T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DOKTORA TEZİ

ÜRETKEN ÇEKİŞMELİ AĞ TABANLI TEK GÖRÜNTÜ ÜRETİM MODELLERİNİN TASARIMI

Eyyüp YILDIZ

DANIŞMAN
Doç. Dr. Selçuk SEVGEN
(danışman imzalı olacaktır)
II. DANIŞMAN
Doç. Dr. Mehmet Erkan YÜKSEL

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği, Doktora Programı

Nisan, 2024

TEZ KABUL VE ONAYI

Eyyüp YILDIZ tarafından, **Doç. Dr. Selçuk SEVGEN** ve **Doç. Dr. Mehmet Erkan YÜKSEL** danışmanlığında hazırlanan **"ÜRETKEN ÇEKİŞMELİ AĞ TABANLI TEK GÖRÜNTÜ ÜRETİM MODELLERİNİN TASARIMI"** başlıklı bu çalışma, jürimiz tarafından **22/04/2024** tarihinde yapılan sınav sonucunda **oy birliği** ile başarılı bulunarak **Doktora Tezi** olarak kabul edilmiştir.

Tez Jürisi

		İmza	**Sonuç**
DANIŞMAN	Doç. Dr. Selçuk SEVGEN İstanbul Üniversitesi - Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☐ Kabul ☐ Ret
ÜYE	Prof. Dr. Ahmet SERTBAŞ İstanbul Üniversitesi - Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☐ Kabul ☐ Ret
ÜYE	Prof. Dr. Rüya ŞAMLI İstanbul Üniversitesi - Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☐ Kabul ☐ Ret
ÜYE	Doç. Dr. Faruk BULUT İstanbul Esenyurt Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı		☐ Kabul ☐ Ret
ÜYE	Dr. Öğr. Üyesi Zeynep TURGUT AKGÜN İstanbul Medeniyet Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı		☐ Kabul ☐ Ret

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve bilimsel etik kuralları içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını ve her türlü hukuki sorumluluğu aldığımı kabul ederim.

Eyyüp YILDIZ

(İmza)

Aileme ithaf ediyorum…

BÜTÇE DESTEKLERİ

ÜRETKEN ÇEKİŞMELİ AĞ TABANLI TEK GÖRÜNTÜ ÜRETİM MODELLERİNİN TASARIMI

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.

TEŞEKKÜR

Bu tez çalışması boyunca yardımları, yönlendirmeleri ve katkılarıyla beni destekleyen değerli tez danışmanı hocalarım sayın Doç. Dr. Selçuk SEVGEN ve sayın Doç. Dr. Mehmet Erkan YÜKSEL'e katkılarından ötürü teşekkür ederim.

Sadece bu süreçte olmayıp hayatın her alanında yardımını ve desteğini hissettiğim çok kıymetli aileme sonsuz minnettarım.

Nisan 2024 **Eyyüp YILDIZ**

İÇİNDEKİLER

Sayfa No

TEZ KABUL VE ONAYI ..... **ii**
BEYAN ..... **iii**
BÜTÇE DESTEKLERİ ..... **v**
TEŞEKKÜR ..... **vi**
İÇİNDEKİLER ..... **vii**
ŞEKİL LİSTESİ ..... **ix**
TABLO LİSTESİ ..... **xi**
SİMGE VE KISALTMA LİSTESİ ..... **xii**
ÖZET ..... **xv**
ABSTRACT ..... **xvii**
1. GİRİŞ **1**
2. KAVRAMSAL ÇERÇEVE **5**
2.1. ÜÇA Modeli 5
2.2. Tek Eğitim Görüntüsü Kullanan ÜÇA Modelleri 11
3. YÖNTEM **23**
3.1. Model 1 (Öz-dikkat ve Yoğun Bağlantılı ÜÇA) 23
3.1.1. Model 1'in Genel Yapısı 23
3.1.2. Eğitim Görüntüsü Ölçeklendirme 24
3.1.3. Model 1'in Eğitim Prosedürü 26
3.1.4. Öz-dikkat Mekanizması 32
3.1.5. Yoğun Bağlantılı Konvolüsyon Blokları 33
3.1.6. Gradyan Cezalı Göreceli Ortalama En Küçük Kareler ÜÇA Kaybı 35
3.1.7. Model 1'in Eğitim Parametreleri ve Optimizasyon Algoritması 36
3.2. Model 2 (Çekişmeli Öğrenilen Çıkarım ÜÇA) 37
3.2.1. Model 2'nin Genel Yapısı 38
3.2.2. Model 2'nin Eğitim Prosedürü 40
3.2.3. Gradyan Cezalı Wasserstein Mesafe ÜÇA Kaybı 42
3.2.4. Model 2'nin Eğitim Parametreleri ve Optimizasyon Algoritması 43

4. BULGULAR **45**
4.1. Kullanılan Veri Kümeleri 45
4.2. Nicel Ölçüm Yöntemleri 46
4.2.1. Tek Görüntü Fréchet Başlangıç Mesafesi 46
4.2.2. Çok Ölçekli Yapısal Benzerlik İndeksi Ölçümü 47
4.2.3. Öğrenilmiş Algısal Görüntü Yama Benzerliği 48
4.3. Kullanıcı Çalışmaları Yöntemleri 48
4.4. Model 1 Sonuçları 51
4.4.1. Görsel Karşılaştırma 51
4.4.2. Nicel Sonuçlar 53
4.4.3. Nitel Sonuçlar 54
4.4.4. Ablasyon Çalışması 55
4.5. Model 2 Sonuçları 61
4.5.1. Görsel Karşılaştırma 61
4.5.2. Nicel Sonuçlar 64
4.5.3. Nitel Sonuçlar 65
4.5.4. Ablasyon çalışması 66
4.6. Uygulamalar 68
4.6.1. Çizimden Görüntü Oluşturma 69
4.6.2. Harmanlama 70
4.6.3. Görüntü Düzenleme 71
4.6.4. Animasyon 72
4.6.5. Değişken Boyutlu Görüntüler 73
5. TARTIŞMA **76**
6. SONUÇ VE ÖNERİLER **78**
KAYNAKLAR **80**
EKLER **88**
İNTİHAL RAPORU İLK SAYFASI **117**
ETİK KURUL İZİN YAZISI **121**
KURUM İZNİ YAZILARI **122**
ÖZGEÇMİŞ **123**

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Klasik ÜÇA mimarisi 6

Şekil 2.2: Gradyan sönümlemesi örneği 8

Şekil 2.3: Mod çökmesi örneği 9

Şekil 2.4: SinGAN mimarisi. 12

Şekil 3.1: Model 1'in genel yapısı 24

Şekil 3.2: Ölçeklendirme sonucu her ölçek için oluşturulan eğitim görüntüleri 26

Şekil 3.3: j. Ölçekteki $Ü_j$ ve A_j ağlarının iç yapısı 28

Şekil 3.4: Konvolüsyon işlemi 29

Şekil 3.5: k. ölçek için $Ü_k$ ve A_k'da kullanılan Öz-dikkat mekanizması 33

Şekil 3.6: Beş yoğun bağlı katman ve bir geçiş katmanından oluşan yoğun bağlantılı blok. ..34

Şekil 3.7: Üç yoğun bağlı katman ve bir geçiş katmanından oluşan iki yoğun bağlantılı blok. 34

Şekil 3.8: Dört yoğun bağlı katman ve bir geçiş katmanından oluşan iki yoğun bağlantılı blok. 34

Şekil 3.9: Model 2'nin genel yapısı. 39

Şekil 3.10: $k-1$ ölçekteki Ü, A, ve K ağlarının ayrıntılı yapısı 40

Şekil 4.1: Eşleştirilmiş çalışma için örnek soru sayfası 49

Şekil 4.2: Eşleştirilmemiş çalışma için örnek soru sayfası 50

Şekil 4.3: Nitel değerlendirme çalışması için örnek soru sayfası 50

Şekil 4.4: Model 1 tarafından oluşturulan rastgele görüntü örnekleri. 51

Şekil 4.5: Model karşılaştırma. 52

Şekil 4.6: Görüntülerin t-SNE ile iki boyutlu uzayda gösterimi 60

Şekil 4.7: Model 2 tarafından oluşturulan rastgele görüntü örnekleri. 61

Şekil 4.8: Model karşılaştırma (Places). 62
Şekil 4.9: Model karşılaştırma (LSUN). 63
Şekil 4.10: Model karşılaştırma (ImageNet) 63
Şekil 4.11: Görüntü ve gürültülerin t-SNE ile iki boyutlu uzayda gösterimi. 68
Şekil 4.12: Görüntü manipülasyonu uygulamalarında işlem süreci. 69
Şekil 4.13: Çizimden Görüntüye uygulama sonuçları. 70
Şekil 4.14: Harmanlama uygulama sonuçları. 71
Şekil 4.15: Görüntü düzenleme uygulama sonuçları. 72
Şekil 4.16: Animasyon uygulama sonuçları. 73
Şekil 4.17: Farklı boyutlarda görüntü üretim sonuçları. 74
Şekil 6.1: Model 1 tarafından üretilen görüntüler (Ek-C). 89
Şekil 6.2: Model 2 tarafından üretilen görüntüler (Ek-D). 90
Şekil 6.3: Çizimden görüntüye uygulama sonuçları (Ek-E) 91
Şekil 6.4: Harmanlama uygulama sonuçları (Ek-F) 92
Şekil 6.5: Görüntü düzenleme uygulama sonuçları (Ek-G) 93
Şekil 6.6: Animasyon görüntüleri (1) (Ek-H) 95
Şekil 6.7: Animasyon görüntüleri (2) (Ek-H) 96
Şekil 6.8: Farklı boyutlarda üretilen görüntüler (1) (Ek-I) 97
Şekil 6.9: Farklı boyutlarda üretilen görüntüler (2) (Ek-I) 98
Şekil 6.10: Tüm modellerin ürettiği görüntüler (Places) (Ek-J) 99
Şekil 6.11: Tüm modellerin ürettiği görüntüler (LSUN) (Ek-K) 104
Şekil 6.12: Tüm modellerin ürettiği görüntüler (ImageNet) (L) 109

TABLO LİSTESİ

Sayfa No

Tablo 3.1: Model 1 eğitim hiper parametre değerleri. 37

Tablo 3.2: Model 2 eğitim hiper parametre değerleri. 44

Tablo 4.1: Places, LSUN ve ImageNet veri kümeleri için nicel sonuçlar. 53

Tablo 4.2: Tek görüntü ÜÇA modellerinin parametre değerleri. 54

Tablo 4.3: Eşleştirilmiş ve eşleştirilmemiş kullanıcı çalışmaları sonuçları 55

Tablo 4.4: Nitel kullanıcı çalışması sonuçları. 55

Tablo 4.5: Farklı konfigürasyonlara sahip Model 1 için nicel sonuçlar. 56

Tablo 4.6: ResNet tabanlı oluşturulan Model 1 için nicel sonuçlar. 57

Tablo 4.7: GoogLeNet (Inception v1) tabanlı oluşturulan Model 1 için nicel sonuçlar. 57

Tablo 4.8: Model 1'in farklı yoğun bağlantılı blok konfigürasyonları için nicel sonuçlar. 58

Tablo 4.9: Places, LSUN ve ImageNet veri kümeleri için nicel sonuçlar. 64

Tablo 4.10: Tek görüntü ÜÇA modellerinin parametre değerleri. 65

Tablo 4.11: Eşleştirilmiş ve eşleştirilmemiş kullanıcı çalışmaları sonuçları. 66

Tablo 4.12: Nitel kullanıcı çalışması sonuçları. 66

Tablo 4.13. Ablasyon çalışmasının nicel sonuçları. 67

Tablo 6.1: Model 1'in farklı konfigürasyonları parametre sayısı ve eğitim süresi (Ek-N) 115

Tablo 6.2: ResNet tabanlı farklı konfigürasyonların parametre sayısı ve eğitim süresi (Ek-N) 116

Tablo 6.3: GoogLeNet tabanlı farklı konfigürasyonların parametre sayısı ve eğitim süresi (Ek-N) 116

SİMGE VE KISALTMA LİSTESİ

Simgeler	Açıklama
$\boldsymbol{z}$	: Klasik ÜÇA gürültü haritası
$\boldsymbol{x}$	: Klasik ÜÇA eğitim (gerçek) verisi
$\tilde{\boldsymbol{x}}$	: Klasik ÜÇA üretilen (sahte) veri
$\boldsymbol{\omega}$	: Üretici ağ ağırlıkları
$\boldsymbol{\theta}$	: Ayırıcı ağ ağırlıkları
$\boldsymbol{\mathcal{L}_2}$	: Karesel yeniden yapılandırma kayıp fonksiyonu
$\boldsymbol{n_s}$	: Ölçek sayısı
$\boldsymbol{h_{sfi}}$	: Başlangıç ölçek oranı
$\boldsymbol{i}$	: Eğitim görüntüsü
$\boldsymbol{i_{min}}$	: Eğitim görüntüsünün minimum boyutu
$\boldsymbol{i_{max}}$	: Eğitim görüntüsünün maksimum boyutu
$\boldsymbol{h^i_{min}}$	: Üretilecek minimun görüntü boyutu
$\boldsymbol{h^i_{max}}$	: Üretilecek maksimum görüntü boyutu
$\boldsymbol{s_f}$	: Ölçek oranı
$\boldsymbol{i^s_k}$	: k. ölçekteki eğitim görüntüsü
$\boldsymbol{\mathcal{L}^ç_k}$	: k. ölçekteki çekişmeli kayıp fonksiyonu
$\boldsymbol{\mathcal{L}^{yy}_k}$	: k. ölçekteki yeniden yapılandırma kayıp fonksiyonu
$\boldsymbol{\lambda_1}$	: Üretici ağ çekişmeli kayıp ağırlığı
$\boldsymbol{\lambda_2}$	: Üretici ağ yeniden yapılandırma kayıp ağırlığı
$\boldsymbol{z^{yy}_k}$	: k. ölçekteki yeniden yapılandırma gürültü haritası
$\boldsymbol{x^k}$	: k. katmandaki konvolüsyon çıktı özellikleri
$\boldsymbol{i^{yy}_k}$	: k. ölçekteki yeniden yapılandırma görüntüsü
$\boldsymbol{\zeta^k}$	: k. normalizasyon katmanı ölçeklendirme değeri
$\boldsymbol{\beta^k}$	: k. normalizasyon katmanı kaydırma değeri
$\boldsymbol{i^g_k}$	: k. ölçekte $Ü_k$ tarafından üretilen görüntü
$\boldsymbol{z^p_k}$	: k. ölçekte K_k tarafından oluşturulan görüntü uzayı
$\boldsymbol{\mathcal{L}^{RALSGAN}_{A_k}}$	: k. ölçekte A_k'nın çekişmeli kayıp fonksiyonu

$\mathcal{L}_{A_k}^{RALSGAN-GP}$: k. Ölçekte A_k'nın toplam kayıp fonksiyonu

$\mathcal{L}^{GP}$: Gradyan cezası kayıp fonksyionu

ψ : Gradyan cezası ağırlığı

β_1 : Adam optimizasyon algoritması parametre değeri

β_2 : Adam optimizasyon algoritması parametre değeri

$\mathcal{L}_{Ü_k}^{RaLSGANrec}$: k. ölçekte $Ü_k$'nın toplam kayıp fonksiyonu

$\hat{x}$: Gradyan cezası için gerçek ve sahte görüntülerin karışımı

lr : Öğrenme oranı

z_k^p : k. ölçekteki eğitim görüntüsünün gürültü haritası

z_k^f : k. ölçekteki üretici ağın girdi gürültü haritası

$\mathcal{L}_{A_k}^{WGAN}$: k. ölçekte A_k'nın çekişmeli kayıp fonksiyonu

$\mathcal{L}_{Ü_k}^{WGANrec}$: k. ölçekte $Ü_k$'nın toplam kayıp fonksiyonu

$\mathcal{L}_{K_k}^{WGAN}$: k. ölçekte K_k'nın çekişmeli kayıp fonksiyonu

φ : Kodlayıcı ağ ağırlıkları

μ_x : x görüntüsünü piksel yoğunluğu

σ_x : x görüntüsünü piksel değerlerinin standart sapması

Tr : Matris izi (trace)

Kısaltmalar	Açıklama
ÜÇA	: Üretken Çekişmeli Ağ (Generative Adversarial Network)
A	: Ayırıcı (Discriminator)
Ü	: Üretici (Generator)
K	: Kodlayıcı (Encoder)
PA	: Yama Ayırıcı (Patch Discriminator)
ÇÖÇ	: Çekişmeli Öğrenilen Çıkarım (Adversally Learned Inference)
RALSGAN	: Göreceli Ortalama En Küçük Kareler ÜÇA (Relativistic Average Least Square GAN)
RALSGAN-GP	: Gradyan Cezalı Göreceli Ortalama En Küçük Kareler ÜÇA (Relativistic Average Least Square GAN with Gradient Penalty)
GP	: Gradyan Cezası (Gradient Penalty)
WGAN	: Wasserstein ÜÇA (Wasserstein GAN)

WGAN-GP : Gradyan Cezalı Wasserstein ÜÇA (Wasserstein GAN with Gradient Penalty)

YN : Yığın Normalizasyon (Batch Normalization)

ÖN : Örnek Normalizasyon (Instance Normalization)

SSIM : Yapısal Benzerlik İndeksi Ölçümü (Structural Similarity Index Measure)

MS-SSIM : Çok Ölçekli Yapısal Benzerlik İndeksi Ölçümü (Multi Scale Structural Similarity Index Measure)

SIFID : Tek Görüntü Frechet Başlangıç Mesafesi (Single Image Fréchet Inception Distance)

LPIPS : Öğrenilmiş Algısal Görüntü Yama Benzerliği (Learned Perceptual Image Patch Similarity

ÖZET

DOKTORA TEZİ

ÜRETKEN ÇEKİŞMELİ AĞ TABANLI TEK GÖRÜNTÜ ÜRETİM MODELLERİNİN TASARIMI

Eyyüp YILDIZ

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği, Doktora Programı

Danışman : Doç. Dr. Selçuk SEVGEN

II. Danışman : Doç. Dr. Mehmet Erkan YÜKSEL

Üretken Çekişmeli Ağ (ÜÇA)'lar kullanılarak tek bir doğal görüntüden görüntü üretimi, ÜÇA modellerinin foto-gerçekçi görüntüler üretmedeki yetenekleri ve bilgisayar görüsündeki potansiyel uygulamaları nedeniyle son zamanlarda yoğun ilgi çekmektedir. Bu tez çalışmasında, tek bir doğal eğitim görüntüsüne dayalı gerçekçi, yüksek kaliteli, farklı görüntü örnekleri üreten iki yeni koşulsuz ÜÇA modeli önerilmiş ve gerçeklenmiştir. Model 1, tek eğitim görüntüsü özelliklerini öğrenebilmek amacıyla hem üretici hem de ayırıcı sinir ağlarında öz-dikkat mekanizması, yoğun bağlantılı blok mimarisi ve gradyan cezalı göreceli ortalama en küçük kareler ÜÇA kayıp fonksiyonu bileşenlerinden yararlanmaktadır. Model 2, tek eğitim görüntüsünün gizli gösterim bilgisini etkili şekilde öğrenmektedir. Öğrenilen gizli gösterim kullanılarak Model 2 için özgün bir eğitim süreci geliştirilmiştir. Önerilen ÜÇA modellerinin performanslarını test etmek ve kapasitelerini göstermek amacıyla literatürde iyi bilinen tek görüntü ÜÇA modelleriyle nicel ve nitel karşılaştırmalar gerçekleştirilmiştir. Karşılaştırmalar, Places, LSUN ve ImageNet veri kümeleri üzerinde, görüntü kalitesini ve çeşitliliğini ölçen nicel ölçüm metrikleri kullanılarak; görüntü gerçekçiliğini ölçek amacıyla nitel kullanıcı çalışmaları

oluşturularak yapılmıştır. Ek olarak önerilen ÜÇA modelleri görüntü manipülasyon uygulamalarında test edilmiştir. Yapılan deneylerde, önerilen ÜÇA modellerinin etkin parametre verimliliğine sahip olduğu, aşırı uyumu önlediği, karmaşık yapı ve dokulara sahip eğitim görüntülerin dahili yama istatistiklerini öğrenebildiği, tek görüntü üretim görevlerinde karşılaştırılabilir performans elde ettiği gözlemlenmiştir. Kullanıcı çalışmaları, önerilen ÜÇA modelleri tarafından sentezlenen görüntülerin genellikle tek eğitim görüntüsüyle karıştırıldığını doğrulamıştır. Bu doğrultuda tez çalışması kapsamında önerilen ÜÇA modelleri, sınırlı eğitim verileriyle uğraşan alanlarda veri artırmanın yanı sıra çeşitli görüntü işleme görevleri için güçlü bir araç olarak kullanılabilir.

Nisan 2024 , 141 sayfa.

Anahtar kelimeler: Tek Görüntü Üretimi, Üretken Çekişmeli Ağlar, Derin Öğrenme, Öz-dikkat mekanizması, DenseNet, ALI, Görüntü manipülasyonu

ABSTRACT

Ph.D. THESIS

DESIGN OF SINGLE IMAGE GENERATION MODELS BASED ON GENERATIVE ADVERSARIAL NETWORKS

Eyyüp YILDIZ

İstanbul University-Cerrahpaşa

Institute of Graduate Studies

Department of Computer Engineering

Programme of Computer Engineering Ph.D.

Supervisor : Assoc. Prof. Dr. Selçuk SEVGEN

Co-Supervisor: Assoc. Prof. Dr. Mehmet Erkan YÜKSEL

Image generation from a single natural image using generative adversarial networks (GANs) has attracted extensive attention recently due to the GANs' practical ability to produce photo-realistic images and their potential applications in computer vision. In this thesis, two novel Unconditional GAN models are developed and implemented to generate diverse, high-quality, realistic image samples from a single natural training image. In both the generator and discriminator neural networks, Model 1 incorporates self-attention mechanisms, densely connected block architecture, and relativistic average least square GAN with gradient penalty loss component to proficiently learn the characteristics of a single training image. Model 2 effectively learns the latent representation information of the single training image. A novel training process has been developed for Model 2 by utilizing the learned latent representation. To test the performance and demonstrate the capabilities of the developed GAN models, quantitative and qualitative comparisons were carried out with well-known single-image GAN

models in the literature. The comparisons were conducted using quantitative metrics measuring image quality and diversity on datasets such as Places, LSUN, and ImageNet, while qualitative user studies were conducted to assess image realism. In addition, the developed GAN models are tested in image manipulation applications. Experimental findings indicate that the developed GAN models exhibit efficient parameter utilization, mitigate overfitting, and successfully learn the internal patch statistics of training images characterized by complex structures and textures, comparable performance in single-image generation tasks. User studies confirm that images synthesized by the developed GAN models are frequently mistaken for the single training image. In this regard, the GAN models proposed within the scope of the thesis can be used as a powerful tool for various image processing tasks as well as data augmentation in areas dealing with limited training data.

April 2024, 141 pages.

Keywords: Single Image Generation, Generative Adversarial Networks, Deep Learning, Self-Attention, DenseNet, ALI, Image Manipulation

1. GİRİŞ

Derin üretken modeller karmaşık ve yüksek boyutlu olasılık dağılımlarını modellemek için eğitilen derin sinir ağlarıdır [1]. Üretken Çekişmeli Ağ (ÜÇA)'lar, bir veri kümesinin yapısal özelliklerini oyun teorisine dayalı olarak öğrenmeyi amaçlayan derin üretken modellerdir [2], [3]. ÜÇA'lar veri üretimine dayalı görevlerde yüksek başarı göstermektedir. ÜÇA'ların, diğer derin üretici modellere oranla daha yüksek başarı sergilemeleri, ÜÇA'larla ilgili araştırmaların sayısını hızlı bir şekilde artırmıştır [3], [4]. Bu araştırmaların bazıları, ÜÇA'ların mimari yapılarını geliştirmeye [5], [6], [7], [8], bazıları ÜÇA'ların eğitimlerini optimize etmeye [9], [10], [11], [12] odaklanmıştır. Diğer kısım araştırmalar ise ÜÇA'ları dil oluşturma [13], görüntü oluşturma [14], [15], [16], görüntüden görüntüye çeviri [17], [18], görüntü iç boyama [19], [20], [21], ses verisi üretme [22], metin açıklamasından görüntü oluşturma [18], insan yüzü yaşlandırma [23], [24] gibi çeşitli problemlerin çözümünde kullanmıştır. Bununla birlikte mevcut ÜÇA modellerin çoğu, eğitim için büyük miktarda veri kümesine ve hesaplama kaynağına ihtiyaç duymaktadır.

Kısıtlı veriler üzerinde çalışan ÜÇA modelleri, kısıtlı sayıda eğitim verisi kullanarak öğrenen ve az veri bulunan senaryolarda daha etkili sonuçlar üretmeyi amaçlayan derin üretken modellerdir. Nadir görülen hastalıkların tıbbî görüntüleri, astronomik olayların görüntüleri (güneş tutulması veya kuyruklu yıldızın geçişi gibi), keşfedilen benzersiz hücre yapıları veya organizmaların görüntüleri ve laboratuvar deneylerine ait özel anların görüntüleri gibi kısıtlı sayıda veri bulunan durumlar üretici model oluşturmada zorluk teşkil etmektedir. Buna bağlı olarak son yıllarda kısıtlı eğitim verileri altında üretken bir modelin öğrenilmesine olan ihtiyaç giderek artmaktadır. Kısıtlı veri üzerinde çalışan bir üretici model için çözülmesi gereken problemler mevcuttur. Kısıtlı sayıdaki veri üzerinde eğitilen bir model, verideki karmaşık desenleri ve varyasyonları öğrenmek için yeterli örneğe sahip değildir. Bu durum modelin düşük performans göstermesine ve sonuçta kalitesiz çıktıların üretilmesine sebep olmaktadır. Başka bir problem ise, modelin aşırı öğrenmeye (modelin veri dağılımını öğrenmek yerine kısıtlı eğitim verisini ezberlemesi) yatkın olmasıdır. Bu, kısıtlı eğitim verilerindeki değişkenliğinin öğrenilememesine ve üretilen örneklerde çeşitlilik eksikliğine sebep olmaktadır. Ayrıca, kısıtlı veri örneklerinin büyük veri kümelerine oranla daha dar bir dağılımı

temsil etmesi eğitim istikrarsızlığına yol açabilmektedir [25]. Araştırmacılar bu problemleri belirli seviyelerde çözerek kısıtlı veriler üzerinde başarıyla çalışan ÜÇA modellerini oluşturmuştur [26], [27].

Tek görüntü üretimi, (Single Image Generation), ön eğitilmiş modellere dayanmadan, sadece tek bir eğitim görüntüsü kullanılarak farklı ve yeni görüntüler oluşturmak için üretken bir modelin eğitilmesidir. Bu bağlamda tek görüntü ÜÇA modelleri, tek eğitim görüntüsünün dahili istatistiklerini öğrenerek yüksek kaliteli, gerçekçi, farklı görüntü örneklerinin üretilmesine odaklanmış ÜÇA modelleridir. Klasik ÜÇA modellerinin tek eğitim görüntüsü kullanarak sentezledikleri yeni görüntüler çok düşük değişkenlik göstermektedir [25]. Bununla birlikte bir görüntüdeki yamaların dahili istatistiklerinin bir üretici modelin başarılı çalışması için yeterli bilgileri taşıdığı gösterilmiştir [28], [29]. Bu kapsamda üretici modeller, tek görüntü üretim probleminde görüntüler arası dağılımı öğrenmek yerine görüntü içi (yama) dağılımını öğrenmeyi amaçlamaktadır ve daha sonra üretici model tek eğitim görüntüsü üzerinden öğrendiği yama dağılımını kullanarak yeni görüntüler üretmektedir. Tek görüntü üretim problemi, kısıtlı veriler üzerinde çalışan üretici modellerin uç noktası olarak görülmektedir ve kısıtlı verilerde bulunan zorluklar tek görüntü üretim problemi içinde geçerlidir. Bu problemdeki temel zorluklar; üretilen görüntülerdeki görsel tutarlılığı ve uyumu sağlamak, aşırı uydurmadan ve eğitim istikrarsızlığından kaçınmaktır. Tek görüntü ÜÇA modelleri, klasik ÜÇA modellerine oranla oldukça yeni olmalarına rağmen birçok yönden geliştirilmektedir [28], [30], [31], [32], [33], [34], [35], [36].

Bu tez çalışmasında, tek görüntü üretim problemi için iki yeni tek görüntü ÜÇA modeli tasarlanmıştır. Önerilen tek görüntü ÜÇA modelleri probleme özgü benzer özellikler taşıyan veri kümelerine gereksinim duymadan, karmaşık yapılar ve dokular içeren doğal tek bir görüntüden öğrenebilmektedir. Oluşturulan modeller tek görüntü üretim problemini iki farklı perspektiften ele almaktadır. Oluşturulan ilk model (Model 1), Hiyerarşik ÜÇA piramit yapısı çerçevesinde çalışan konvolüsyon ağıdır. Model 1'in yüksek kapasiteli ağ yapısına sahip olması ve daha etkili bir kayıp fonksiyonu kullanması amaçlamıştır. Model 1, tek görüntü üretim probleminde aşırı uydurma durumunu önlemekte ve istikrarlı eğitim süreci sağlamaktadır. Model 1, bir görüntünün olası yeni kompozisyonlarını, içerik ve düzen farklılıklarıyla sentezleyebilmekte ve eğitim görüntüsünün genel yapısını ve ince dokularını korurken gerçekçi, yüksek kaliteli, farklı görüntüler üretebilmektedir. Model 1'in temel katkıları aşağıdaki gibidir:

- Görüntü oluşturmayı iyileştirmek için öz-dikkat (self-attention), yoğun bağlantılı bloklardan (densely connected netwoks) ve gradyan cezalı göreceli ortalama en küçük kareler ÜÇA (Relativistic Average Least Square GAN with Gradient Penalty-RaLSGAN-GP) kayıp fonksiyonunu kullanan yeni bir tek görüntü ÜÇA modeli oluşturulmuştur. Model 1 farklı mimari özellikler oluşturularak analiz edilmiştir. Model 1'in görüntü düzenleme, uyumlaştırma, görüntü boyama ve animasyon gibi görüntü işleme görevlerindeki etkinliğini göstermek için uygulamalar yapılmıştır.

- Öz-dikkat [11], [37] mekanizması kullanılarak, görüntü üretim kalitesi için dikkat odaklı, uzun menzilli bağımlılık modellemesi sağlanmıştır. Öz-dikkat, görüntü içerisindeki yamaların farklı seviyede önemlerde değerlendirerek karmaşık yapısal ve dokusal desenlere sahip görüntü sınıflarının öğrenilmesine yardımcı olmaktadır.

- Yoğun bağlantılı bloklardan [38] yararlanılarak, üretici ve ayırıcı ağların katmanları arasında bilgi akışı artırılmıştır. Yoğun bağlantılı blokların sağladığı yeniden özellik kullanımıyla Model 1'in potansiyelinden daha etkili bir şekilde yararlanılmıştır. Ayrıca yoğun bağlantılı bloklar kullanılarak hesaplama yükünde yüksek performans gösterilmiş, eğitimi stabilize edilmiş ve aşırı uyumu azaltılmıştır.

- İstikrarlı eğitim sürecini sağlamak için göreceli olarak çalışan RaLSGAN [39] kayıp fonksiyonu, gradyan cezası (gradient penalty) [40] ile birlikte kullanılarak tek görüntü üretim problemine uyarlanmıştır.

Tez çalışması kapsamında geliştirilen ikinci model (Model 2) ise gürültü uzayı bilgilerinden yararlanarak tek görüntü üretim probleminde özgün bir eğitim süreci oluşturmayı amaçlamaktadır. Bu amaçla Çekişmeli Öğrenilen Çıkarım (Adversarially Learned Inference-ÇÖÇ) [41], [42] modeli tek görüntü ÜÇA modeline uyarlanmıştır. Model 2'nin literatüre temel katkıları aşağıdaki gibidir:

Model 2 temel katkıları aşağıdaki gibidir:

- ÇÖÇ mimarisi tek görüntü üretim problemine uygulanarak tek eğitim görüntüsünün yama dağılımı etkili şekilde öğrenilmektedir.

- Tek eğitim görüntüsünün gizli uzay gösterim bilgisinden yararlanmak için kodlayıcı ağından (yama düzeyinde çalışan konvolüsyon ağ) yararlanılmıştır. Kodlayıcı ağı tek

eğitim görüntüsünden elde edilen bilgileri üretici ağa aktararak Model-2'nin görüntü üretme kapasitesini artırmaktadır.

- Model 2'in görüntü düzenleme, uyumlaştırma, görüntüye boyama ve animasyon gibi görüntü işleme görevlerindeki etkinliğini göstermek için uygulamalar yapılmıştır.

Tez çalışmasının kalan bölümleri şu şekilde düzenlenmiştir; ikinci bölümde klasik ÜÇA modelinin çalışma prensibi ayrıntılı olarak anlatılmıştır. ÜÇA alanında yapılan yenilikler, gelişmeler açıklanmış ve tez çalışmasında yararlanılan çalışmalara değinilmiştir. Ardından Tek görüntü üretim problemi için yayımlanan çalışmaların incelendiği kapsamlı literatür taraması verilmiştir. Üçüncü bölümde tek görüntü üretim problemi için geliştirilen iki ÜÇA modeli ayrıntılı olarak açıklanmıştır. Model-1'de kullanılan öz-dikkat, yoğun bağlantılı blok ve kayıp fonksiyonu anlatılmıştır. Model-2'de yararlanılan ÇÖÇ mimarisi ve eğitim süreci açıklanmıştır. Dördüncü bölümde oluşturulan tek görüntü ÜÇA modellerinin bulgularına yer verilmiştir. Modeller var olan diğer modellerle nicel ve nitel olarak karşılaştırılmıştır. Ek olarak her model bileşenlerinin modele katkıları incelenmiştir. Ayrıca farklı görüntü manipülasyonu uygulamaları için deneyler yapılarak modellerin sağlamlığı test edilmiştir. Beşinci bölümde oluşturulan modellerin üstünlükler ve zayıflıkları objektif olarak diğer modellere karşılaştırmalı bir şekilde değerlendirilmiştir. Altıncı bölümde oluşturulan modellerin ve tez sürecinin genel değerlendirilmesi yapılmıştır. Ek olarak yapılması planlanan çalışmalar hakkında bilgilendirme yapılmıştır.

2. KAVRAMSAL ÇERÇEVE

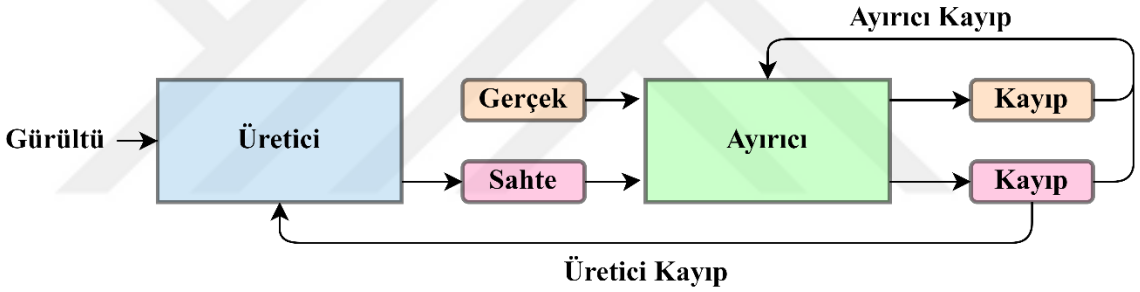
Veri dağılımını öğrenmek denetimsiz öğrenmenin (unsupervised learning) zor problemlerinden birisidir. Bu zorluğun temel sebebi veri dağılımını öğrenmek için optimize edilecek bir amaç/kayıp fonksiyonunun bulunmamasıdır. Klasik makine öğrenmesi yöntemlerinin kısıtlı başarı gösterdiği bu problem için derin üretici modeller yüksek başarı göstermektedir. Derin üretken model ailesinin bir üyesi olan ÜÇA, bu problem için öne çıkan yenilikçi bir derin sinir ağı modelidir. I. Goodfellow ve diğ. (2014) [2] tarafından geliştirilen ÜÇA gösterim öğrenme (represantation learning) alanına yeni bir bakış açısı sunarak derin öğrenme alanında çığır açmıştır. ÜÇA, verinin geldiği dağılımı bulmak için optimize edilecek kayıp fonksiyonu yerine öğrenilebilir parametreleri olan bir ayırıcı ağ kullanmak fikrine dayanmaktadır. Bu ayırıcı ağ sınıflandırma görevi yapmaktadır. Böylelikle veri dağılımı öğrenme problemi sınıflandırma problemine dönüştürülmüş olmaktadır. Bu durum ÜÇA modellerinin belirli matematiksel bir kayıp fonksiyonu kullanılarak çözülemeyen veya bu kayıp fonksiyonunun henüz bulunmadığı durumlarda dolaylı olarak bir çözüm sunması açısından çığır açıcı etki göstermektedir. Derin öğrenme alanında öncü bilim adamı Yann LeCun ÜÇA modeli için *"makine öğreniminde son 10 yılın en ilginç fikri"* tanımlamasını yapmıştır [43].

2.1. ÜÇA Modeli

ÜÇA modellerinin çalışması ile ilgili kullanılan analojik örneklere bir yenisi şu şekilde eklenebilir; "*Bir ağın, üretici (generator- Ü) olarak bilinen bir şef ve bir ağın ayırıcı (discriminator- A) olarak adlandırılan seçici bir yemek eleştirmeni olarak düşünülen iki sinir ağı vardır. Şef (Ü), yalnızca görsel olarak çekici değil aynı zamanda özgün tada sahip yemekler üretmeyi amaçlayan yeni tarifler yaratmakla görevlendirilmiştir. Diğer taraftan, yemek eleştirmeni (A), hem şef tarafından hazırlanan yemekleri hem de var olan yemek tarifleri arasından ayrım yapmak amacıyla değerlendirmeden sorumludur. Eğitim süreci boyunca şef ve yemek eleştirmeni eş zamanlı ve çekişmeli bir öğrenme süreci geçirirler. Şef, yenilik yapmaya ve daha gerçekçi tarifler oluşturmaya, yemek eleştirmeni de şefin yemekleri ile var olan yemek tarifleri arasında ayrım yapma yeteneğini geliştirmeye çalışmaktadır. Bu etkileşimli eğitim süreci sonunda, şef benzersiz ve gerçeğe yakın tarifler oluşturma konusunda daha*

becerikli hale gelmektedir. Benzer şekilde yemek eleştirmeni ise oluşturulan yemekler ile mevcut yemek arasında ayrım yapmada uzmanlaşmaktadır."

Şekil 2.1'de genel yapısı verilen ÜÇA, Ü ve A sinir ağlarından oluşmaktadır. Ü ve A, mimari farklılıklar göstermekle birlikte genellikle konvolüsyon katmanlarından oluşan derin sinir ağlarıdır. Ü, girdi olarak bir gürültü vektörü z'yi almaktadır. z, genellikle düşük boyutludur ve çok değişkenli gauss dağılımı gibi belirli olasılık dağılımı $p(z)$ 'den örneklenmektedir. Ü, z girdisinden sahte *(fake)* olarak adlandırılan $\tilde{x}$ verilerini üretmeyi amaçlamaktadır. Öte yandan A'ya girdi olarak, gerçek *(real)* veri kümesi x ve sahte $\tilde{x}$ verileri verilmektedir. A, $\tilde{x}$ ve x verilerini doğru şekilde ayırt etmeyi amaçlamakta ve klasik sınıflandırıcı olarak çalışmaktadır. Sınıflandırma işlemi $\tilde{x}$ ve x verileri üzerinde yapıldığı için bu veriler iki farklı etiketle (*genellikle 0 ve 1 olarak*) belirtilmektedir. ÜÇA'nın eğitim süreci $p(z)$ 'den üretilen $\tilde{x}$'in gerçek veri kümesi x'ten ayırt edilemediğinde sonlanmaktadır.



Şekil 2.1: Klasik ÜÇA mimarisi.

