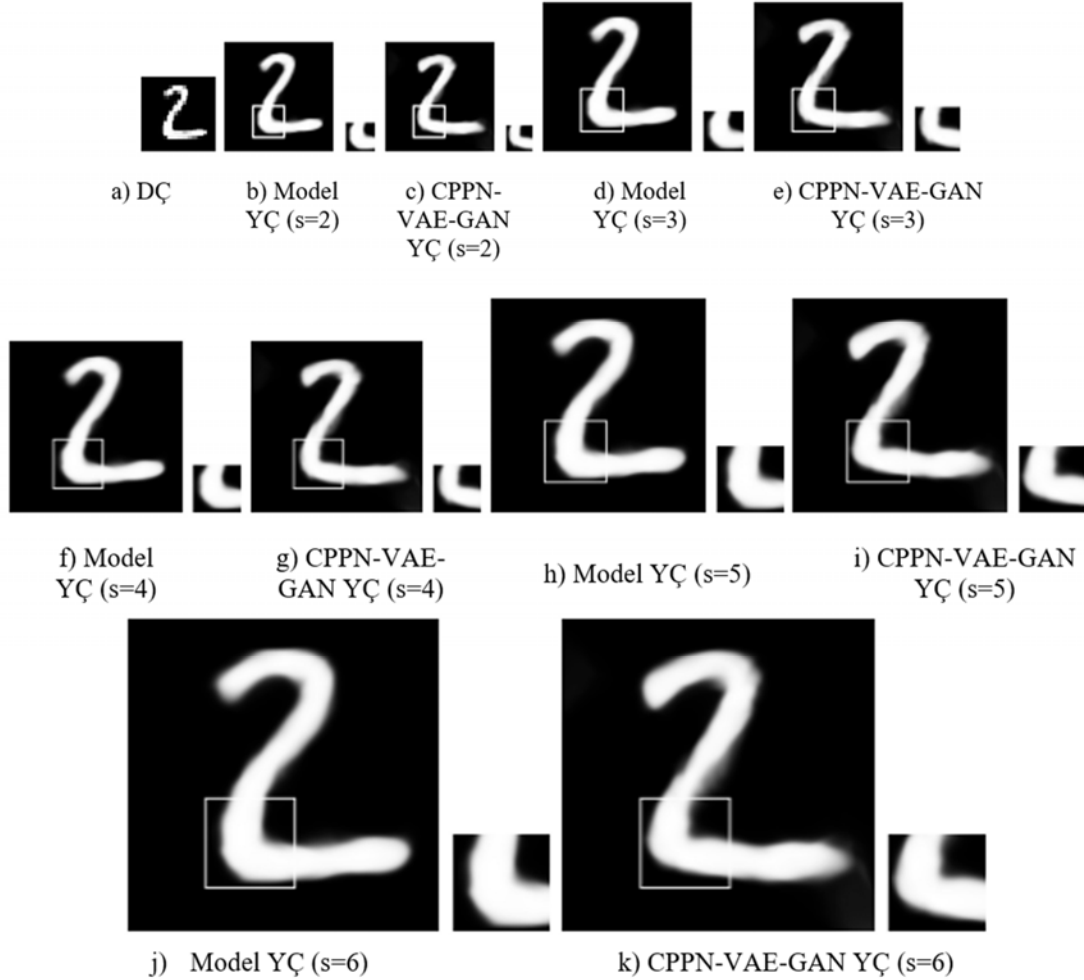


Şekil 5. a) VAE tabanlı modellerin iterasyon sayısına göre yeniden yapılandırma hata eğrileri b) GAN tabanlı modellerin iterasyon sayısına göre üretici ağ hata eğrileri

((a) The reconstruction loss over iterations curves of VAE based models (b) The generative loss over iterations curves of GAN based models)



Şekil 6. Düşük çözünürlüklü (DÇ) orijinal görüntüler ile model ve NN interpolasyonu ile elde edilen süper çözünürlüklü (YÇ) görüntü karşılaştırmaları (The comparisons of low resolution (LR) and super resolution (SR) images for different scales)