ÜÇA modelinin eğitim aşamasında Ü ve A ağları çekişmeli oyun (adversarial game) olarak adlandırılan süreçten geçmektedir. Bu süreçte ortak kayıp fonksiyonu Ü ve A tarafından farklı yönlerde optimize etmeye çalışılmaktadır. A, gerçek veri kümesi örnekleriyle Ü'nün ürettiği örnekleri yüksek başarıyla sınıflandırmayı amaçlamaktadır. Diğer yandan Ü, A'nın gerçek verilerden ayırt edemeyeceği sahte görüntüleri üretmeye çalışmaktadır. Her iki ağ bu zıt hedeflere ulaşmak için aynı anda sıralı olarak eğitilmektedir. Klasik ÜÇA modelinin kayıp fonksiyonu (2.1)'de verilmiştir. Ü, (2.1)'i maksimize etmeyi, A minimize etmeyi amaçlamaktadır.

$$\mathcal{L}_{\text{ÜÇA}} = \min_{\text{Ü}} \max_{A} \mathbb{E}_{x^i \sim p_d(X)}[logA(x)] + \mathbb{E}_{z^i \sim p(z)}\left[1 - A\left(\text{Ü}\left(z^i\right)\right)\right] \qquad (2.1)$$

(2.1)'de X gerçek verileri, x^i ve z^i sırasıyla gerçek ve sahte veri mini yığınlarını temsil etmektedir. A, (2.1) için gerçek verilerin etiket değerini 1 olarak, sahte verilerin etiket değerini

0 olarak belirlemeye çalışmaktadır. Bu bakımdan klasik ÜÇA için A ağının kayıp fonksiyonu ikili çapraz entropi (binary cross entropy) olarak kullanılmaktadır. (2.1)'de gerçek verilerin sınıflandırıldığı kısma Ü'nün bir etkisi olmamaktadır. Dolayısıyla Ü, eğitimde (2.1)'in sahte verileri sınıflandırdığı kısmı kayıp fonksiyonu olarak kullanılmaktadır. Fakat burada Ü ağının ürettiği veriler, gerçek veri olarak kabul edilir ve 1 olarak etiketlenir. Amaç A ağının 1 ile etiketlenmiş olan sahte verileri 1 sınıfına atamasıdır. Diğer bir deyişle Ü ağının ürettiği sahte verilerle A ağını kandırmayı başarmasıdır. (2.2) ise Ü ağının sadeleştirilmiş kayıp fonksiyonunu göstermektedir. X gerçek veri kümesinin bir p_d dağılımından geldiği kabul edilmektedir. z sahte veri kümesi ise $\mathrm{p}(z)$ gürültü dağılımından oluşturulan p_g olarak temsil edilmektedir. (2.1)'in amacı p_d ve p_g dağılımlarını olabildiğince birbirlerine yaklaştırmaktır. Bu sebeple (2.1) Kullback–Leibler ıraksamasına (Kullback-Leibler divergence- KL) dayanmaktadır.

$$\mathcal{L}_{Ü} = \min_{Ü} \mathbb{E}_{z^i \sim p(z)} \left[1 - A\left(Ü\left(z^i\right)\right)\right] \tag{2.2}$$

Algoritma 2.1: Klasik ÜÇA Modeli [2]

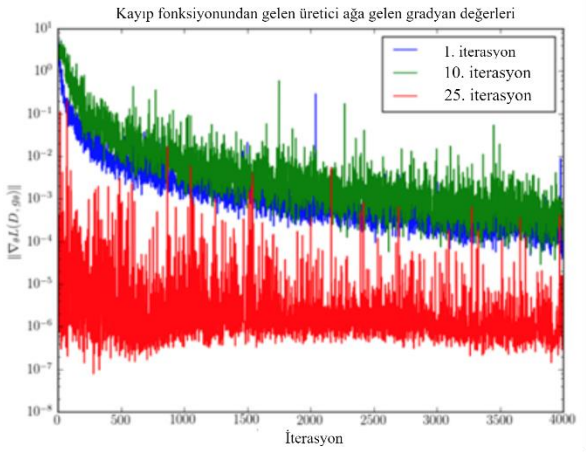
Gerekenler: ω, Üretici ağ ağırlıkları.
θ, Ayırıcı ağ ağırlıkları.
for $iterasyon\ sayısı$ **do**
  for $k\ adım$ **do**
    $p(z)$ öncü dağılımından, m adet veri içeren $z_0, z_1, \dots, z_m$ mini yığınını oluştur.
    $p_d(X)$ eğitim verisi dağılımından, m adet veri içeren $x_0, x_1, \dots, x_m$ mini yığınını oluştur.
    Ayırıcı ağı optimizasyon algoritması stokastik gradyanı artırarak güncelle.
      $\nabla_\omega \left[\mathbb{E}[logA(x)] + \mathbb{E}\left[1 - A\left(Ü(z)\right)\right]\right]$
  end for
  $p(z)$ öncü dağılımından, m adet veri içeren $z_0, z_1, \dots, z_m$ mini yığınını oluştur.
  Üretken ağı optimizasyon algoritması stokastik gradyanı azaltarak güncelle.
    $\nabla_\theta \left[\mathbb{E}\left[1 - A\left(Ü(z)\right)\right]\right]$
end for

Algoritma 2.1. klasik ÜÇA modelinin eğitim aşamasının sözde kodunu göstermektedir. Bahsedildiği üzere ÜÇA'nın eğitim süreci çekişmeli oyun olarak adlandırılmaktadır. Bu sebeple ÜÇA eğitimi klasik konvolüsyon sinir ağlarından daha farklıdır. Ü ve A ağları aynı kayıp fonksiyonu üzerinden farklı yönlerde eğitilmektedir. Eğitim sürecinin ileri aşamasında Ü girdi olarak aldığı z^i gürültü vektörlerinden $\tilde{x}^i$ sahte verilerini üretmektedir. A'ya x^i ve $\tilde{x}^i$ verileri girdi olarak verilmekte ve A'dan sınıflandırma işlemi yapması beklenmektedir. Sınıflandırma işlemi sonucu (2.1)'den elde edilen kayıp değeri ile ilk olarak A ağının parametre değerleri güncellenmektedir. Ardından A ağının parametreleri sabitlenerek $\tilde{x}^i$ sahte verileri gerçek veri etiketiyle A'ya gönderilir. (2.1) üzerinden elde edilen kayıp değeriyle Ü ağının parametre değerleri güncellenir. Eğitim süreci bu şekilde çekişmeli olarak devam eder. ÜÇA

kayıp değerinin Nash equilibrium dengesine ulaştığı durumda (üretilen sahte verilerin gerçek verilerden ayırt edilemediği) eğitim sonlanır [44], [45].

Makine öğrenmesi algoritmalarının kayıp fonksiyonların konveks yapıda olması hedeflenmektedir. Fakat bu durum ÜÇA modellerinde mümkün değildir çünkü ÜÇA'ların kayıp fonksiyonlarında mutlak bir minimum bulunmamaktadır. Şekil 2.1'den görüldüğü üzere Ü'nin eğitimde *A*'dan gelecek kayıp değerine bağımlı olması, eğitim sürecini dengeli hale getirmeyi daha da zorlaştırmaktadır. Sonuç olarak ÜÇA'ların eğitim süreci son derece dengesiz ve optimize edilmesi zor yapıdadır. ÜÇA'ların eğitim sürecini iyileştirmek için çözülmesi gereken iki temel sorun ortaya çıkmaktadır; gradyan sönümlemesi (gradient vanishing) ve mod çökmesi (mode collapse).

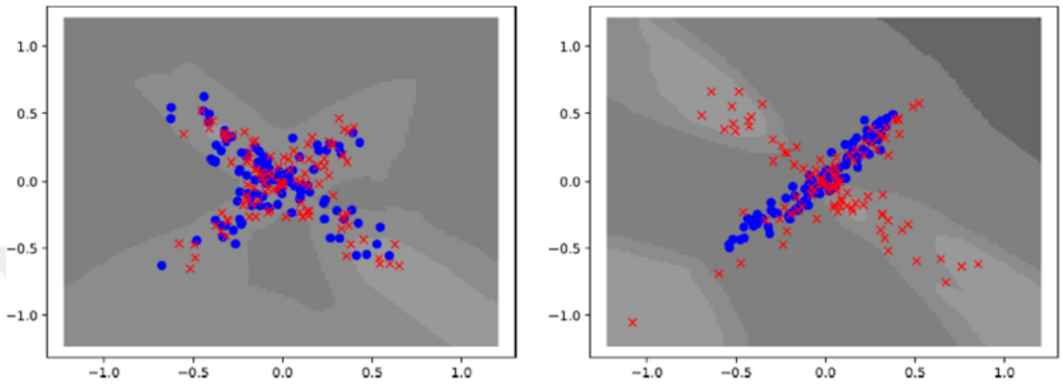
Gradyan sönümlemesi: Eğitim sürecinde *A* ağı çok üst seviyeye çıkarak gerçek ve sahte verileri yüksek başarıyla doğru sınıflandırır hale gelebilmektedir. Bu durumda Ü'nün *A*'yı kandırması imkansızlaşmaktadır. Sonuç olarak, kayıp fonksiyonu sıfıra yaklaşır ve öğrenme yinelemeleri sırasında kaybı güncellemek için gradyan değeri çok küçülür (bkz. Şekil 2.2). Gradyan sönümlemesi olarak adlandırılan durum sonucu Ü, öğrenme yapamaz ve ÜÇA başarısız sonuçlar ortaya çıkarır.



Şekil 2.2: Gradyan sönümlemesi örneği [10]

Mod çökmesi: Eğitim sırasında, Ü sürekli aynı çıktıları ürettiği bir yere doğru yönelebilir. Mod çökmesi olarak adlandırılan bu problem ÜÇA'lar için yaygın bir başarısızlık durumudur. Mod çökmesinde birbirine çok benzer ama başarılı çıktılar üreten Ü, *A*'yı kandırabilmektedir. Fakat Ü karmaşık gerçek dünya veri dağılımını temsil etmeyi öğrenemez ve son derece düşük çeşitliliğe sahip küçük bir alanda sıkışıp kalır. Şekil 2.3'te örnek bir mod çökmesi durumu

görülmektedir. Burada Ü çok sınıflı veya tek sınıflı veri kümesi için tek bir sınıf veya alana sıkışmış verileri üretmektedir. Mod çökmesi gradyan sönümlemesinin bir sonucudur. Mod çökmesinin temel sebebi, *A*'nın bazı verileri çok başarılı sınıflandırması sonucu, o verilerle alakalı kayıp değerinin sıfıra yaklaşmasıdır. Bu durumda Ü ağına gradyan akışı olmamaktadır. Dolayısıyla Ü bazı veri özelliklerini öğrenemez hale gelmektedir.



a) Yakınsama a) Mod çökmesi

Şekil 2.3: Mod çökmesi örneği [10]

ÜÇA'nın eğitim sürecini kararlı hale getirmek ve ÜÇA'yı farklı problemlere uygulayabilmek için birçok yenilik önerilmiştir. Model yapısını geliştiren çalışmalar ÜÇA'nın öğrenme kapasitesini geliştirmeye odaklanmıştır.

Deep Convolutional Generative Adversarial Networks (DCGANs), Radford ve diğ. (2015) tarafından konvolüsyon katmanlarının bulunduğu ÜÇA modeli olarak önerilmiştir. Model, konvolüsyon filtrelerini kullanarak verilerden anlamlı temsiller öğrenmenin zorluğunu çözmeye çalışmaktadır. DCGANs'nın yüksek başarısı klasik ÜÇA'daki [2] tamamen bağlantılı katmanların ortadan kaldırılarak konvolüsyon katmanlarının kullanılmasıyla sağlanmıştır. Bu sayede DCGANs'ın, hiyerarşik ve kullanışlı özellikleri denetimsiz bir şekilde öğrenme becerisini göstermesi daha sonra geliştirilen ÜÇA modellerinin konvolüsyon sinir ağı tabanlı olmasını sağlamıştır [5].

Karras ve diğ. (2018) tarafından önerilen StyleGAN görüntü içeriği ve stilinin kontrolünü ayırmaya odaklanarak daha tutarlı ve gerçekçi sonuçlar üretmektedir. StyleGAN'ın temel yeniliği, stil eşleme ağında yatmaktadır. Bu ağ, gizli uzay vektörünü, oluşturulan görüntünün özelliklerini kontrol eden bir stil vektörüne dönüştürmektedir. Bu şekilde gürültü vektörünün hangi özelliğinin oluşturulan görüntülerde neleri etkilediği kontrol edilebilmiştir [6].

CycleGAN [46] ve DiscoGAN [18] modelleri yenilikçi mimari yapılarıyla görüntü transferi problemlerini oldukça başarılı bir şekilde ele almıştır. Bir veri kümesine ait özelliklerin korunarak başka bir veri kümesine aktarılmasını sağlayan CycleGAN ve DiscoGAN, yüksek çözünürlüklü görüntü ve video üretebilmektedir.

Conditional GANs [8], [23], [47] ve Semi-Supervised GAN [48] veri üretimi için verilerin etiket bilgilerini de kullanmaktadır. Bu ek bilgi kullanımından gelen farklı gradyan değerleri Ü ‘nün üretim kapasitesini olumlu yönde etkilemektedir. Ayrıca Semi-Supervised ÜÇA’da kullanılan yardımcı kayıp fonksiyonları (auxiliary loss) ÜÇA’ların daha spesifik özellikleri öğrenmesine veya eğitim sürecini istikrarlı hale getirmeye yardımcı olabilmektedir.

Klasik ÜÇA’nın [2] kayıp fonksiyonu temel olarak Kullback-Leibler ıraksamasına dayalıdır. KL ıraksamasının matematiksel yapısından ve eğitim sürecinin istikrarsızlıklarından kaynaklı olarak ÜÇA eğitimi için yeterli bilginin sağlanamadığı durumlar olabilmektedir. Bu durumu ele alan M. Arjovsky ve diğ. (2017) klasik ÜÇA [2] yapısındaki kayıp fonksiyonunu iyileştirmek amacıyla, Wasserstein kayıp fonksiyonu kullanılan ÜÇA modelini (Wasserstein Generative Adversarial Network- WGAN) önermiştir [49]. Gerçek ve sahte veri dağılımlarının birbirlerine olan mesafesinin azaltılarak benzerliklerinin artırılması hedeflenmektedir. Earth Mover's uzaklığı (Earth Mover's Distance) kayıp fonksiyonu temelinde çalışan WGAN, mod çökmesi ve kararsızlık ile ilgili sorunları etkili bir şekilde azaltmaktadır.

Progressive GAN [50] ve Laplacian Pyramid Adversarial Networks [51] iteratif olarak küçük boyutlu görüntülerden büyük boyutlu görüntüler üreterek ilerleyen ÜÇA modelleridir. Büyük boyutta ve yüksek kalitede görüntü üretme konusundaki sıkıntılara karşı önerilen Progressive GAN ve Laplacian Pyramid Adversarial Networks oldukça başarılı sonuçlar elde etmiştir. Progressive GAN yalnızca 4×4 boyutundan başlayarak 1024×1024 boyutunda yüksek kalitede görüntüler üretebilmektedir [50].

H. Zhang ve diğ. (2018) tarafından oluşturulan Self-Attention GAN (SAGAN), dikkat mekanizması (attention mechanism) kullanarak yüksek kalitede görüntü üretimi yapabilmesi sebebiyle öne çıkmaktadır [11]. Dikkat mekanizması, her bir pikselin diğer piksellerle olan ilişkisini hesaplamaktadır. Bu durum görüntü içerisindeki uzak mesafelerdeki bağlantıları keşfetmeye ve daha geniş bağlamı anlamaya yardımcı olmaktadır.

ÇÖÇ modelleri eğitim veri kümesi ile gizli uzay arasındaki bağlantıyı eğitim aşamasında öğrenmeyi amaçlayan ÜÇA modelleridir. ÇÖÇ, eğitim veri kümesi ve gizli uzay arasında yalnızca tersine çevirme değil, iki yönlü haritalama imkânı sağlayarak yorumlanabilir ve denetimsiz özellik öğrenme yapmaktadır. ÇÖÇ, ayırıcı ağların veri, gizli uzay çiftlerini incelediği bir çıkarım mekanizmasına dayanmaktadır. ALI [41] ve BiGAN [42] aynı anda ve bağımsız olarak önerilen ÇÖÇ modelleridir. Çalışmalarda ALI'nin denetimsiz bir şekilde ayrımcı özellikler öğrenebileceği, gerçek örneklerin gizli uzaya kodlanabilmesi ve gizli uzaydan çeşitli örneklerin üretilebilmesi gösterilmiştir. Farklı ALI modelleri [52], [53], [54], [55], [56], veri uzayı ile gizli uzay arasındaki ilişkiyi analiz etmekle kalmayıp, aynı zamanda görüntü oluşturma [57], [58], [59], gizli uzayın ayrıştırılması [53], [60] gibi farklı sorunlara çözüm olarak önerilmiştir. Bununla birlikte, bu çalışmalardan hiçbiri özellikle tek bir görüntü üzerinde eğitim yapmak üzere tasarlanmamıştır.

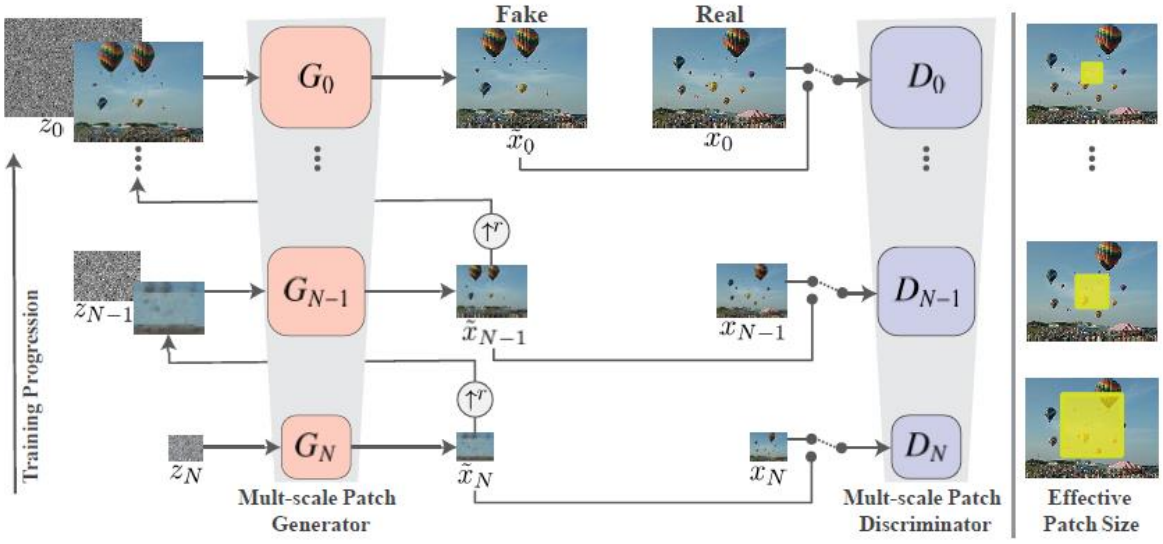
2.2. Tek Eğitim Görüntüsü Kullanan ÜÇA Modelleri

Tek görüntünün eğitim verisi olarak kullanılması ve bu görüntü üzerinden eğitim yapılarak yeni görüntüler üretmek güncel olarak çalışılan problemler arasındadır. Genellikle üretken modeller büyük veri kümeleri üzerinde çalışmaktadır. Bununla birlikte Ulyanov ve diğ. (2019) oluşturdukları Derin iç Öncül (Deep Internal Prior- DIP) modeliyle rastgele ağırlıklarla eğitime başlanan ağların tek eğitim görüntüsünden gerekli ve yeterli bilgileri öğrenebildiklerini ortaya koymuşlardır [61].

Tek eğitim görüntüsüyle çalışan üretken modellerin öncüsü olarak Shocher ve diğ. (2019) tek bir görüntü üzerinde eğitim yapan ve görüntünün dahili istatistiklerini yakalayan denetimsiz bir koşullu ÜÇA olan Internal GAN (InGAN)'yı sunmuştur. InGAN, tümü eğitim görüntüsüyle aynı dahili yama dağılımına sahip, farklı boyutlarda, şekillerde görüntü örneklerini sentezleyebilmektedir [62]. Bununla beraber model görüntü veya görüntü içerisindeki nesneler hakkındaki semantik bilgiye odaklanmamaktadır. InGAN, yalnızca giriş görüntüsünün çok ölçekli yama dağılımını yakalama ve yeniden eşleme işlemlerini yapmaktadır. Bu durum modelin doğal olmayan sonuçlar üretmesine sebep olmaktadır.

Tek görüntü üzerinde yapılan çalışmalarda ilk başarılı koşulsuz ÜÇA modeli T. Shaham ve diğ. (2019) tarafından önerilmiştir [28]. Çalışma tek eğitim görüntüsünü bir bütün olarak ele almak yerine görüntü içerisindeki yamaların farklı boyutlarda öğrenilmesi fikrine dayanmaktadır. Single Image GAN (SinGAN) olarak adlandırılan model konvolüsyon katmanlı ÜÇA'ların

piramit yapıda birbirine bağlanmasıyla oluşturulmuştur. SinGAN mimarisi Şekil 3'te verilmiştir. Model belirli oranda ölçeklenmiş eğitim görüntüsünü girdi olarak alan ÜÇA'ların ölçekli olarak eğitilmesi şeklinde çalışmaktadır. Her ölçekte üretken model bir alt ölçekten gelen üretilmiş görüntüye gürültü eklenmiş girdiyi alır ve yeni boyuttaki görüntüyü üretir. Yalnızca en alt katmanda (en küçük boyutta görüntü üreten katman) bulunan üretken model sadece gürültü girdisiyle görüntü üretir. Modelde her katmanda üretilen görüntüler ile boyutlandırılmış gerçek görüntüler yama ayırıcı (Patch Discriminator- PA) [47] kullanılarak ayırt edilmeye çalışılmaktadır. Şekil 2.4 SinGAN'ın genel yapısını göstermektedir. SinGAN'daki Ü ve A ağları 5 konvolüsyon katmanından oluşmaktadır. Her konvolüsyon katmanı 3x3 filtre, yığın normalizasyonu ve Leaky ReLU aktivasyon fonksiyonundan oluşmaktadır. Ü ağında giriş verisini çıktı verisiyle toplayan artık bağlantı (residual connection) kullanılmıştır. Kayıp fonksiyonu olarak ağırlıklandırılmış WGAN-GP [40] ve $\mathcal{L}_2$ toplamı kullanılmıştır. SinGAN, WGAN-GP'nin ağırlığını 1, $\mathcal{L}_2$ 'nin ağırlığını 10 olarak belirlemiştir. Çalışmada tek bir doğal görüntü içindeki bölümlerin/yamaların (patches) dahili istatistiklerinin, bir üretken modeli eğitmek için yeterli bilgiyi taşıdığı gösterilmiştir [7]. Ek olarak SinGAN, harmonizasyon, görüntü düzenleme, çizimi görüntüye çevirme, yüksek çözünürlük vb. farklı uygulamalarda da kullanılmış ve başarılı sonuçlar elde edilmiştir.



Şekil 2.4: SinGAN mimarisi.

T. Hinz ve diğ. (2021), SinGAN modelini temel alarak oluşturdukları ConSinGAN modelini oluşturmuştur. SinGAN modeline benzer olarak piramit yapıda olan ConSinGAN, eğitim şekli ve üretken ağ içyapısı farklılığıyla öne çıkmaktadır. ConSinGAN, her ölçekte üretilen görüntü yerine daha fazla bilgi içeren özellik haritalarını bir sonraki ölçeğe girdi olarak gönderir.

Bununla birlikte çalışmada Progressive GAN [50] modeline benzer şekilde yalnızca son katmanı eğitilmemektedir. ConSinGAN belirli sayıda ölçeği farklı öğrenme oranı hiper parametresi kullanarak eğitmektedir [32]. ConSinGAN küçük boyutlu görüntülere daha fazla önem verecek bir ölçeklendirme yöntemi kullanılmıştır. Ölçeklendirme oranının daha küçük olması model eğitiminin daha hızlı olmasını sağlamıştır. Çalışmada ulaşılan nicel ve nitel sonuçlar ConSinGAN'ın başarılı olarak tek görüntü verisinden öğrenebildiğini göstermektedir [32].

S. Mahendren ve diğ. (2023) tarafından yapılan çalışmada SinGAN modeline öz-dikkat mekanizması [11] ve çekişmeli geri bildirim (Adversarial Feedback- AF) [63] eklemeleri yapılmıştır. Çalışmada modelin öğrenme kapasitesinin artırılması ve daha gerçekçi görünen görüntüler üretmeye odaklanılmıştır. Görüntü içerisindeki yerel ve global bağımlılıkların tespit edilmesi için öz-dikkat mekanizması kullanılmıştır. Bununla birlikte AF mekanizmasıyla ayırıcı ağın çıktı özellik haritası üretken ağa girdi olarak verilerek modelin gerçekçi görüntü üretme kapasitesi artırılmaya çalışılmıştır [33].

Zhang ve diğ. (2021) yaptıkları çalışmada ExSinGAN modelini önermiştir, Çalışmada tek eğitim görüntüsünün yapısal (Structure), anlamsal (Semantic) ve dokusal (Texture) olmak üzere üç parçadan oluştuğu kabul edilmiştir. Bu sebeple modelde yapı, anlam ve doku ile ilgili dağılımların ardışık öğrenilmesi yoluyla koşullu dağılımın modellenmesini kolaylaştıran hiyerarşik bir çerçeve kullanılmıştır. ExSinGAN'ın kullandığı üç ÜÇA: yapısal ÜÇA, anlamsal ayrıntıları öğrenen anlamsal ÜÇA ve ince doku ayrıntılarını ekleyen dokusal ÜÇA şeklindedir. Buna ek olarak, ExSinGAN kayıp fonksiyonu olarak WGAN-GP ve $\mathcal{L}_2$'ye ek olarak ÜÇA ters çevirme'den yararlanmıştır. Çalışmada ExSinGAN, görüntü sentezi, görüntü düzenleme, harmanlama ve resme boyama görevlerinde, niteliksel ve niceliksel olarak DGP, SinGAN, ConSinGAN ve HP-VAEGAN ile karşılaştırılmıştır [35].

Sushko ve diğ. (2021), yaptıkları çalışmada eğitim görüntüsü bağlamını korurken tek bir görüntüden veya videodan çeşitli örnekler oluşturan One-Shot GAN'ı tanıtmıştır. Görüntü üretimi için piramit şeklinde çoklu ÜÇA kullanmak yerine tek ÜÇA kullanılan çalışmada, üretken ağ yapısı geliştirilerek daha kaliteli görüntü üretimi hedeflenmiştir. Ayrıca fazla uydurmayı önlemek ve üretken ağa daha iyi bilgi akışı sağlamak için eğitim görüntüsünün içerik dağılımını ve gerçekçiliğini ayrı ayrı değerlendiren iki dallı bir ayırıcı mimari oluşturulmuştur. One-Shot GAN, tek görüntü ve video üretme senaryolarında SinGAN ve FastGAN modelleriyle niceliksel ve görsel olarak karşılaştırılmıştır [64].

Chen ve diğ. (2023) MOGAN modelini önerdikleri çalışmalarını yayınladılar. MOGAN, bir resimden farklı resimler üretirken nesnelerin orijinal yapılarını doğru bir şekilde korumaya odaklanmaktadır. MOGAN, hiyerarşik olarak düzenlenmiş ancak farklı özelliklere sahip iki paralel ağ içermektedir. Bu ağlar sırasıyla ilgi alanlarının (Region of interest- ROI) ve görüntü arka planın üretilmesinden sorumludur. ROI ağı için model, artırılmış eğitim resminden morfolojik bilgi kullanmaktadır. MOGAN, artırılmış eğitim görüntüsünü yapı ve görünüş bilgisine dönüştüren bir stil ekleyici kullanmakta ve nesne üretim stilini affine dönüşümlerle kontrol etmektedir. Ayrıca görüntü arka plan ağı için MOGAN, eğitim görüntüsü arka planını elde etmek için eğitim görüntüsüne ikili bir maske uygular. Böylece ROI ağının odaklandığı nesneler korunur ve maskelenmiş arka plan görüntüsü üretimine odaklanılır. Son olarak, MOGAN, bu ağlar tarafından üretilen sonuçları birleştirerek tam bir resim oluşturur. Çalışmada, model ROI tabanlı rastgele resim üretimi, resim düzenleme ve tek resim animasyonu gibi uygulamalarda test edilmiştir. Ayrıca görüntü üretim sonuçları nitel ve nicel olarak değerlendirmiş ve modelin kalite ve çeşitlilik açısından diğer modellerle karşılaştırılmıştır [30].

X. Chen ve diğ. (2021) yayınladıkları çalışmada, SA-SinGAN'ı önermiş ve üretken ağın farklı bölgelerine bir öz-dikkat mekanizması eklenmesi durumunda modeldeki değişiklikleri tartışmışlardır. Eğitim istikrarsızlığı sorununu daha iyi çözmek amacıyla üretken ve ayırıcı ağlarda için spektral normalizasyon kullanmıştır. Çalışmada bununla birlikte ağın öğrenme oranının ağ üzerindeki etkisi analiz edilmiştir. SA-SinGAN nitel ve nicel olarak değerlendirme yöntemleriyle üç farklı veri kümesinde test edilmiş ve sonuçlar SinGAN ve ConSinGAN ile karşılaştırılmıştır [36].

Zhang ve diğ. (2022) tek bir resim üretim problemine düzenlenmiş bir gizli değişken modeli ile formülize ettikleri PetsGAN modelini önermiştir. PetsGAN, SinGAN modelinin piramit yapısı yerine tek ölçekli eğitim prosedürünü kullanması açısından öne çıkmaktadır. Bu amaçla PetsGAN tek eğitim görüntüsünün dahili (internal) ve harici (external) öncü bilgilerini daha iyi öğrenmek amacıyla Deep External Prior Network (DEPNet), Deep Internal Prior Network (DIPNet) ve Co-learning ağlarını içermektedir. Bu ağların sıralı eğitilmesi, yüksek kaliteli görüntülerin üretilmesine kolektif olarak katkı sağlamaktadır. PetsGAN'da tek ölçekli bir yapı kullanılmasına rağmen DIPNet alt ağının içerisinde ölçeklendirme için piksel karıştırma yöntemi [65] kullanılmaktadır. Bu durum eğitim görüntüsünün DIPNet içyapısı içinde yeniden boyutlandırmak için kullanıldığını ifade etmektedir. Ayrıca modelde, ön eğitilmiş BigGAN

[66] modeli kullanılarak ÜÇA tersine çevirmeden ve eğitim görüntüsü çoğaltmadan (jitter) yararlanılmıştır. Böylelikle eğitim başlangıcında daha iyi bir gürültü verisinin kullanılması hedeflenmiştir. Çalışmada, modelin resim kalitesi, çeşitlilik ve eğitim hızı açısından avantajını göstermek için üç veri kümesinde nitel ve nicel deneyler gerçekleştirilmiş ve model diğer çalışmalarla karşılaştırılmıştır. Ayrıca, PetsGAN resimden resme boyama, stil transferi, resim düzenleme ve uyumlaştırma uygulamalarında da test edilmiştir [67].

Wang ve diğ. (2022) oluşturdukları IMAGINE, görüntü üretimi sırasında semantik özgünlük kısıtlamalarını öğrenmeye, üretken ağ eğitimi olmadan gerçekçi görüntüler üretmeye ve kullanıcılara üretim süreci üzerinde sezgisel kontrol sağlamasına olanak tanımaktadır. IMAGINE kayıp fonksiyonu olarak çekişmeli kaybına ek olarak toplam varyans (total variation) kaybı [68] ve özellik dağılımı eşleştirme düzenleyicisi (feature distribution matching regularizer) [69] içermektedir. Bu ek kayıp fonksiyonları, IMAGINE'nin eğitim sürecine katkıda bulunmakta ve üretilen çıktının geliştirilmesine yardımcı olmaktadır. Üretilen görüntülerde belirli oranda kontrol sağlamak amacıyla IMAGINE, ön eğitilmiş BigGAN ağını kullanarak ÜÇA tersine çevirmeyle eğitim görüntüsünün öncül bilgilerinden yararlanmaktadır. Yapılan deneysel çalışmalarda, IMAGINE'ın nesneler, sahneler ve dokular gibi üç farklı görüntü alanında, ÜÇA tabanlı ve tersine çevirme tabanlı yöntemlerle nitel ve nicel olarak karşılaştırılmıştır [31].

Gur ve diğ. (2020) tek görüntü eğitimi için değişimsel otokodlayıcı (Variational Autoencoder-VAE) ve ÜÇA mimarilerini birleştirdikleri HP VAE-GAN modelini ortaya koymuştur. HP VAE-GAN, görüntü üretimiyle birlikte yeni bir hiyerarşik video üretimi de yapabilmektedir. Bunun için modelde yüksek çeşitlilikte çıktı elde etmek amacıyla kaba ölçeklerde, yama-VAE (Patch-VAE), yüksek ölçeklerde yama-ÜÇA modeli kullanılmıştır. Oluşturulan bu hibrit yöntem tek görüntüden gerçekçi videolar üretme açısından öne çıkmaktadır. Çalışmada elde edilen deneysel sonuçlar HP VAE-GAN modelinin başarılı şekilde çalıştığını ortaya koymaktadır. Bununla beraber gözetimsiz olarak eğitim yapılan modelin üretilen görüntüler için semantik anlayışı bulunmamaktadır [70].

Granot ve diğ. (2021) yaptıkları çalışmada ÜÇA modeli kullanmadan tek eğitim görüntüsü kullanarak yeni görüntü üreten Generative Patch Nearest-Neighbor (GPNN)'ni önermiştir. GPNN, eğitim görüntüsü içerisindeki yama görüntülerinin benzerlik oranlarını en yakın komşuluk aramasını kullanarak basitçe kaba-ağdan-inceye ölçmektedir. Daha sonra benzer yamaların görüntü içerisine yeniden yerleştirilmesiyle görüntü üretimi yapılmaktadır. GPNN

bu yaklaşım sayesinde eğitim aşamasında herhangi bir parametre optimizasyonu yapmamaktadır. Bu durum GPNN'in oldukça hızlı rastgele yeni görüntüler üretmesine olanak sağlamaktadır. Çalışmada GPNN, resim düzenleme ve yeniden düzenleme, farklı boyutlara yeniden boyutlandırma, yapısal benzetmeler, resim kolajı ve koşullu dolgu uygulama problemlerinde denenmiştir [71]. Bununla beraber GPNN'in herhangi bir üretken eğitim yapısına sahip olmaması (ÜÇA'larda olduğu gibi sürekli yama dağılımı öğrenmek yerine ayrık yama dağılımı öğrenmesi) üretilen yeni görüntülerin yapısal yeniliklerden yoksun olmasına ve kısıtlı çeşitlilik göstermesine sebep olmuştur.

Gandelsman ve diğ. (2020) önerdikleri Double-DIP modeli ile tek eğitim görüntüsü kullanarak görüntü segmentasyonu (ön plan ve arka plan katmanlarına ayrılması); şeffaf katman ayrımı (yansıma ve iletim katmanlarına ayrılması); görüntü netleştirme (net bir görüntü ve sis haritasına ayrılması) gibi görüntü manipülasyonu uygulamalarına odaklanmışlardır. Yapılan deneysel çalışmalarda modelin görüntü manipülasyon uygulamalarında başarılı şekilde çalıştığı ortaya konulmuştur [72]. Bununla beraber modelin koşulsuz görüntü üretimi için uygun olmaması, çalışmanın genellenebilirliği ve farklı problemlere uygulanabilirliğini kısıtlamaktadır.

Asano ve Diğ. (2020) yaptıkları çalışmada kendi kendine gözetimli (self-supervised) öğrenme ve veri artırma ile birleştirildiğinde, sadece tek bir görüntünün bile, standart derin ağların ilk birkaç katmanını milyonlarca görüntü ve tam denetim kullanarak öğrenmeye yeterli olduğunu göstermişlerdir. Çalışmada derin ağların ilk katmanlarının ağırlıklarının doğal görüntü istatistikleri hakkında sınırlı bilgi içermekte olduğu ve bu düşük seviye istatistiklerin güçlü denetimle olduğu kadar kendiliğinden denetimle de öğrenilebilir olduğu ortaya konulmuştur. Ayrıca düşük seviye istatistiklerin büyük bir görüntü veri kümesi yerine sentetik dönüşümlerle yakalanabildiği deneysel çalışmalarla gösterilmiştir. Çalışmada ayrıca konvolüsyon sinir ağlarının derin katmanlarının milyonlarca görüntü eğitimi kullanılsa bile, kendi kendine gözetimli eğitimden geride kaldığı belirtilmiştir [29].

Vinker ve diğ. (2021) tek bir görüntüye dayalı koşullu görüntü manipülasyonu için DeepSIM adlı bir üretken model oluşturdular. Bunun için veri çoğaltma yöntemlerinden olan thin-plate spline yöntemi kullanılarak model performansı artırılmıştır. DeepSIM, görüntünün ilkel bir temsili (tek eğitim görüntüsünün çizimi) ile görüntü arasında bir eşlemeyi öğrenir. İlkel bir temsilin seçimi, manipülasyonların kolaylığını sağlamaktadır. Test aşamasında manipülasyon için üretken ağ ilkel giriş temsilini değiştirerek ve eşleme yaparak karmaşık görüntü

değişiklikleri yapabilmektedir. Ayrıca çalışmanın koşulsuz görüntü üretimi problemine çözüm getirmemesi, modelin kullanabilirliğini kısıtlamaktadır [73].

Vinker ve diğ. (2021) oluşturdukları AugurOne modeli, girdi görüntüsünün ölçeklendirilmiş bir versiyonunu yüksek çözünürlükte yeniden yapılandırarak (tek resim süper çözünürlüğü gibi) yükseltmeyi öğrenmektedir. Çalışmada tek bir görüntünün yeterli çeşitliliği sunmadığı iddia edilerek, eğitim görüntüsü doğrusal ve doğrusal olmayan arttırmalarla çoğaltılmıştır. Çalışmada model eğitimi için çekişmeli kayıp yerine algısal yeniden yapılandırma (perceptual reconstraction) kayıp fonksiyonu kullanılmıştır. AugurOne görüntü üretimi aşamasında ise eğitim görüntüsünün gizli uzay özelliğinden yararlanmak amacıyla değişimsel otokodlayıcı kullanmaktadır. Model koşulsuz görüntü üretiminde ve harmanlama, görüntü animasyonu, çizimden görüntüye gibi uygulamalarda başarılı sonuçlar ortaya koymuştur [74].

ÜÇA'ları çok az veri kullanarak eğitmek, genellikle aşırı uydurma problemine yol açmaktadır. Kerras ve diğ. (2020) çalışmalarında sınırlı veri koşullarında eğitimi önemli ölçüde kararlı hale getiren adaptif bir ayırıcı veri arttırma mekanizması önermişlerdir. Fakat önerilen yaklaşım tek eğitim görüntüsü kullanmak yerine sınırlı sayıda eğitim görüntüsü kullanmaktadır. Bununla beraber elde edilen sayısal sonuçlar modelin başarılı şekilde çalıştığını ortaya koymaktadır. Bu durum çalışmadaki yaklaşımın tek eğitim görüntüsü içinde kullanılabilmesinin araştırılması için umut vadetmektedir [26].

Denetimsiz görüntüden görüntüye çeviri (image to image translation) görevi, eşleşmiş görüntüler içermeyen iki alan arasında bir eşleme öğrenme işlemiyle ilgilenmektedir. Lin ve diğ. (2020) tek eğitim görüntüsü çifti kullanarak görüntüden görüntüye çeviri probleminin çözümü için TuiGAN modelini önermişlerdir. TuiGAN simetrik iki adet koşullu üretken ve ayırıcı ağı piramit yapıda kullanarak çalışmaktadır. Bu bağlamda TuiGAN, CycleGAN'ın tek görüntü için SinGAN'a benzer şekilde değiştirilmiş hali olarak görülebilir. Çalışmada kayıp fonksiyonu olarak dört kayıp fonksiyonun ağırlıklandırılmış toplamı kullanılmıştır; ÜÇA eğitimi için WGAN-GP kayıp fonksiyonu, mod çökmesini minimize etmek amacıyla Cycle-Consistency kayıp fonksiyonu, renk ve doku uyumunu sağlamak için Identity kayıp fonksiyonu ve düzenlileştirme için toplam varyans kayıp fonksiyonu. Çalışmada TuiGAN, nicel, nitel ve kullanıcı deneyleri yapılarak değerlendirilmiş ve modelin başarılı şekilde çalıştığı ortaya konulmuştur [75].

Haim ve diğ. (2022) tek eğitim videosundan yeni videolar üretme problemi için yeni bir model önermişlerdir. Bu amaçla çalışmada klasik uzay-zaman yama-en-yakın-komşular yaklaşımı kullanılmış ve öğrenme olmadan ölçeklenebilir bir koşulsuz üretken model oluşturulmuştur. Bu amaçla video-yama-yakın-komşu (Video-Patch-Nearest-Neighbor) modülleri piramit yapıda oluşturulmuş ve eğitim videosu girdi ölçeklendirilmiş olarak kullanılmıştır. Çalışmada farklı video üretimi dışında, aynı çerçeve kullanılarak video benzetimleri ve uzaysal-zamansal yeniden boyutlandırma gibi diğer uygulamalar da gerçeklenmiştir. Önerilen yaklaşım Full-HD videolara ölçeklenebilmektedir. Yapılan sayısal ve kullanıcı çalışmaları sonuçları, derin ağlar içermeyen klasik yaklaşımların doğru bir şekilde uyarlama yapıldığında önemli ölçüde başarı gösterdiğini ortaya koymaktadır. Bununla beraber çalışmada önerilen model yerel yamaların üretilmesinde global geometrik tutarlılık eksikliğine yol açması modelin geliştirilmesi gerektiğini göstermektedir [76].

Bensadoun ve diğ. (2021) tek eğitim görüntüsünün dahili istatistiklerini daha etkili bir şekilde modellemek için bir meta-öğrenme (meta-learning) yaklaşımı önermişlerdir. Sunulan çalışma meta-öğrenme yaklaşımının kullanılması açısından öne çıkmaktadır. Önerilen modelin nicel ve nitel deney sonuçları meta-öğrenme yaklaşımının tek eğitim görüntüsü için başarılı olabileceğini ortaya koymaktadır. Çalışmada ayrıca önerilen yaklaşım çeşitli uygulamalar üzerinde test edilmiş ve başarılı sonuçlara ulaşılmıştır. Bununla beraber önerilen modelin tek eğitim görüntüsü üzerinde çalışan mevcut ÜÇA modellerine oranla çok fazla parametre sayısına sahip olması ve yüksek hafıza kapasitesine gereksinim duyması kullanılabilirliğini kısıtlamaktadır [77].

Yang ve diğ. (2023) GenDA modelini önererek büyük veri kümesi üzerinde önceden eğitilmiş bir ÜÇA modelini sadece bir hedef görüntü kullanarak yeni bir alana aktarmayı amaçlamaktadır. Bunun için ön eğitilmiş üretken ağa hedef görüntülerin özniteliklerini çıkarmak için eğitilebilir basit bir sinir ağı eklenmiştir. Benzer şekilde ayırıcı ağın üretken ağdan gelen görüntülerin uygun karakterlerini yakalamasını sağlamak için öznitelik sınıflandırıcı eklenmiştir. GenDA modelinde, geleneksel ÜÇA eğitimine benzer şekilde üretken ve ayırıcı ağa eklenen bu iki hafif ağı çekişmeli şekilde eğitilmektedir. GenDA'nın girdi görüntüsünün farklı hedef alanlara adaptasyonunu başarılı bir şekilde sağladığı deneysel sonuçlarla gösterilmiştir. Çalışmada önerilen modelin derin görüntü sahteciliğinin tespiti için kullanılabileceği vurgulanmıştır [78].

Zheng ve diğ. (2021) üretken model ailesi içerisinde yer alan enerji tabanlı (energy based) modellerin tek eğitim görüntüsü için tasarlanmasını sağlamışlardır. Çalışmada önerilen model SinGAN'a benzer olarak piramit yapıda eğitilen enerji tabanlı ağlardan oluşmaktadır. Her enerji ağı piramidi, farklı çözünürlüklerdeki yamaların dağılımlarını yakalamak için kullanılmaktadır. Önerilen model, eğitimi desteklemek için bir yardımcı model (örneğin, ayırıcı) gerektirmez ve iç istatistikleri öğrenmeyi ve resim üretmeyi tek bir çerçevede birleştirir. Çalışmada süper çözünürlük, resim düzenleme, stil transferi vb. gibi çeşitli görüntü oluşturma ve manipülasyon görevlerinde sunulan deneysel sonuçlar, önerilen modelin etkinliğini göstermektedir. Bununla beraber model çıkarım aşamasında yeni örnekler oluşturmak için aynı alanda çalışan ÜÇA modellerine oranla çok daha uzun sürede çalışmaktadır [79].

Xu ve diğ. (2021) konvolüsyon ağların doldurma (padding) işleminin tek eğitim görüntüsündeki özellikleri öğrenmedeki etkisini geniş çapta incelemişlerdir. Bu bağlamda SinGAN modeli farklı doldurma işlemleriyle çalıştırılmış ve sonuçları ortaya konulmuştur. Buna göre sıfır doldurmanın pozisyonel kodlamayı tanıttığı gösterilmiştir. Çalışmada SinGAN modelinin çok yönlülüğünü ve sağlamlığını artırmak için cartesian grid ve spatial positional encoding olmak üzere iki yeni doldurma yöntemi gösterilmiştir [80].

Wu ve diğ. (2022) gerçek görüntünün var olmadığı, görüntüleri iç boyama yapan ECycleGAN modelini önermiştir. Kör görüntü iç boyama olarak adlandırılan problemin çözümü için CycleGAN mimarisini temel alan model elde ettiği sonuçlarla başarılı çıktılar oluşturmuştur. ECycleGAN'da kullanılan doldurma tekniklerinin etkinliğini doğrulamak için gerçek gürültü ve süper çözünürlük veri kümeleri üzerinde kapsamlı deneyler yapılmıştır [81].

Li ve diğ. (2022) Ganimator adlı bir üretken modeli oluşturmuşlardır. Bu model tek ve kısa bir hareket dizisinden yeni hareketleri sentezlemeyi öğrenmektedir. SinGAN modelinin yapısına sadık kalan Ganimator, girdi verisinin görüntü yerine hareket dizgisi olması sebebiyle öne çıkmaktadır. Çalışmadaki strateji, veri toplama sürecini büyük ölçüde basitleştirirken, önerilen model gerçekçi varyasyonlar oluşturabilmekte ve bireysel hareket dizisinin detaylarını başarılı bir şekilde yakalayabilmektedir. Ganimator, aynı temel hareketle birlikte olası varyasyonlarla daha uzun hareket dizilerinin sentezlenmesi, kontrol edilebilir hareket sentezi, anahtar kare düzenleme ve interpolasyon ve stil transferi gibi çeşitli uygulamalarda başarılı sonuçlar elde etmiştir [82].

Rodriguez-Pardo ve Garces (2022) önerdikleri SeamlessGAN modeliyle, tek eğitim görüntüsünden döşenebilir (tileable) doku haritalarını otomatik olarak oluşturmaktadırlar. Sadece üretme sorununu çözmeye odaklanan benzer yöntemlerin aksine çalışmada hem üretim hem de döşenebilirlik sorunları aynı anda ele alınmıştır. SeamlessGAN temelde, karşıt genişleme teknikleri kullanılarak eğitilmiş bir üretken ağ içindeki gizli bir alanı döşemenin, daha sonra merkezi alanı kırparak döşenebilir görüntülere dönüştürülebilen dikiş kesişiminde sürekliliğe sahip çıktılar üretmektedir. SeamlessGAN benzer yöntemlerle karşılaştırılmış farklı türdeki dokular için çalıştığı niteliksel ve niceliksel olarak gösterilmiştir [34].

Bell-Kligler ve diğ. (2019) sadece düşük çözünürlüklü test görüntüsü üzerinde eğitilerek dahili istatistikleri öğrenen KernelGAN modelini önermişlerdir. KernelGAN'de çapraz-ölçek tekrarlanma özelliğinin derin dahili öğrenme alanına uyarlanması gösterilmiştir. Böylelikle yüksek çözünürlüklü görüntüsü bulunmayan düşük çözünürlüklü görüntülerin kör öğrenme yapılarak yüksek çözünürlüğe çevrilmesi mümkün kılınmıştır [83].

Benaim ve diğ. (2021) görüntüden görüntüye çevirme problemi için tek eğitim görüntüsü çifti için çalışan bir model önermişlerdir. Önerilen model çevirme için eğitim görüntüsü çiftlerinin farklı ölçeklerdeki yamalarını eşleştirmeye çalışmaktadır. Eşleştirilen bu yapısal benzer yamalar çekişmeli ve yeniden yapılandırma kayıp fonksiyonlarıyla beraber döngü (cycle) kayıp fonksiyonu yardımıyla çevirme işlemine tabi tutulur. Önerilen model SinGAN'ın görüntüden görüntüye çevirme problemine uyarlaması şeklindedir. Elde edilen test sonuçları ve yapılan farklı uygulamalar modelin başarılı şekilde çalıştığını ortaya koymaktadır [84].

Schubert ve diğ. (2021) tek simge (token) eğitim verisi kullanarak video oyunu içeriği üreten TOAD-GAN modelini önerdiler. Girdi simge verisini hiyerarşik bir düzen içerisinde işleyen model her ölçekte oyun içeriğinin farklı bileşenlerini önceki ölçekten gelen çıktılara eklemektedir. TOAD-GAN, Super Mario Bros. oyunu için verilen girdi simge verisi üzerinde test edilmiş ve gerçekçi oyun içeriği üretmeyi başarmıştır. Çalışmanın bu yönüyle video oyunları sektörü için kullanılabilecek hafif boyutlu modellere öncü olma potansiyeli taşımaktadır [85].

Barahona-Ríos ve Collins (2022) tek birimlik ses eğitim verisinden yeni sesler üreten SpecSinGAN modelini tanıtmışlardır. SpecSinGAN tek bir ses efektini (örneğin, bir adım atma veya karakterin zıplaması gibi) eğitim verisi olarak kullanarak bu ses efektinin farklı versiyonlarını üretmektedir. SpecSinGAN, ConSinGAN mimarisini temel alarak

oluşturulmuştur. Ses inandırıcılığı ve çeşitliliği açısından gerçek kayıtlar ve dijital işaret işleme ses modelleri ile karşılaştıran kullanıcı çalışması ve çok kanallı spektrogramlar ile yapılan nicel ölçümler SpecSinGAN'ın diğer prosedürel ses modellerinden daha gerçekçi ve çeşitli olduğunu göstermektedir [86].

Zhong ve Zhao (2021) düşük çözünürlüklü tek eğitim görüntüsünden yüksek çözünürlüklü çıktı elde eden GAN modelini tanıttılar. Model SinGAN'ı modifiye ederek yüksek çözünürlüklü görüntü üretmede başarı göstermiştir. Görüntü içerisindeki bağımlılıkları kullanmak amacıyla kanal ilgi mekanizması üretken ağda kullanılmıştır. Kayıp fonksiyonu olarak çekişmeli kayıp ile birlikte ön eğitilmiş VGG ağından elde edilen özellik haritalarının $\mathcal{L}_1$ farkı kullanılmıştır [87].

Abu-Hussein ve diğ. (2022) görüntü restorasyonu için tek eğitim görüntüsü kullanan P-GSURE modelini önermişlerdir. Yaklaşım, gerçek hatayı tahmin eden genelleştirilmiş SURE adlı önyargısız bir MSE tahmincisine dayanmaktadır. Elde edilen sonuçlar P-GSURE'nin, en küçük kareler kaybına dayalı olarak ağları eğiten DIP stratejisinden daha iyi sonuçlar verdiğini göstermiştir [88].

Shocher ve diğ. (2020) önceden eğitilmiş bir sınıflandırma modeli tarafından öğrenilen derin özellikleri kullanan yeni bir ÜÇA tabanlı model sunmuşlardır. Piramit yapıda birleştirilen ön eğitilmiş sınıflandırıcı ve ÜÇA'lardan oluşan model, bir referans görüntüden çıkarılmış özellik kümesi verildiğinde, her birinin sınıflandırma modelinin her bir semantik düzeyinde eşleşen özelliklere sahip olan çeşitli görüntü örnekleri üretir. Model eğitiminde çekişmeli ve yeniden yapılandırma kayıplarının yanında üretilen görüntülerin çeşitliliğini kontrol etmek için yeni bir kayıp fonksiyonu daha kullanılmıştır [89].

Arora ve Lee (2021) gif verisi üreten SinGAN-GIF modelini oluşturmuşlardır. Çalışmada görüntü tabanlı SinGAN modeli gif veya kısa video oluşturmak için geliştirilmiştir. SinGAN-GIF, hem gif içindeki görüntü yamalarının dağılımını hem de hareket desenlerini modellemektedir. Bu amaçla, temporal bilgiyi modellemek için piramit yapıdaki 3D ve 2D konvolüsyon ağları beraber kullanılmıştır. Ulaşılan sonuçlar SinGAN-GIF'in, tek eğitim verisinden için farklı mekânsal çözünürlüklerde veya hızlarda benzer görünümlü video örnekleri üretebildiğini, hareket transferi gibi video uygulamalarında kullanılabildiğini göstermiştir [90].

Wu ve Zheng (2022) çalışmalarında tek eğitim verisinden 3D görüntü üretmek için SinGAN modelini modifiye etmişlerdir. Önerilen model giriş şeklinin geometrik özelliklerini çeşitli mekânsal ölçeklerde yakalamak üzere tasarlanmış çok ölçekli bir GAN tabanlı modeldir. Çalışmada 3D verinin hesaplama maliyetini azaltmak amacıyla girdi verisi 2D düzlemlerde ifade edilerek 2D konvolüsyon filtreleriyle işlem yapılmıştır [91].

Son ve diğ. (2023) yaptıkları çalışmada yalnızca birkaç tane 3D giriş görüntüsüyle eğitilerek farklı 3D görüntüler üreten SinGRAF modelini oluşturmuşlardır. Bu çerçevede model, tek bir sahnenin birkaç görüntüsünü giriş olarak alır ve StyleGAN tabanlı ÜÇA modeli kullanılarak 3D çıktı üretmektedir [92].

3. YÖNTEM

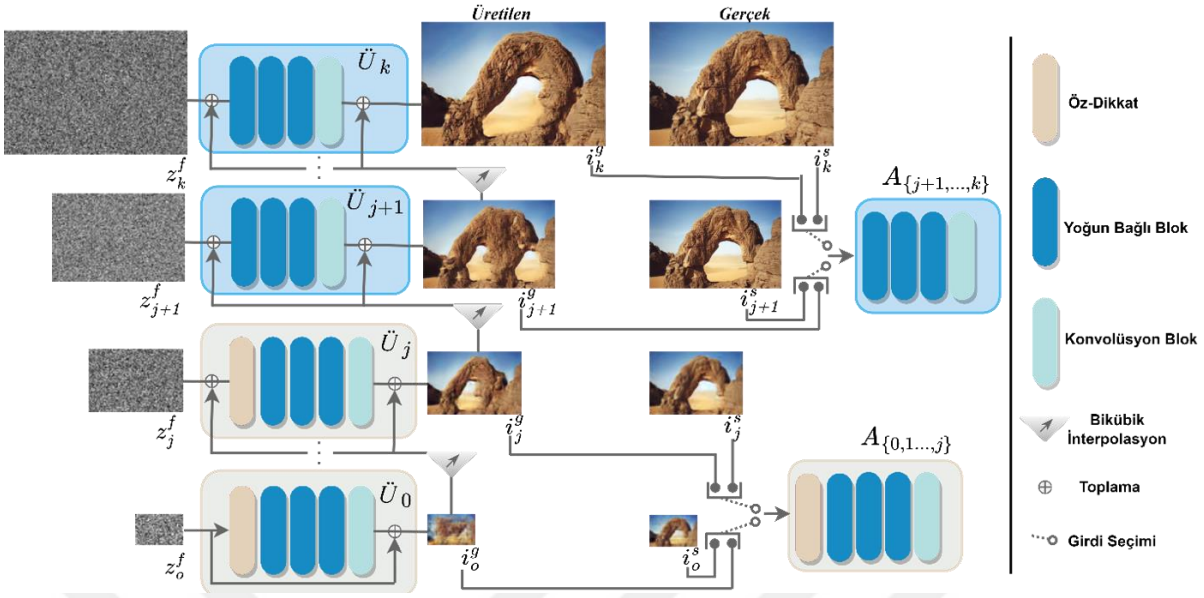
Bu bölümde tez çalışması kapsamında tek görüntü üretim problemi için geliştirilen iki tek görüntü ÜÇA modeli tanıtılmıştır. Geliştirilen tek görüntü ÜÇA modelleri, Bölüm 3.1 ve Bölüm 3.2'de ayrıntılı olarak açıklanmıştır.

3.1. Model 1 (Öz-dikkat ve Yoğun Bağlantılı ÜÇA)

Model 1, tek görüntü ÜÇA modeli olarak mimari yapıya ve kayıp fonksiyonuna odaklanmaktadır. Amaç, bu iki bileşeni daha etkili şekilde kullanarak gelişmiş bir tek görüntü ÜÇA modeli ortaya koymaktır. Tek görüntü ÜÇA modellerinin üretici ağının katman sayısını veya katmanların konvolüsyon filtre sayısını artırmak, aşırı uyum (ezberleme) sorununa neden olabilmektedir. Bu soruna çözüm olarak Model 1'de yoğun bağlantılı blok katmanları kullanılmıştır. Yoğun bağlantılı bloklar kullanmak hem aşırı uyumu önlemekte hem de tek eğitim görüntüsünden öğrenilen özellik haritalarının farklı katmanlarda kullanılmasını sağlamaktadır. Bu şekilde, ayırıcı ağdan gelen gradyan akışını sürekli ve daha bilgilendirici hale getirmek amaçlanmaktadır. Üretilen görüntülerdeki piksel bağımlılıklarının dikkate alınmaması durumunda nesnelerin bütünlük bilgisinden yararlanılamaz ve gerçekçi görüntüler elde edilemez. Bu sorunu çözmek için, Model 1'de öz-dikkat mekanizmasından yararlanılarak tek eğitim görüntüsündeki yerel ve global bağımlılıkların öğrenilmesi sağlanmıştır. Son olarak Model 1'de daha verimli bir gradyan akışı sağlayabilmek amacıyla A. Jolicoeur-Martineau (2018) [39] tarafından önerilen göreceli (relativistic) kayıp fonksiyonu kullanılmıştır.

3.1.1. Model 1'in Genel Yapısı

Tek görüntü ÜÇA modellerinin görüntü üretimi için temel fikri, tek eğitim görüntüsünün dahili yama dağılımlarını (internal patch distribution) öğrenmektir. Model 1, çok ölçekli bir çekişmeli eğitim mimarisi kullanarak tek eğitim görüntüsünün dahili yama dağılımlarını öğrenmektedir. Model 1'deki Ü ve *A* ağlar tamamen konvolüsyon katmanlardan oluşmaktadır ve piramit yapıda çalışmaktadır. Model 1'in yapısı sebebiyle hem eğitim hem de çıkarım işlemi kaba ölçekten (küçük) ince ölçeğe (büyük) doğru stratejisiyle gerçekleştirilmektedir. Model 1'in genel yapısı Şekil 3.1'de gösterilmiştir. Model 1 tek eğitim görüntüsünün küresel bağlam/yapısını korurken foto-gerçekçi, yüksek kaliteli ve çeşitli örnekler üretebilmektedir.



Şekil 3.1: Model 1'in genel yapısı.

Model 1, iki alt bölümden oluşmaktadır. Kaba ölçekli katmanlarda, küçük ölçekli görüntü üretiminde tutarlılığı desteklemek için öz-dikkat mekanizması ve yoğun bağlantılı bloklar beraber kullanılmaktadır. İnce ölçekli katmanlar, sadece yoğun bağlantılı bloklardan oluşmaktadır. Ayrıca, tüm ölçeklerde çekişmeli kayıp fonksiyonu olarak RaLSGAN-GP'den ve yeniden yapılandırma kayıp fonksiyonu olarak $\mathcal{L}_2$'den yararlanılmıştır. Eğitim, aşağıdan yukarı doğru sıralı ölçeklendirilmiş alt ÜÇA'ların eğitilmesi şeklinde yapılmaktadır. Birim zamanda sadece bir ölçekteki ÜÇA eğitilmektedir. Bir ölçekteki eğitim bittiği zaman o ölçekte çalışan ÜÇA'daki öğrenebilen ağırlıklar sabitlenmekte ve bir üst ölçekteki ÜÇA'nın eğitimine başlanmaktadır. İlk ölçek haricindeki her bir ölçekte Ü, bir önceki ölçekte üretilen görüntünün büyütülmüş halini ve rastgele gürültüyü giriş olarak kabul etmektedir. İlk ölçekteki üretici Ü, yalnızca rastgele gürültüyü girdi olarak almaktadır. Eğitim en ince ölçekteki ÜÇA'nın eğitilmesiyle sonlanmaktadır.

3.1.2. Eğitim Görüntüsü Ölçeklendirme

Model 1 ardışıl ölçeklerde çalışan alt ÜÇA'lardan oluşmaktadır ve eğitime başlamadan önce tek eğitim görüntüsünde alt ölçeklerde çalışan ÜÇA'lar için ölçeklendirme işlemi yapmaktadır. Model 1 tek eğitim görüntüsünü belirlenen ölçek değerine göre bikübik interpolasyon yöntemiyle ölçeklendirmektedir [28].

İnterpolasyon, bir görüntünün görsel aslına uygunluğunu korurken onun yeniden boyutlandırılmasını sağlamaktadır. Görüntü keskinliği ve ayrıntısında bozulmayı önlemek için

yüksek kaliteli interpolasyon teknikleri kullanmak önemlidir. İnterpolasyon yönteminin seçimi, yeniden boyutlandırılmış görüntülerin nihai görünümünü önemli ölçüde etkilemektedir. Bikübik interpolasyon, çevreleyen piksellerin ağırlıklı ortalamasına dayalı olarak piksel değerlerini hesaplamaktadır. Her komşu pikselin katkısının ağırlıklarını belirlemek için kübik polinomları kullanarak, hedef pikselin etrafındaki 4x4 piksel komşuluğunu dikkate almaktadır. Bu yaklaşım, pikseller arasında daha yumuşak geçişlere ve ölçekleme sırasında görüntü ayrıntılarının korunmasına yardımcı olmaktadır. Bikübik interpolasyonun güçlü yönlerinden biri, çift doğrusal veya en yakın komşu gibi daha basit interpolasyon yöntemlerinde yaygın olarak görülen görüntü netliği ve pürüzlü kenarları daha iyi oluşturmasıdır. Bu durum bikübik interpolasyon'u özellikle keskinliği ve ince ayrıntıları korumanın çok önemli olduğu görüntülerin ölçeğini yükseltmek için uygun hale getirmektedir.

$$n_s = \left[ceil\left(\log_{h_{sfi}}\left(\frac{h^i_{min}}{i_{min}}\right)\right) + 1\right] - \left[ceil\left(\log_{h_{sfi}}\left(\frac{\min\left(h^i_{max}, i_{max}\right)}{i_{max}}\right)\right)\right] \quad (3.1)$$

$$s_f = \left(\frac{h^i_{min}}{i_{min}}\right)^{\frac{1}{n_s}} \quad (3.2)$$

$$f_{\text{ölçek}}: \left(i^s_k = bik\ddot{u}bik^{\downarrow}\left(i, {s_f}^{n_s-k}\right) \quad for\ (k, 0, n_s + 1)\right) \quad (3.3)$$

(3.1) seçilen tek eğitim görüntüsü için ölçek sayısını hesaplamaktadır. Burada eğitim görüntüsünün, maksimum ve minimum boyutlarına göre başlangıç ölçek oranı (h_{sfi}, hiper parametre) tabanında logaritması alınmaktadır. (3.2), eğitim görüntüsünün her ölçekte sahip olacağı boyutları belirleyen oran çarpanını belirtmektedir. Bu değer hiper parametre olan üretilecek minimum görüntü boyutunun (h^i_{min}) eğitim görüntüsünün minimum boyutuna (i_{min}) oranının ters ölçeğidir ($\frac{1}{n_s}$). Burada elde edilen ölçek oranı (s_f), h_{sfi} oranına yakın olmaktadır. Amaç her ölçekte olabildiğince h_{sfi} oranına yakın ölçeklenmiş eğitim görüntüsü elde etmektir. (3.3), her ölçekte eğitim için kullanılacak ölçeklenmiş eğitim görüntüsünü (i^s_k) oluşturmaktadır. Bunun için ölçek sayısı (n_s) kadar oluşturulan bir döngüde, her iterasyon için ölçek oranı (s_f) kullanılarak eğitim görüntüsü (i) yeniden boyutlandırılmaktadır. Şekil 3.2 $h_{sfi} = 0{,}7$ ve $i = (248{,}186)$ olan örnek bir eğitim görüntüsü için ölçeklendirme sonucu oluşturulan görüntüleri göstermektedir.



i_0^s i_1^s i_2^s i_3^s i_4^s i_5^s i_6^s $i_{k=7}^s = i$

(43,33) (56,42) (71,54) (91,69) (117,88) (151,113) (193,145) (248,186)

Şekil 3.2: Ölçeklendirme sonucu her ölçek için oluşturulan eğitim görüntüleri

3.1.3. Model 1'in Eğitim Prosedürü

Görüntü üretimi, en kaba (ilk) ölçekte rastgele gürültü ile başlamaktadır ($Ü_0$ gürültüyü bir görüntü örneğine eşlemektedir). Her ölçekte, $Ü_k$, gerçekçi ve çeşitli görüntüler üretmeyi amaçlamaktadır. Eğer $Ü_k$ yüksek başarıyla eğitilirse ürettiği görüntüleri A_k, eğitim görüntüsünden ayırt edemez hale gelmektedir. i eğitim görüntüsü Bölüm 3.1.2'de açıklandığı üzere $\{i_0^s, i_1^s, \cdots, i_k^s\}$ şeklinde ölçeklenerek her alt ÜÇA için eğitim görüntüleri (bkz. Şekil 3.2) oluşturulur. $k = 0$ ölçek için $Ü_0$ sadece $z_{k=0}^f = N(0,1)$ gürültü haritasını girdi olarak almaktadır. Bu ölçekteki ÜÇA tamamen koşulsuz şekilde çalışmaktadır. Diğer ölçeklerdeki z_k^f gürültü haritaları her ölçekteki girdi görüntüleriyle aynı boyutlu olarak Gauss dağılımından ($N(0,1)$) üretilmektedir. Her ölçekte $Ü_k$ tarafından üretilen görüntü i_k^g ile gösterilmektedir. $Ü_k$, (3.4)'da verildiği üzere bir alt (kaba) ölçekte $Ü_{k-1}$ tarafından üretilen görüntüyü ve oluşturulan rastgele gürültü verisinin (z_k^f) toplamını girdi olarak almaktadır. Ardışıl iki ölçek arasında boyut farkı bulunması sebebiyle $k-1$ ölçekte üretilen görüntü bikübik interpolasyon yöntemiyle k ölçekteki boyuta boyutlandırılmaktadır. $k = n_s$ ölçekte $A_{k=n_s}$'nın ürettiği görüntü tüm modelin üretmeyi amaçladığı çıktı görüntüsüdür. A_k, i_k^g ile i_k^s 'yi girdi olarak alarak görüntülerdeki yamaları sahte ve gerçek olmak üzere ayırt etmeye çalışmaktadır.

$$i_k^g = Ü_k\left(z_k^f + bik\ddot{u}bik^{\uparrow}\left(i_{k-1}^g\right)\right) \tag{3.4}$$

Model 1'in genel eğitim prosedürü Algoritma 3.1'de sözde kod (pseudo code) olarak verilmiştir. Kayıp fonksiyonu bölümü hariç olmak üzere, Algoritma 3.1 [28], [33], [35] [36] tek görüntü ÜÇA modelleri için de geçerlidir.

Algoritma 3.1: Model 1

Gerekenler: ω_0, Üretici ağ başlangıç ağırlıkları.
θ_0, Ayırıcı ağ başlangıç ağırlıkları.
λ_1, Üretici ağ çekişmeli kayıp katsayısı (varsayılan=2).
λ_2, Üretici ağ yeniden yapılandırma kayıp katsayısı (varsayılan=10).
ψ, Gradyan ceza katsayısı (varsayılan=0.1).
N, Her ölçekteki iterasyon sayısı (varsayılan=2000).

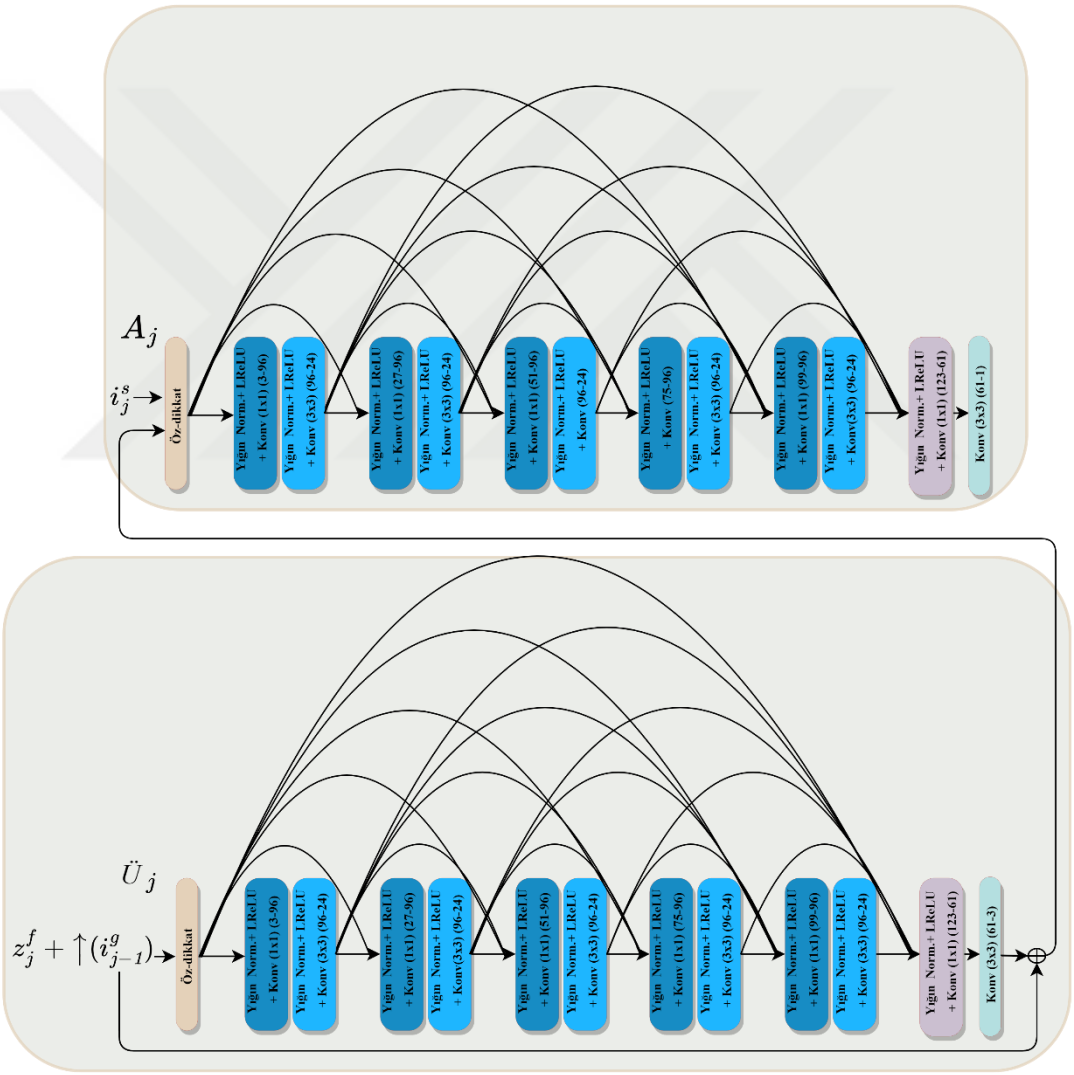
$i^s_{0\dots k}, n_s \leftarrow f_{\text{ölçek}}(i)$
$i^g_{0\dots k} \leftarrow 0, i^{yy}_{0\dots k} \leftarrow 0$
Yeniden yapılandırma kaybı için sabit gürültü vektörleri $z^{yy}_{0\dots k} \sim N(0,1)$
for $k = 0, \dots, n_s$ **do**
  for $n = 0, \dots, N$ **do**
    if $k == 0$ **then**
      Gürültü vektörü $z^f_k \sim N(0,1)$
      $i^g_k \leftarrow \text{Ü}_k\left(z^f_k\right)$
      $i^{yy}_k \leftarrow \text{Ü}_k\left(z^{yy}_k\right)$
    else
      $i^g_0 \leftarrow 0$
      for $s = 0, \dots, k$ **do**
        Gürültü vektörü $z^f_s \sim N(0,1)$
        $i^g_s \leftarrow \text{Ü}_s\left(i^g_s + z^f_s\right)$
        $i^{yy}_s \leftarrow \text{Ü}_s\left(i^{yy}_s + z^{yy}_s\right)$
        $i^g_{s+1} \leftarrow bik\ddot{u}bik\left(i^g_s\right)$
        $i^{yy}_{s+1} \leftarrow bik\ddot{u}bik\left(i^{yy}_s\right)$
      end for
    end if
    Rastgele Sayı $\epsilon \sim [0,1]$
    $\hat{x} \leftarrow \epsilon \times i^s_k + (1-\epsilon) \times i^g_k$
    $\mathcal{L}^{RALSGAN}_{A_k} \leftarrow \mathbb{E}_{i^s_k \sim \mathbb{P}}\left[\left(A_k(i^s_k) - \mathbb{E}_{i^g_k \sim \mathbb{Q}} A_k\left(i^g_k\right) - 1\right)^2\right] + \mathbb{E}_{i^g_k \sim \mathbb{Q}}\left[\left(A_k\left(i^g_k\right) - \mathbb{E}_{i^s_k \sim \mathbb{P}} A_k(i^s_k) + 1\right)^2\right]$
    $\mathcal{L}^{GP}_k \leftarrow \left(\|\nabla_{\hat{x}} A_k(\hat{x})\|_2 - 1\right)^2$
    $\mathcal{L}^{RALSGAN-GP}_k \leftarrow \mathcal{L}^{RALSGAN}_{A_k} + \psi \mathcal{L}^{GP}_k$
    $\omega_k \leftarrow \text{Adam}\left(\nabla_{\omega_k} \mathcal{L}^{RALSGAN-GP}_k, \omega_k\right)$
    $\mathcal{L}^{RALSGAN}_{\text{Ü}_k} = \mathbb{E}_{i^s_k \sim \mathbb{P}}\left[\left(A_k(i^s_k) - \mathbb{E}_{i^g_k \sim \mathbb{Q}} A_k\left(i^g_k\right) + 1\right)^2\right] + \mathbb{E}_{i^g_k \sim \mathbb{Q}}\left[\left(A_k\left(i^g_k\right) - \mathbb{E}_{i^s_k \sim \mathbb{P}} A_k(i^s_k) - 1\right)^2\right]$
    $\mathcal{L}^{RALSGANrec}_{\text{Ü}_k} = \lambda_1\left(\mathcal{L}^{RALSGAN}_{\text{Ü}_k}\right) + \lambda_2\left(\left\|\text{Ü}_k\left(z^{yy}_k\right) - i^s_k\right\|^2_2\right)$
    $\theta_k \leftarrow \text{Adam}\left(\nabla_{\theta_k} \mathcal{L}^{RALSGANrec}_{\text{Ü}_k}, \theta_k\right)$
  end for
  ω_k ve θ_k ağırlık değerlerini sabitle.
end for

Her ölçekte çalışan bir alt ÜÇA vardır. Bir ölçekteki eğitim süresi boyunca sadece o ölçekte çalışan alt ÜÇA (Ü_k ve A_k çifti) eğitilmektedir. Buradaki eğitim klasik ÜÇA eğitimidir. (3.5), her ölçekte kullanılan toplam kayıp fonksiyonunu göstermektedir. Model 1 eğitimi için çekişmeli kayıp fonksiyonuyla ($\mathcal{L}^{ç}_k$) beraber yeniden yapılandırma kayıp fonksiyonu da ($\mathcal{L}^{yy}_k$) kullanılmaktadır. $\mathcal{L}^{ç}_k$, bölüm 3.1.6'da açıklanmıştır.

$$\mathcal{L}_k = \lambda_1 \times \mathcal{L}^{ç}_k + \lambda_2 \times \mathcal{L}^{yy}_k \tag{3.5}$$

$$\mathcal{L}_k^{yy} = \left\| \ddot{\mathrm{U}}_k\left(z_k^{yy}\right) - i_k^s \right\|_2^2 \tag{3.6}$$

Sonsuz gürültü uzayından örneklemle çalışan ÜÇA'yı tek eğitim görüntüsü kullanarak sadece çekişmeli kayıpla eğitmek modelin başarısız çalışmasına sebep olmaktadır. Tek görüntü ÜÇA modelleri, eğitim görüntüsünü üretebildiğini garanti altına almak amacıyla yeniden yapılandırma kayıp fonksiyonu kullanmaktadır [28], [30], [31], [32], [33], [34], [35], [36]. Bu sebeple Model 1'in yeniden yapılandırma kaybı olarak kullandığı $\mathcal{L}_k^{yy} = \mathcal{L}_2$ kayıp fonksiyonu (3.6)'da verilmiştir.



Şekil 3.3: j. Ölçekteki $\ddot{\mathrm{U}}_j$ ve $\boldsymbol{A_j}$ ağlarının iç yapısı

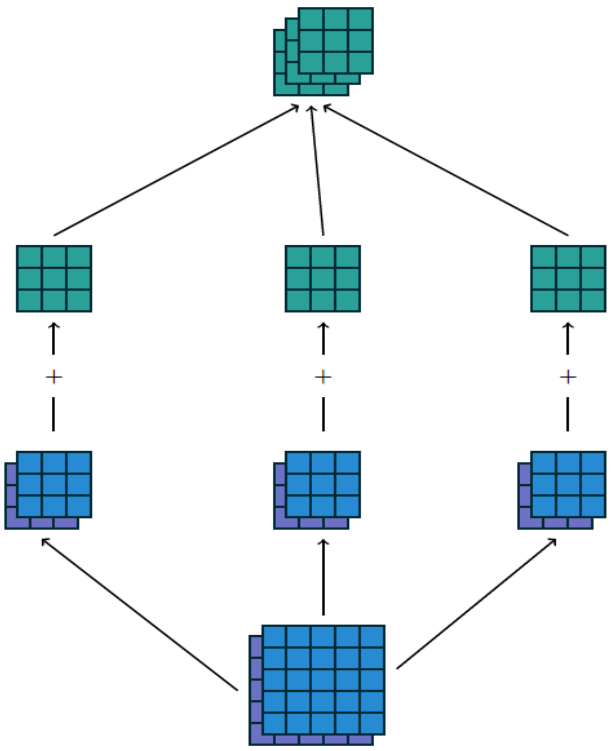
$\mathcal{L}_k^{yy}$ kayıp fonksiyonu Model 1'in eğitim sırasında tek eğitim görüntüsünü üreten sabit bir gürültü etrafında işlem yapmasını sağlamaktadır. Tek eğitim görüntüsünün üretilebildiğini

garantilemek için eğitim öncesi sabit bir gürültü kümesi $z = (z_0^{yy}, z_2^{yy}, \dots, z_k^{yy})$ belirlenmekte ve eğitim boyunca $\mathcal{L}_2$ kayıp fonksiyonu bu sabit gürültü kümesi $(z_0^{yy}, z_2^{yy}, \dots, z_k^{yy})$'den üretilen görüntü $i_{0\dots k}^{yy} = \text{Ü}_{0\dots k}(z_{0\dots k}^{yy})$'yi kullanmaktadır.

Ü ve A ağları tamamen konvolüsyon katmanlarından oluşmaktadır. Şekil 3.3'te j. ölçek için Ü_j ve A_j ağlarını gösterilmektedir. Öz-dikkat mekanizması harici her katman 1x1 veya 3x3 konvolüsyon blok, yığın normalizasyon ve Leaky ReLU aktivasyon fonksiyonundan oluşmaktadır. Her Ü, girdi verisini çıktı verisiyle toplayan artık bir bağlantı içermektedir.

3.1.3.1. Konvolüsyon Blok

Matematiksel açıdan konvolüsyon, iki fonksiyonun birinin diğerinin üzerinden geçerken ne kadar örtüştüğünü ölçen integraldir. Formal olmayan bir tanım olarak, konvolüsyon işlemi, bir fonksiyonun diğer fonksiyon üzerinde kaydırılması ve bu iki fonksiyonun ortak bir çıktı üretmesidir.



Şekil 3.4: Konvolüsyon işlemi [93]

Konvolüsyon, bir matris (genellikle bir filtre veya çekirdek olarak adlandırılır) ile bir giriş matrisi (ör. bir görüntü) arasında yapılan işlemdir. Filtre, belirli bir örüntüyü veya özellikleri vurgulamak veya algılamak için tasarlanmıştır. Bunun için boyutları sabit olan filtreler giriş

matrisi üzerinde gezdirilerek karşılık gelen yamalarla çarpılıp toplanır. Böylelikle filtrenin veri üzerindeki temsil durumu çıktı olarak elde edilir [94], [95], [96]. Şekil 3.4, konvolüsyon işlemini 3 boyutlu veri üzerinde örnek olarak göstermektedir.

Konvolüsyon işleminde filtre boyutu, filtre sayısı, doldurma işlemi (padding), atlama durumu (stride) gibi hiper parametreler kullanıcı tarafından belirlenmektedir. Bu parametrelerin değerlerine göre çıktı boyutu ve değerleri değişmektedir. Filtre değerleri konvolüsyon sinir ağlarından eğitilebilir parametre değerleridir. Bu durum klasik sabit filtrelere oranla çok daha başarılı sonuç vermektedir. Ayrıca konvolüsyon sinir ağlarının gezen filtre yapısı daha az parametre ile örüntü öğrenme imkânı sağlamaktadır. Bu avantajları sebebiyle konvolüsyon sinir ağları günümüzde en çok kullanılan modellerdir.

3.1.3.2. Yığın Normalizasyon

Yığın Normalizasyon (batch normalization- YN), derin öğrenme modellerinde kullanılan bir tekniktir. YN derin sinir ağının daha hızlı ve stabil bir şekilde eğitmeye yardımcı olmaktadır. YN'de, tüm örneklerde ve her iki uzamsal boyutta her bir kanal için ortalama ve varyans hesaplanmaktadır. (3.7) YN işlemini göstermektedir. İlk olarak k katmanından elde edilen x^k özellikleri normalleştirilmektedir. Burada μ ve σ^2 sırasıyla x^k mini yığın verisinin ortalama ve varyansını ifade etmektedir. Ardından normalleştirilen özellikler ölçeklendirilmekte (ζ^k) ve kaydırılmaktadır (β^k) [97].

$$YN(x^k) = \zeta^k \times \left(\frac{x^k - \mu(x^k)}{\sqrt{\sigma^2(x^k) + \epsilon}} \right) + \beta^k \tag{3.7}$$

YN'nin temel amacı, her bir eğitim mini yığınlarındaki özellik haritalarını normalleştirmektir. Derin sinir ağlarını eğitmek zor olabilmektedir, çünkü her katmandaki ağırlıkları başlangıçta uygun bir şekilde ayarlamak zordur. Yığın normalizasyonu, özelliklerin ortalamasını sıfır ve standart sapmasını bir yaparak ağı daha toleranslı hale getirmektedir. Derin ağlarda, gradyan değerleri çok küçük veya çok büyük olması nedeniyle eğitim zorlukları oluşmaktadır. Yığın normalizasyonu, özelliklerin belirli bir aralıkta kalmasını sağlayarak bu sorunları hafifletmektedir [97].

3.1.3.3. Örnek Normalizasyon

Örnek normalizasyon (Instance Normalization- ÖN), sinir ağlarında, her bir örneğin (verinin) özelliklerini mini yığından bağımsız olarak normalleştirmek için kullanılan bir tekniktir. Bu, her girdi örneği için normalizasyonun, gruptaki diğer örneklerin istatistikleri dikkate

alınmaksızın ayrı ayrı uygulandığı anlamını taşımaktadır. ÖN'de, uzamsal boyutta her bir örnek ve her bir kanal için ortalama ve varyans hesaplanmaktadır. Burada μ ve σ^2 sırasıyla x^k mini yığın verisinin uzamsal boyutta ortalama ve varyansını ifade etmektedir. (3.8), örnek normalizasyon formülünü göstermektedir [98].

$$IN(x^k) = \zeta^k \times \left(\frac{x^k - \mu(x^k)}{\sqrt{\sigma^2(x^k) + \epsilon}} \right) + \beta^k \tag{3.8}$$

ÖN, her örnek için ayrı öğrenilebilir ölçek (ζ^k) ve kaydırma (β^k) parametreleri bulundurmaktadır. Bu, modelin her örnek için normalleştirilmiş değerleri bağımsız olarak nasıl ölçeklendireceğini ve kaydıracağını göstermektedir. Uzamsal değişmezliğin (Spatial Invariance) sağlanması konusundaki başarısı ÖN'nin, görüntü oluşturma modellerinde sıklıkla kullanılmasını sağlamıştır. ÖN her örnekte bağımsız olarak çalıştığı için yığın istatistiklere bağlı değildir. Bu, eğitim verisinin küçük olduğu durumlarda veya eğitim ve çıkarımın tek bir örnekle gerçekleştirildiği durumlarda ÖN'nin kullanımını öne çıkarmaktadır [98].

3.1.3.4. Aktivasyon Fonksiyonu

Aktivasyon fonksiyonları, sinir ağlarının katmanları arasındaki çıktıları hesaplamak için kullanılan matematiksel işlevlerdir. Bu fonksiyonlar, ağın öğrenme kapasitesini artırmak ve lineerliği bozmak amacıyla kullanılmaktadır. Aktivasyon fonksiyonları, ağın doğrusal olmayan özellikler öğrenmesine olanak tanıyarak, ağın karmaşıklığını artırmaktadır. Aktivasyon fonksiyonları, ağın geri yayılım algoritması sırasında gradyanların doğru bir şekilde aktarılmasına yardımcı olmaktadır. Tez kapsamında oluşturulan sinir ağlarında Leaky ReLU (Leaky Rectified Linear Unit) [99] ve Tanh aktivasyon fonksiyonları kullanılmıştır.

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3.9}$$

$$Leaky\ ReLU(x) = \begin{cases} x, & x \geq 0 \\ \ell \times x, & x < 0, \ell = 0.01 \end{cases} \tag{3.10}$$

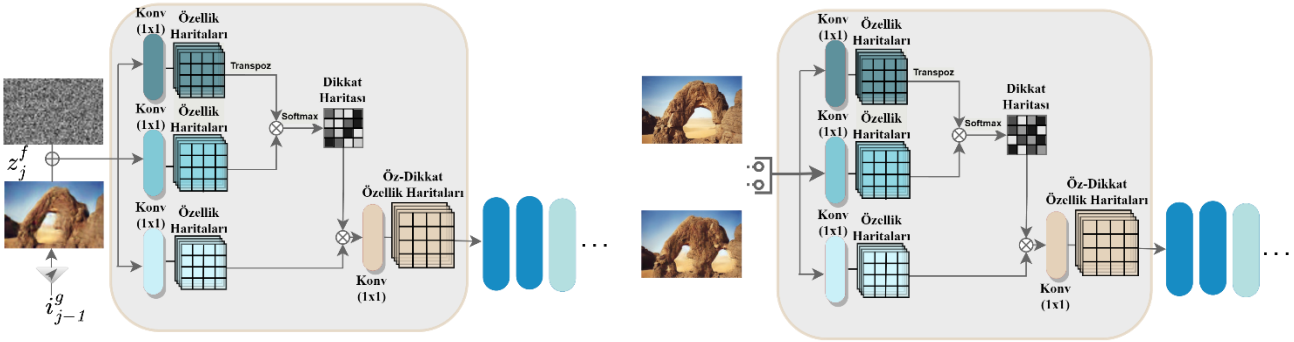
(3.9)'da verilen Tanh aktivasyon fonksiyonu özellik haritalarını $(-1,1)$ aralığında bir değere eşlemektedir. Fakat bu fonksiyon hesaplama maliyeti ve gradyan sönümlemesi gibi sorunlara sebep olabilmektedir. Günümüzde oluşturulan sinir ağlarında Tanh aktivasyon fonksiyonu genellikle çıktı katmanlarında kullanılmaktadır.

(3.10)'da verilen Leaky ReLU, ReLU aktivasyon fonksiyonunun negatif değerlerden yararlanılarak genişletilmesidir. Leaky ReLU birçok uygulamada yüksek performans göstermektedir. Bununla beraber Leaky ReLU ile ilgili önemli problemlerden biri negatif girdiler için doğrusal fonksiyonda doğru eğimin bulunmasıdır. Farklı eğimler, farklı problemler ve farklı ağlar için uygun olabilmektedir.