Şekil 7. Düşük çözünürlüklü orijinal görüntüler ile CPPN tabanlı modeller ile elde edilen süper çözünürlüklü görüntülerin karşılaştırmaları (The comparison of CPPN based models SR images at different scales)

Süper çözünürlük analizi için elimizde MNIST veri kümesinin süper çözünürlüklü yer gerçekliği verileri olmadığı için elde edilen YÇ görüntülerinin performansını

değerlendirmek üzere farklı metrikler incelenmiştir. Bu nedenle, bulanıklık analizi yapabilmek üzere laplaların varyans değeri tabanında netlik skorları ile eşik değeri 100

için netlik oranları hesaplanmıştır. Elde edilen sonuçlar Tablo 3'te verilmiştir. Tabloda verilen sonuçlara bakıldığında ölçek sayısı arttıkça beklenildiği şekilde görüntünün netlik değeri düşmektedir. Bu nedenle, farklı ölçeklerde elde edilen netlik skorları arasında önemli farklar her iki model için de söz konusu olmaktadır. Ölçek 2 ve 3 için netlik oranları %100 olarak hesaplanmış iken ölçek 4'te netlik oranının daha düşük kaydedildiği görülmektedir. Tabloya göre, piksel tabanlı yeniden yapılandırmaya dayalı CPPN modeli ile biraz daha net fakat gürültülü olması sebebi ile gerçekten uzak veriler elde edildiği değerlendirilmektedir. Diğer bir deney el yazısı stillerinin gizli kodlarla aktarılabilmesini analiz etmek üzere

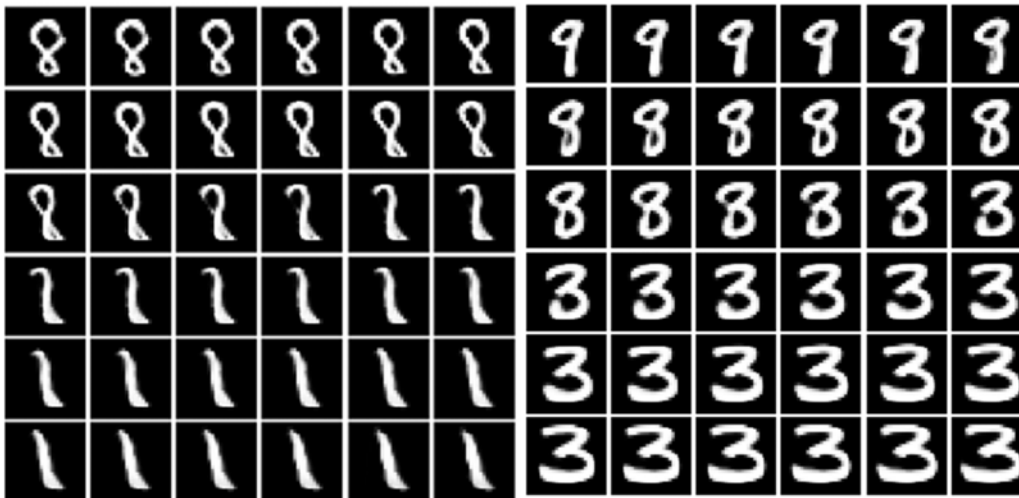
gerçekleştirilmiştir. Bu amaçla, gizli kodlara dayalı aritmetik ile hedeflenen stile yönelik gizli kod elde edilmeye çalışılarak bu gizli kod aracılığıyla stile yönelik görüntüler elde edilmeye çalışılmıştır. Bu analiz ile ilgili bir kaç örnek Şekil 8'de gösterilmiştir. Şekilde, hedeflenen stilin başka bir rakama ait görüntüye nasıl transfer edildiği gösterilmiştir. Şekildeki ilk örnekte gizli kod aritmetiği ile 7 rakamı daha italik bir yazım stili ile üretilebilmiştir. Stil transferi işlemine benzer olarak son deney çalışması olarak gizli kodların doğrusal interpolasyonu ile elde edilen yeni örnekler Şekil 9'da verilmiştir. Bu örneklerde görülebileceği gibi 8 rakamının 1 rakamına gizli kodların interpolasyonu ile nasıl dönüştürülebildiği gösterilmiştir.

Tablo 3. Süper çözünürlük görüntülerinin bulanıklık analizi (The clearness analysis of super resolution images)

Metot		CPPN-VAE-GAN	VAE/CPGAN
Ölçek (s) = 2	Netlik skoru	1647,0256±432,4643	1538,5852±428,1420
	Netlik oranı	1,0000	1,0000
Ölçek (s) = 3	Netlik skoru	509,7392 ±137,2906	450,7175±129,2560
	Netlik oranı	1,0000	1,0000
Ölçek (s) = 4	Netlik skoru	203,0867±55,8707	175,8523±52,6528
	Netlik oranı	0,9901	0,9586



Şekil 8. Gizli kod aritmetiği örnekleri (Examples of latent code arithmetic)



Şekil 9. Gizli kod interpolasyonu örnekleri (Examples of latent code interpolations)

5. SONUÇLAR VE TARTIŞMALAR (RESULTS AND DISCUSSIONS)

Günümüzde, derin öğrenme temelli modeller ile bilgisayarlı görü ve diğer alanlardaki uygulamalarında en yüksek performansın elde edildiği görülmektedir. Özellikle çekişmeli eğitim önde olmak üzere üretici modellerin, derin öğrenme alanının en ilgi çekici konularından bir olarak kaydedildiği görülmüştür. Son yıllarda sunulan modellerin ağırlıklı olarak çekişmeli eğitime dayalı üretici modellerden oluştuğu görülmektedir. Bu yaklaşımlarla yeniden dikkat çeken otokodlayıcılara göre üstünlüklerine rağmen otokodlayıcı yaklaşımlarında bulunan çıkarım mekanizmalarının bulunmamasından dolayı son yıllarda özellikle VAE ve GAN hibrit modellerinde artış eğilimi söz konusu olmaktadır. Bu çalışmalardan etkilenerek bu çalışmada da modellerin avantajlarını bir arada kullanarak sentetik görüntü oluşturabilmek üzere VAE ve GAN tabanlı bir modele odaklanılmıştır. CPPN modelinin ölçeklenebilirlik sağlamasından etkilenilerek bu çalışmada CPPN benzeri bir üretici model tercih edilmiştir. Bu üretici ağ ile yalnızca veri dağılımının modellenmesini sağlanmaz, aynı zamanda düşük boyutlu görüntünün gizli bir kodundan süper çözünürlük görüntülerinin oluşturulmasını sağlayan görüntü yoğunluğu da öngörülebilmektedir. Bu yaklaşıma dayalı çıkarsama mekanizması ile çekişmeli eğitim yaklaşımı ve öznitelik tabanlı bir objektif bir araya getirilerek VAE/CPGAN olarak adlandırılan yeni bir model sunulmuştur. Böylece daha çeşitli, net, gerçekçi ve yüksek çözünürlüklü görüntüler önerilen hibrit VAE/GAN modeli ile elde edilebilmiştir.

Önerilen model performansı VAE, VAE/GAN, DCGAN, WGAN, DRAGAN modelleri ve CPPN benzeri fakat farklı objektife dayalı oluşturulan CPPN-GAN-VAE modeli ile karşılaştırılmıştır. Karşılaştırmalar için benzer çalışmalarda da olduğu gibi inception skoru tercih edilmiştir. Deney sonuçlarına bakıldığında, önerilen model ile diğer VAE ve GAN tabanlı modellerden daha iyi performansın elde edildiği görülmektedir. Deney sonuçları incelendiğinde daha çeşitli, gerçekçi örneklerin daha tutarlı bir eğitim yaklaşımı ile önerilen model kullanılarak elde edilebildiği değerlendirilmiştir. Önerilen modelin ölçeklenebilirliği sayesinde farklı ölçeklerde eğitim yapılmaksızın elde edilen süper çözünürlük görüntüler, NN interpolasyon ve karşılaştırmalarda kullanılan diğer CPPN benzeri model ile karşılaştırıldığında modelin başarısı dikkat çekmektedir. Ayrıca, gizli kod aritmetiği ve interpolasyon örnekleri ile önerilen model ile veri dağılımının başarı ile modellenildiği gösterilmiştir. Devam eden çalışmalarda, modelin daha karmaşık yapıları görüntüler için geliştirilmesi hedeflenmektedir.