3.1.4. Öz-dikkat Mekanizması

Çoğu ÜÇA tabanlı görüntü üretim yöntemi, farklı görüntü bölgeleri arasındaki bağımlılıkları modellemek için konvolüsyon sinir ağlarını kullanmaktadır. Bu yöntemler, düşük çözünürlüklü özellik haritalarında yalnızca mekânsal olarak yerel noktalara dayanarak yüksek çözünürlüklü detaylar üretmeyi amaçlamaktadır. Konvolüsyon işlemi, bir bölgesel pikseli komşu piksellerle ilişkilendirmekte ve sabit bir konvolüsyon filtre boyutu kullanarak işlem yapmaktadır. Bu nedenle, uzun menzilli bağımlılıklar ancak birkaç konvolüsyon katmanından geçtikten sonra yakalanabilmektedir. Bu sebeple sadece konvolüsyon katmanlarının kullanılması, görüntülerde uzun menzilli bağımlılıkları yakalamak için hesaplama açısından verimsizdir ve bu katmanların optimizasyonu zahmetlidir.

Öz-dikkat, uzun menzilli/global düzeyde bağımlılıkları modellemek için daha iyi bir denge sağlamaktadır. Öz-dikkat, tek görüntü ÜÇA modellerinde konvolüsyon işlemlerine tamamlayıcı olarak kullanılmaktadır ve görüntü bölgeleri arasında uzun menzilli, çok düzeyli bağımlılıkları etkin bir şekilde yakalamayı amaçlamaktadır. Ayrıca öz-dikkat, ağ hesaplama performansının iyileştirilmesine, kararsız eğitim davranışının azaltılmasına, daha fazla küresel bilgi elde edilmesine ve oluşturulan görüntülerin kalitesinin iyileştirilmesine olanak tanımaktadır [11], [37]. Dolayısıyla öz-dikkat'in ÜÇA'da kullanılma amacı daha fazla global bilgi elde etmeye ve üretilen görüntülerin kalitesini artırmaya çalışmaktır. Öz-dikkat, görüntü içerisindeki her konum için, üretilen görüntünün bölgesel tutarlılığını sağlamayı amaçlamaktadır. Bu sebeplerden dolayı Model 1'de, uzun menzilli bağımlılıkları verimli bir şekilde öğrenebilmek için hem Ü hem de A'da Şekil 3.5'te verilen öz-dikkat mekanizması kullanılmıştır. Öz-dikkat ile Ü, her konumdaki ince ayrıntıların, görüntünün uzak kısımlarındaki ince ayrıntılarla dikkatli bir şekilde koordine edildiği görüntüler üretmeyi ve A, küresel görüntü yapısı üzerinde karmaşık geometrik kısıtlamaları daha doğru bir şekilde öğrenmeyi amaçlamaktadır [11].



a) Ü$_k$'daki öz-dikkat **b)** A_k'daki öz-dikkat

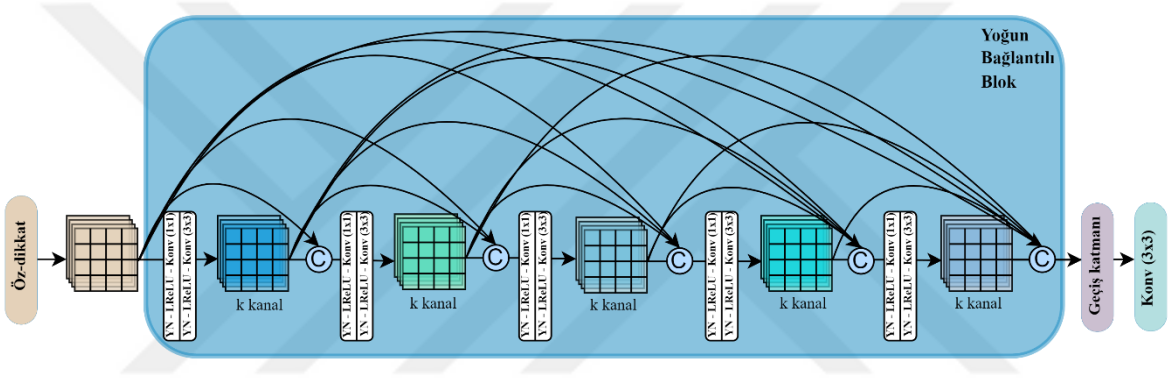
Şekil 3.5: k. ölçek için Ü$_k$ ve A_k'da kullanılan Öz-dikkat mekanizması

Model 1'de, tek eğitim görüntü boyutunun yarısından küçük olduğu ölçeklerde ($boyut(i_k^s) \leq boyut(i)/2$ for $(k, 0, n_s)$) çalışan Ü$_k$ ve A_k'larında öz-dikkat mekanizması ilk katman olarak kullanılmıştır. Ardından üç adet yoğun bağlantılı konvolüsyon bloğu ve bir adet klasik konvolüsyon blok bulunmaktadır. Öz-dikkat içeren Ü$_k$'nin girişi, Ü$_{k-1}$'den elde edilen yeniden boyutlandırılmış görüntünün ve rastgele gürültünün toplamıdır ve bu giriş öz-dikkat mekanizmasına beslenmektedir (Şekil 3.5: a). A_k'nın girişi ise Ü$_k$ tarafından üretilen görüntü ve ölçeklendirilmiş eğitim görüntüsüdür (Şekil 3.5: b). Öz-dikkat mekanizmasını tüm ölçeklerde kullanmak, görüntü içinde daha fazla uzun menzilli bağımlılığı yakalamaya yardımcı olabilir; ancak bu, görüntü üretiminde çeşitliliği azaltmaktadır. Ayrıca, öz-dikkat mekanizması hesaplama karmaşıklığı açısından (ör. matris çarpımı büyük bellek gereksinimi duymaktadır) zaman alıcıdır.

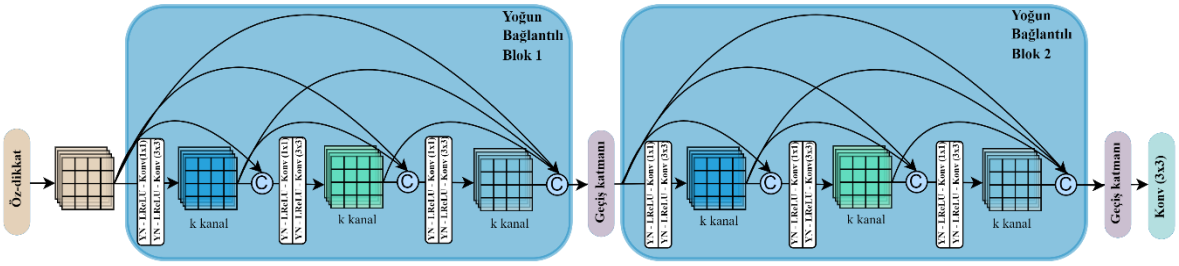
3.1.5. Yoğun Bağlantılı Konvolüsyon Blokları

Tek bir eğitim görüntüsünden öğrenilen koşulsuz üretken modellerdeki konvolüsyon yapılar ve konvolüsyon katmanların sayısı, modeli aşırı uyum veya ezberlemeye sokabilecek önemli parametrelerdir. Konvolüsyon katmanların sayısının arttırılması, çıkarılan özelliklerin karmaşıklığını azaltır; ancak ağın aşırı uyum riskini ve hesaplama yükünü/maliyetini artırır, modelin hızla ezberlemesine neden olur ve görüntü oluşturma performansını etkili bir şekilde iyileştiremez. Model 1'de, her katmanı diğer katmanlara ileri beslemeli olarak bağlayan (her katman, önceki tüm katmanlardan bilgi alır) bir yoğun bağlantılı blok mimarisi kullanılmıştır. Yoğun bağlantılı blok, aynı özellik haritası boyutuna sahip herhangi iki katman arasında doğrudan bağlantılar sunmakta ve Ü ile A'da daha iyi bilgi ve gradyan akışı sağlayarak modelin eğitilmesini kolaylaştırmaktadır. Ek olarak yoğun bağlantılı blok kaybolan gradyan sorununu hafifletmekte ve özelliklerin yeniden kullanılmasını sağlamaktadır [38], [100], [101].

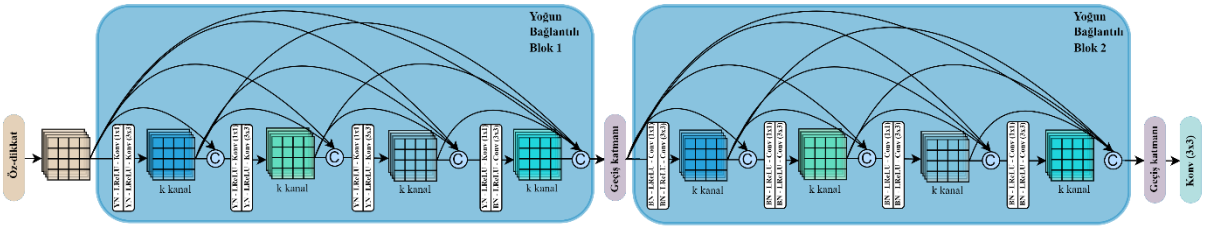
Model 1'deki temel yoğun bağlantılı blok mimarisi, beş yoğun bağlı katman ve bir geçiş katmanından oluşmaktadır (Şekil 3.6). Her yoğun katman, Ü ile *A*'da ön aktivasyon stratejisi ile tasarlanmış YN, Leaky ReLU ve konvolüsyondan oluşmaktadır. Her yoğun katman, önceki tüm katmanlardan ek girdiler almakta ve özellik haritalarını sonraki tüm katmanlara aktarmaktadır. Geçiş katmanı YN ve 1×1 konvolüsyon ve ardından opsiyonel olarak 2×2 ortalama havuzlamadan oluşmaktadır. Geçiş katmanı, çıkarılan tüm özellik bilgilerini sıkıştırmak, özellik haritası boyutunu, yoğun blok parametrelerini ve boyutluluğu azaltmak, model kompaktlığını ve hesaplama verimliliğini artırmak için kullanılmaktadır. Şekil 3.6, Şekil 3.7 ve Şekil 3.8, Model 1'deki yoğun bağlantılı blok mimarisinin farklı konfigürasyonlarını göstermektedir.



Şekil 3.6: Beş yoğun bağlı katman ve bir geçiş katmanından oluşan yoğun bağlantılı blok.



Şekil 3.7: Üç yoğun bağlı katman ve bir geçiş katmanından oluşan iki yoğun bağlantılı blok.



Şekil 3.8: Dört yoğun bağlı katman ve bir geçiş katmanından oluşan iki yoğun bağlantılı blok.

Tek görüntü oluşturma bağlamında eğitim verileri sınırlı olduğundan, oluşturulan her özellik haritası daha da önemli hale gelmektedir. Yoğun bağlantılı blok, sonraki tüm katmanların herhangi bir yoğun bağlı katmanı tarafından öğrenilen özellik haritalarına erişmesine izin vermekte ve özelliğin oluşturucu ve ayırıcı ağlarda yeniden kullanılmasına olanak tanımaktadır. Model 1, yoğun bağlantılı blok mimarisi sayesinde tek bir eğitim görüntüsünden diğer tek görüntü ÜÇA modellerine göre daha efektif özellik kullanımı yapmaktadır.

3.1.6. Gradyan Cezalı Göreceli Ortalama En Küçük Kareler ÜÇA Kaybı

Koşulsuz üretken modelde, görüntü üretim performansı kritik olarak ayırıcı ağ kalitesine bağlıdır. Tek eğitim görüntüsünün A tarafından sınıflandırılması gradyan sönümlemesine sebebiyet vermesi sebebiyle yama ayırıcı ağ kullanmak model eğitimi için gereklidir [28]. Yama ayırıcı, görüntünün tamamı yerine küçük yerel yamaları değerlendirmeye odaklanır. Bu yaklaşım, ayırıcının üreticiye bir görüntüdeki farklı bölgelerin gerçekçiliği hakkında daha ayrıntılı geri bildirim sağlamasına olanak tanımaktadır. Ayırıcı, yerel yamaları değerlendirerek görüntülerdeki yüksek çözünürlüklü ayrıntılara, dokulara ve yerel varyasyonlara odaklanabilir. Bu, ayırıcının tüm görüntüyü bir kerede dikkate alırken gözden kaçabilecek küçük bilgileri yakalamasını sağlar. Yama ayırıcı sayesinde üretici, bir görüntünün hangi alanlarının iyileştirilmesi gerektiği hakkında daha spesifik geri bildirim alır ve böylelikle gerçekçi ayrıntılar oluşturmaya odaklamasına olanak tanır. Bu sebeplerden dolayı tek görüntü üretimi için A ağı yama ayırıcı olarak belirlendi [28].

Yama ayırıcı kullanılan modelde görüntüler göreceli kayıp fonksiyonu kullanılarak değerlendirilmiştir. Model 1'de yama A'nın çıktı yamaları mini yığın verisi olarak ele alınmıştır. Buna göre gerçek ve sahte görüntülerin yama mini yığın verileri Gradyan Cezalı Göreceli Ortalama En Küçük Kareler ÜÇA (Relativistic Average Least Square GAN with Gradient Penalty- RaLSGAN-GP) kayıp fonksiyonu kullanılarak işlem yapılmıştır. RaLSGAN-GP, her ölçekte modeli optimize etmek için kullanılmıştır. RaLSGAN-GP, A'da aşırı büyüyen veya kaybolan gradyanları önlemek için gradyan cezası kullanmaktadır.

$$\mathcal{L}_{A_{k=0\dots n_S}}^{RALSGAN} = \mathbb{E}_{i_k^s \sim \mathbb{P}}\left[\left(A_k(i_k^s) - \mathbb{E}_{i_k^g \sim \mathbb{Q}} A_k(i_k^g) - 1\right)^2\right] + \mathbb{E}_{i_k^g \sim \mathbb{Q}}\left[\left(A_k(i_k^g) - \mathbb{E}_{i_k^s \sim \mathbb{P}} A_k(i_k^s) + 1\right)^2\right] \quad (3.11)$$

$$\hat{x} = \epsilon \times i_k^s + (1-\epsilon) \times i_k^g \quad (3.12)$$

$$\mathcal{L}_k^{GP} = (\|\nabla_{\hat{x}} A_k(\hat{x})\|_2 - 1)^2 \quad (3.13)$$

$$\mathcal{L}_{A_{k=0\dots n_S}}^{RALSGAN-GP} = \mathcal{L}_{A_{k=0\dots n_S}}^{RALSGAN} + \psi \mathcal{L}_k^{GP} \quad (3.14)$$

$$\mathcal{L}_{Ü_{k=0\ldots n_S}}^{RALSGAN} = \mathbb{E}_{i_k^s \sim \mathbb{P}}\left[\left(A_k(i_k^s) - \mathbb{E}_{i_k^g \sim \mathbb{Q}} A_k(i_k^g) + 1\right)^2\right] + \mathbb{E}_{i_k^g \sim \mathbb{Q}}\left[\left(A_k(i_k^g) - \mathbb{E}_{i_k^s \sim \mathbb{P}} A_k(i_k^s) - 1\right)^2\right] \quad (3.15)$$

$$\mathcal{L}_{Ü_{k=0\ldots n_S}}^{RaLSGANrec} = \lambda_1\left(\mathcal{L}_{Ü_{k=0\ldots n_S}}^{RASLGAN}\right) + \lambda_2\left(\mathcal{L}_k^{yy}\ \right) \quad (3.16)$$

RaLSGAN yama ayırıcı, gerçek görüntünün ortalama olarak sahte görüntüden daha gerçekçi olma durumunu tahmin etmektedir. Öte yandan sahte görüntünün ortalama olarak gerçek görüntüye benzerlik oranını bulmaktadır. Bu şekilde, yama ayırıcı ilk önce gerçek görüntüden elde edilen yamaların sahte görüntüden elde edilen yamalardan ortalama olarak nasıl daha iyi performans gösterdiğini ve ardından sahte görüntüden elde edilen yamaların gerçek görüntüden elde edilen yamalara ne kadar yakın olduğunu değerlendirmektedir [39], [102]. Ek olarak, yama ayırıcıya [40]'de önerilen gradyan cezası dahil edilmiştir. RaLSGAN-GP için gradyan cezalı ayırıcı kayıp fonksiyonu ve üretici kayıp fonksiyonları (3.14) ve (3.16)'da verilmiştir. Burada (3.11), A_k için RaLSGAN denklemini göstermektedir. (3.13) gradyan cezasını formülize etmekte ve bu denklemde kullanılan sahte ve gerçek görüntülerin rastgele karışımı $\hat{x}$, (3.14)'te verilmektedir. (3.15), $Ü_k$, için RaLSGAN denklemini ifade etmektedir. (3.16)'da bulunan yeniden yapılandırma kayıp fonksiyonu $\mathcal{L}_k^{yy}$, Bölüm 3.1.3'te açıklanmıştır.

3.1.7. Model 1'in Eğitim Parametreleri ve Optimizasyon Algoritması

Tablo 3.1, Model 1'in eğitiminde kullanılan hiper parametre değerlerini listelemektedir. Model 1 görüntü üretim ölçekleri için $h_{min}^i = 25\,px$ ve $h_{max}^i = 250\,px$ görüntü boyutu kullanmaktadır. Ölçeklendirme oranı $h_{sfi} = 0.7$ olarak belirlenmiştir. Tüm ölçeklerdeki üretici ve ayırıcı ağlar, 2000 iterasyon eğitilmiştir. Üretici ve ayırıcı ağların eğitilebilir parametreleri her iterasyon sonrası güncellenmektedir. İlk ölçekteki üretici ve ayırıcı ağlar rastgele belirlenen ağırlık değerleriyle eğitime başlamaktadır. Diğer ölçeklerdeki üretici ve ayırıcı ağlar, bir önceki ölçekte son iterasyon sonucu hesaplanan ağırlık değerleriyle eğitime başlamaktadır [28]. Üretici ve ayırıcı ağlar, öz-dikkat mekanizması, yoğun bağlantılı konvolüsyon ve çıktı konvolüsyon bloklarından oluşmaktadır. Öz-dikkat, girdi görüntüsü üzerinde işlem yapmaktadır. Yoğun bağlantılı bloklar, öz-dikkat çıktısını girdi olarak almaktadır. Model 1'in ilk yoğun bağlantılı blok konfigürasyonunda bir adet yoğun bağlantılı blok kullanılmıştır. Bu yoğun bağlantılı blok 5 yoğun bağlı katmandan ve bir tane geçiş (transition) katmanından oluşmaktadır (bkz. Şekil 3.6). Yoğun bağlantılı blokta büyüme oranı (grow rate) 24 olarak belirlenmiştir. Yoğun bağlantılı blok içerisindeki geçiş katmanı çıktıyı oluşturmaktadır. Geçiş katmanda sıkıştırma (compression) oranı 0.5 olarak kullanılmıştır. Çıktı konvolüsyon bloğu,

yoğun bağlantılı blok çıktısını girdi olarak almakta ve nihai çıktıyı oluşturmaktadır. Üretici ağdaki çıktı konvolüsyon bloğunun aktivasyon fonksiyonu Tanh'tır. Diğer ağların tüm katmanlarında Leaky ReLU aktivasyon fonksiyonu kullanılmıştır. Öğrenme oranı (Learning rate- lr) $5e-4$ olarak alınmıştır. Her 1600 iterasyon sonrası 0.1 oranında ağırlık düşümü (weight decay) yapılmıştır. Çekişmeli ve yeniden yapılandırma kayıp fonksiyonları katsayıları sırasıyla $\lambda_1 = 2$ ve $\lambda_2 = 10$ olarak belirlenmiştir. Model 1, NVIDIA TITAN X PASCAL GPU, Intel(R) Core (TM) i7-10700KF CPU @ 3.80GHz sistem üzerinde çalıştırılmıştır. Model 1 eğitimi için Adam [103] optimizasyon algoritması kullanılmıştır.

Tablo 3.1: Model 1 eğitim hiper parametre değerleri.

Hiper parametre	Değer
Minimum Görüntü Boyutu (h^i_{min}):	*25*
Maksimum Görüntü Boyutu (h^i_{max}):	*250*
Öğrenme Oranı (lr):	$5e-4$
Ölçeklendirme Oranı (h_{sfi}):	*0.7*
İterasyon Sayısı:	*Her ölçekte 2000*
Her İterasyondaki Güncelleme Sayısı:	*1*
Ağırlık Düşümü:	*0.1 (Her 1600 iterasyondan sonra)*
Çekişmeli Kayıp Ağırlığı (λ_1):	*2*
Yeniden Yapılandırma Ağırlığı (λ_2):	*10*
Gradyan Cezası Ağırlığı (ψ):	*0.1*
Optimizasyon Algoritması:	*Adam* ($\beta_1 = 0.5, \beta_2 = 0.9$)
Büyüme Oranı:	*24*
Aktivasyon Fonksiyonu:	*Leaky ReLU (Ü çıktı aktivasyonu: Tanh)*

3.2. Model 2 (Çekişmeli Öğrenilen Çıkarım ÜÇA)

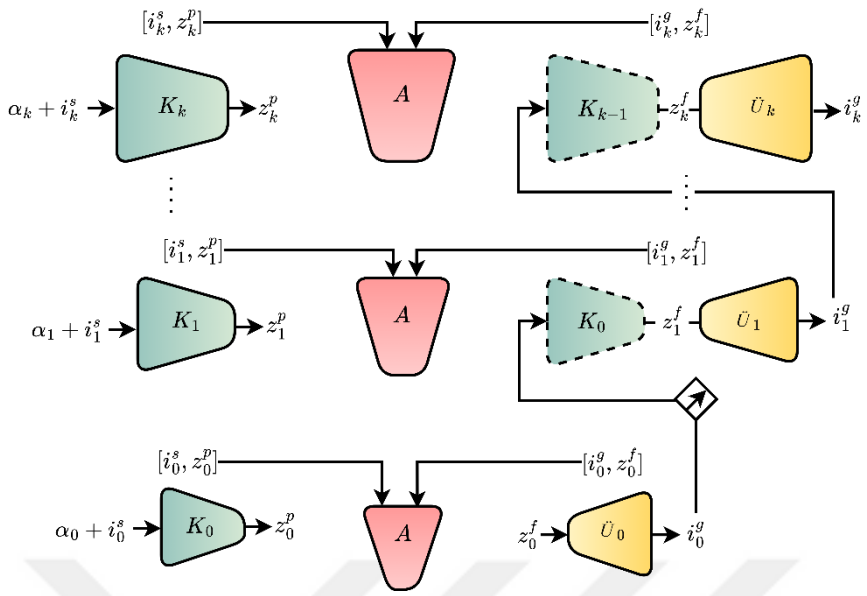
Tek eğitim görüntüsünden görüntü oluşturma problemi Model 2'de mevcut tek görüntü ÜÇA modellerinden farklı bir perspektifle ele alınmaktadır. Mevcut tek görüntü ÜÇA modelleri, model kapasitesi, yardımcı bileşenler, farklı kayıp fonksiyonları kullanma vb. üzerine yoğunlaşmıştır [28], [30], [31], [32], [33], [34], [35], [36]. Bununla beraber üretilen görüntülerin gürültü uzayı, görüntü üretiminde önemli rol oynamaktadır [6]. Bu doğrultuda, Model 2'de eğitim görüntüsünün gürültü uzayının (gizli gösteriminin) etkili şekilde kullanılması amaçlanmıştır.

Model 2, tek eğitim görüntüsünün gürültü uzayına odaklanmaktadır. Amaç bu bileşeni etkili şekilde kullanarak sade ve gelişmiş bir model ortaya koymaktır. Model 2'de, gürültü uzayından yararlanabilmek amacıyla ÇÖÇ [41], [42] mimarisi tek görüntü ÜÇA modeli için uyarlanarak özgün bir model ortaya konulmuştur. ÇÖÇ mimarisini tek görüntü üretim problemine uyarlaması açısından Model 2, diğer tek görüntü ÜÇA modellerinden farklı bir mimari yapıya sahiptir. Model 2'de üretici ve ayırıcı ağlara ek olarak bir kodlayıcı ağ bulunmaktadır. Kodlayıcı iki farklı amaç için kullanılarak model 2 için özgün bir eğitim süreci oluşturulmuştur. Kodlayıcı ağın ilk amacı, eğitim görüntüsünden gürültü uzayına bir eşleme bulmak, ikinci amacı üretici ağa daha iyi gürültü girdisi sağlamaktır. Taranan literatür sonucu edinilen bilgi dahilinde Model 2, tek eğitim görüntüsünden öğrenen, yüksek kaliteli, gerçekçi ve çeşitli örnekler üreten ilk ÇÖÇ temelli ÜÇA modelidir. [31], [35], [67] çalışmaları, Model 2'ye benzer tek görüntü ÜÇA modelleridir, ancak bu modellerin ana odak noktası, ön eğitilmiş ÜÇA'lardan yararlanarak eğitim görüntüsünün gizli gösterimini (gizli uzay değil) elde etmektir. Bu nedenle, bu modeller eğitim aşamasında ÇÖÇ mimarisi temeline (çift yönlülüğe) sahip değildir.

3.2.1. Model 2'nin Genel Yapısı

Model 2'de tek eğitim görüntüsü gizli uzaya eşlenmektedir. Üretici ağ, tek eğitim görüntüsünün gizli uzay bilgisini kullanarak yüksek kaliteli ve çeşitli görüntüler üretebilmektedir. Model 2, üç konvolüsyon sinir ağından oluşmaktadır; üretici Ü, ayırıcı A, ve kodlayıcı K. Model 2'de Ü, A, ve K ağları beraber eğitilmektedir. Ü, gürültü uzayından gelen girdileri yeni görüntülere eşlemektedir. A, gerçek ve üretilmiş görüntüleri ayırt etmeye çalışmaktadır. K eğitim görüntüsünün (i) gürültü uzayını (gizli uzayı – z^p) bulmaktadır. Model 2'de, {gürültü uzayı-eğitim görüntüsü} ve {gürültü uzayı-üretilmiş görüntüler} arasındaki bağlantı öğrenilmektedir.

Şekil 3.9, Model 2'nin genel şemasını içermektedir. Model 2 piramit yapıda çalışan $\{(\text{Ü}_0, K_0, A), (\text{Ü}_1, K_1, A), \cdots, (\text{Ü}_k, K_k, A)\}$, ağlarının hiyerarşik şekilde birleştirilmesiyle oluşturulmuştur. Her ölçekteki Ü, A, ve K ağlarının eğitimi ÇÖÇ [41], [42] eğitimine benzer yapıdadır. Model 2 tek A ağına sahiptir ve tüm katmanların eğitimi bu ağ üzerinden yapılmaktadır. İlk ölçekte (k_0), Ü_0, A, ve K_0 ağları klasik ÇÖÇ mimarisidir. Model 2'nin özgün yanlarından birisi (3.17)'de verilen ağ yapısıdır. Burada $k = 1, 2, \cdots, n_s$ ölçeklerinde Ü_k ağı (3.17)'de verildiği üzere Ü_{k-1} ve K_{k-1} ağlarının birleşiminden oluşmaktadır. Böylelikle kodlayıcı ağlar üretici ağlarda kullanılarak kodlayıcı-kod çözücü (*encoder-decoder*) ağ yapısı elde edilmiştir. K_{k-1} ağı $k-1$ ölçeğindeki i_{k-1}^s görüntüsünü girdi olarak almakta ve çıktı olarak gürültü haritası oluşturmaktadır. Şekil 3.9 ve (3.17)'de verildiği üzere Ü_k'deki K_{k-1} ağı

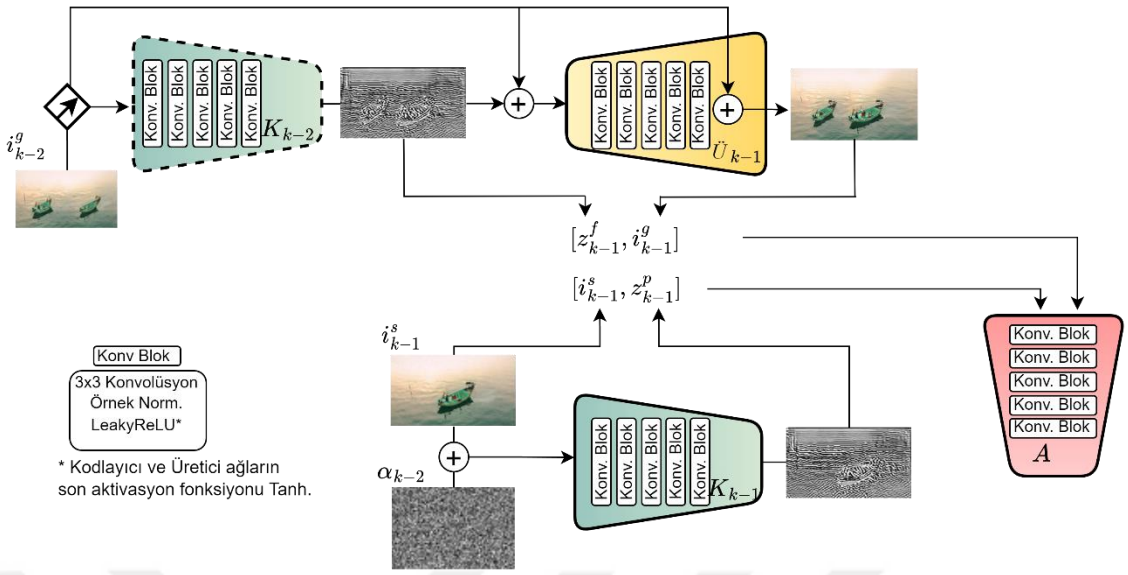


Şekil 3.9: Model 2'nin genel yapısı.

kodlayıcı gibi davranmakta ve girdi verisini gürültü haritasına eşlemektedir. K_{k-1} tek eğitim görüntüsü (i_{k-1}^S) ile eğitildiği için çıktı olarak verdiği gürültü haritası (z_{k-1}^p) eğitim görüntüsü özelliklerini taşımaktadır ve rastgele üretilen gürültüye oranla daha işlevseldir. Bu bağlamda $\ddot{U}_k$'deki $\ddot{U}_{k-1}$ ağı kod çözücü gibi davranarak, K_{k-1}'den gelen gürültü haritasından nihai sahte görüntüyü oluşturmaktadır. Sonuç olarak model 2'deki kodlayıcı-kod çözücü yapısı, $(G_1, \cdots G_n)$ ağlarının girdilerinin eğitim görüntüsünden öğrenilmiş gizli gösterimler olmasına olanak sağlamaktadır. Böylelikle üretici ağa gerçek görüntülerden güçlü bir şekilde bilgi akışı sağlanmaktadır.

$$\ddot{U}_k = (K_{k-1}, \ddot{U}_{k-1}) \tag{3.17}$$

Şekil 3.10, $k-1$ ölçekteki model 2'nin ayrıntılı yapısını göstermektedir. Her ölçekteki ağların içyapısı aynı şekildedir. Ü, A, ve K ağları ardışık çalışan 5 konvolüsyon katmanından oluşmaktadır. Her konvolüsyon katmanı sırasıyla 3x3'lük konvolüsyon filtreleri, ÖN ve Leaky ReLU aktivasyon fonksiyonundan oluşmaktadır. İstisnai olarak Ü ve K ağlarının çıktı konvolüsyon katmanlarında aktivasyon fonksiyonu olarak Tanh kullanılmıştır. Ek olarak Ü ağının kod çözücü bileşeninin ($\ddot{U}_{k-1}$) girdi gürültü haritasını (z_{k-1}^f) ve çıktı görüntüsünü, kodlayıcı bölümünün girdi verisiyle toplayan artık bağlantı bulunmaktadır.



Şekil 3.10: $k-1$ ölçekteki Ü, ***A***, ve ***K*** ağlarının ayrıntılı yapısı

3.2.2. Model 2'nin Eğitim Prosedürü

Model 2'nin eğitim süreci öncesinde ilk olarak eğitim görüntüsü i, $\{i_0^s, i_1^s, \cdots, i_k^s\}$ olarak ölçeklendirilmektedir. Ölçeklendirme yöntemi olarak Bölüm 3.1.2'de açıklanan bikübik interpolasyon ölçeklendirme kullanılmaktadır. Her $(\ddot{U}_k, K_k)$ ağı karşılık gelen ölçekte eğitim görüntüsü i_k^s ile eğitilmektedir. İlk boyutta çalışan $\ddot{U}_0$ girdi olarak uzamsal beyaz Gauss gürültüsünü $(z_0 = N(0, I))$ girdi almaktadır. (3.18)'de görüleceği üzere $k = 1, 2, \cdots, n_s$ arasındaki ölçeklerde önceki katmanda $\ddot{U}_{k-1}$ tarafından üretilen görüntü bikübik interpolasyon ile yeniden boyutlandırılır. Boyutlandırılan görüntü K_{k-1} ağına girdi olarak verilerek görüntünün gürültü haritası oluşturulur. Bu gürültü haritasına, rastgele oluşturulan gauss $z_k^f = N(0, I)$ gürültüsü eklenerek $\ddot{U}_k$ ağına girdi olarak verilir. Elde edilen i_k^g görüntüsü, eğitim görüntüsünün gizli uzayına yakın bir konumdan gelmektedir.

$$i_k^g = \ddot{U}_k\left(z_k^f + K_{k-1}\left(bicubic^{\uparrow}(i_{k-1}^g)\right)\right) \tag{3.18}$$

Sadece tek eğitim görüntüsünden gizli uzaya eşleme yapmak basit bir işlemdir. Bu işlem eğitim gerektirmeksizin ön eğitilmiş bir ÜÇA üzerinden yapılabilmektedir. Fakat Model 2'deki amaç eğitim görüntüsünün yama uzayında birden fazla noktaya eşleme yaparak daha fazla gürültü alanını kapsamaktır. Bu doğrultuda (3.19)'de verildiği üzere $(K_0, \cdots K_k)$ ağlarının girdi görüntülerine $(i^s_{k=0,1,\cdots,n_s})$ çok küçük değerlere sahip $\alpha_{k=0,1,\cdots,n_s}$ matrisi eklenerek bozulur. Böylelikle eğitim süresi boyunca K, sabit tek eğitim görüntüsü ile eğitilmek yerine, eğitim

Algoritma 3.2: Model 2

Gerekenler: ω_0, Üretici ağ başlangıç ağırlıkları.
θ_0, Ayırıcı ağ başlangıç ağırlıkları.
φ_0, Kodlayıcı ağ başlangıç ağırlıkları.
λ_1, Üretici ağ çekişmeli kayıp katsayısı (varsayılan=1).
λ_2, Üretici ağ yeniden yapılandırma kayıp katsayısı (varsayılan=10).
ψ, Gradyan ceza katsayısı (varsayılan =0.1).
N, Her ölçekteki iterasyon sayısı (varsayılan=2000).

$i^s_{0\dots k}, n_s \leftarrow f_{\text{ölçek}}(i)$
$i^g_{0\dots k} \leftarrow 0, i^{yy}_{0\dots k} \leftarrow 0$
Yeniden yapılandırma kaybı için sabit gürültü vektörleri $z^{yy}_{0\dots k} \sim N(0,1)$
for $k = 0, \dots, n_s$ **do**
  for $n = 0, \dots, N$ **do**
    Kodlayıcı ağ gürültü vektörü $\alpha_k \sim N(0,1)$
    $z^p_k \leftarrow K_k(i^s_k + \alpha_k)$
    if $k == 0$ **then**
      Üretici ağ gürültü vektörü $z^f_k \sim N(0,1)$
      $i^g_k \leftarrow \ddot{\text{U}}_k(z^f_k)$
      $i^{yy}_k \leftarrow \ddot{\text{U}}_k(z^{yy}_k)$
    else
      $i^g_0 \leftarrow 0$
      for $s = 0, \dots, k$ **do**
        Gürültü vektörü $z^f_s \sim N(0,1)$
        if s == 0 **then**
          $i^g_s \leftarrow \ddot{\text{U}}_s(i^g_s + z^f_s)$
        else
          $i^g_s \leftarrow \ddot{\text{U}}_s(K_{s-1}(i^g_s) + z^f_s)$
          $i^{yy}_s \leftarrow \ddot{\text{U}}_s(i^{yy}_s + z^{yy}_s)$
          $i^g_{s+1} \leftarrow bik\ddot{u}bik(i^g_s)$
          $i^{yy}_{s+1} \leftarrow bik\ddot{u}bik(i^{yy}_s)$
        end if
      end for
    end if
    Rastgele Sayı $\epsilon \sim [0,1]$
    $\hat{x} \leftarrow \epsilon \times i^s_k + (1-\epsilon) \times i^g_k$
    $\mathcal{L}^{WGAN}_{A_k} \leftarrow \mathbb{E}\left(A(i^g_k, z^f_k)\right) - \mathbb{E}\left(A(i^s_k, z^p_k)\right)$
    $\mathcal{L}^{GP}_k \leftarrow (\|\nabla_{\hat{x}} A_k(\hat{x})\|_2 - 1)^2$
    $\mathcal{L}^{WGAN-GP}_k \leftarrow \mathcal{L}^{WGAN}_{A_k} + \psi \mathcal{L}^{GP}_k$
    $\omega_k \leftarrow \text{Adam}(\nabla_{\omega_k} \mathcal{L}^{WGAN-GP}_k, \omega_k)$
    $\mathcal{L}^{WGAN}_{\ddot{\text{U}}_k} = \left(-\mathbb{E}\left(A(i^g_k, z^f_k)\right)\right)$
    $\mathcal{L}^{WGANrec}_{\ddot{\text{U}}_k} = \lambda_1\left(\mathcal{L}^{WGAN}_{\ddot{\text{U}}_k}\right) + \lambda_2\left(\left\|\ddot{\text{U}}_k(z^{yy}_k) - i^s_k\right\|^2_2\right)$
    $\theta_k \leftarrow \text{Adam}\left(\nabla_{\theta_k} \mathcal{L}^{WGANrec}_{\ddot{\text{U}}_k}, \theta_k\right)$
    $\mathcal{L}^{WGAN}_{K_k} = \mathbb{E}\left(A(i^s_k, z^p_k)\right)$
    $\varphi_k \leftarrow \text{Adam}(\nabla_{\varphi_k} \mathcal{L}^{WGAN}_{K_k}, \varphi_k)$
  end for
  for $j = 0, \dots, (k-3)$ **do**
    φ_j ve θ_j ağırlık değerlerini sabitle.
  end for
  ω_k ağırlık değerlerini sabitle.
end for

görüntüsünün küçük oranda gürültü eklenmiş versiyonlarıyla eğitilmektedir. Bu durum K ağının eğitim süresince öğrenmesini sağlamaktadır.

$$z_k^p = K_k(i_k^s + \alpha_k) \tag{3.19}$$

Tüm ölçeklerde Ü, A, ve K ağları ÇÖÇ mimarisi prensibi temelinde çekişmeli ($\mathcal{L}^{ç}_{k=0,1,\cdots,n_s}$) ve yeniden yapılandırma ($\mathcal{L}^{yy}_{k=0,1,\cdots,n_s}$) kayıp fonksiyonları ile eğitilmektedir. (3.5), her ölçekte kullanılan toplam kayıp fonksiyonunu göstermektedir. $\mathcal{L}_k^{yy}$, Bölüm 3.1.3'te açıklanmıştır. Model 2'de model 1'de olduğu gibi eğitim görüntüsünün üretilmesini garanti altına almak amacıyla yapılandırma kayıp fonksiyonu kullanılmaktadır. Bu doğrultuda Model 2'nin yeniden yapılandırma kayıp fonksiyonu $\mathcal{L}_k^{yy} = \mathcal{L}_2$ (3.6)'da verilmiştir. Model 2'de $\mathcal{L}_k^{ç}$ çekişmeli kayıp fonksiyonu olarak Gradyan Cezalı Wasserstein Mesafe ÜÇA Kaybı (Wasserstein GAN with Gradient Penalty – WGAN-GP) [40] kayıp fonksiyonu belirlenmiştir [10]. $\mathcal{L}_k^{ç}$, Bölüm 3.2.3'te açıklanmaktadır. A, yama ayırıcı yapısındadır. Böylelikle her $(K_k, Ü_k)$ kendi girdi verilerinin yama dağılımlarını öğrenmektedir.

Model 2, Algoritma 3.2'de verildiği üzere en kabadan en ince ölçeğe kadar sıralı bir şekilde eğitilmektedir. Model 2'de eğitim sırasında sadece bir katmanı eğitmenin, ölçekler arasında zayıf bağlılık oluşturduğu gözlemlenmiştir. Bu durum ölçek sayısının artırılması veya birden fazla katmanın aynı anda eğitilmesiyle çözülebilmektedir. Ölçek sayısının artırılması hem parametre sayısı hem de zaman açısından modele ek maliyet getirmektedir. Dolayısıyla birden fazla ölçekteki ağların aynı anda eğitilmesi sadece eğitim süresi açısından maliyet oluşturmaktadır. Bu sebeple Model 2 eğitiminde birden fazla ölçeğin aynı anda eğitilmesi kararlaştırılmıştır. Model 2'de, [32] olduğu gibi aynı anda 3 ölçek $(k, k-1, k-2)$ eğitilmektedir. Her ölçekte farklı değere sahip öğrenme oranı kullanılmıştır. Öğrenme oranı büyüklüğü eğitilen katmanlar arasında ince ölçekten kabaya doğru (3.20)'de belirtilen oranda azaltılmaktadır.

$$lr_{k-1} = lr_k \times 5e^{-2} \times (2^k) \tag{3.20}$$

3.2.3. Gradyan Cezalı Wasserstein Mesafe ÜÇA Kaybı

Model 2'nin eğitiminde (3.5)'te verilen kayıp fonksiyonu yapısı kullanılmıştır. Burada çekişmeli kayıp ($\mathcal{L}_k^{ç}$) için WGAN-GP [40] kayıp fonksiyonu belirlenmiştir [10]. WGAN-GP kaybı, Wasserstein uzaklık ölçüsünü ve gradyan ceza tekniğini kullanan başarılı bir kayıp fonksiyonudur. WGAN-GP kaybının, çekişmeli öğrenmeyi teşvik etmede ve yüksek kaliteli

sonuçlar üretmede etkili olduğu kanıtlanmıştır. WGAN-GP kaybı, eğitim kararsızlığı ve modun çökmesi gibi geleneksel ÜÇA'larda yaygın olarak karşılaşılan sorunlara çözüm sunmaktadır. Wasserstein uzaklık metriği, Jensen-Shannon veya Kullback-Leibler sapmalarına kıyasla daha bilgilendirici ve optimize edilmesi daha kolay bir ölçüm sunarak, gerçek ve üretilmiş verilerin dağılımları arasındaki farklılığı ölçmeye çalışmaktadır [40]. Model 2'de çekişmeli kayıp olarak kullanılan WGAN-GP kaybı, Ü için (3.21)'de, A için (3.22)'de, K için (3.23)'te verilmiştir.

$$\mathcal{L}^{WGAN}_{\text{Ü}_{k=0\dots n_S}} = \lambda_1 \left(-\mathbb{E}\left(A\left(i_k^g , z_k^f \right) \right) \right) + \lambda_2 \left(\mathcal{L}_k^{yy} \right) \tag{3.21}$$

$$\mathcal{L}^{WGANgp}_{A_{k=0\dots n_S}} = \mathbb{E}\left(A\left(i_k^g , z_k^f \right) \right) - \mathbb{E}\left(A\left(i_k^s , z_k^p \right) \right) + \psi \mathcal{L}^{\mathcal{GP}} \tag{3.22}$$

$$\mathcal{L}^{WGAN}_{K_{k=0\dots n_S}} = \mathbb{E}\left(A\left(i_k^s , z_k^p \right) \right) \tag{3.23}$$

Model 2'de A, tek eğitim görüntüsü (i^s) ve K aracılığıyla elde edilen α değeri kadar bozulmuş tek eğitim görüntüsünün gürültü verisini (z^p), gerçek veri olarak kabul etmektedir. A ağı Ü tarafından üretilen sahte görüntüyü (i^g) ve girdi gürültüsünü (z^f) sahte veri olarak kabul etmektedir. A'nın amacı hem gerçek ve üretilen görüntüleri hem de gerçek ve üretilen görüntülerin gürültü verilerini ayırt etmektir. Böylelikle Ü ve K ağlarının eğitiminde görüntü ve gürültü değerlerine göre ilerleme sağlanmaktadır.