KAYNAKLAR (REFERENCES)

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y., Generative adversarial nets, *Advances in neural information processing systems*, 2014.
- Goodfellow, I. J., Jonathon S., Christian S., Explaining and harnessing adversarial examples, *arXiv preprint arXiv:1412.6572*, 2014.
- Gauthier, J., Conditional generative adversarial nets for convolutional face generation, *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition*, Winter semester, 2014:5, 2014.
- Radford, A., Metz, L., Chintala, S., Unsupervised representation learning with deep convolutional generative adversarial networks, *arXiv preprint arXiv:1511.06434*, 2015.
- Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H., Generative adversarial text to image synthesis. In *Proceedings of The 33rd International Conference on Machine Learning*, 2016.
- Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X. and Metaxas, D.N., Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks, In *Proceedings of the IEEE International Conference on Computer Vision*, 5907-5915, 2017.
- Yeh, R.A., Chen, C., Yian Lim, T., Schwing, A.G., Hasegawa-Johnson, M. and Do, M.N., Semantic image inpainting with deep generative models, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5485-5493, 2017.
- Iizuka, S., Simo-Serra, E. and Ishikawa, H., Globally and locally consistent image completion, *ACM Transactions on Graphics (ToG)*, 36 (4), 2017.
- Liu, M.Y., Breuel, T. and Kautz, J., Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, 700-708, 2017.
- Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A, Image-to-image translation with conditional adversarial networks, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125-1134, 2017.
- Kingma, D.P. and Welling, M., Auto-encoding variational bayes, *arXiv preprint arXiv:1312.6114*, 2013.
- Stanley, K.O., Compositional pattern producing networks: A novel abstraction of development. *Genetic programming and evolvable machines*, 8 (2), pp.131-162, 2007.
- Mirza, M. and Osindero, S., Conditional generative adversarial nets, *arXiv preprint arXiv:1411.1784*, 2014.
- Denton, E.L., Chintala, S. and Fergus, R., Deep generative image models using a laplacian pyramid of adversarial networks, In *Advances in neural information processing systems*, 1486-1494, 2015.
- Radford, A., Metz, L. and Chintala, S., Unsupervised representation learning with deep convolutional generative adversarial networks, *arXiv preprint arXiv:1511.06434*, 2015.
- Hinton, G.E. and Zemel, R.S., Autoencoders, minimum description length and Helmholtz free energy, In *Advances in neural information processing systems*, 3-10, 1994.

17. Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I. and Frey, B., Adversarial autoencoders, arXiv preprint arXiv:1511.05644, 2015.
18. Larsen, A.B.L., Sønderby, S.K., Larochelle, H. and Winther, O., Autoencoding beyond pixels using a learned similarity metric, arXiv preprint arXiv:1512.09300, 2015.
19. Lamb, A., Dumoulin, V. and Courville, A., Discriminative regularization for generative models, arXiv preprint arXiv:1602.03220, 2016.
20. Dosovitskiy, A. and Brox, T., Generating images with perceptual similarity metrics based on deep networks, In Advances in Neural Information Processing Systems, 658-666, 2016.
21. Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M. and Courville, A., Adversarially learned inference, arXiv preprint arXiv:1606.00704, 2016.
22. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C., Improved training of wasserstein gans, In Advances in neural information processing systems, 5767-5777, 2017.
23. Kodali, N., Abernethy, J., Hays, J., & Kira, Z. On convergence and kararlıty of gans. arXiv preprint arXiv:1705.07215, 2017.
24. Stanley, K.O. and Miikkulainen, R., Evolving neural networks through augmenting topologies. Evolutionary computation, 10 (2), 99-127, 2002.
25. Dong, C., Loy, C.C., He, K. and Tang, X., Learning a deep convolutional network for image super-resolution, In European conference on computer vision, 184-199, 2014.
26. Dong, C., Loy, C.C. and Tang, X., Accelerating the super-resolution convolutional neural network, In European conference on computer vision, 391-407, 2016.
27. Kim, J., Kwon Lee, J. and Mu Lee, K., Deeply-recursive convolutional network for image super-resolution, In Proceedings of the IEEE conference on computer vision and pattern recognition, 1637-1645, 2016.
28. Yamanaka, J., Kuwashima, S. and Kurita, T., Fast and accurate image super resolution by deep CNN with skip connection and network in network, In International Conference on Neural Information Processing, 217-225, 2017.
29. Lai, W.S., Huang, J.B., Ahuja, N. and Yang, M.H., Fast and accurate image super-resolution with deep laplacian pyramid networks, IEEE transactions on pattern analysis and machine intelligence, 2018.
30. Lim, B., Son, S., Kim, H., Nah, S. and Mu Lee, K., Enhanced deep residual networks for single image super-resolution, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 136-144, 2017.
31. Johnson, J., Alahi, A. and Fei-Fei, L., Perceptual losses for real-time style transfer and super-resolution, In European conference on computer vision, 694-711, 2016.
32. Simonyan, K. and Zisserman, A., Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.
33. Kim, J., Kwon Lee, J. and Mu Lee, K., Accurate image super-resolution using very deep convolutional networks, In Proceedings of the IEEE conference on computer vision and pattern recognition, 1646-1654, 2016.
34. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., Photo-realistic single image super-resolution using a generative adversarial network, In Proceedings of the IEEE conference on computer vision and pattern recognition, 4681-4690, 2017.
35. He, K., Zhang, X., Ren, S. and Sun, J., Deep residual learning for image recognition, In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778, 2016.
36. Sajjadi, M.S., Scholkopf, B. and Hirsch, M., Enhancenet: Single image super-resolution through automated texture synthesis, In Proceedings of the IEEE International Conference on Computer Vision, 4491-4500, 2017.
37. Turhan, C. G., Bilge, H. S., Single image super resolution using deep convolutional generative neural networks, In 2018 26th Signal Processing and Communications Applications Conference (SIU), 1-4, 2018.
38. Wang, Y., Perazzi, F., McWilliams, B., Sorkine-Hornung, A., Sorkine-Hornung, O., & Schroers, C., A fully progressive approach to single-image super-resolution, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 864-873, 2018.
39. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C., Esrgan: Enhanced super-resolution generative adversarial networks, In Proceedings of the European Conference on Computer Vision (ECCV), 2018.
40. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y., Image super-resolution using very deep residual channel attention networks, In Proceedings of the European Conference on Computer Vision (ECCV), 286-301, 2018.
41. Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., & Wu, W., Feedback Network for Image Super-Resolution, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3867-3876, 2019.
42. Dai, T., Cai, J., Zhang, Y., Xia, S. T., & Zhang, L., Second-order Attention Network for Single Image Super-Resolution, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 11065-11074, 2019.
43. Secretan, J., Beato, N., D Ambrosio, D.B., Rodriguez, A., Campbell, A. and Stanley, K.O., Picbreeder: evolving pictures collaboratively online, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1759-1768, 2008.

44. LeCun, Y., Cortes, C. and Burges, C.J., MNIST handwritten digit database, AT&T Labs. <http://yann.lecun.com/exdb/mnist>, 02.18.2010, 03.05.2019.
45. Maas, A.L., Hannun, A.Y. and Ng, A.Y., Rectifier nonlinearities improve neural network acoustic models, In Proc. Icml, 30 (1), 3, 2013.
46. Kingma, D.P. and Ba, J., A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.
47. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X, Improved techniques for training gans, In Advances in neural information processing systems, 2234-2242, 2016.