3.2.4. Model 2'nin Eğitim Parametreleri ve Optimizasyon Algoritması

Tablo 3.2, Model 2'nin eğitiminde kullanılan hiper parametre değerlerini listelemektedir. Model 2, görüntü üretim ölçekleri için $h^i_{min} = 20px$ ve $h^i_{max} = 250px$ görüntü boyutunu kullanmaktadır. Ölçeklendirme oranı $h_{sfi} = 0.5$ olarak belirlenmiştir. Her ölçek, 2000 iterasyon eğitilmiştir. Ağların eğitilebilir parametreleri her iterasyon sonucu güncellenmektedir. Üretici ve kodlayıcı ağlar 32 konvolüsyon filtresi kullanan 5 bloklu yapıya sahiptir. Ayırıcı ağ 64 konvolüsyon filtresi kullanan 5 bloklu yapıya sahiptir. Model 2'deki ağların tüm katmanlarında normalizasyon yöntemi olarak ÖN kullanılmaktadır. Üretici ve kodlayıcı ağlardaki, çıktı konvolüsyon bloğunun aktivasyon fonksiyonu Tanh'tır. Diğer ağların tüm katmanlarında Leaky ReLU aktivasyon fonksiyonu kullanılmıştır. Öğrenme oranı $5e-4$ olarak alınmıştır. Her 1600 iterasyon sonrası 0.1 oranında ağırlık düşümü yapılmıştır. Çekişmeli ve yeniden yapılandırma kayıp fonksiyonları katsayıları sırasıyla $\lambda_1 = 1$ ve $\lambda_2 = 10$ olarak belirlenmiştir. Model 2, NVIDIA TITAN X PASCAL GPU, Intel(R) Core (TM) i7-

10700KF CPU @ 3.80GHz sistem üzerinde çalıştırılmıştır. Model eğitimi için Adam [103] optimizasyon algoritması kullanılmıştır.

Tablo 3.2: Model 2 eğitim hiper parametre değerleri.

Hiperparametre	Değer
Minimum Görüntü Boyutu (h^i_{min}):	*20*
Maksimum Görüntü Boyutu (h^i_{max}):	*250*
Öğrenme Oranı (lr):	$5e-4$
Ölçeklendirme Oranı (h_{sfi}):	*0.5*
İterasyon Sayısı:	*Her ölçekte 2000*
Her İterasyondaki Güncelleme Sayısı:	*1*
Ağırlık Düşümü:	*0.1 (Her ölçekte 1600 iterasyondan sonra)*
Çekişmeli Kayıp Ağırlığı (λ_1):	*1*
Yeniden Yapılandırma Ağırlığı (λ_2):	*10*
Gradyan Cezası Ağırlığı (ψ):	*0.1*
Optimizasyon Algoritması:	*Adam* ($\beta_1 = 0.5, \beta_2 = 0.9$)
Ü ve K için Konvolüsyon Filtre Sayısı:	*32*
A için Konvolüsyon Filtre Sayısı:	*64*
Aktivasyon Fonksiyonu:	*Leaky ReLU (Ü ve K çıktı aktivasyonu: Tanh)*

4. BULGULAR

Bu bölümde tez kapsamında önerilen modellerin, doğa, şehir manzaraları, sanatsal resimler gibi çok çeşitli sahneleri kapsayan görüntüler üzerinde nicel analiz, nitel analiz ve ablasyon çalışması yoluyla performansları mevcut tek görüntü ÜÇA modelleri ile karşılaştırmalı olarak kapsamlı bir şekilde değerlendirilmiştir. Nicel değerlendirme yöntemleri olarak tek görüntü ÜÇA modellerinin değerlendirilmesi için oluşturulan Tek Görüntü Fréchet Başlangıç Mesafesi (Single Image Fréchet Inception Distance- SIFID) [28] ölçümü ile görüntü benzerlik metrikleri Çok Ölçekli Yapısal Benzerlik İndeksi Ölçümü (Multi-Scale Structural Similarity Index Measure- MS-SSIM) [104] ve Öğrenilmiş Algısal Görüntü Yama Benzerliği (Learned Perceptual Image Patch Similarity- LPIPS) [105] kullanılmıştır. Nitel değerlendirme için 3 farklı kullanıcı çalışması oluşturulmuştur. Ayrıca model bileşenlerinin modele olan etkilerinin incelendiği ablasyon çalışması yapılmıştır. Tüm bulgular NVIDIA TITAN X PASCAL GPU, Intel(R) Core (TM) i7-10700KF CPU @ 3.80GHz sistem üzerinde çalıştırılan tek görüntü ÜÇA modellerinden elde edilmiştir. Karşılaştırılan ÜÇA modellerinin kod linkleri Ek-M bölümünde belirtilmiştir.

4.1. Kullanılan Veri Kümeleri

Tek görüntü ÜÇA modellerin sonuçlarını değerlendirmek için literatürde var olan üç veri kümesinden yararlanılmıştır. Bu veri kümeleri Places205 [106], LSUN [107] ve ImageNet [108] veri kümeleridir.

Places205 veri kümesi olarak da bilinen Places veri kümesi 205 farklı sınıfa ayrılmış iki milyondan fazla görüntüden oluşmaktadır. Her sınıf, ormanlar, mutfaklar, yatak odaları ve ofisler gibi farklı türde bir ortamı veya sahneyi temsil etmektedir. Veri kümesi, sahne tanıma için kullanılan önceki veri kümelerinden daha büyük ve daha çeşitlidir. Bu özelliği Places veri kümesini derin öğrenme modellerinin eğitimi ve değerlendirilmesi için öne çıkarmaktadır. Tez kapsamında oluşturulan ÜÇA modellerinin karşılaştırmalı değerlendirilmesi için Places veri kümesi içerisinden seçilen 50 adet görüntü belirlenmiştir [28].

The Large-Scale Scene Understanding (LSUN) veri kümesi, bilgisayarla görme alanında araştırma ve geliştirme çalışmalarına kaynak olması amacıyla oluşturulmuş bir görüntü veri kümesidir. LSUN, kapsamlı olarak ve farklı sahne türlerinin hem iç hem de dış ortamlarını kapsayacak şekilde çok çeşitli sınıfları barındırmaktadır. LSUN veri kümesinin içerdiği sınıflar, yatak odaları, mutfaklar, ofisler, oturma odaları ve daha fazlası dahil olmak üzere gerçek dünya görüntüleridir. Tez kapsamında oluşturulan ÜÇA modellerinin karşılaştırmalı değerlendirilmesi için LSUN veri kümesinden 50 adet görüntü seçilmiştir [32].

ImageNet, bilgisayarlı görü, derin öğrenme ve görüntü tanıma modellerini karşılaştırma üzerindeki etkisinin de dahil olduğu birçok çalışmada kullanılmaktadır. ImageNet 20.000'den fazla sınıfı kapsayan 14 milyondan fazla görüntüden oluşmaktadır. ImageNet, 2010'dan 2017'ye kadar her yıl düzenlenen bir yarışma olan ILSVRC'de birçok derin öğrenme modelinin başarı ölçümünde kullanılmıştır. Tez kapsamında oluşturulan ÜÇA modellerinin karşılaştırmalı değerlendirilmesi için ImageNet veri kümesinden 50 adet görüntü model değerlendirmesi için seçilmiştir [35].

4.2. Nicel Ölçüm Yöntemleri

Nesnel görüntü kalitesi değerlendirme yöntemleri, algısal görüntü kalitesini otomatik olarak tahmin edebilen nicel ölçümler geliştirmeyi amaçlamaktadır [109]. Tek bir doğal görüntüden foto-gerçekçi, yüksek kaliteli, çeşitli görüntüler üreten koşulsuz ÜÇA modellerinin sonuçlarını nicel olarak analiz etmek için yaygın olarak kabul edilen ölçüm yöntemleri henüz bulunmamaktadır. Bu durum çözülmeyi bekleyen bir problem olmakla birlikte, literatürde bu modellerin etkinliğini ölçmek ve karşılaştırmak için farklı ölçüm metrikleri önerilmiştir. Bu bağlamda, Tek Görüntü Fréchet Başlangıç Mesafesi, tek görüntü ÜÇA modellerinin görüntü oluşturma kalitesini ölçmek için kullanılmaktadır [28]. Ayrıca, çeşitlilik değerlendirmeleri için Çok Ölçekli Yapısal Benzerlik İndeksi Ölçümü ve Öğrenilmiş Algısal Görüntü Yama Benzerliği kullanılmaktadır.

4.2.1. Tek Görüntü Fréchet Başlangıç Mesafesi

Fréchet Başlangıç Mesafesi (Fréchet Inception Distance- FID), üretici bir model tarafından oluşturulan görüntülerin kalitesini değerlendirmek için kullanılan sayısal bir ölçüm metriğidir. Fréchet mesafesi, bir metrik uzaydaki iki olasılık dağılımı arasındaki benzerliğin ölçüsüdür [9].

Fréchet mesafesi dağılımların hem konumunu hem de yayılmasını dikkate almaktadır. (4.1) iki dağılım arasındaki Fréchet mesafesini ifade etmektedir.

$$\mathrm{F}(\mathcal{N}_1, \mathcal{N}_2) = \|\mu_1 - \mu_2\| + \mathrm{Tr}\left(\Sigma_1 + \Sigma_2 - 2 \times \left(\Sigma_1^{1/2} \times \Sigma_2 \times \Sigma_1^{1/2}\right)^{\frac{1}{2}}\right) \quad (4.1)$$

FID'de gerçek görüntülerin dağılımı ile oluşturulan görüntülerin dağılımı arasındaki benzerlik, ön eğitilmiş bir derin ağdan elde edilen özellik gösterimleri arasındaki Fréchet mesafesini hesaplayarak ölçülmektedir. Bu ölçüm için ön eğitilmiş Inception-v3 [110] ağından yararlanılmaktadır. Inception-v3'ün son havuzlama katmanı, bir görüntünün yüksek düzeyde özellik temsilini sağladığı için buradaki özellikler kullanılmaktadır. Bu özellik gösterimlerinin ortalama ve kovaryans matrisi hem gerçek hem de üretilen görüntüler için hesaplanmaktadır. Böylelikle gerçek ve üretilen görüntüleri temsil eden çok değişkenli Gauss dağılımı elde edilmektedir. Bu çok değişkenli Gauss dağılımları arasındaki Fréchet mesafesi, (4.1) kullanılarak hesaplanır.

Tek Görüntü ÜÇA modellerinde görüntü kalitesini ölçek için FID metriğinin tek görüntüye uyarlanmış hali Single Image Fréchet Inception Distance (SIFID) kullanılmaktadır [28]. Düşük SIFID değeri, ÜÇA modelinin gerçekçi, yüksek kaliteli görüntüler üretme yeteneğini ve üretilen örnek ile eğitim görüntüsü arasında ayrım yapmanın daha zor olduğunu göstermektedir.

4.2.2. Çok Ölçekli Yapısal Benzerlik İndeksi Ölçümü

Çok Ölçekli Yapısal Benzerlik İndeksi Ölçümü (MS-SSIM), tek ölçekli Yapısal Benzerlik İndeksinin (Structural Similarity Index Measure- SSIM) gelişmiş bir görüntü kalitesi değerlendirme metriğidir. SSIM, yapısal bilgi, parlaklık ve kontrast açısından iki görüntü arasındaki benzerliği etkili bir şekilde ölçerken, insanın görsel algısının çok ölçekli doğasını yakalamada yetersiz kalmaktadır. MS-SSIM, görüntüleri birden çok ölçekte analiz ederek bu sınırlamanın üstesinden gelmektedir. MS-SSIM, bir Gauss piramidi kullanılarak gerçek ve sahte görüntülerin çoklu ölçeklere ayrıştırılmasıyla çalışmaya başlamaktadır. Piramidin her seviyesi, değişen detay seviyeleri içeren görüntünün farklı bir ölçeğini temsil etmektedir. Her ölçek seviyesinde, karşılık gelen görüntü yamaları için SSIM hesaplanmaktadır. SSIM hesaplaması üç ana bileşen içermektedir: *parlaklık bozulması, kontrast bozulması ve yapısal benzerlik*. Belirli bir yama için bu bileşenler, gerçek ve sahte görüntüler arasındaki piksel yoğunlukları ve gradyanları karşılaştırılarak hesaplanmaktadır. Parlaklık bozulması, gerçek ve sahte yamalar arasındaki ortalama piksel yoğunluklarındaki farkı ölçmektedir. Kontrast

bozulması, gerçek ve sahte yamalar arasındaki kontrast veya piksel yoğunluklarının standart sapması arasındaki farkı değerlendirmektedir. Yapısal benzerlik bileşeni, gerçek ve sahte yamalar arasındaki dokular açısından benzerliği değerlendirmektedir [104]. (4.2) ve (4.3) sırasıyla SSIM ve MS-SSIM denklemlerini göstermektedir. c_1 ve c_2 sıfıra bölmeye engellemek için kullanılan birer sabittir. μ görüntülerin piksel yoğunluğunu (ortalamasını), σ standart sapmasını göstermektedir. M, MS-SSIM için piramit sayısını, ∂^i her SSIM değerinin ağırlığını ifade etmektedir. Düşük/yüksek MS-SSIM, görüntü yamalarının algısal olarak farklı/benzer olduğu anlamına gelmektedir.

$$\mathrm{SSIM}(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{4.2}$$

$$\text{MS-SSIM}(\mathrm{x},\ \mathrm{y}) = \frac{1}{\mathrm{M}} \sum_{\mathrm{i}=1}^{\mathrm{M}} (\ \mathrm{SSIM}(x_i,\ y_i)\ )^{\partial^i} \tag{4.3}$$

4.2.3. Öğrenilmiş Algısal Görüntü Yama Benzerliği

Öğrenilmiş Algısal Görüntü Yama Benzerliği (Learned Perceptual Image Patch Similarity-LPIPS), görüntü işlemede, insan algısal yargılarına dayalı görüntü yamaları arasındaki benzerliği ölçmek için derin sinir ağlarını kullanan bir ölçüm yöntemidir. LPIPS, görüntü benzerliğinin hesaplamalı ölçümleri ile insanların benzerliği algılama biçimi arasındaki boşluğu kapatmayı amaçlamaktadır. Bu yöntemde büyük bir görüntü veri kümesi üzerinde ön eğitilmiş derin bir sinir ağı, özellik çıkarıcı olarak kullanılmaktadır. (4.4), LPIPS için kullanılan denklemi göstermektedir. Burada ön eğitilmiş VGG ağına gönderilen görüntülerden elde edilen özellik haritalarından alınan P adet yamanın ortalama farkı hesaplanmaktadır. Düşük/yüksek LPIPS, görüntü yamalarının algısal olarak benzer/farklı olduğu anlamına gelmektedir [105]. Bu sebeple tek görüntü ÜÇA modellerinin yüksek çeşitlilik göstermesi için yüksek LPIPS değerine sahip olması gerekmektedir.

$$\mathrm{LPIPS}(x, y) = \frac{1}{P}\sum_{i=1}^{P}\left(\mathrm{VGG}(x_i) - \mathrm{VGG}(y_i)\right)^2 \tag{4.4}$$

4.3. Kullanıcı Çalışmaları Yöntemleri

Önerilen ÜÇA modelleri tarafından oluşturulan görüntülerdeki gerçekçilik, görsel kalite, çeşitliliği değerlendirmek; diğer tek görüntü ÜÇA modelleri tarafından üretilen görüntülerle

karşılaştırmalı analizler yapmak; modellerin bir eğitim görüntüsünün içsel yama istatistiklerini ne kadar iyi yakaladığını incelemek amacıyla, Google Forms kullanılarak 50 katılımcıyla kullanıcı çalışmaları gerçekleştirilmiştir. Her çalışma için Places, LSUN ve ImageNet veri kümelerinden önerilen ve karşılaştırma yapılan tek görüntü ÜÇA modellerinin ürettiği görüntüler kullanılmıştır. Kullanıcı çalışmaları için üç farklı veri kümesi üzerinde [28], [32], [33] tarafından sunulana benzer bir değerlendirme prosedürü takip edilmiştir. Tüm ÜÇA modellerinin kullanıcı çalışması sonuçlarında performansının düşük olabileceğini dikkate alarak [28], [32] farklı olarak kullanıcı çalışmalarında, kullanıcılara herhangi bir zaman kısıtlaması getirilmemiştir. Önerilen tek görüntü ÜÇA modellerinin kullanıcı çalışmalarının linkleri Ek-A ve Ek-B bölümlerinde sunulmuştur. Kullanıcı çalışmaları için üç farklı senaryo oluşturulmuştur.

1. *Eşleştirilmiş çalışma (gerçek-ve-sahte görüntü çiftleri gösterilir):* Katılımcılara zaman sınırlaması olmadan, her ÜÇA modeli için üretilen sahte bir görüntü ve ilgili gerçek görüntü içeren 50 görüntü dizisi sunulmaktadır (Şekil 4.1). Katılımcılardan sahteyi (veya gerçeği) belirlemeleri istenmektedir. Burada, sahte görüntünün gerçek olanla ne kadar benzer olduğunu ölçmek amaçlanmıştır.



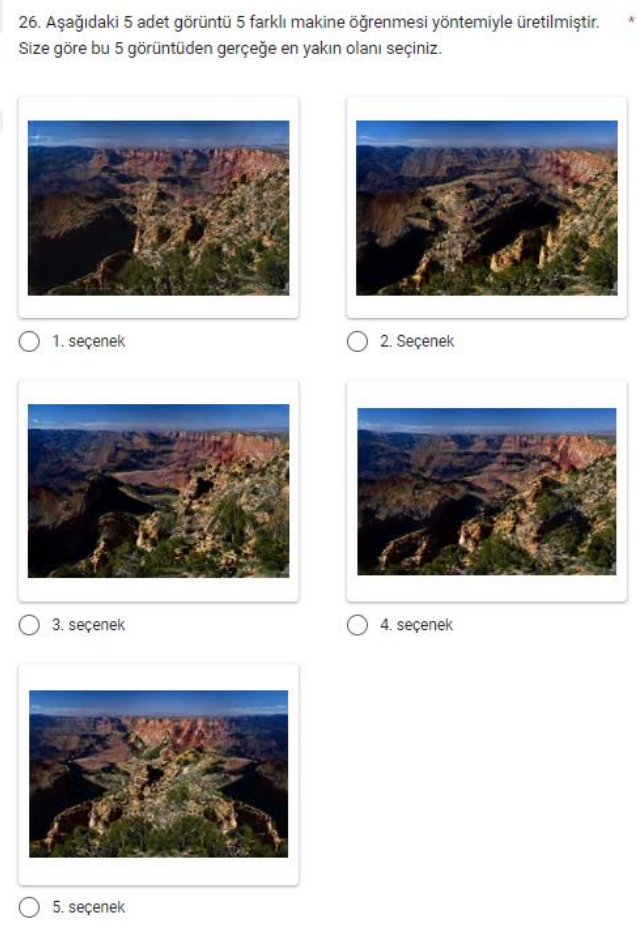
Şekil 4.1: Eşleştirilmiş çalışma için örnek soru sayfası.

2. *Eşleştirilmemiş çalışma (gerçek veya sahte bir görüntü gösterilir):* Katılımcılara zaman sınırlaması olmadan tek bir görüntü sunulmakta ve bu görüntünün gerçek veya sahte olup olmadığı sorulmaktadır (Şekil 4.2). Katılımcılara her veri kümesi için rastgele sırayla 50 gerçek görüntü ve 50 sahte görüntüden oluşan farklı bir kümeler dizisi gösterilmektedir.



Şekil 4.2: Eşleştirilmemiş çalışma için örnek soru sayfası.

3. *Nitel değerlendirme çalışması (tüm modeller tarafından üretilen görüntüler bir arada gösterilir):* Katılımcılara zaman sınırlaması olmadan tek görüntü ÜÇA modelleri tarafından üretilen görüntülerin bulunduğu 50 görüntü dizisi sunulmaktadır (Şekil 4.3). Katılımcılardan hangi görüntünün en gerçekçi ve yüksek kaliteli olduğunu belirlemeleri istenmektedir.

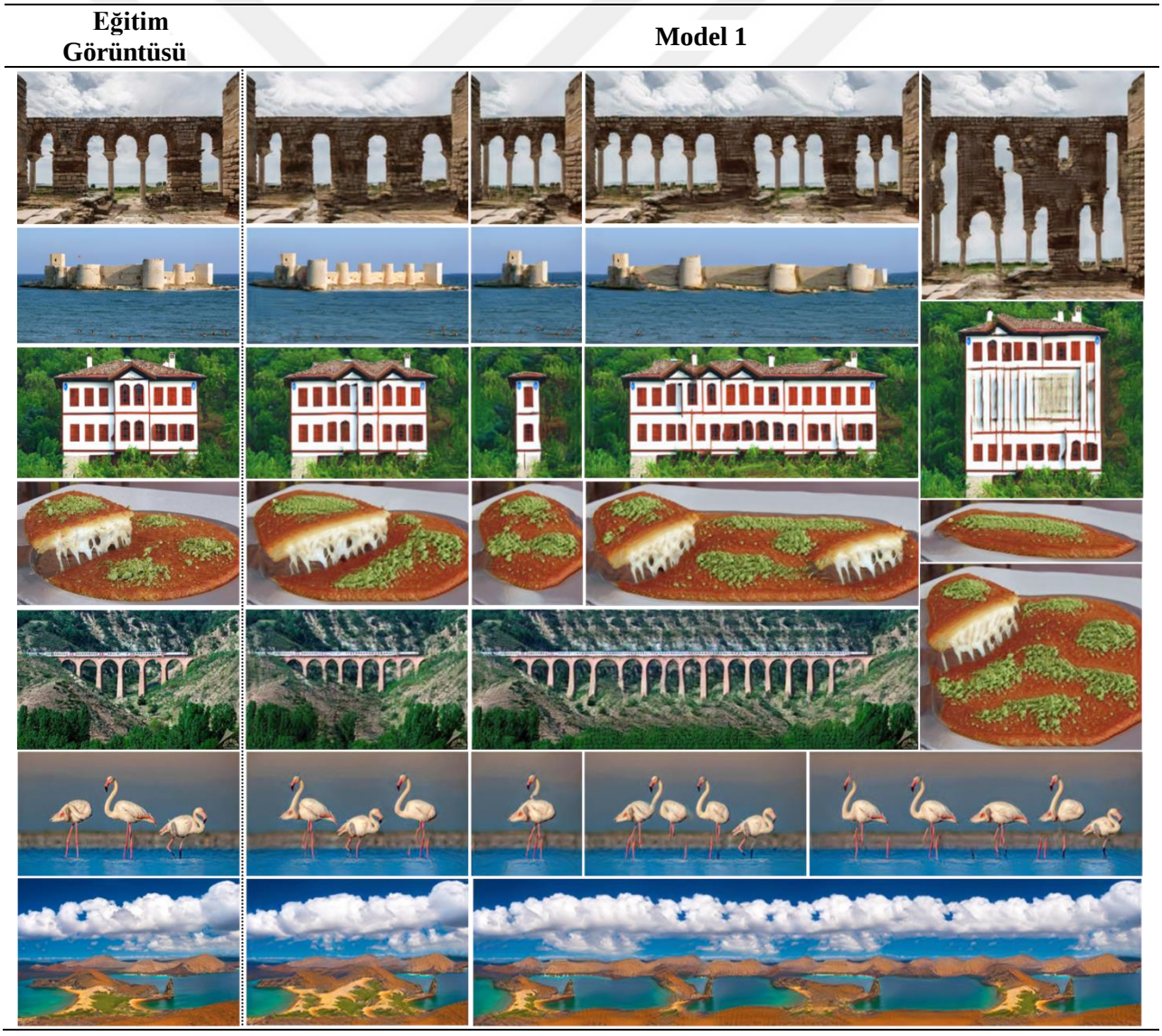


Şekil 4.3: Nitel değerlendirme çalışması için örnek soru sayfası.

4.4. Model 1 Sonuçları

4.4.1. Görsel Karşılaştırma

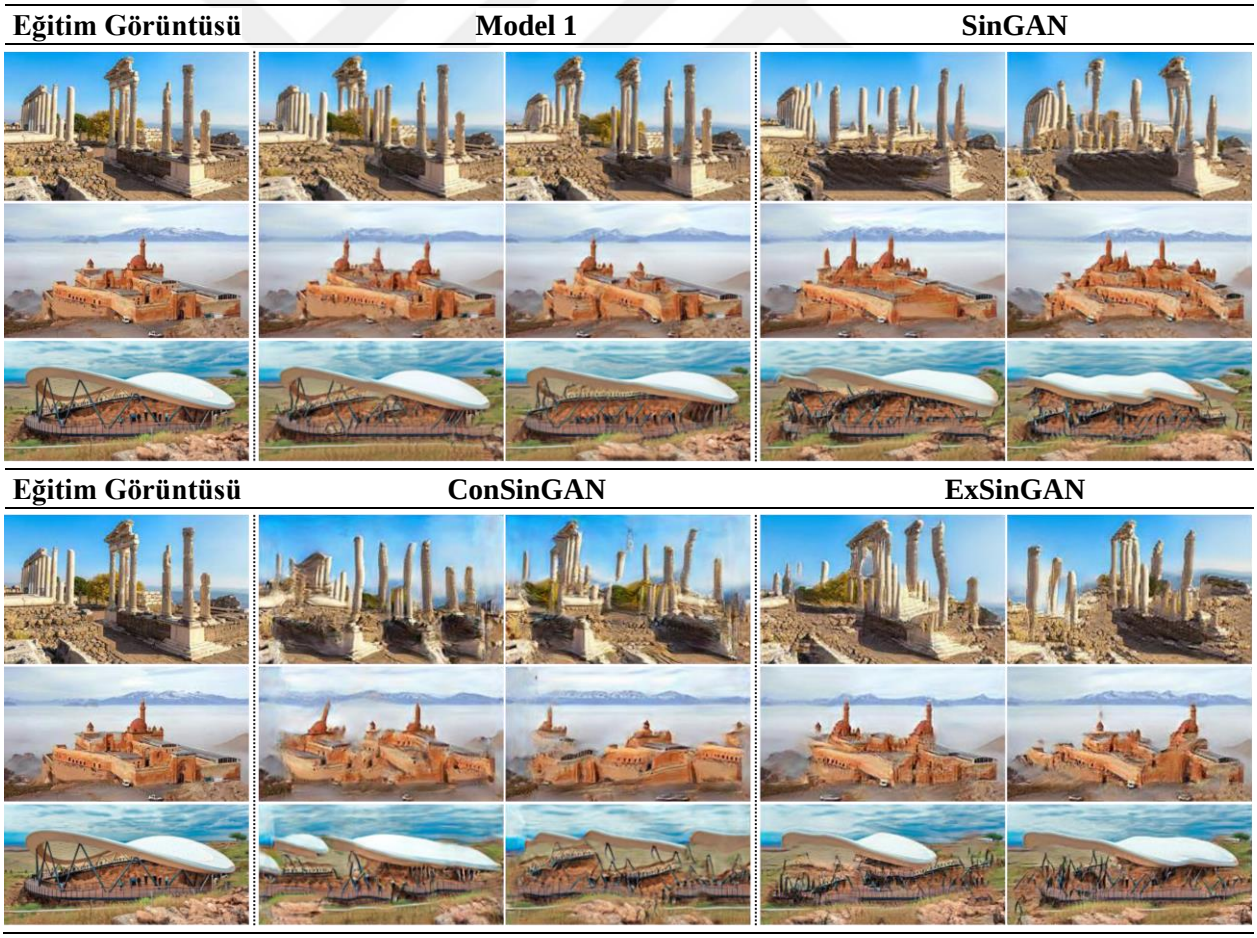
Şekil 4.4, karmaşık yapı ve doku içeren çeşitli doğal görüntüler için Model 1 tarafından üretilen rastgele örnekleri göstermektedir. Eğitim görüntüleri dokusal (*ör. tatlı, ev, kale*), anlamsal (*ör. kuşlar, köprü*) ve yapısal (*ör. tarihi yapı, bungalov evler*) varyasyonları içerecek şekilde seçilmiştir. İlk satırdaki tarihi yapı incelendiğinde Model 1'in gerçekçi bir şekilde kolonları farklı konumlarda ve şekillerde üretebildiği görülmektedir. Ayrıca ikinci satırdaki kale örnekleri için Model 1 kalenin yapısını ve konumunu koruyarak gerçekçi farklı örnekler üretmiştir. Bu durum, Model 1'in görüntü içerisindeki pikseller arasındaki bağımlılıkları başarıyla öğrenebildiğini göstermektedir.



Şekil 4.4: Model 1 tarafından oluşturulan rastgele görüntü örnekleri.

Model 1'deki Ü ve *A* ağları tamamen konvolüsyon katmanlardan oluşmaktadır. Bu durum modele eğitim görüntünün küresel bağlamını korurken girdi olarak verilen gürültü haritalarının boyutlarını değiştirerek çıkarım zamanında istenilen boyutlarda yeni görüntüler üretebilme esnekliği sağlamaktadır. Şekil 4.4, farklı boyutlara sahip rastgele üretilmiş görüntülerin niteliksel örneklerini göstermektedir. Bu uygulama daha ayrıntılı ve çok örnekle Bölüm 4.6.5'te açıklanmıştır.

Model 1'in birçok farklı görüntü üzerinde genel olarak başarılı çalıştığını daha çok örnekle göstermek için internet üzerinden rastgele görüntüler seçilmiştir. Bu görüntülerle yapılan eğitim sonucu üretilen örnekler Ek-C Şekil 6.1'de gösterilmektedir. Oluşturulan yeni örnekler, eğitim görüntüsünün görsel içeriğini korurken yeni nesne konfigürasyonları ve karmaşık dokular içermektedir. Sonuç olarak Model 1, nesnelerin genel yapılarını/düzenlemelerini, ayrıca ince ayrıntıları ve doku bilgilerini başarıyla koruduğu görülmektedir.



Şekil 4.5: Model karşılaştırma**.**

Şekil 4.5, SinGAN, ConSinGAN, ExSinGAN ve Model 1'in görsel karşılaştırmasını göstermektedir. İnternet üzerinden rastgele seçilen üç doğal görüntü üzerinde yapılan eğitim sonucunda, tüm tek görüntü ÜÇA modelleri tarafından oluşturulan iki rastgele görüntü örneği verilmiştir. Bu örnekler, eğitim görüntüsünün görsel içeriğini korurken yeni düzenleri ve nesne yapılandırmalarını göstermektedir. Model 1 *Bergama Antik Kenti* için global ve yerel özellikleri rakiplerinden daha iyi öğrenerek gerçekçi, yüksek kaliteli görüntüler üretmiştir. Ayrıca diğer iki görselde de (*İshak Paşa Sarayı ve Göbekli Tepe*) Model 1 nesnelerin global yapısını ve ince doku bilgilerini diğer tek görüntü ÜÇA modellerine oranla daha iyi korumuştur.

4.4.2. Nicel Sonuçlar

Model 1'in SIFID, MS-SSIM ve LPIPS metriklerini kullanarak Places, LSUN ve ImageNet veri kümeleri üzerinde ulaştığı nicel sonuçlar diğer tek görüntü ÜÇA modelleriyle karşılaştırılmıştır. Tablo 4.1, Places, LSUN ve ImageNet veri kümeleri için Model 1 ve diğer tek görüntü ÜÇA modellerinin nicel sonuçlarını göstermektedir.

Places veri kümesi sonuçlarına bakıldığında SIFID metriği açısından, Model 1, SinGAN ve ExSinGAN'dan daha başarılı ve ConSinGAN'a benzer sonuca ulaşmıştır. Buradaki sonuç, Model 1'in ürettiği görüntülerin piksel değerlerini orijinal görüntünün piksel değerlerine rakiplerinden daha iyi veya benzer oranda yakınsadığını göstermektedir. ConSinGAN, MS-SSIM ve LPIPS dikkate alındığında çeşitlilik açısından diğer tek görüntü ÜÇA modellerinden daha iyi performans sergilemiştir. Ancak çeşitlilik, gerçekçi çeşitli görüntüler üretmekle rastgele piksel değerleri oluşturmak arasında ayrım yapmaya yardımcı olmamaktadır. SIFID değerleriyle birlikte MS-SSIM ve LPIPS değerleri göz önüne alındığında, tüm tek görüntü ÜÇA modellerinin ürettiği görüntülerde yeterli çeşitliliğin olduğu görülmektedir.

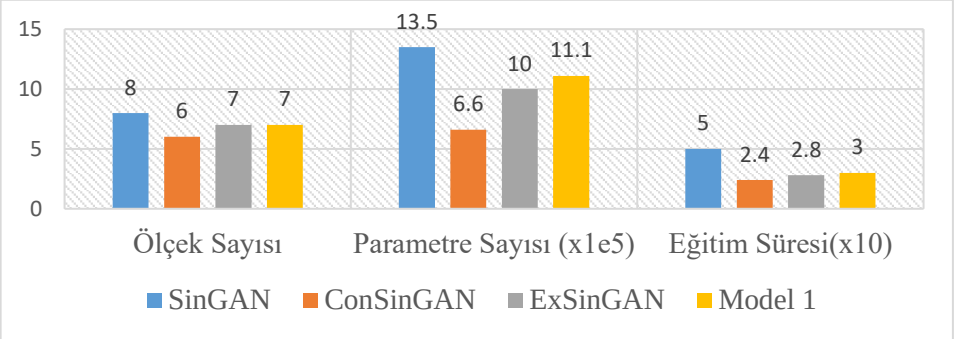
Tablo 4.1: Places, LSUN ve ImageNet veri kümeleri için nicel sonuçlar.

	Places			LSUN			ImageNet		
Model	SIFID (↓)	MSSSIM (↓)	LPIPS (↑)	SIFID (↓)	MSSSIM (↓)	LPIPS (↑)	SIFID (↓)	MSSSIM (↓)	LPIPS (↑)
SinGAN	0.09	0.50	0.46	0.23	**0.42**	0.48	0.13	0.48	0.50
ConSinGAN	**0.06**	0.47	**0.53**	0.11	0.48	0.49	0.11	**0.43**	0.51
ExSinGAN	0.10	0.56	0.40	0.11	0.46	0.48	**0.05**	0.50	0.47
Model 1	**0.06**	**0.45**	0.51	**0.08**	0.44	**0.50**	**0.05**	0.45	**0.52**

LSUN veri kümesi, Places ve ImageNet veri setlerinden daha karmaşık sahnelere, yapılara ve dokulara sahip görüntüler içermektedir. Bu nedenle, LSUN veri kümesinde modellerin görüntü oluşturma performanslarını değerlendirmek önemlidir. LSUN veri kümesi için Model 1 SIFID

değerinde rakiplerinden daha iyi performans göstermektedir. Tek görüntü ÜÇA LSUN veri kümesinden oluşturduğu görüntülerdeki çeşitlilik birbirine yakındır. ImageNet veri kümesi için LPIPS değeri açısından Model 1, SinGAN, ConSinGAN ve ExSinGAN'a göre daha başarılı olduğu görülmektedir. Ek olarak Model 1 ve ExSinGAN ile benzer bir SIFID'e sahip olarak diğer SinGAN ve ConSinGAN'dan daha üstün performans sergilemiştir.

Tablo 4.2: Tek görüntü ÜÇA modellerinin parametre değerleri.



Tablo 4.2 Model 1 ve diğer tek görüntü ÜÇA modellerinin yapısal özelliklerinin karşılaştırmasını içermektedir. Görüldüğü üzere, Model 1, eğitim süresi ve parametre sayısı SinGAN'dan az, ConSinGAN'dan fazla ve ExSinGAN'dan fazla; Ölçek sayısı SinGAN'dan az, ExSinGAN ile aynı ve ConSinGAN'dan fazladır. Model 1'deki öz-dikkat ve yoğun bağlantılı blok, eğitim süresini ve parametre sayısını etkilemektedir. Model başarısı ve parametre sayısı arasında denge durumu vardır. Öz-dikkat ve yoğun bağlantılı blok parametre sayısını artırmasına rağmen eğitim süresinin kısaltmakta ve modelin başarılı sonuçlar ortaya koymasına yardımcı olmaktadır.

4.4.3. Nitel Sonuçlar

Tablo 4.3, sırasıyla Places, LSUN ve ImageNet veri kümeleri için Model 1'in ve diğer tek görüntü ÜÇA modellerinin kullanıcı çalışmaları sonuçlarını sunmaktadır. Beklenildiği üzere tüm veri kümelerinde eşleştirilmemiş kullanıcı çalışmalarında sonuçlar, eşleştirilmiş kullanıcı çalışmalarına oranla daha yüksektir. Bunun temel nedeni eşleştirilmiş kullanıcı çalışmalarında gerçek görüntüyle eşleştirilen üretilmiş görüntüyü kullanıcıların aynı anda inceleyebilmesidir. Bu durumda kullanıcılar sahte görüntüleri daha kolay ayırt etmektedir. Model 1 her iki kullanıcı çalışmasında da diğer modellere üstünlük sağlamaktadır. Bu durum Model 1'in gerçekçi görüntü üretme konusunda mevcut rakipleriyle yarışabilecek düzeyde olduğunu

göstermektedir. Ek olarak kullanıcı çalışmalarının sonuçları tüm tek görüntü ÜÇA modellerinin gerçekçi görüntü üretme konusunda gelişmeye açık olduğunu ortaya çıkarmıştır.

Tablo 4.3: Eşleştirilmiş ve eşleştirilmemiş kullanıcı çalışmaları sonuçları

Model	Places		LSUN		ImageNet	
	Eşleştirilmiş (↑)	Eşleştirilmemiş (↑)	Eşleştirilmiş (↑)	Eşleştirilmemiş (↑)	Eşleştirilmiş (↑)	Eşleştirilmemiş (↑)
SinGAN	11%	22%	7%	14%	3%	6%
ConSinGAN	12%	18%	3%	9%	2%	5%
ExSinGAN	17%	26%	3%	13%	8%	13%
Model 1	**19%**	**27%**	**9%**	**15%**	**9%**	**21%**

Tablo 4.4, katılımcıların her veri kümesinin kullanıcı çalışması için belirli bir model tarafından üretilen görüntüleri ne sıklıkta seçtiğini göstermektedir. Bu çalışmada katılımcılara aynı anda dört model tarafından oluşturulan dört görüntü sunulmaktadır. Katılımcılardan (gerçekçilik, kalite ve çeşitlilik açısından) hangi görüntünün daha iyi olduğuna karar vermeleri istenmektedir (bir görüntü seçmesi istendi). Tablo 4.4'te görüldüğü gibi Model 1 karşılaştırma yapılan diğer tek görüntü ÜÇA modellerinden daha iyi performans göstermiştir

Tablo 4.4: Nitel kullanıcı çalışması sonuçları.

Veri Kümesi	**SinGAN**	**ConSinGAN**	**ExSinGAN**	**Model 1**
Places	24%	18%	28%	**30%**
LSUN	24%	20%	22%	**34%**
ImageNet	13%	15%	30%	**42%**
Ortalama	20%	18%	27%	**35%**

Sonuç olarak tüm kullanıcı çalışmalarında katılımcıların Model 1 tarafından oluşturulan görselleri daha çok tercih ettikleri görülmektedir. Deneyler, Model 1'in tek bir eğitim görüntüsünden başarılı bir şekilde fotoğraf gerçekliğinde, yüksek kaliteli, çeşitli görüntüler oluşturduğunu ve karmaşık görüntülerdeki genel özellikleri, ince ayrıntıları ve doku bilgilerini öğrenebildiğini göstermektedir. Places, LSUN ve ImageNet veri kümelerinden Model 1 ve karşılaştırma yapılan tek görüntü ÜÇA modelleri tarafından üretilen görüntüler Ek-J Şekil 6.10, Ek-K Şekil 6.11 ve Ek-L Şekil 6.12'de verilmiştir.

4.4.4. Ablasyon Çalışması

Model 1 bileşenleri (öz-dikkat, yoğun bağlantılı blok, ResNet, GoogLeNet, RaLSGAN-GP, WGAN-GP) ile görüntü oluşturma performansı arasındaki etkileşimi incelemek için çalışma gerçekleştirilmiştir. Çalışmada her bileşenin görüntü kalitesini, çeşitliliğini, gerçekçiliğini, eğitim süresini ve eğitim istikrarını etkilediği gözlemlenmiştir. Tek bir doğal görüntüden

görüntüler oluşturan Model 1 üç farklı konvolüsyon sinir ağı mimarisiyle tasarlanmıştır; *yoğun bağlantılı blok, ResNet ve GoogLeNet (Inception-v1)*. Places, LSUN ve ImageNet veri kümeleri üzerinde eğitim yapılarak bu mimarilerin görüntü oluşturma performansı üzerindeki etkilerini değerlendirilmiştir. Çıkarım aşamasında her bir eğitim görüntüsünden oluşturulan görüntüler için SIFID, MS-SSIM ve LPIPS değerleri hesaplanmıştır.

Tablo 4.5: Farklı konfigürasyonlara sahip Model 1 için nicel sonuçlar.

Places			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + Yoğun bağlantılı blok * + RaLSGAN-GP	**0.06**	0.45	**0.51**
Öz-dikkat + Yoğun bağlantılı blok * + WGAN-GP	0.10	0.46	0.50
Yoğun bağlantılı blok * + RaLSGAN-GP	0.11	0.47	0.49
Yoğun bağlantılı blok * + WGAN-GP	0.13	0.48	0.49
Öz-dikkat + RaLSGAN-GP	0.14	**0.42**	0.47
Öz-dikkat + WGAN-GP	0.15	0.43	0.48
LSUN			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + Yoğun bağlantılı blok * + RaLSGAN-GP	**0.08**	0.44	**0.50**
Öz-dikkat + Yoğun bağlantılı blok * + WGAN-GP	0.11	0.45	0.50
Yoğun bağlantılı blok * + RaLSGAN-GP	0.10	0.46	0.49
Yoğun bağlantılı blok * + WGAN-GP	0.12	0.47	0.48
Öz-dikkat + RaLSGAN-GP	0.15	**0.40**	0.54
Öz-dikkat + WGAN-GP	0.18	0.42	0.54
ImageNet			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + Yoğun bağlantılı blok * + RaLSGAN-GP	**0.05**	0.45	**0.52**
Öz-dikkat + Yoğun bağlantılı blok * + WGAN-GP	0.09	0.47	0.51
Yoğun bağlantılı blok * + RaLSGAN-GP	0.12	0.46	0.52
Yoğun bağlantılı blok * + WGAN-GP	0.13	0.48	0.50
Öz-dikkat + RaLSGAN-GP	0.14	**0.43**	0.54
Öz-dikkat + WGAN-GP	0.16	0.44	0.53

* Yoğun bağlantılı blok beş adet yoğun bağlantılı katmandan ve bir geçiş katmanından oluşan tek bloktur. (Şekil 3.6)

Tablo 4.5, Tablo 4.6 ve Tablo 4.7, Model 1'in farklı konfigürasyonlardaki nicel sonuçlarını sunmaktadır. Deneysel sonuçlar öz-dikkat, yoğun bağlantılı blok ve RaLSGAN-GP'nin fotogerçekçi, yüksek kaliteli ve çeşitli görüntüler oluşturmada pozitif etkisi olduğunu göstermektedir. Test edilen üç konvolüsyon sinir ağı mimarisi arasında yoğun bağlantılı blok, görüntü kalitesi ve çeşitliliği açısından ResNet ve GoogLeNet'ten daha iyi performans göstermiştir. Ayrıca ResNet ve GoogLeNet'in aşırı uyum sağlamaya daha yatkın olduğu ortaya

çıkmıştır. Ek-N Tablo 6.1, Tablo 6.2 ve Tablo 6.3 yoğun bağlantılı blok, ResNet ve GoogLeNet tabanlı Model 1'in farklı konfigürasyonlarının parametre sayısı ve eğitim süresini göstermektedir.

Tablo 4.6: ResNet tabanlı oluşturulan Model 1 için nicel sonuçlar.

Places			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + ResNet + RaLSGAN-GP	0.01	0.97	0.13
Öz-dikkat + ResNet + WGAN-GP	0.01	0.97	0.16
ResNet + RaLSGAN-GP	0.01	0.97	0.13
ResNet + WGAN-GP	0.02	0.97	0.16
LSUN			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + ResNet + RaLSGAN-GP	0.01	0.98	0.08
Öz-dikkat + ResNet + WGAN-GP	0.02	0.97	0.14
ResNet + RaLSGAN-GP	0.01	0.98	0.08
ResNet + WGAN-GP	0.02	0.97	0.14
ImageNet			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + ResNet + RaLSGAN-GP	0.006	0.98	0.08
Öz-dikkat + ResNet + WGAN-GP	0.02	0.97	0.13
ResNet + RaLSGAN-GP	0.006	0.98	0.07
ResNet + WGAN-GP	0.01	0.97	0.12

Tablo 4.7: GoogLeNet (Inception v1) tabanlı oluşturulan Model 1 için nicel sonuçlar.

Places			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + GoogLeNet + RaLSGAN-GP	0.03	0.97	0.15
Öz-dikkat + GoogLeNet + WGAN-GP	0.06	0.95	0.21
GoogLeNet + RALSGAN-GP	0.03	0.97	0.14
GoogLeNet + WGAN-GP	0.07	0.96	0.21
LSUN			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + GoogLeNet + RaLSGAN-GP	0.02	0.97	0.11
Öz-dikkat + GoogLeNet + WGAN-GP	0.05	0.96	0.17
GoogLeNet + RaLSGAN-GP	0.02	0.97	0.11
GoogLeNet + WGAN-GP	0.04	0.96	0.18
ImageNet			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + GoogLeNet + RaLSGAN-GP	0.02	0.98	0.12
Öz-dikkat + GoogLeNet + WGAN-GP	0.05	0.96	0.20

Tablo 4.7 (devam): GoogLeNet (Inception v1) tabanlı oluşturulan Model 1 için nicel sonuçlar.

GoogLeNet + RaLSGAN-GP	0.03	0.96	0.13
GoogLeNet + WGAN-GP	0.04	0.96	0.19

Model 1'in yoğun bağlantılı blok mimarisi için farklı konfigürasyonlar oluşturularak bu konfigürasyonların görüntü üretimine etkisi incelenmiştir. Tablo 4.8'de yoğun bağlantılı blok mimarisinin farklı konfigürasyonları için Model 1'in performans sonuçları sunulmaktadır.

Tablo 4.8: Model 1'in farklı yoğun bağlantılı blok konfigürasyonları için nicel sonuçlar.

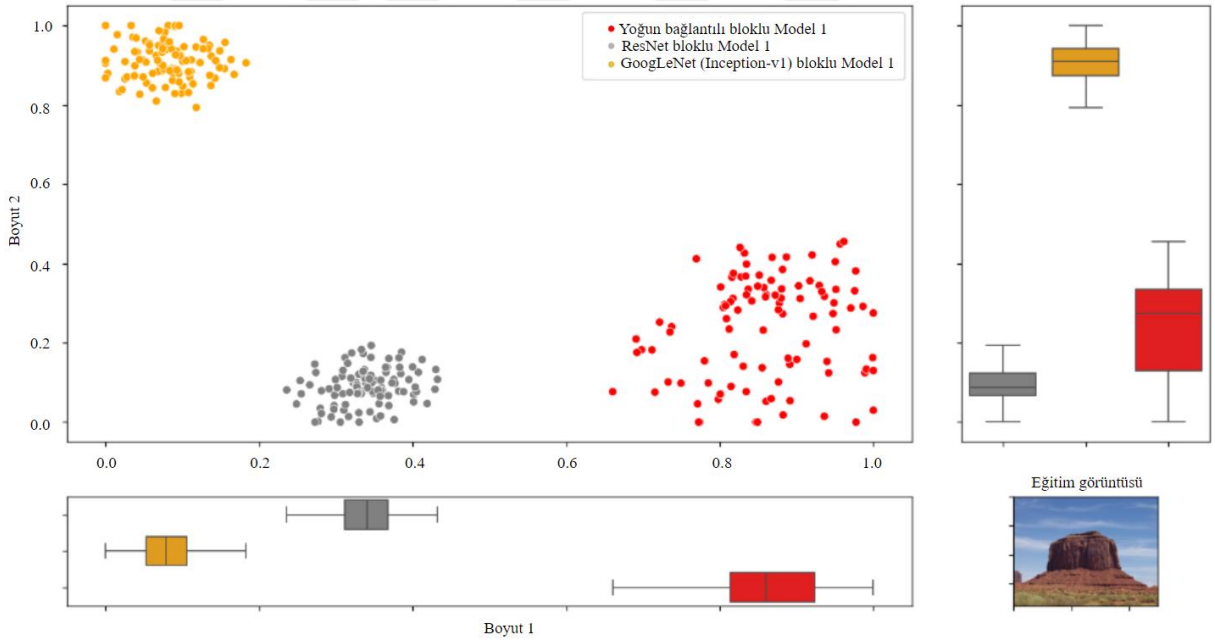
Places			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü'de ortalama havuzlama var) + RaLSGAN-GP	0.10	0.47	0.48
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.08	0.56	0.40
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	**0.06**	**0.45**	**0.51**
Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	0.09	0.80	0.32
Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.08	0.85	0.30
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	0.04	0.96	0.17
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.04	0.96	0.18
LSUN			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü'de ortalama havuzlama var) + RaLSGAN-GP	0.09	0.46	0.48
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.08	0.48	0.45
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	**0.08**	**0.44**	**0.50**

Tablo 4.8 (devam): Model 1'in farklı yoğun bağlantılı blok konfigürasyonları için nicel sonuçlar.

Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	0.08	0.65	0.35
Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.05	0.80	0.30
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	0.02	0.95	0.15
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.02	0.95	0.15
ImageNet			
Model 1:	**SIFID** (↓)	**MS-SSIM** (↓)	**LPIPS** (↑)
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü'de ortalama havuzlama var) + RaLSGAN-GP	0.10	0.47	0.50
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.10	0.49	0.48
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	**0.05**	**0.45**	**0.52**
Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	0.09	0.65	0.35
Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.09	0.75	0.32
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama yok) + RaLSGAN-GP	0.03	0.92	0.17
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve *A*'da ortalama havuzlama var) + RaLSGAN-GP	0.02	0.96	0.16

Öz-dikkat: Tablo 4.5'te görüldüğü gibi Places, LSUN ve ImageNet veri kümeleri için öz-dikkat bulunan Model 1, bulunmayana göre daha düşük bir SIFID'e sahiptir. Çeşitlilik açısından da öz-dikkat bulunan Model 1, bulunmayana göre daha başarılıdır. Sonuç olarak öz-dikkat, Model 1'in performansını anlamsal çeşitlilik ve görüntü oluşturma kalitesi açısından pozitif ölçüde artırmaktadır. Ayrıca, öz-dikkat'in ağın farklı yerlerinde kullanılmasının, oluşturulan görüntülerin çeşitliliği ve aslına uygunluğu arasında iyi bir denge sağladığı gözlemlenmiştir.

Yoğun Bağlantılı Blok: Model 1 tarafından farklı konvolüsyon sinir ağı mimarileri (Yoğun bağlantılı blok, ResNet ve GoogLeNet) ile üretilen görüntülerin yapısal özelliklerini (örneğin anlamsal çeşitlilik) analiz etmek için t-SNE veri görselleştirme tekniği [111] kullanılmıştır. Oluşturulan görüntülerin veri noktalarının düşük boyutlu uzaya yansıtıldığında nasıl davrandığını incelenmiştir. Şekil 4.6, yoğun bağlantılı blok, ResNet ve GoogLeNet mimarilerini çeşitlilik açısından karşılaştırmak için oluşturulan görüntülerimizin t-SNE görselleştirmesini göstermektedir. Buna göre ResNet ve GoogLeNet mimarileri ile oluşturulan Model 1'in ürettiği görüntüler daha dar bir alanda yer alırken (bu, örtüşen çok sayıda yama olduğu ve dolayısıyla ezberleme olduğu anlamına gelmektedir.), yoğun bağlantılı blok ile oluşturulan Model 1'in ürettiği görüntüler daha geniş bir alana yayılmıştır. Dolayısıyla yoğun bağlantılı blok mimarisine sahip Model 1 çeşitlilik açısından diğer mimarilere göre daha iyi sonuçlar vermektedir.



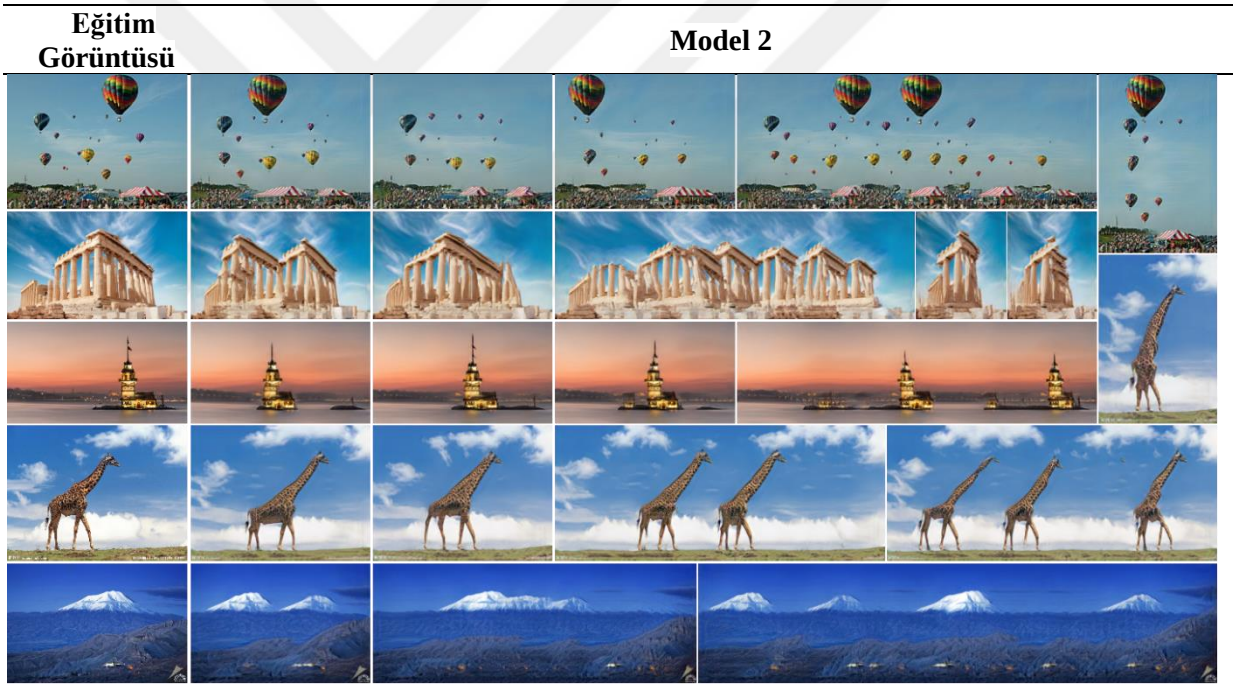
Şekil 4.6: Görüntülerin t-SNE ile iki boyutlu uzayda gösterimi

RaLSGAN-GP: Kayıp fonksiyonunun etkinliğini analiz etmek için Model 1 RaLSGAN-GP ve WGAN-GP ile eğitilmiştir. Tablo 4.5, Places, LSUN ve ImageNet veri kümeleri için Model 1'in performans sonuçlarını göstermektedir. Tüm bu veri kümeleri için RaLSGAN-GP'li Model 1'in SIFID değeri, WGAN-GP'li Model 1'İn SIFID değerine göre daha düşüktür. SIFID, MS-SSIM ve LPISP sonuçlarını birlikte değerlendirildiğinde RaLSGAN-GP Model 1'in başarısında daha etkilidir. Ayrıca genel olarak çeşitlilik açısından değerlendirildiğinde RaLSGAN-GP'li Model 1, WGAN-GP'li Model 1'e göre daha iyi sonuçlar vermiştir.

4.5. Model 2 Sonuçları

4.5.1. Görsel Karşılaştırma

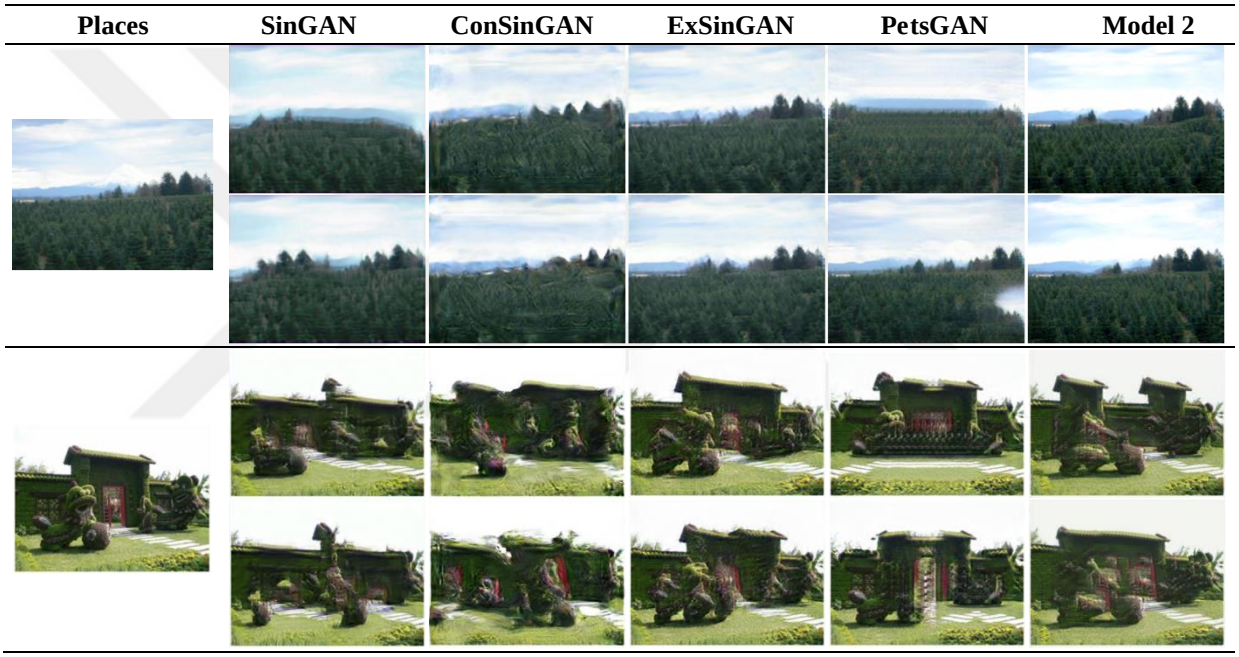
Şekil 4.7, farklı özellikler içeren görüntüler üzerinde eğitilen Model 2 tarafından üretilen görüntüler arasından seçilen rastgele örnekleri göstermektedir. Eğitim görüntüleri dokusal (*ör. dağ, tarihi eser*), anlamsal (*ör. zürafa, kız kulesi*) ve yapısal (*ör. balonlar*) varyasyonları içerecek şekilde seçilmiştir. Sonuçlara göre Model 2, çıkarım aşamasında eğitim görüntülerinden hem yerel hem de küresel bağımlılıkları koruduğu gözlemlenmektedir. Üretilen görüntüler tek eğitim görüntüsü ile aynı örüntülere sahiptir ve yüksek düzeyde bir çeşitlilik sergilemektedir. Örneğin zürafa eğitim görüntüsünden üretilen örneklerde Model 2, zürafa nesnesinin yapısını koruyarak farklı konumlarda zürafa içeren yeni görüntüleri başarılı bir şekilde üretmiştir.



Şekil 4.7: Model 2 tarafından oluşturulan rastgele görüntü örnekleri.

Konvolüsyon ağlardan oluşan Model 2'den değişken boyutlu gürültü girdilerinden, çıkarım aşamasında farklı boyutlarda görüntüler oluşturmak mümkündür. Şekil 4.7, değişken girdi boyutlarına dayalı olarak Model 2 tarafından üretilen yüksek gerçekçiliğe ve çeşitliliğe sahip görüntüleri de içermektedir. İlk satırdaki balon görüntüleri için Model 2, farklı boyutlarda, sayıda ve konumda nesneleri başarılı ve gerçekçi olarak üretmeyi başarmıştır. Yapılan eğitim sonucu Model 2 tarafından üretilen daha fazla örnek Ek-D Şekil 6.2'de gösterilmektedir.

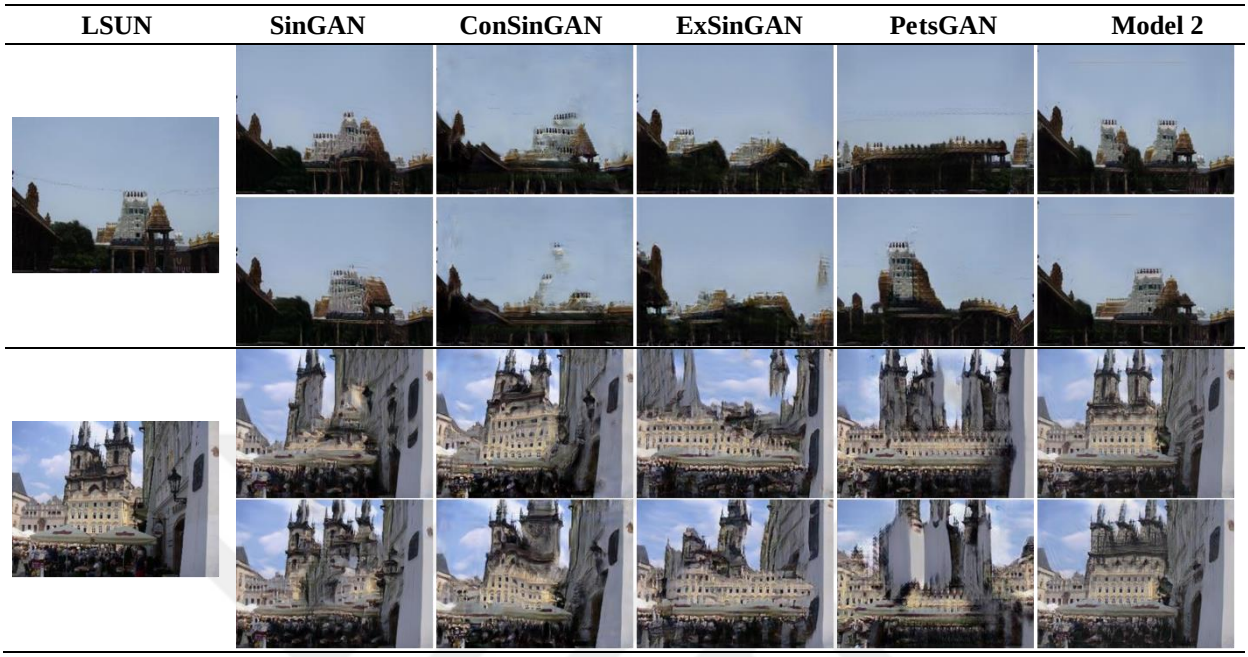
Görsel performansını karşılaştırmak amacıyla, Model 2 ile diğer tek görüntü ÜÇA modelleri arasında görsel bir analiz yapılmıştır. Şekil 4.8, Şekil 4.9 ve Şekil 4.10, sırasıyla Places, LSUN ve ImageNet veri kümeleri seçilen ikişer eğitim görüntüsü için Model 2, SinGAN, ConSinGAN, ExSinGAN ve PetsGAN modelleri tarafından oluşturulan görüntüleri içermektedir. Şekil 4.8'de yinelenen dokulara sahip görüntülerden oluşan Places veri kümesindeki tüm tek görüntü ÜÇA modelleri, diğer veri kümelerine kıyasla daha gerçekçi görüntüler oluşturmuştur. Places veri kümesindeki görüntüler için Model 2, üretilen görüntülerde doku uyumluluğu ve çeşitlilik açısından diğer tek görüntü ÜÇA modellerine göre üstünlük göstermektedir.



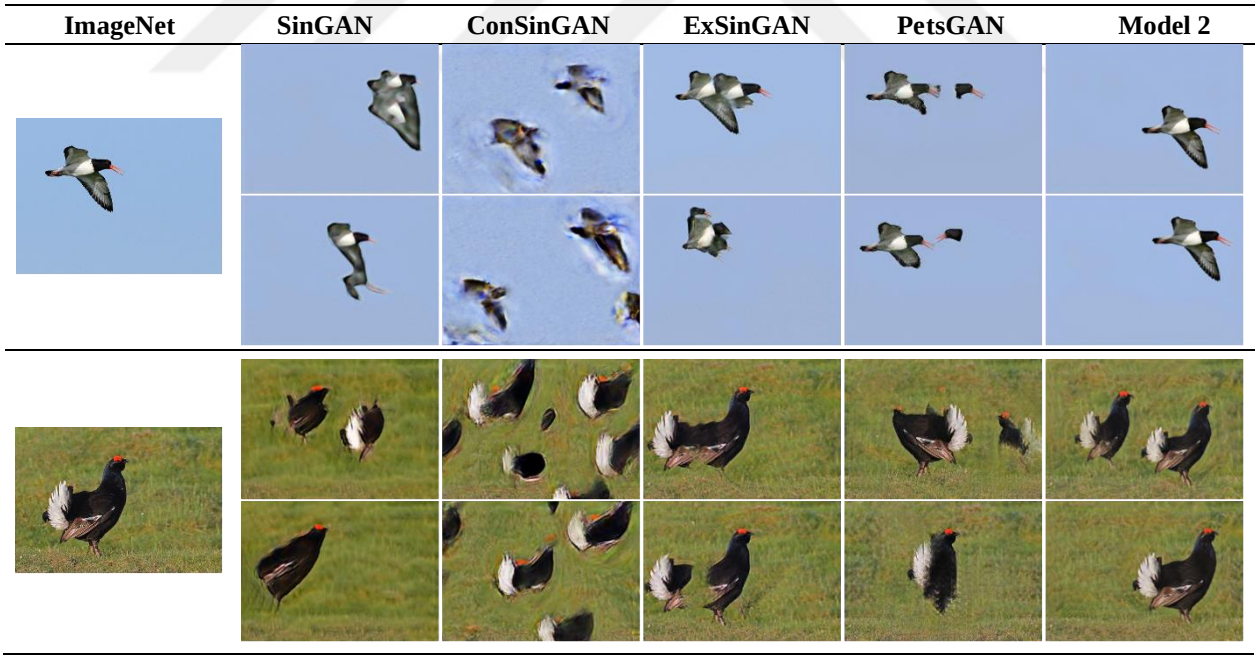
Şekil 4.8: Model karşılaştırma (Places).

Model 2, karmaşık veri kümesi olan LSUN veri kümesinde de daha iyi görüntüler üretmeyi başarmıştır (bkz. Şekil 4.9). Özellikle Model 2'nin, *The Church of Our Lady before Týn* eğitim görüntüsünün (bkz. Şekil 4.9- 3. ve 4. satır) yerel ve global doku bağımlılıklarını başarılı bir şekilde öğrendiği görülmektedir. ImageNet veri kümesi, çoğunlukla görüntü içerisinde tek bir büyük nesne içeren görüntülerden oluşmaktadır. Bu bağlamda ImageNet veri kümesinde tek görüntü ÜÇA'ları için nesnelerin bütünlüğünü korumak, üretilen görüntülerin gerçekçilik sergilemesi açısından daha önemlidir. Yapılan deneylerde tek görüntü ÜÇA modellerinin ImageNet veri kümesinde Places ve LSUN veri kümelerine oranla daha az çeşitlilik sergileyen görüntüler ürettiği gözlemlenmiştir. Model 2 bu zorluğu başarılı şekilde aşarak Şekil 4.10'da görüldüğü üzere diğer modellere oranla çok daha gerçekçi ve çeşitli görüntüler üretmeyi

başarmıştır. Özellikle *tavuk* nesnesinin bulunduğu eğitim görüntüsünde Model 2 diğer modellere kıyasla gerçekçi nesne çoğullamayı ve konumlamayı başarmıştır.



Şekil 4.9: Model karşılaştırma (LSUN).



Şekil 4.10: Model karşılaştırma (ImageNet)

SinGAN ve ConSinGAN, görüntü yamalarını değiştirmeye odaklandığından eğitim görüntülerinde nesne bütünlüğünü doğru şekilde koruyamamaktadır. ExSinGAN ve PetsGAN, ÜÇA tersine çevirme ve veri artırma araçlarını kullanarak, nesne bütünlüğünü SinGAN ve ConSinGAN'a oranla daha fazla koruyan görüntüler üretebilmektedir. Ancak, bu tek görüntü

ÜÇA'ları nesneyi yeniden konumlandırma (yer değiştirme) açısından yetersiz kalabilmektedir. Özellikle PetsGAN'ın eğitim sürecinde kullanılan eğitim görüntüsünü dikey ters çevirme ve PetsGAN'ın çıkarım aşamasındaki izlediği proseüdür, nesnelerin yapısının korunmasında olumsuz etkiye sahip olabilmektedir. Şekil 4.8, Şekil 4.9 ve Şekil 4.10'da oluşturulan görüntüler incelendiğinde Model 2'nin hem görüntü içindeki nesne yapısını hem de bağımlılıkları ile nesne konumlarını değiştirebilme yeteneğini başarılı bir şekilde öğrendiği görülmektedir. Bu sonuçlar, Model 2'nin eğitim görüntüsünden K ağı aracılığıyla elde edilen gizli uzay bilgisini etkili bir şekilde kullandığını göstermektedir. Ayrıca Model 2'nin tek görüntü ÜÇA problemi için etkili bir çözüm olduğu ortaya çıkmaktadır.

4.5.2. Nicel Sonuçlar

Model 2, SinGAN, ConSinGAN, ExSinGAN ve PetsGAN Places, LSUN ve ImageNet veri kümeleri üzerinde eğitilmiştir. Tüm tek görüntü ÜÇA modelleri çıkarım aşamasında her eğitim görüntüsünden oluşturulan örneklerden ÜÇA modellerinin SIFID, MS-SSIM ve LPIPS değerleri hesaplanmıştır.

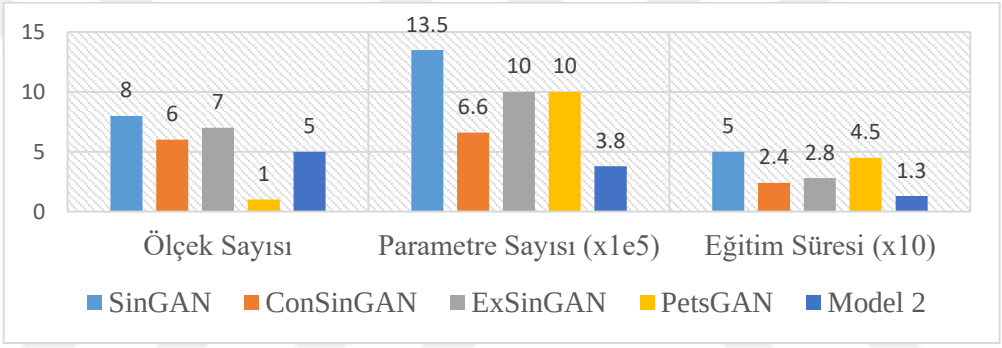
Tablo 4.9, Places, LSUN ve ImageNet veri kümeleri üzerinde eğitilen tek görüntü ÜÇA modelleri tarafından elde edilen SIFID, MS-SSIM ve LPIPS değerlerini içermektedir. SIFID değerlerine göre Model 2 diğer tek görüntü ÜÇA'larından daha iyi performans göstermekte ve tüm veri kümelerinde en iyi skorları elde etmektedir. Çeşitlilik açısından bazı tek görüntü ÜÇA modelleri Model 2'ye göre daha iyi değerlere sahiptir. Fakat bu ÜÇA modellerinin SIFID değerleri veya görsel sonuçları daha zayıf olabilmektedir. Bu durum yüksek çeşitlilik değerleri ile görsel olarak gerçekçi görüntüler oluşturma arasındaki dengeyi göstermektedir. Bunun sonucu olarak çeşitlilik değerleri açısından tüm modellerin benzer sayılabilecek sonuçlara ulaştığı çıkarımı yapılabilir. Model 2 tarafından elde edilen SIFID ve çeşitlilik değerleri, Model 2'nin gerçekçi dokusal özelliklere sahip farklı görüntüler üretme yeteneğini göstermektedir.

Tablo 4.9: Places, LSUN ve ImageNet veri kümeleri için nicel sonuçlar.

	Places			LSUN			ImageNet		
Model	SIFID(↓)	MSSSIM(↓)	LPIPS(↑)	SIFID(↓)	MSSSIM(↓)	LPIPS(↑)	SIFID(↓)	MSSSIM(↓)	LPIPS(↑)
SinGAN	0.09	0.50	0.46	0.23	0.42	0.48	0.13	0.48	0.50
ConSinGAN	0.06	0.47	**0.53**	0.11	0.48	**0.49**	0.11	0.43	**0.51**
ExSinGAN	0.10	0.56	0.40	0.11	0.46	0.48	***0.05***	0.50	0.47
PetsGAN	0.08	***0.46***	0.23	0.09	***0.35***	0.26	0.07	**0.37**	0.25
Model 2	***0.05***	0.65	0.35	***0.07***	0.52	0.43	***0.05***	0.58	0.35

Tablo 4.10, Model 1 ve diğer tek görüntü ÜÇA modellerinin yapısal özelliklerinin karşılaştırmasını içermektedir. PetsGAN, ölçek sayısı açısından lider gibi görünse de model dahili olarak piksel karıştırma [65] ile ölçek ayarlaması yapmaktadır. Bu doğrultuda PetsGAN eğitim aşamasında 4 ölçekli bir yapıya sahip olmasına rağmen PetsGAN bir bütün olarak eğitilmesi sebebiyle ölçek sayısı 1 olarak kabul edilmiştir. PetsGAN istisnai durumu haricinde, Model 2 eğitim süresi ve öğrenilebilir parametre sayısı açısından rakiplerine oranla üstünlük sergilemektedir. Model 2'nin daha kısa eğitim süresi ve eğitilebilir parametre ile başarılı sonuçlara ulaşması, Model 2'nin eğitim aşamasında gürültü verisi bilgisinden etkili şekilde yararlanan özgün bir model olduğunu göstermektedir.

Tablo 4.10: Tek görüntü ÜÇA modellerinin parametre değerleri.



Nicel ve performans metriklerin analizi, Model 2'yi gerçekçi dokulara sahip yüksek kaliteli görüntüler oluşturmak için oldukça etkili ve verimli bir tek görüntü ÜÇA modeli olarak ortaya koymaktadır. Model 2'nin SIFID değerleri açısından diğer modellerden daha iyi performans göstermesi, görsel olarak gerçekçi görüntüler oluşturma konusundaki yeterliliğini göstermektedir. Ek olarak, Model 2 rekabetçi çeşitlilik değerleri sergilemektedir ve eğitim verimliliği, parametre kullanımı açısından üst seviyededir.

4.5.3. Nitel Sonuçlar

Model 2, Bölüm 4.3'te açıklandığı üzere eşleştirilmiş, eşleştirilmemiş ve tüm modellerle karşılaştırma olmak üzere 3 adet kullanıcı çalışmasına tabi tutulmuştur. Amaç Model 2 için nicel olarak elde edilen sonuçları nitel olarak desteklemektir. SinGAN, ConSinGAN ve ExSinGAN modellerinin kullanıcı çalışmaları Bölüm 4.4.3'te açıklanmıştır. Model 2 için PetsGAN ÜÇA modeli için de kullanıcı çalışmaları yapılmıştır. Tablo 4.11, karşılaştırma yapılan 5 tek görüntü ÜÇA modeli için eşleştirilmiş ve eşleştirilmemiş kullanıcı çalışmaları sonuçlarını listelemektedir. Tüm veri kümelerinde Model 2 en iyi sonuçlara ulaşmıştır.

Katılımcılar beklendiği üzere ExSinGAN ve PetsGAN'ı Model 2'ye en yakın tek görüntü ÜÇA modelleri olarak değerlendirmiştir. ConSinGAN'ın sonuçları incelendiğinde, ConSinGAN'ın MS-SSIM ve LPIPS çeşitlilik ölçümlerinde gösterdiği başarılı sonuçların görüntü gerçekçiliğine negatif etkisi olduğu görülmektedir. Bu durum yüksek çeşitlilik göstermek ile gerçekçilik arasındaki dengenin önemini vurgulamaktadır. LSUN veri kümesi üzerinde yapılan iki kullanıcı çalışması sonucunda da tüm tek görüntü ÜÇA modelleri en düşük değerleri elde etmiştir. Nicel ölçüm sonuçları ardından kullanıcı çalışmalarında da LSUN veri kümesinin zorluğu ortaya çıkmıştır.

Tablo 4.11: Eşleştirilmiş ve eşleştirilmemiş kullanıcı çalışmaları sonuçları.

Model	Places		LSUN		ImageNet	
	Eşleştirilmiş (↑)	Eşleştirilmemiş (↑)	Eşleştirilmiş (↑)	Eşleştirilmemiş (↑)	Eşleştirilmiş (↑)	Eşleştirilmemiş (↑)
SinGAN	11%	22%	7%	14%	3%	6%
ConSinGAN	12%	18%	3%	9%	2%	5%
ExSinGAN	17%	**26%**	3%	13%	8%	13%
PetsGAN	13%	18%	7%	13%	11%	15%
Model 2	**22%**	**26%**	**10%**	**15%**	**18%**	**25%**

Tablo 4.12, nitel değerlendirme kullanıcı çalışması sonuçlarını göstermektedir. Burada Model 2 ve PetsGAN modeli diğer modellere üstünlük kurmuştur. Bununla beraber tüm veri kümelerinin ortalamaları ve LSUN veri kümesi haricinde Model 2 en başarılı olan modeldir. Dolayısıyla, eğitim görüntüsünün görüntü uzayının kullanılmasının ve Model 2'nin özgün eğitim şemasının tek görüntü oluşturma problemine etkili bir çözüm getirdiği görülmektedir. Places, LSUN ve ImageNet veri kümelerinden Model 2 ve karşılaştırma yapılan tek görüntü ÜÇA modelleri tarafından üretilen görüntüler Ek-J Şekil 6.10, Ek-K Şekil 6.11 ve Ek-L Şekil 6.12'de verilmiştir.

Tablo 4.12: Nitel kullanıcı çalışması sonuçları.

Veri Kümesi	**SinGAN**	**ConSinGAN**	**ExSinGAN**	**PetsGAN**	**Model 2**
Places	13%	11%	13%	24%	**39%**
LSUN	9%	8%	10%	**41%**	32%
ImageNet	3%	3%	10%	32%	**46%**
Ortalama	8%	7%	11%	33%	**39%**

4.5.4. Ablasyon çalışması

Model 2'nin performansını etkileyen faktörleri daha iyi anlamak ve model bileşenlerinin katkısını analiz etmek amacıyla ablasyon çalışması gerçekleştirilmiştir. Bunun için üç farklı model konfigürasyonu hazırlanmıştır. Hazırlanan modeller Places, LSUN ve ImageNet veri

kümeleri üzerinde eğitilmiştir. Eğitim sonrası üretilen örnekler üzerinde SIFID, MS-SSIM ve LPIP ölçümleri yapılmıştır. Tablo 4.13 elde edilen sonuçları göstermektedir.

İlk model konfigürasyonunda, çok görüntü eğitim kümesi için önerilen klasik ÇÖÇ [41], [42] modeli, SinGAN modelinin eğitim prosedürü kullanılarak tek bir eğitim görüntüsü kullanarak eğitilmiştir. Klasik ÇÖÇ modelinin yüksek SIFID değerleri sergilediğini ve bunun da gerçekçi görüntüler oluşturma konusundaki sınırlı becerisini gösterdiği görülmektedir. Ayrıca klasik ÇÖÇ'ün sahip olduğu yüksek çeşitlilik değerleri, oluşturulan görüntülerdeki yamaların dağılımında tutarlılık eksikliği bulunduğunu göstermektedir. Bu bulgular, klasik ÇÖÇ modelinin karmaşık ayrıntıları yakalama ve eğitim görüntüsünü aslına sadık bir şekilde temsil etme konusundaki eksikliklerini vurgulamaktadır.

Tablo 4.13. Ablasyon çalışmasının nicel sonuçları.

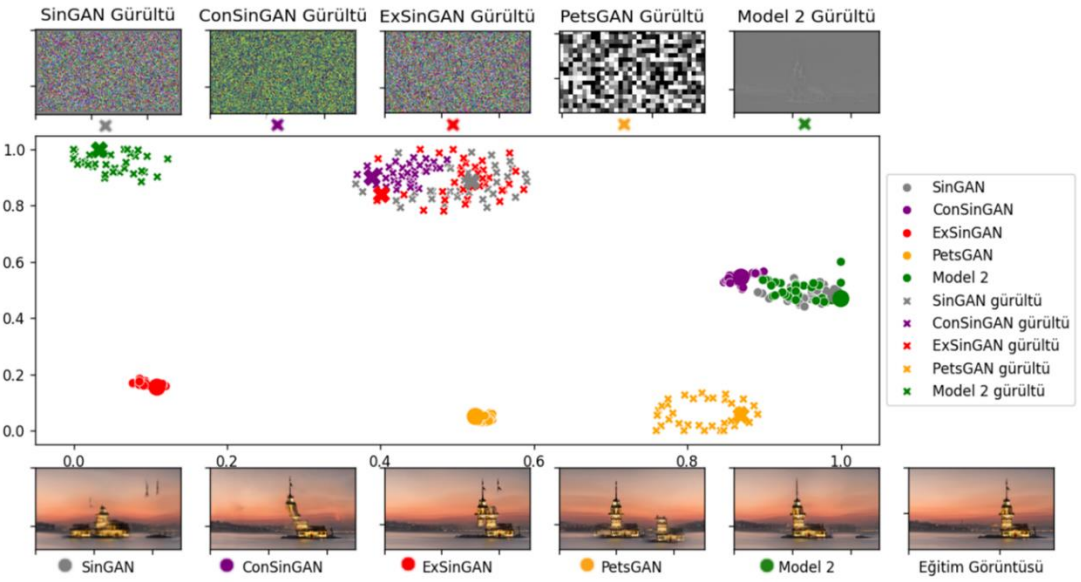
Model	Places			LSUN			ImageNet		
	SIFID (↓)	MSSSIM (↓)	LPIPS (↑)	SIFID (↓)	MSSSIM (↓)	LPIPS (↑)	SIFID (↓)	MSSSIM (↓)	LPIPS (↑)
Klasik ÇÖÇ	2.8	**0.38**	**0.58**	1.27	**0.34**	**0.57**	3.3	**0.30**	**0.61**
Model 2 (gürültü yok)	**0.05**	0.66	0.35	0.08	0.50	0.42	0.06	0.55	0.36
Model 2 (h_{sfi}=0.5)	**0.05**	0.65	0.35	0.07	0.52	0.43	**0.05**	0.58	0.35
Model 2 (h_{sfi}=0.6)	**0.05**	0.58	0.39	**0.06**	0.63	0.30	0.08	0.51	0.40
Model 2 (h_{sfi}=0.4)	0.06	0.54	0.42	0.10	0.40	0.48	0.11	0.32	0.49

İkinci model konfigürasyonunda, model 2, tek eğitim görüntüsüne küçük değerli α gürültü haritasını eklemeden ($z_k^p = K_k(i_k^s)$) eğitilmiştir. Elde edilen nicel sonuçlar Model 2 ile karşılaştırılabilir olsa da eğitim görüntülerine gürültü eklemenin modelin performansı üzerinde olumlu bir etkisi olduğu görülmektedir. Bu durum tek eğitim görüntüsüne küçük değerli α gürültüsü eklenerek K ağına gönderilmesi, K ağının öğrenme sürecinin durmadan devam etmesini ve Model 2'nin etkili çalışmasını sağladığını ortaya çıkarmaktadır. Ayrıca K_{k-1} ağının $Ü_k$ ağında kodlayıcı olarak kullanılması eğitim görüntüsü ve üretilen görüntüler arasında güçlü ilişki kurulmasını sağlamaktadır

Üçüncü model konfigürasyonunda, model performansı ile eğitim süresi arasında en uygun dengeyi bulma amaçlanmıştır. Bu doğrultuda Model 2 farklı ölçeklendirme faktörleriyle (s_f) eğitilmiştir. Buna göre $h_{sfi} = 0.4, 0.5$ ve 0.6 olmak üzere Model 2, üç farklı ölçek değeriyle eğitilmiştir. SIFID ve çeşitlilik değerleri incelendiğinde $h_{sfi} = 0.5$ ölçekleme değerinin Model 2 için en uygun değer olduğu görülmektedir. $h_{sfi} = 0.5$ ölçeklendirme faktörü görüntü

kalitesi ve çeşitlilik arasında iyi bir denge kurarak Model 2'nin eğitim süresini en aza indirirken yüksek kaliteli görüntüler oluşturmasını sağlamaktadır.

Girdi gürültüsü haritasının Model 2'ye etkisi açısından tüm tek görüntü ÜÇA modellerinin karşılaştırması, t-SNE [111] görselleştirmesi kullanılarak Şekil 4.11'de sunulmaktadır. Tek görüntü ÜÇA modellerinin girdi gürültüleri ve bunlara karşılık oluşturulan görüntüler, iki boyutlu uzayda t-SNE ile yansıtılmıştır.



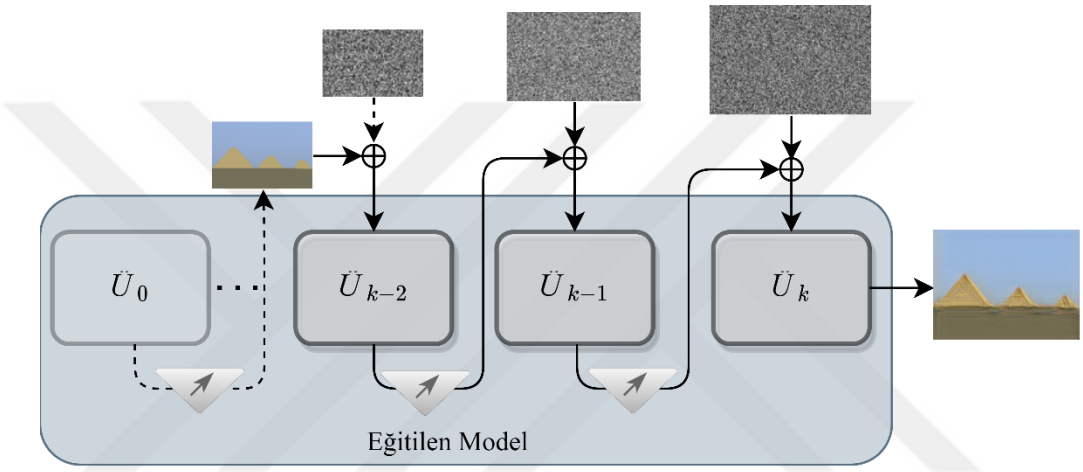
Şekil 4.11: Görüntü ve gürültülerin t-SNE ile iki boyutlu uzayda gösterimi.

Şekil 4.11'de PetsGAN haricinde, diğer modellerin girdi gürültüleri veya oluşturulan görüntüleri arasında bir kesişme olduğu görülmektedir. Bu kesişme durumuna rağmen Model 2'nin nicel ve nitel ölçümlerde daha başarılı olması, Model 2'nin eğitim görüntüsünün gürültü uzayını başarılı olarak kullandığını göstermektedir. Diğer bir deyişle Model 2'nin başarısının, eğitim prosedürüne ve çıkarım aşamasında eğitim görüntüsünden çıkarılan gizli uzayın kullanımına bağlı olduğu görülmektedir.

4.6. Uygulamalar

Bu bölümde *Görüntü düzenleme, harmanlama, çizimden görüntü oluşturma, animasyon ve farklı boyutlarda görüntü oluşturma* olmak üzere 5 görüntü işleme uygulaması yapılmıştır. Amaç tez kapsamında önerilen iki modelin etkinliğini, verimliliğini ve sağlamlığını

göstermektedir. Uygulamalar, herhangi bir ek bilgi veya eğitim görüntüsü kullanılmadan, farklı yapıda eğitim olmadan gerçekleştirildi. Her uygulama için model tek eğitim görüntüsüyle koşulsuz olarak eğitilmektedir. Eğitim sonrası çıkarım aşamasında belirlenen bir ölçekte uygulama görüntüsü modele girdi olarak verilmektedir. $(k-2)$. Ölçek örneği Şekil 4.12'de görüldüğü üzere eğitim sonrası çıkarım aşamasında $(k-2)$. ölçekte $k-2 \leq n_s$ olacak şekilde uygulama görüntüsü modele girdi olarak verilmektedir. $k = n_s$ ölçekte elde edilen çıktı nihai görüntüyü oluşturmaktadır [28]. Uygulamalar için çıkarım aşamasında model üzerinde herhangi bir mimari değişiklik yapılmamaktadır.

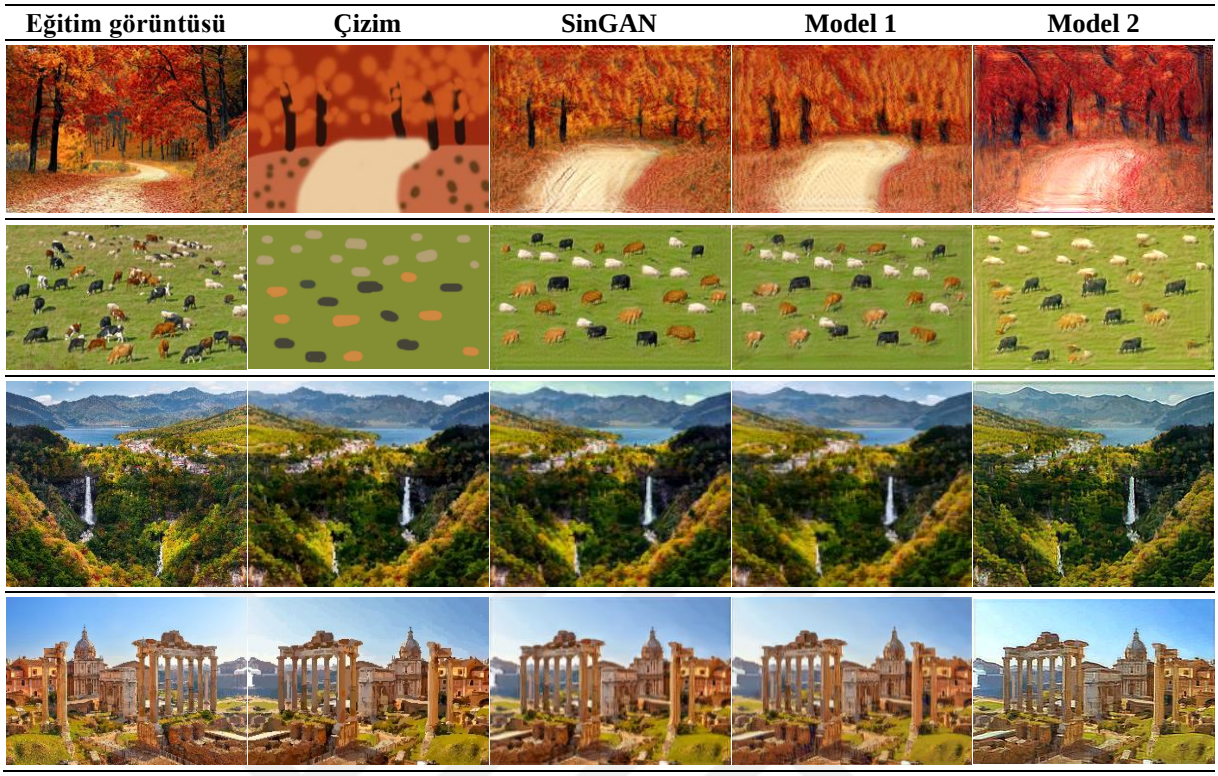


Şekil 4.12: Görüntü manipülasyonu uygulamalarında işlem süreci.

4.6.1. Çizimden Görüntü Oluşturma

Çizimden Görüntü Oluşturma uygulamasında, eğitim görüntüsüne benzer olarak çizilen bir resmi gerçeklemek amaçlamaktadır. Çizilen görüntüyü girdi olarak alan Ü, bu görüntüyü gerçek görüntüye çevirmektedir. Belirlenen eğitim görüntüsü kullanılarak ÜÇA eğitilmektedir. Çıkarım zamanında, çizilen görüntünün ölçeklendirilmiş sürümü $k \leq n_s$ olmak üzere $Ü_k$ ağına gönderilmektedir. Ü hiyerarşik olarak $k = n_s$ ölçeğine gelene kadar görüntü üretmektedir. $k = n_s$ ölçekteki çıktı görüntüsü çizimden oluşturulan görüntüyü göstermektedir.

Şekil 4.13'te görüldüğü gibi önerilen tek görüntü ÜÇA modelleri, çizimleri foto-gerçekçi bir görüntüye başarıyla aktarmaktadır. Oluşturulan görüntülerde, görüntünün düzeni ve genel yapısı korurken eğitim görüntüyle eşleşen gerçekçi dokular üretilebilmektedir. Önerilen tek görüntü ÜÇA modellerinin ürettiği görüntüler, rakiplerinden daha üç boyutlu (ör. piramitler) ve bağlam tabanlıdır. Çizimden Görüntü Oluşturma uygulaması için daha fazla sayıda örnek Ek-E Şekil 6.3'te bulunmaktadır.

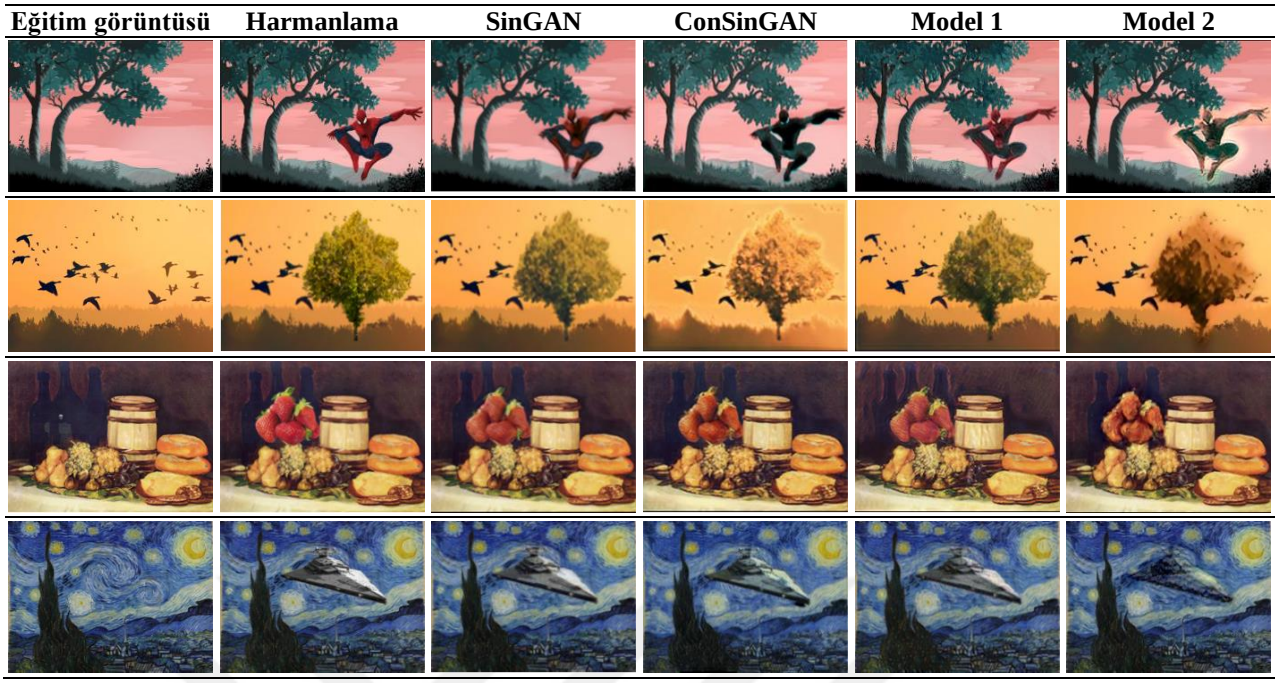


Şekil 4.13: Çizimden Görüntüye uygulama sonuçları.

4.6.2. Harmanlama

Harmanlama uygulaması eğitim görüntüsü üzerine eklenen bir nesnenin, eğitim görüntüsünün dokusal yapısına uygun olarak değiştirilmesidir. Eklenen nesnenin arka plan görüntüsündeki özellikleri taşıma oranı gerçekçiliği göstermektedir. Belirlenen eğitim görüntüsü kullanılarak model eğitilmektedir. Çıkarım zamanında, eklenen nesneyi içeren görüntünün ölçeklendirilmiş sürümü $k \leq n_s$ olmak üzere $Ü_k$ ağına gönderilmektedir. Ü hiyerarşik olarak $k = n_s$ ölçeğine gelene kadar görüntü üretmektedir. $k = n_s$ ölçekteki çıktı görüntüsü eklenen nesnenin eğitim görüntüsüne harmanlanmış halini göstermektedir.

Şekil 4.14, önerilen tek görüntü ÜÇA modellerinin harmanlama uygulaması için yeteneğini göstermektedir. Önerilen tek görüntü ÜÇA modellerinin oluşturduğu görüntülerde global yapıları ve ince dokuları korurken makul renklendirme ile gerçekçi bir bileşik görüntü üretmek için farklı nesneleri ve renk dağılımlarını arka plan görüntüsüyle nasıl uyumlu hale getireceğini başarıyla öğrenebilmektedir. Ayrıca, oluşturulan görüntüler yakınlaştırıp/uzaklaştırıldığında, diğer modellerin görüntü kalitesinin genellikle zayıf olduğunu (ör. yapısal bozulmalar ve bulanık dokular), bileşik görüntüdeki düzenlenen bölgenin ayırt edilebildiğini ve eklenen nesnelerin kenarlarında bulanık görüş olduğu görülebilmektedir. Harmanlama uygulaması için daha fazla sayıda örnek Ek-F Şekil 6.4'te bulunmaktadır.

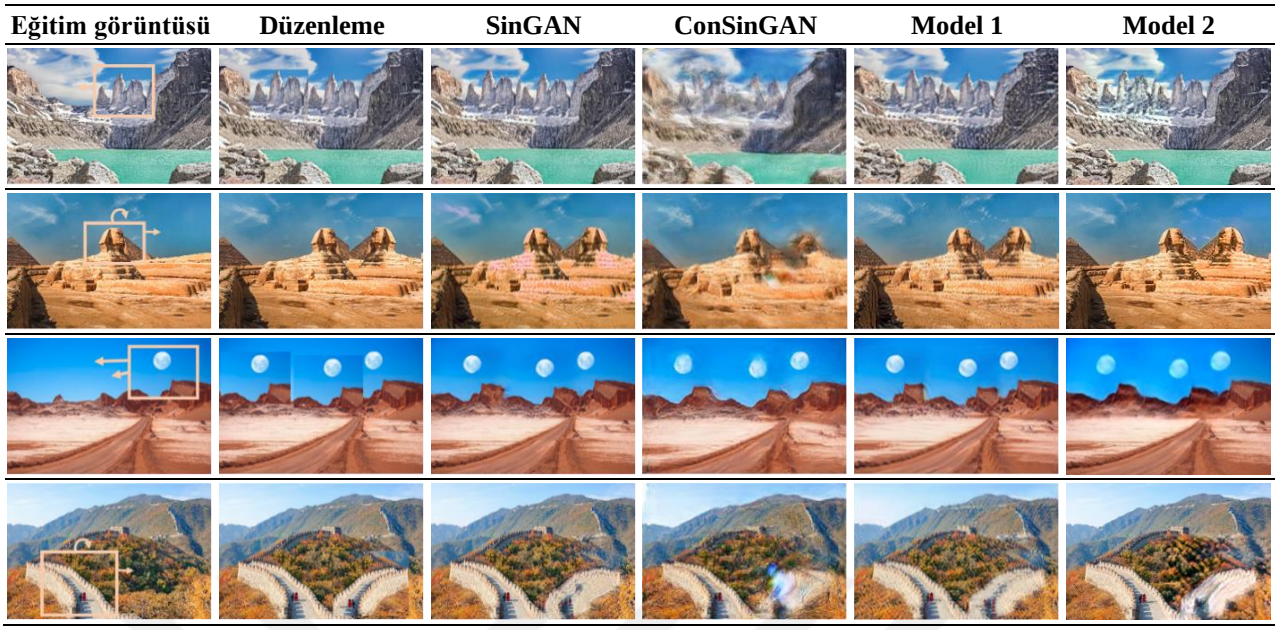


Şekil 4.14: Harmanlama uygulama sonuçları.

4.6.3. Görüntü Düzenleme

Görüntü düzenleme uygulaması kullanıcı yönlendirmesine bağlı olarak görsel bilgileri düzenlemeyi ve kompozit görüntü oluşturmayı amaçlamaktadır. Belirlenen eğitim görüntüsü kullanılarak model eğitilmektedir. Ardından eğitim görüntüsü üzerinde *yamaların yerlerini değiştirmek, yamaları döndürmek, yamaları ölçeklendirmek, yamaları kopyala/yapıştır yapmak* gibi bazı değişiklikler yapılmaktadır. Çıkarım zamanında, düzenlenen görüntünün ölçeklendirilmiş sürümü $k \leq n_s$ olmak üzere $Ü_k$ ağına gönderilmektedir. Ü hiyerarşik olarak $k = n_s$ ölçeğine gelene kadar görüntü üretmektedir. $k = n_s$ ölçekteki çıktı görüntüsü düzenlenen görüntünün son halini göstermektedir. Bu görevde görsel bütünlük gerekli değildir. Düzenlenen görüntüler, eğitim için artırılmış veri olarak kullanılabilir.

Şekil 4.15, önerilen tek görüntü ÜÇA modellerinin görüntü düzenleme görevi için diğer modellerle karşılaştırmalı olarak yeteneğini göstermektedir. Önerilen tek görüntü ÜÇA modellerinin düzenlemeleri başarılı bir şekilde tutarlı ve gerçekçi yapılara dönüştürebilmektedir. Ayrıca nesnelerin genel düzenlemesini ve şeklini iyi kontrol etmekte ve düzenlenen bölgelerin kenarlarında daha canlı ve bulanık olmayan efektler üretmektedir. Ayrıca diğer modeller tarafından üretilen görüntülerde, bazen arka planı birleştirmeyi, bileşik görüntüdeki nesneleri/yamaları yeniden düzenlemeyi veya düzenlenen görüntünün kalan kısımlarını tutarlı bir şekilde yeniden oluşturmayı başaramadığı görülmektedir. Görüntü düzenleme uygulaması için daha fazla sayıda örnek Ek-G Şekil 6.5'te bulunmaktadır.



Şekil 4.15: Görüntü düzenleme uygulama sonuçları.

4.6.4. Animasyon

Tek görüntü ÜÇA modelleri eğitim sırasında yeniden yapılandırma kayıp fonksiyonu kullanmaktadır [30], [32], [33]. Böylelikle modeller eğitim görüntüsünü yeniden oluşturabilmektedirler. Yeniden yapılandırma kayıp fonksiyonu için eğitim öncesi oluşturulan sabit $z = (z_0^{yy}, z_2^{yy}, \dots, z_k^{yy})$ vektörleri kullanılmaktadır. Çıkarım aşamasına gelindiğinde eğitim görüntüsünün gürültü uzayındaki vektörlerine ulaşılmış olur. Öğrenilen bu bilgi doğrultusunda z_k^{yy} gürültü vektörlerinin küçük oranda bozulmasıyla elde edilecek görüntüler birbirlerine oldukça yakın piksel değerlerinden oluşmaktadır. Sonuç olarak bu görüntülerin birleştirilmesiyle animasyon görüntüsü elde edilmektedir.

$$z_k(t+1) = \rho^1 \times z_k^{yy} + (1-\rho^1) \times \Big(z_k(t) + \rho^2 \times \big(z_k(t) - z_k(t-1)\big) + (1-\rho^2) \times u_k(t)\Big) \quad (4.5)$$

t zamanındaki k'ıncı ölçekli gürültü vektörü $z_k(t)$ ile bozulma süreci (4.5)'te verilmiştir. Burada $z_k(0) = z_0^{yy}$'dir ve $u_k(t)$ rastgele gürültü haritasıdır. Burada ρ^1, bozulmaların eğitim görüntüsüne yakınlığını ρ^2, animasyondaki değişim hızını kontrol etmektedir. Çıkarım aşamasında farklı ölçeklerde süreci başlatarak farklı gerçekçilikte animasyonlar elde edilmektedir [28]. Şekil 4.16, elde oluşturulan modeller için elde edilen animasyon sonuçlarını göstermektedir. Animasyon uygulaması için daha fazla sayıda örnek Ek-H Şekil 6.6 ve Şekil 6.7'de bulunmaktadır.

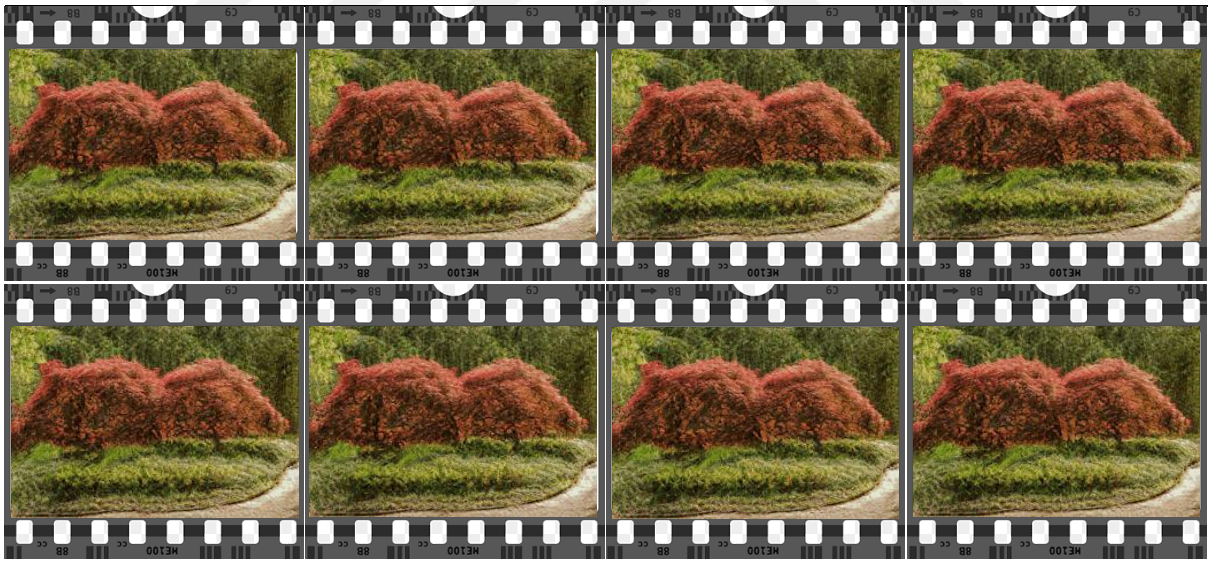
Eğitim Görüntüsü



Model 1



Model 2

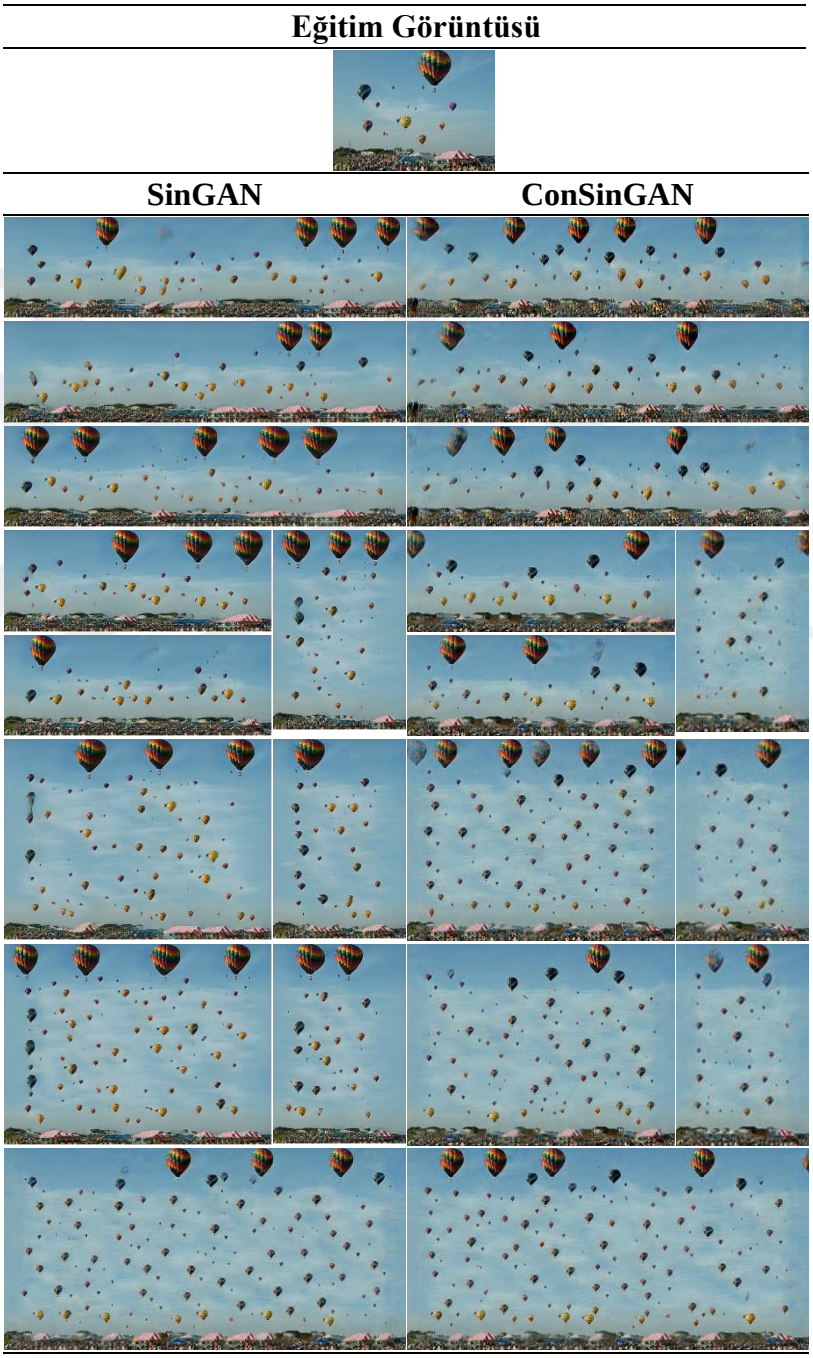


Şekil 4.16: Animasyon uygulama sonuçları.

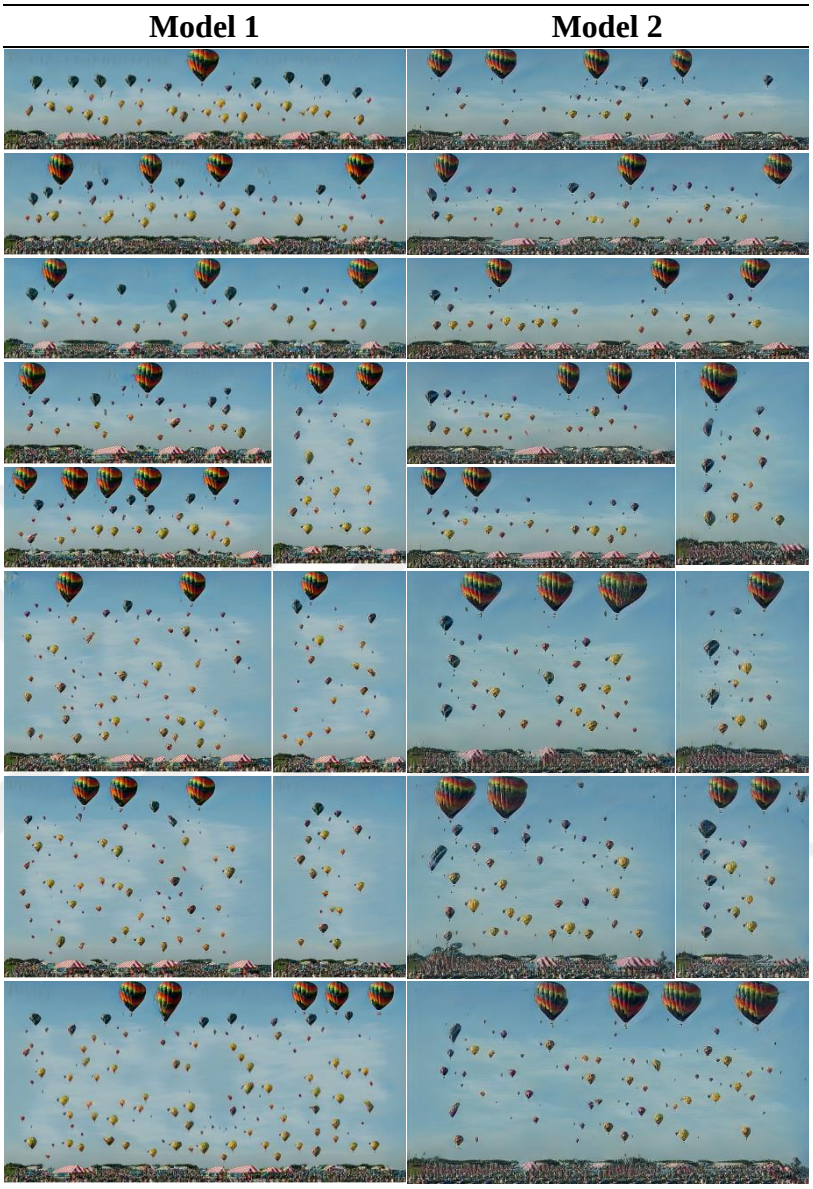
4.6.5. Değişken Boyutlu Görüntüler

Konvolüsyon işlemleri için kullanılan filtreler, görüntülerin genel boyutunu hesaba katmaz. Bu nedenle, konvolüsyon tabanlı üretken ağlarda, girdi boyutu değiştirilerek farklı boyutlarda çıktılar üretilebilir. Önerilen modeller konvolüsyon tabanlı oldukları için farklı boyutlarda görüntüler oluşturmak için de kullanılabilir. Bu kapsamda Şekil 4.17'de verildiği üzere farklı

boyutlarda görseller oluşturulmuştur. Oluşturulan görüntüler SinGAN ve ConSinGAN modelleriyle karşılaştırmalı olarak sunulmuştur. Görsellerin eğitim görüntüsüne göre boyut değişim oranları yukarıdan aşağıya, soldan sağa şu şekildedir: (1,3), (1,2), (2,1), (2,2), (2,3). Değişken boyutlu görüntüler uygulaması için daha fazla sayıda örnek Ek-I Şekil 6.8 ve Şekil 6.9'da bulunmaktadır.



Şekil 4.17: Farklı boyutlarda görüntü üretim sonuçları.



Şekil 4.17 (devam): Farklı boyutlarda görüntü üretim sonuçları.

5. TARTIŞMA

Dahili öğrenme, harici öğrenmeye kıyasla semantik çeşitlilik ve düşük veri sayısı açısından doğası gereği sınırlıdır. Tek bir doğal görüntünün dahili yama istatistikleri, genellikle güçlü bir üretken modelin öğrenmesi için yeterli bilgiyi taşımaktadır. Bu doğrultuda bu tez çalışmasında dahili öğrenme için tek eğitim görüntüsü kullanan ÜÇA modelleri üzerinde araştırmalar yapılmıştır. Bunun sonucunda tek görüntü üretim problemine iki farklı açıdan yaklaşarak çözüm üreten iki tek görüntü ÜÇA modeli geliştirilmiştir.

Model 1 için, tek görüntü ÜÇA modelleri incelenmiş ve bu modelleri eğitmenin zorluklarını anlamak için çeşitli deneyler yapılmıştır. Bu deneyler sonucu elde edilen bilgiler dahilinde öz-dikkat, yoğun bağlantılı blok ve RaLSGAN-GP'yi kullanan yeni bir koşulsuz ÜÇA modeli oluşturulmuştur. Model 1, tek bir doğal görüntü üzerinde eğitilmekte ve eğitim görüntüsüne anlamsal olarak benzeyen gerçekçi, yüksek kaliteli, çeşitli örnekler üretmektedir. Bunun için tek eğitim görüntüsü içindeki yamaların dahili dağılımı öğrenilmektedir. Model 1'de global bağlam/yapı bilgilerinin eklenme seviyesini başarılı bir şekilde kontrol etmek, daha fazla anlamsal bütünlük sağlamak ve görüntü oluşturma kalitesini iyileştirmek için Ü ve A'da öz-dikkat mekanizması kullanılmıştır. Ayrıca daha küçük eğitim seti boyutlarına sahip görevlerde aşırı uyumu önlemek ve optimizasyon zorluklarını aşmak için yoğun bağlantılı blok mimarisinden yararlanılmıştır. Yoğun bağlantılı bloğun öğrenilen özellik haritalarının yeniden kullanılmasına olanak sağlama, kaybolan gradyan problemini hafifletme ve parametreleri daha verimli kullanma gibi avantajları vardır. Deneysel sonuçlar, RaLSGAN-GP'nin herhangi bir ek hesaplama maliyetine maruz kalmadan veri oluşturma kalitesini ve modelimizin kararlılığını artırdığını göstermiştir. Ayrıca, RaLSGAN-GP kayıp fonksiyonu ile çalışan A'nın, Ü'ye daha iyi geri bildirim sağlayarak daha yüksek kaliteli görüntüler üretmede çok önemli bir rol oynadığını gözlemlenmiştir. Model 1'in performansını üç veri kümesi üzerinde analiz etmek için nicel ve nitel ölçüm yöntemleri kullanılarak, Model 1 var olan tek görüntü ÜÇA modelleriyle karşılaştırıldı. Model 1'in sağlamlığını göstermek için koşulsuz görüntü oluşturma, görüntü düzenleme, harmanlama ve çizimden görüntü oluşturma gibi farklı uygulamalar gerçekleştirildi. Deneyler, Model 1'in görüntü oluşturma performansının rakip benzer veya daha iyi seviyede olduğunu göstermektedir.

Model 2'de, tek görüntü üretim problemi için ÇÖÇ mimarisine yoğunlaşılmıştır. Bu çerçevede eğitim görüntüsünün gizli gösterim bilgisini etkili şekilde kullanan özgün ÜÇA modeli önerilmiştir. Araştırılan literatüre bilgisi dahilinde Model 2, tek görüntüden görüntü üretme problemini ÇÖÇ çerçevesini kullanarak ele alan ilk yaklaşımdır. Model 2'nin başarısına katkıda bulunan en önemli faktörlerden biri, çıkarım aşamasında eğitim görüntüsünün gizli gösteriminin etkili bir şekilde kullanılmasıdır. Gizli gösterimin kullanılması, Model 2'nin eğitim görüntüsünün temel yapısını ve özelliklerini daha iyi anlamasını sağlayarak hedef dağılıma çok benzeyen görüntüler üretmesine olanak tanımaktadır. Modelde kodlayıcı ağ eğitim görüntüsünün gizli gösterimini öğrenmektedir. Ayrıca kodlayıcı ağa girdi verilen eğitim görüntüsünün küçük değerli gürültüler eklenerek bozulması sağlanmıştır. Dolayısıyla kodlayıcı ağın tek eğitim görüntüsünün gizli gösterimini ezberlenmesi engellenmiştir. Buna ek olarak kodlayıcı ağ gerçek görüntüye çok yakın gizli gösterimleri de öğrenerek üretilen görüntülerin çeşitliliğine katkı sağlamıştır. Yapılan nicel ve nitel ölçüm sonuçları Model 2'nin tek eğitim görüntüsünden öğrenerek yeni gerçekçi görüntüler üretmekte en etkili modellerden birisi olduğunu göstermiştir. Ayrıca Model 2 eğitilebilir parametre, eğitim zamanı, kaynak kullanımı gibi kıstaslarda da diğer modellerden daha iyidir. Bu bulgular, görüntü sentezi alanında artan bilgi birikimine katkıda bulunmakta ve Model 2'nin çeşitli görüntü oluşturma uygulamaları için sağlam ve çok yönlü bir model olarak potansiyelini göstermektedir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, tek görüntü üretim problemine odaklanan iki yeni üretken çekişmeli ağ modeli oluşturulmuştur; Model 1 ve Model 2.

Model 1, mimari yapısında öz-dikkat mekanizması ve yoğun bağlantılı bloklardan yararlanmaktadır. Öz-dikkat mekanizması, görüntü içindeki yamaların önemini farklı seviyelerde değerlendirerek karmaşık yapısal ve dokusal desenlerin öğrenilmesine yardımcı olmaktadır. Yoğun bağlantılı bloklar, ağın katmanları arasında bilgi akışını artırarak etkili bir özellik öğrenme süreci sunmaktadır. RaLSGAN-GP kayıp fonksiyonu, Model 1'in eğitim kararlılığını ve görüntü oluşturma kalitesini artırmada etkili olmuştur. Model 1, farklı uygulamalarda üstün performans sergilemiş ve rakip modellerle karşılaştırıldığında rekabetçi sonuçlar elde etmiştir.

Model 2, çekişmeli öğrenen çıkarım mimarisini kullanarak tek görüntü üretim problemine özgün bir yaklaşım sunmaktadır. Model 2, tek eğitim görüntüsünün gizli gösterimini etkili bir şekilde kullanarak tek eğitim görüntüsüne benzeyen farklı görüntüler üretebilmektedir. Model 2'de kodlayıcı ağın kullanımı, modelin tek eğitim görüntüsünün temel yapısını anlamasını sağlayarak üretimde çeşitlilik ve kalite sağlamaktadır. Model 2'nin diğer modellere kıyasla daha iyi performans gösterdiği ve eğitim süreci, parametre sayısı ve kaynak kullanımı gibi kriterlerde avantajlı olduğu gösterilmiştir.

Bu tez çalışması sonucunda elde edilen bulgulara dayanarak, gelecek çalışmalar için öneriler şu şekildedir;

- Kısıtlı veri bulunana araştırma alanlarında veri çoğaltma yöntemleri ve tek görüntü ÜÇA modelleri kullanılarak artırılan verilerin kalitesi araştırılabilir. Böylelikle tek görüntü ÜÇA modellerinin veri çoğaltma yöntemlerine göre zayıflık ve üstünlükleri belirlenir. Tek görüntü ÜÇA modellerinin bu alanlarda kullanımı değerlendirilir.

- Var olan tek görüntü ÜÇA modellerinin kayıp fonksiyonu yapısı yüksek benzerlik göstermektedir. Tek görüntü üretim problemi için daha yüksek başarı sağlayabilecek kayıp fonksiyonları araştırılabilir.

- Tek görüntü üretim problemi için önerilen ÜÇA modellerinin genel eğitim prosedürü aynıdır. Daha etkili bir eğitim prosedürü için araştırmalar yapılabilir.
- Tek görüntü ÜÇA modelleri, tek veri kullanılarak derinlik tahmini, uzun süreli video üretimi, metin veya ses üretimi vb. problemlere uyarlanabilirlikleri araştırılabilir.

KAYNAKLAR

[1] Ruthotto, L. ve Haber, E. (2021). An introduction to deep generative modeling. GAMM-Mitteilungen, 44(2), e202100008.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville A. ve Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[3] Iglesias, G., Talavera, E. ve Díaz-Álvarez, A. (2023). A survey on GANs for computer vision: Recent research, analysis and taxonomy. Computer Science Review, 48, 100553.

[4] Wang, Z., She, Q. ve Ward, T. E. (2021). Generative adversarial networks in computer vision: A survey and taxonomy. ACM Computing Surveys (CSUR), 54(2), 1-38.

[5] A. Radford, L. Metz ve S. Chintala. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. *4th International Conference on Learning Representations, ICLR 2016- Conference Track Proceedings*, 1–16.

[6] Karras, T., Laine, S. ve Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401-4410.

[7] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I. ve Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. Advances in neural information processing systems, 29.

[8] Mirza, M. ve Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

[9] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. ve Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems, 30.

[10] Mutlu, U. ve Alpaydın, E. (2020). Training bidirectional generative adversarial networks with hints. Pattern Recognition, 103, 107320.

[11] Zhang, H., Goodfellow, I., Metaxas, D. ve Odena, A. (2019). Self-attention generative adversarial networks. *International conference on machine learning PMLR*, 7354-7363.

[12] Che, T., Li, Y., Jacob, A. P., Bengio, Y. ve Li, W. (2019). Mode regularized generative adversarial networks. *5th International Conference on Learning Representations ICLR 2017*.

[13] De Rosa, G. H. ve Papa, J. P. (2021). A survey on text generation using generative adversarial networks. Pattern Recognition, 119, 108098.

[14] Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J. ve Aila, T. (2021). Alias-free generative adversarial networks. Advances in neural information processing systems, 34, 852-863.

[15] Bau, D., Zhu, J. Y., Strobelt, H., Zhou, B., Tenenbaum, J. B., Freeman, W. T. ve Torralba, A. (2018). GAN Dissection: Visualizing and Understanding Generative Adversarial Networks. *International Conference on Learning Representations*.

[16] Gauthier, J. (2014). Conditional generative adversarial nets for convolutional face generation. Sınıf Ödevi Stanford CS231N: convolutional neural networks for visual recognition, 5(2).

[17] Isola, P., Zhu, J. Y., Zhou, T. ve Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125-1134.

[18] Kim, T., Cha, M., Kim, H., Lee, J. K. ve Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. *International conference on machine learning PMLR*, 1857-1865.

[19] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. ve Huang, T. S. (2018). Generative image inpainting with contextual attention. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5505-5514.

[20] Yan, Z., Li, X., Li, M., Zuo, W. ve Shan, S. (2018). Shift-net: Image inpainting via deep feature rearrangement. *Proceedings of the European conference on computer vision (ECCV)*,1-17.

[21] Hedjazi, M. A. ve Genc, Y. (2021). Efficient texture-aware multi-GAN for image inpainting. Knowledge-Based Systems, 217, 106789.

[22] Dong, H. W., Hsiao, W. Y., Yang, L. C. ve Yang, Y. H. (2018). Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34-41.

[23] Antipov, G., Baccouche, M. ve Dugelay, J. L. (2017). Face aging with conditional generative adversarial networks. *2017 IEEE international conference on image processing (ICIP)*, 2089-2093.

[24] Huang, Z., Chen, S., Zhang, J. ve Shan, H. (2020). PFA-GAN: Progressive face aging with generative adversarial network. IEEE Transactions on Information Forensics and Security, 16, 2031-2045.

[25] Yang, M. ve Wang, Z. (2023). Image Synthesis under Limited Data: A Survey and Taxonomy. arXiv preprint arXiv:2307.16879.

[26] Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J. ve Aila, T. (2020). Training generative adversarial networks with limited data. Advances in neural information processing systems, 33, 12104-12114.

[27] Zhao, S., Liu, Z., Lin, J., Zhu, J. Y. ve Han, S. (2020). Differentiable augmentation for data-efficient gan training. Advances in neural information processing systems, 33, 7559-7570.

[28] Shaham, T. R., Dekel, T. ve Michaeli, T. (2019). Singan: Learning a generative model from a single natural image. *Proceedings of the IEEE/CVF international conference on computer vision*, 4570-4580.

[29] Asano, Y. M., Rupprecht, C. ve Vedaldi, A. (2019). A critical analysis of self-supervision, or what we can learn from a single image. *International Conference on Learning Representations*.

[30] Chen, J., Xu, Q., Kang, Q. ve Zhou, M. (2023). Mogan: Morphologic-structure-aware generative learning from a single image. IEEE Transactions on Systems, Man, and Cybernetics: Systems.

[31] Wang, P., Li, Y., Singh, K. K., Lu, J. ve Vasconcelos, N. (2021). Imagine: Image synthesis by image-guided model inversion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3681-3690.

[32] Hinz, T., Fisher, M., Wang, O. ve Wermter, S. (2021). Improved techniques for training single-image gans. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1300-1309.

[33] Mahendren, S., Edussooriya, C. U. ve Rodrigo, R. (2023). Diverse single image generation with controllable global structure. Neurocomputing, 528, 97-112.

[34] Rodriguez-Pardo, C. ve Garces, E. (2022). Seamlessgan: Self-supervised synthesis of tileable texture maps. IEEE Transactions on Visualization and Computer Graphics.

[35] Zhang, Z., Han, C. ve Guo, T. (2021). Exsingan: Learning an explainable generative model from a single image. *Proceedings of the 32nd British Machine Vision Conference*.

[36] Chen, X., Zhao, H., Yang, D., Li, Y., Kang, Q. ve Lu, H. (2021). SA-SinGAN: self-attention for single-image generation adversarial networks. Machine Vision and Applications, 32.

[37] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. ve Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[38] Huang, G., Liu, Z., Van Der Maaten, L. ve Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700-4708.

[39] Jolicoeur-Martineau, A. (2018). The relativistic discriminator: a key element missing from standard GAN. *International Conference on Learning Representations*.

[40] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. ve Courville, A. C. (2017). Improved training of wasserstein gans. *Advances in neural information processing systems*, 30.

[41] Dumoulin, V., Belghazi, I., Poole, B., Lamb, A., Arjovsky, M., Mastropietro, O. ve Courville, A. (2016). Adversarially Learned Inference. *International Conference on Learning Representations*.

[42] Donahue, J., Krähenbühl, P. ve Darrell, T. (2016). Adversarial Feature Learning. *International Conference on Learning Representations*.

[43] Miller, A. I. (2019). *Ian Goodfellow's Generative Adversarial Networks: AI Learns to Imagine*. The Artist in the Machine: The World of AI-Powered Creativity, Bölüm: 10. MIT Press, 87–98.

[44] Harshvardhan, G. M., Gourisaria, M. K., Pandey, M. ve Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. Computer Science Review, 38, 100285.

[45] Bond-Taylor, S., Leach, A., Long, Y. ve Willcocks, C. G. (2021). Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. IEEE transactions on pattern analysis and machine intelligence, 44(11), 7327-7347.

[46] Zhu, J. Y., Park, T., Isola, P. ve Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*, 2223-2232.

[47] Isola, P., Zhu, J. Y., Zhou, T. ve Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125-1134.

[48] Odena, A. (2016). Semi-supervised learning with generative adversarial networks. arXiv preprint arXiv:1606.01583.

[49] M. Arjovsky, S. Chintala ve L. Bottou, "Wasserstein generative adversarial networks," *34th International Conference on Machine Learning, ICML 2017*, 298–321.

[50] Karras, T., Aila, T., Laine, S. ve Lehtinen, J. (2018). Progressive Growing of GANs for Improved Quality, Stability, and Variation. *International Conference on Learning Representations*.

[51] Denton, E. L., Chintala, S. ve Fergus, R. (2015). Deep generative image models using a laplacian pyramid of adversarial networks. *Advances in neural information processing systems*, 28.

[52] Zhu, J., Zhao, D., Zhou, B. ve Zhang, B. (2019). Lia: Latently invertible autoencoder with adversarial learning. arXiv preprint arXiv:1906.08090.

[53] Lang, O., Gandelsman, Y., Yarom, M., Wald, Y., Elidan, G., Hassidim, A., ... ve Mosseri, I. (2021). Explaining in style: Training a gan to explain a classifier in stylespace. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 693-702.

[54] Pidhorskyi, S., Adjeroh, D. A. ve Doretto, G. (2020). Adversarial latent autoencoders. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14104-14113.

[55] Feng, R., Lin, Z., Zhu, J., Zhao, D., Zhou, J. ve Zha, Z. J. (2021). Uncertainty principles of encoding GANs. *International Conference on Machine Learning PMLR*, 3240-3251.

[56] Ulyanov, D., Vedaldi, A. ve Lempitsky, V. (2018). It takes (only) two: Adversarial generator-encoder networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

[57] Shocher, A., Gandelsman, Y., Mosseri, I., Yarom, M., Irani, M., Freeman, W. T. ve Dekel, T. (2020). Semantic pyramid for image generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7457-7466.

[58] Deja, K., Dubiński, J., Nowak, P., Wenzel, S., Spurek, P. ve Trzcinski, T. (2020). End-to-end sinkhorn autoencoder with noise generator. IEEE Access, 9, 7211-7219.

[59] H. Huang, Z. Li, Z. Sun ve T. Tan, "IntroVAE: Introspective Variational Autoencoders for Photographic Image Synthesis," in *n Proc. 32nd Int. Conf. Neural Inf. Process. Syst*, 2018, pp. 52–63.

[60] Huang, H., He, R., Sun, Z. ve Tan, T. (2018). Introvae: Introspective variational autoencoders for photographic image synthesis. *Advances in neural information processing systems*, 31.

[61] Ulyanov, D., Vedaldi, A. ve Lempitsky, V. (2018). Deep image prior. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 9446-9454.

[62] Shocher, A., Bagon, S., Isola, P. ve Irani, M. (2019). Ingan: Capturing and retargeting the" dna" of a natural image. *Proceedings of the IEEE/CVF international conference on computer vision*, 4492-4501.

[63] Shama, F., Mechrez, R., Shoshan, A. ve Zelnik-Manor, L. (2019). Adversarial feedback loop. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3205-3214.

[64] Sushko, V., Gall, J. ve Khoreva, A. (2021). One-shot gan: Learning to generate samples from single images and videos. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2596-2600.

[65] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... ve Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *Proceedings of the IEEE conference on computer vision and pattern recognition*,1874-1883.

[66] Brock, A., Donahue, J. ve Simonyan, K. (2018). Large Scale GAN Training for High Fidelity Natural Image Synthesis. *International Conference on Learning Representations*.

[67] Zhang, Z., Liu, Y., Han, C., Shi, H., Guo, T. ve Zhou, B. (2022). Petsgan: Rethinking priors for single image generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(3), 3408-3416.

[68] Gottlieb, S. ve Shu, C. W. (1998). Total variation diminishing Runge-Kutta schemes. Mathematics of computation, 67(221), 73-85.

[69] Yin, H., Molchanov, P., Alvarez, J. M., Li, Z., Mallya, A., Hoiem, D., ... ve Kautz, J. (2020). Dreaming to distill: Data-free knowledge transfer via deepinversion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8715-8724.

[70] Gur, S., Benaim, S. ve Wolf, L. (2020). Hierarchical patch vae-gan: Generating diverse videos from a single sample. *Advances in Neural Information Processing Systems 33*, 16761-16772.

[71] Granot, N., Feinstein, B., Shocher, A., Bagon, S. ve Irani, M. (2022). Drop the gan: In defense of patches nearest neighbors as single image generative models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13460-13469.

[72] Gandelsman, Y., Shocher, A. ve Irani, M. (2019). " Double-DIP": unsupervised image decomposition via coupled deep-image-priors. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11026-11035.

[73] Vinker, Y., Horwitz, E., Zabari, N. ve Hoshen, Y. (2021). Image shape manipulation from a single augmented training sample. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13769-13778.

[74] Y. Vinker, N. Zabari ve Y. Hoshen. (2021) Training End-to-end Single Image Generators without GANs. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13769–13778.

[75] Lin, J., Pang, Y., Xia, Y., Chen, Z. ve Luo, J. (2020). Tuigan: Learning versatile image-to-image translation with two unpaired images. *Computer Vision–ECCV 2020: 16th European Conference*, 18-35.

[76] Haim, N., Feinstein, B., Granot, N., Shocher, A., Bagon, S., Dekel, T. ve Irani, M. (2022). Diverse generation from a single video made possible. *European Conference on Computer Vision*, 491-509.

[77] Bensadoun, R., Gur, S., Galanti, T. ve Wolf, L. (2021). Meta internal learning. *Advances in Neural Information Processing Systems 34*, 20645-20656.

[78] Yang, C., Shen, Y., Zhang, Z., Xu, Y., Zhu, J., Wu, Z. ve Zhou, B. (2023). One-shot generative domain adaptation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7733-7742.

[79] Zheng, Z., Xie, J. ve Li, P. (2021). Patchwise generative convnet: Training energy-based models from a single natural image for internal learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2961-2970.

[80] Xu, R., Wang, X., Chen, K., Zhou, B. ve Loy, C. C. (2021). Positional encoding as spatial inductive bias in gans. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13569-13578.

[81] Wu, S., Dong, C., ve Qiao, Y. (2022). Blind image restoration based on cycle-consistent network. IEEE Transactions on Multimedia, 25, 1111-1124.

[82] Li, P., Aberman, K., Zhang, Z., Hanocka, R. ve Sorkine-Hornung, O. (2022). Ganimator: Neural motion synthesis from a single sequence. ACM Transactions on Graphics (TOG), 41(4), 1-12.

[83] Bell-Kligler, S., Shocher, A. ve Irani, M. (2019). Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems, 32*.

[84] Benaim, S., Mokady, R., Bermano, A. ve Wolf, L. (2021). Structural analogy from a single image pair. *Computer Graphics Forum*, 40(1), 249-265.

[85] Awiszus, M., Schubert, F. ve Rosenhahn, B. (2020). TOAD-GAN: Coherent style level generation from a single example. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 16(1), 10-16.

[86] Barahona Rios, A. ve Collins, T. (2022). SpecSinGAN: Sound Effect Variation Synthesis Using Single-Image GANs. *Proceedings of the Sound and Music Computing Conference*.

[87] Zhong, G. ve Zhao, H. (2021). "Zero-Shot" Super-Resolution Based on SinGAN. *IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 883-888.

[88] Abu-Hussein, S., Tirer, T., Chun, S. Y., Eldar, Y. C. ve Giryes, R. (2022). Image restoration by deep projected GSURE. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3602-3611.

[89] Shocher, A., Gandelsman, Y., Mosseri, I., Yarom, M., Irani, M., Freeman, W. T. ve Dekel, T. (2020). Semantic pyramid for image generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7457-7466.

[90] Arora, R. ve Lee, Y. J. (2021). Singan-gif: Learning a generative video model from a single gif. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1310-1319.

[91] Wu, R. ve Zheng, C. (2022). Learning to Generate 3D Shapes from a Single Example. ACM Transactions on Graphics, 41(6).

[92] Son, M., Park, J. J., Guibas, L. ve Wetzstein, G. (2023). Singraf: Learning a 3d generative radiance field for a single scene. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8507-8517.

[93] Dumoulin, V. ve Visin, F. (2018). A guide to convolution arithmetic for deep learning. Stat, 1050, 11.

[94] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. ve Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.

[95] Khan, A., Sohail, A., Zahoora, U. ve Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial intelligence review, 53, 5455-5516.

[96] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... ve Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, 1-74.

[97] Ioffe, S. ve Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning PMLR*, 448-456.

[98] Ulyanov, D., Vedaldi, A. ve Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.

[99] Maas, A. L., Hannun, A. Y. ve Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *International conference on machine learning*, 30(1).

[100] Kuang, P., Ma, T., Chen, Z. ve Li, F. (2019). Image super-resolution with densely connected convolutional networks. Applied Intelligence, 49, 125-136.

[101] Wang, Z., Lu, Y., Li, W., Wang, S., Wang, X. ve Chen, X. (2021). Single image super-resolution with attention-based densely connected module. Neurocomputing, 453, 876-884.

[102] Jolicoeur-Martineau, A. (2020). On relativistic f-divergences. *International Conference on Machine Learning PMLR*, 4931-4939.

[103] Kingma, D. P. ve Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[104] Wang, Z., Simoncelli, E. P. ve Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers IEEE*, 1398-1402.

[105] Zhang, R., Isola, P., Efros, A. A., Shechtman, E. ve Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586-595.

[106] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A. ve Oliva, A. (2014). Learning deep features for scene recognition using places database. *Advances in neural information processing systems, 27*.

[107] Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T. ve Xiao, J. (2015). Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv preprint arXiv:1506.03365.

[108] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. ve Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *IEEE conference on computer vision and pattern recognition*, 248-255.

[109] Wang, Z., Bovik, A. C., Sheikh, H. R. ve Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

[110] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... ve Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.

[111] Van der Maaten, L. ve Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

EKLER

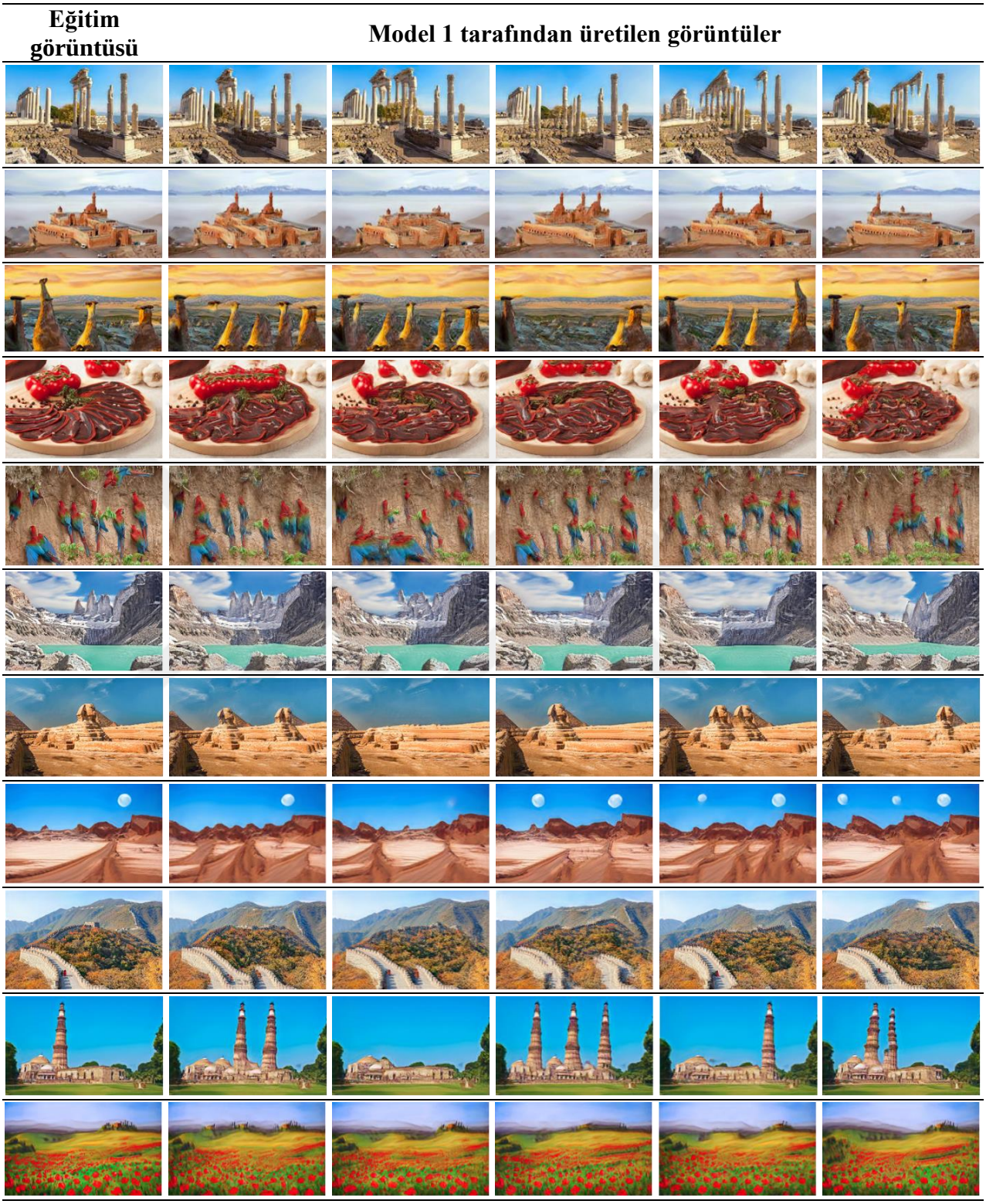
A. Model 1'in kullanıcı çalışması linkleri

1. Places veri kümesi:
    - ✓ *Eşleştirilmiş*: https://forms.gle/t3dHorLE4P3DgUjN7
    - ✓ *Eşleştirilmemiş*: https://forms.gle/7JgCxQbFugzY8GGM9
2. LSUN veri kümesi:
    - ✓ *Eşleştirilmiş*: https://forms.gle/ZNqFhNdLBxgLsscZ9
    - ✓ *Eşleştirilmemiş*: https://forms.gle/PbjcydvQUHkExFJ59
3. ImageNet veri kümesi:
    - ✓ *Eşleştirilmiş*: https://forms.gle/V1vhkAQujULoywdx9
    - ✓ *Eşleştirilmemiş*: https://forms.gle/AiZ7KTfcE8z5nsar8
4. Görüntü karşılaştırma:
    - ✓ *Places*: https://forms.gle/KGE4NPQg6HRvzWM47
    - ✓ *LSUN*: https://forms.gle/gwkHkM3QkqT1CMtTA
    - ✓ *ImageNet*: https://forms.gle/iiiyTcYbDWcPHnmC7

B. Model 2'nin kullanıcı çalışması linkleri

1. Places veri kümesi:
    - ✓ *Eşleştirilmiş*: https://forms.gle/eP75oPBxGVrbSJvu9
    - ✓ *Eşleştirilmemiş*: https://forms.gle/6gvw8xXssdptWJW77
2. LSUN veri kümesi:
    - ✓ *Eşleştirilmiş*: https://forms.gle/wFX2LSXGvc1Pb8vW9
    - ✓ *Eşleştirilmemiş*: https://forms.gle/A83YPPtBak6mnFDC9
3. ImageNet veri kümesi.
    - ✓ *Eşleştirilmiş*: https://forms.gle/CBUtE9G81a4fgBiy9
    - ✓ *Eşleştirilmemiş*: https://forms.gle/PBoqDwm6yiQRFxNK9
4. Görüntü karşılaştırma:
    - ✓ *Places*: https://forms.gle/kSo318fXZfbnNqYE6
    - ✓ *LSUN*: https://forms.gle/nsaCnfvsdj45c2q69
    - ✓ *ImageNet*: https://forms.gle/Ej9U5HinNgoPzgxJ8

C. Model 1 tarafından üretilen rastgele görüntüler



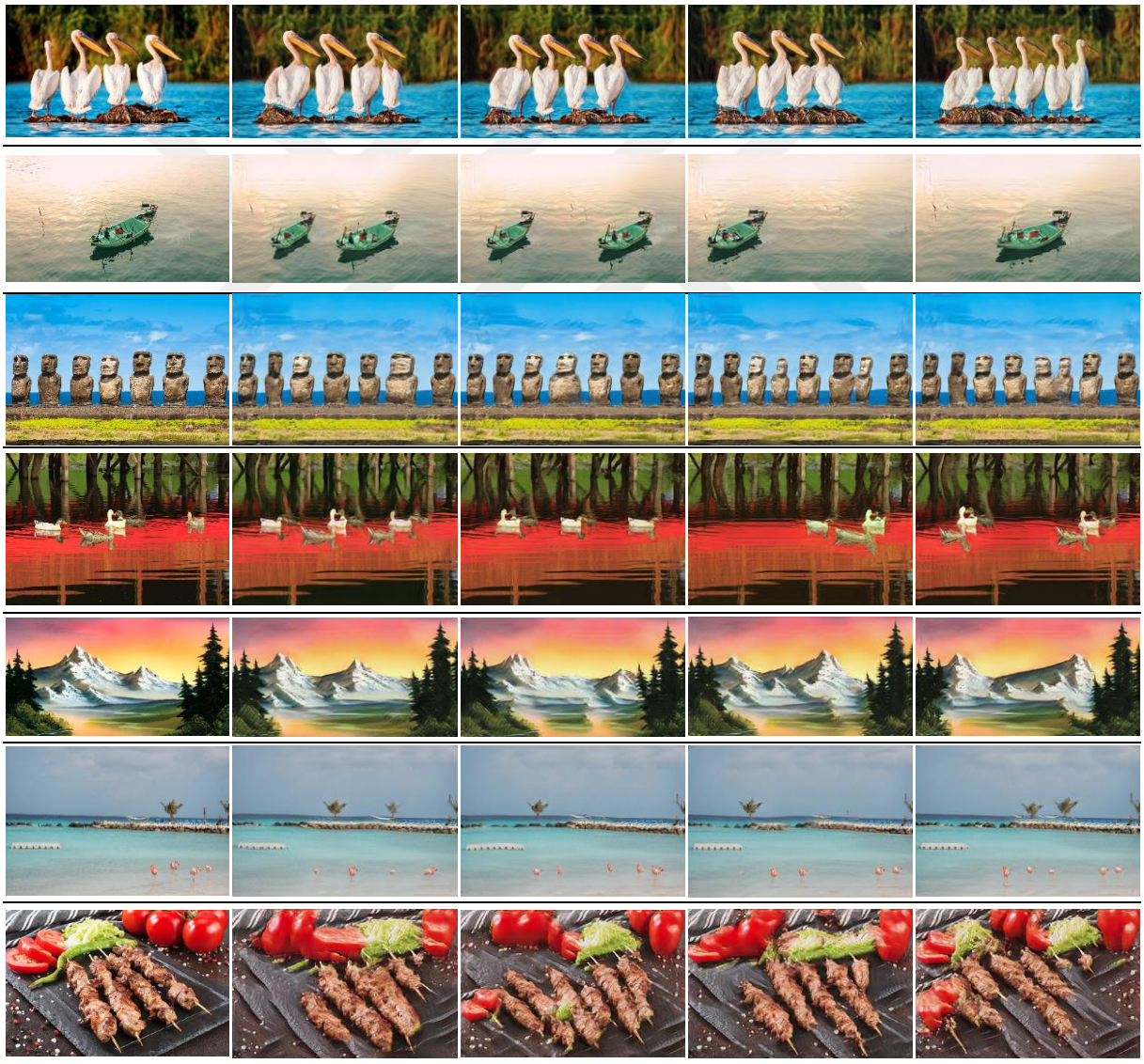
Şekil 6.1: Model 1 tarafından üretilen görüntüler.



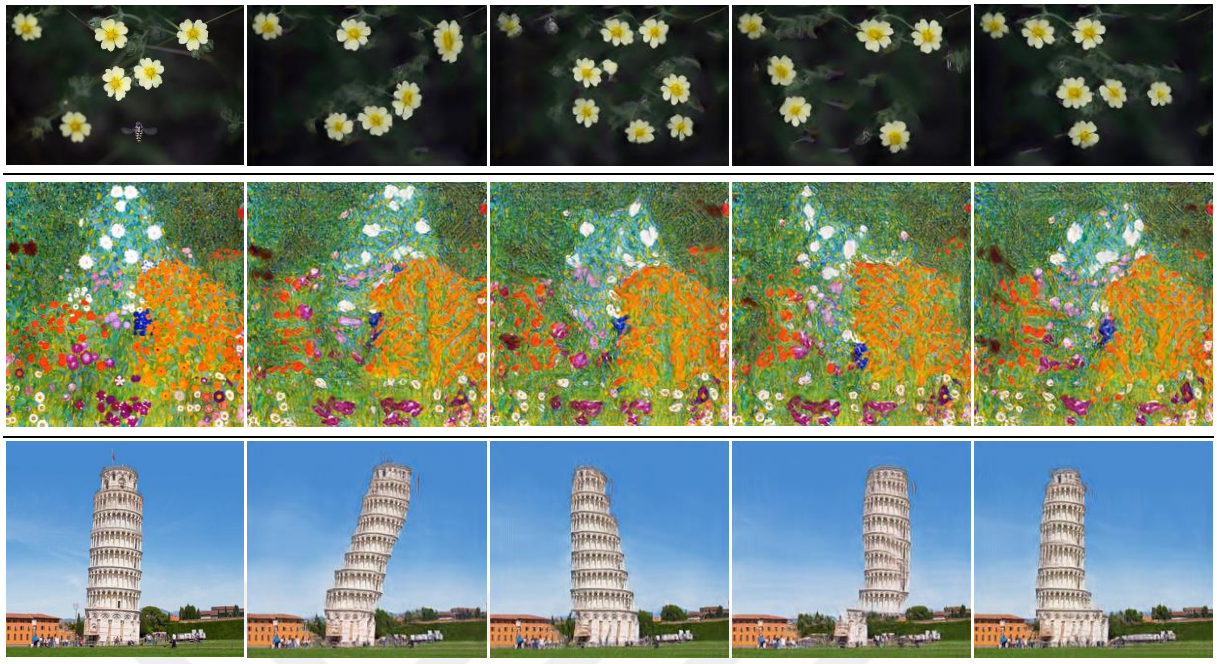
Şekil 6.1 (devam): Model 1 tarafından üretilen görüntüler.

D. Model 2 tarafından üretilen rastgele görüntüler

Eğitim görüntüsü	Model 2 tarafından üretilen görüntüler

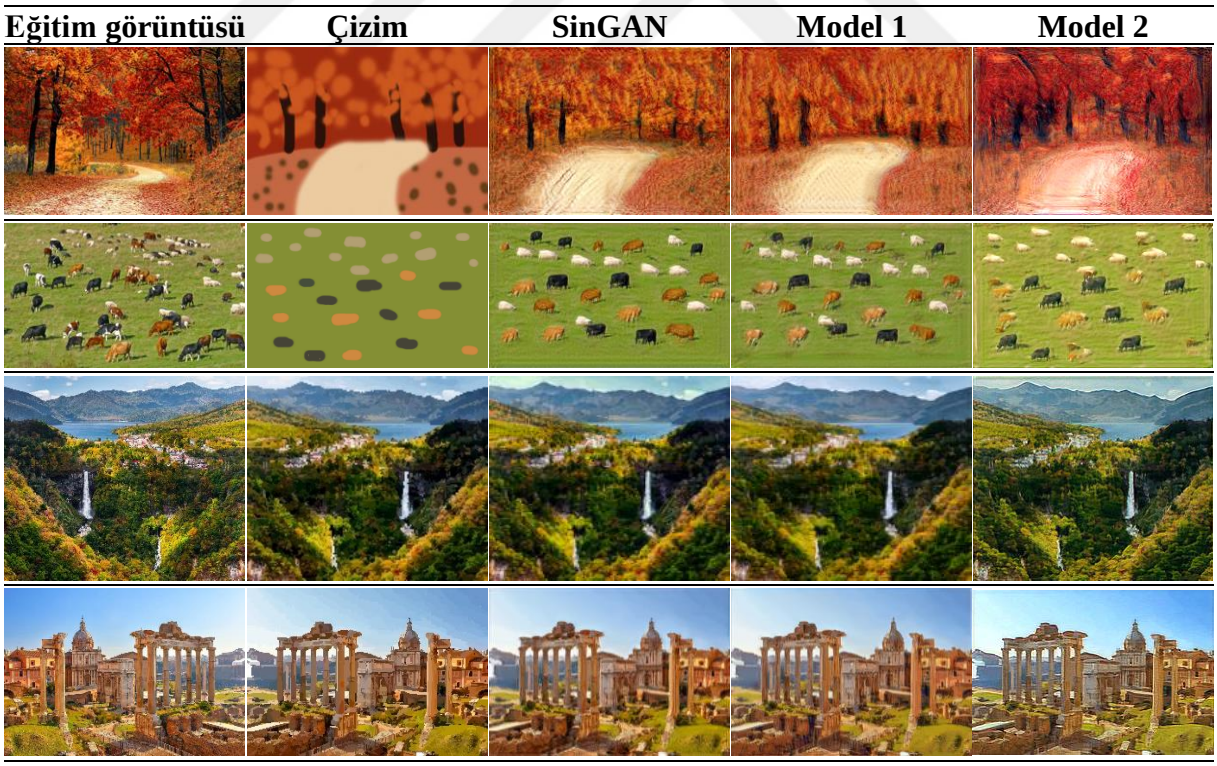


Şekil 6.2: Model 2 tarafından üretilen görüntüler.

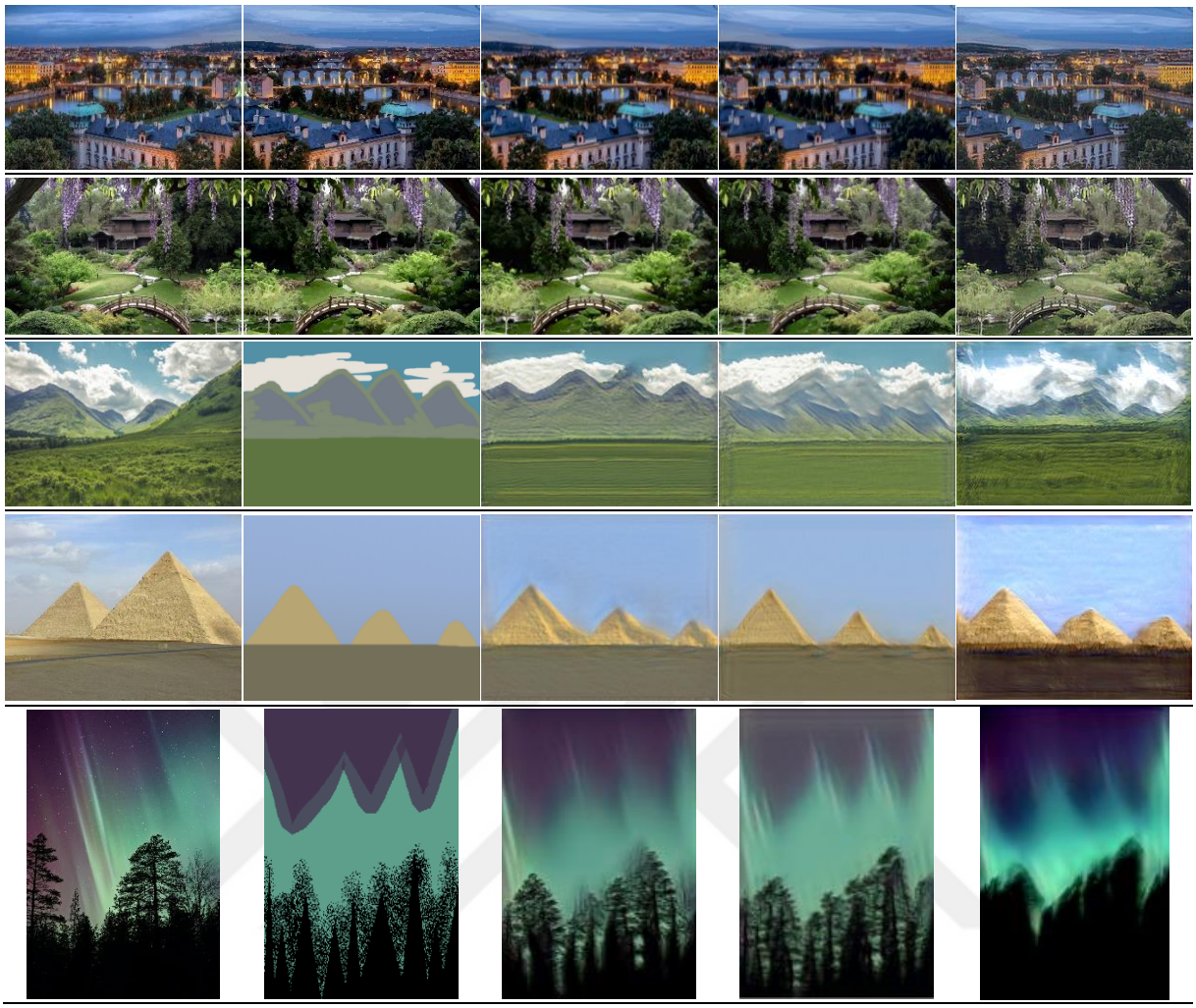


Şekil 6.2 (devam): Model 2 tarafından üretilen görüntüler.

E. Çizimden görüntüye uygulaması örnekleri

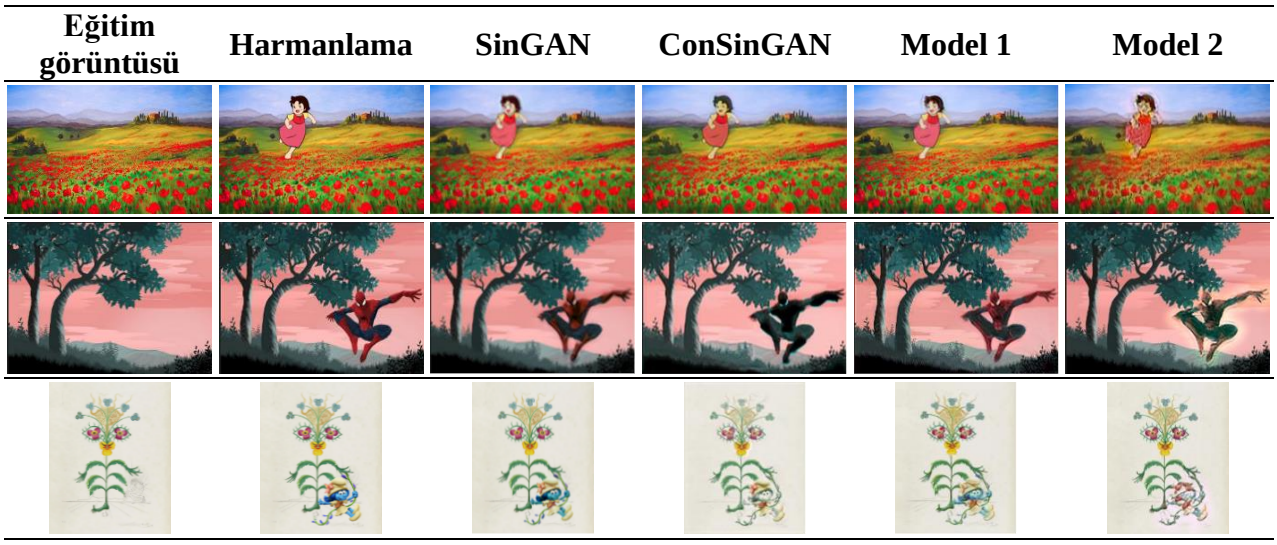


Şekil 6.3: Çizimden görüntüye uygulama sonuçları.



Şekil 6.3 (devam): Çizimden görüntüye uygulama sonuçları.

F. Harmanlama uygulaması örnekleri

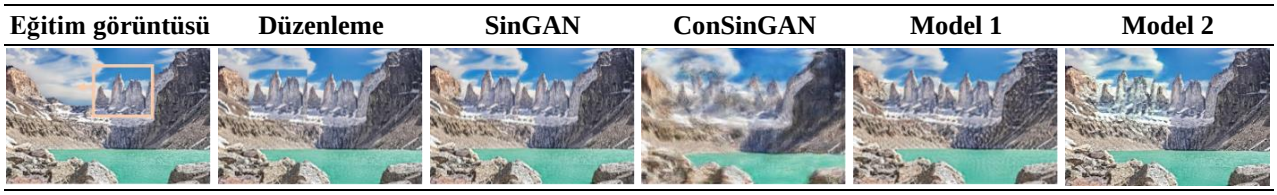


Şekil 6.4: Harmanlama uygulama sonuçları.

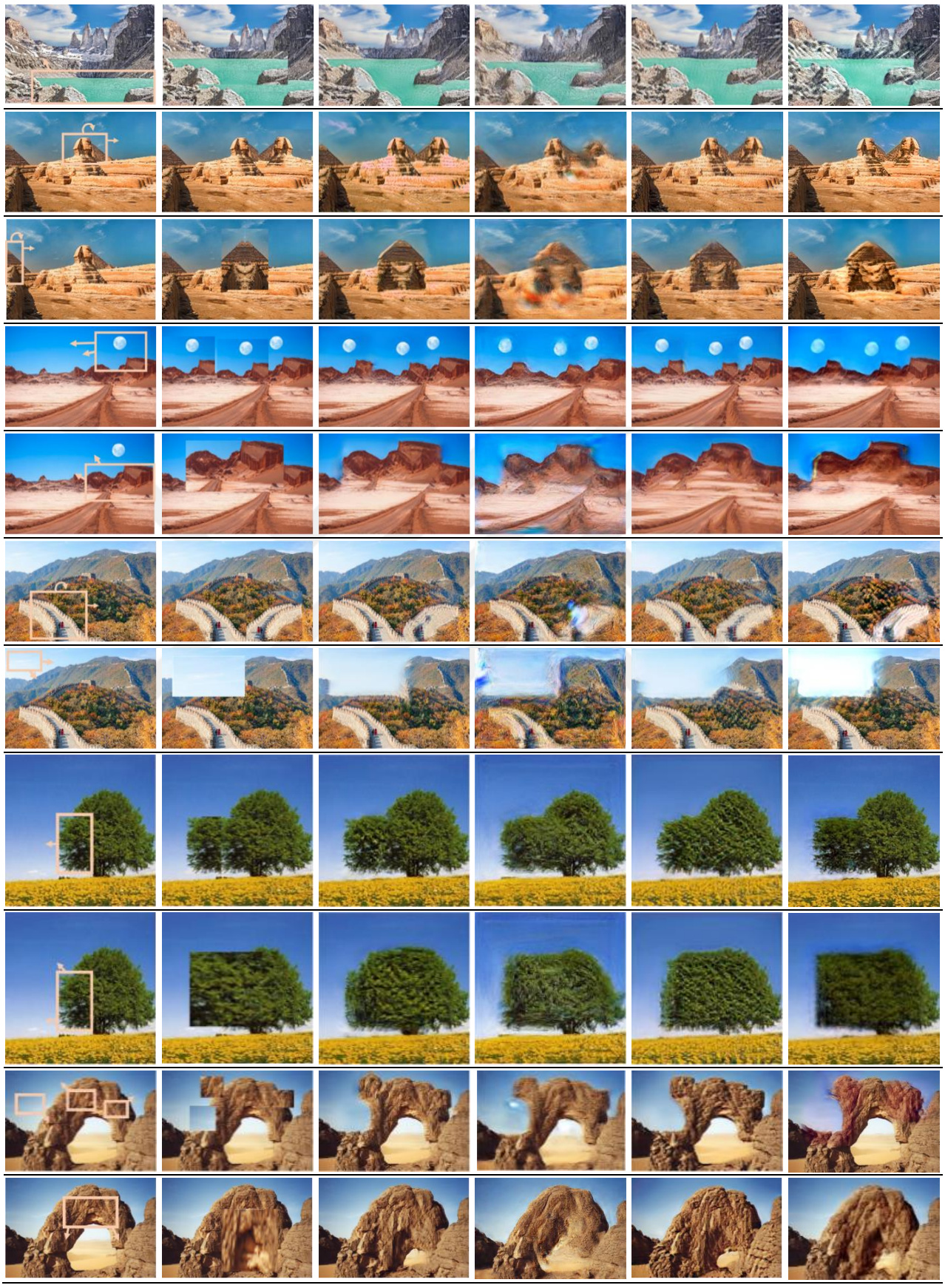


Şekil 6.4 (devam): Harmanlama uygulama sonuçları.

G. Görüntü düzenleme uygulaması örnekleri



Şekil 6.5: Görüntü düzenleme uygulama sonuçları.



Şekil 6.5 (devam): Görüntü düzenleme uygulama sonuçları.

H. Animasyon uygulaması örnekleri

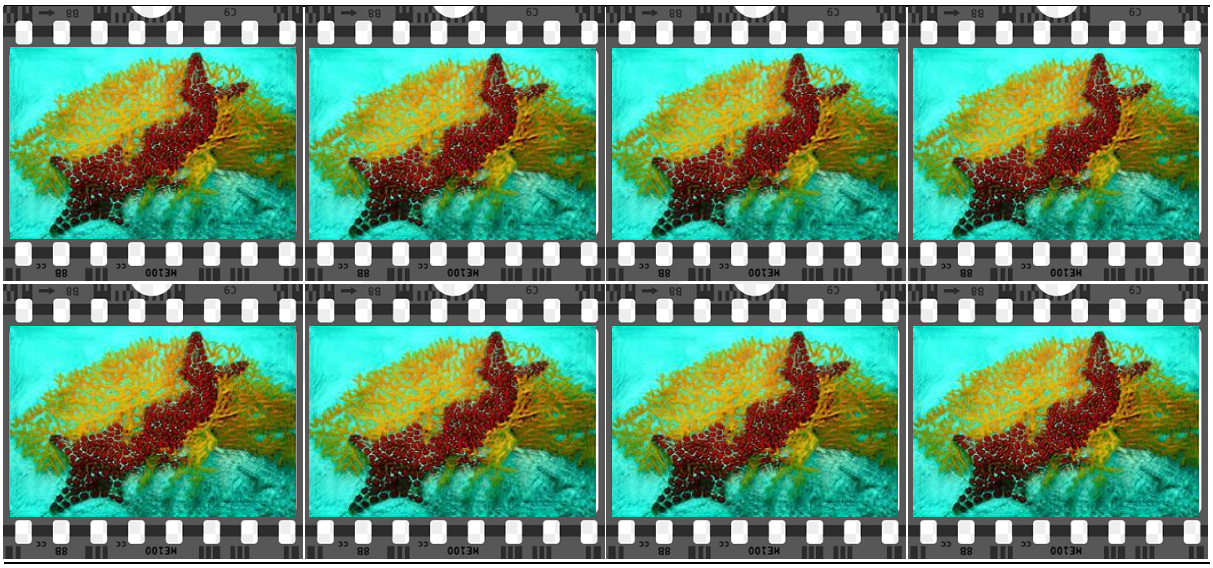
Animasyon uygulaması YouTube linkleri;

- ✓ **Model 1:** https://www.youtube.com/watch?v=Kh8BWtvMNsc
- ✓ **Model 2:** https://www.youtube.com/watch?v=GS96qUEj9x4

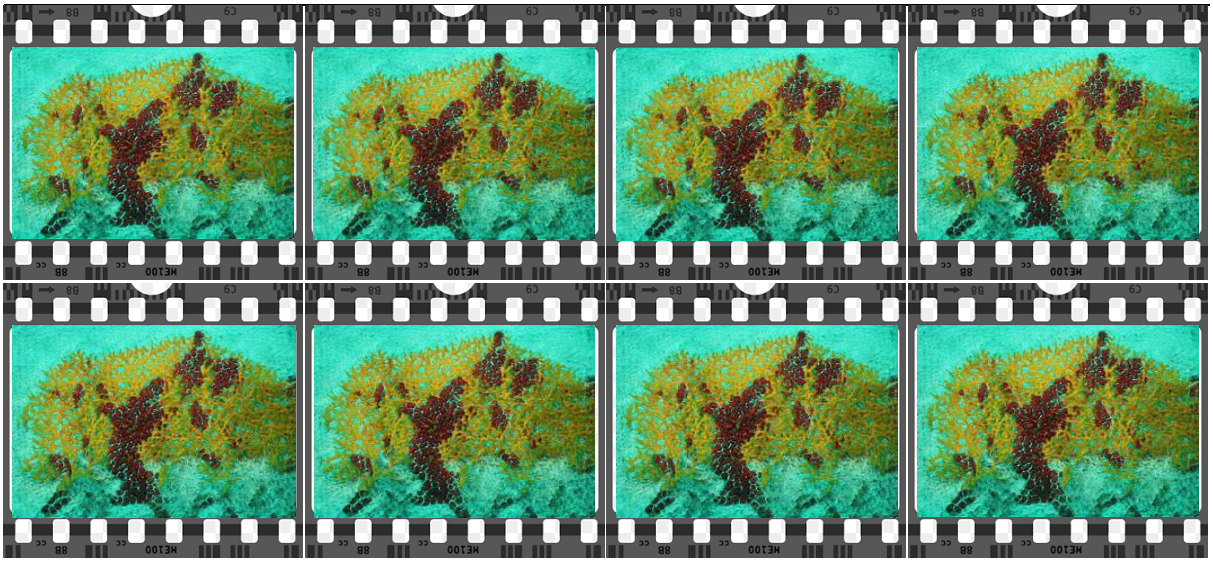
Eğitim Görüntüsü



Model 1



Model 2



Şekil 6.6: Animasyon görüntüleri (1).

Eğitim Görüntüsü



Model 1

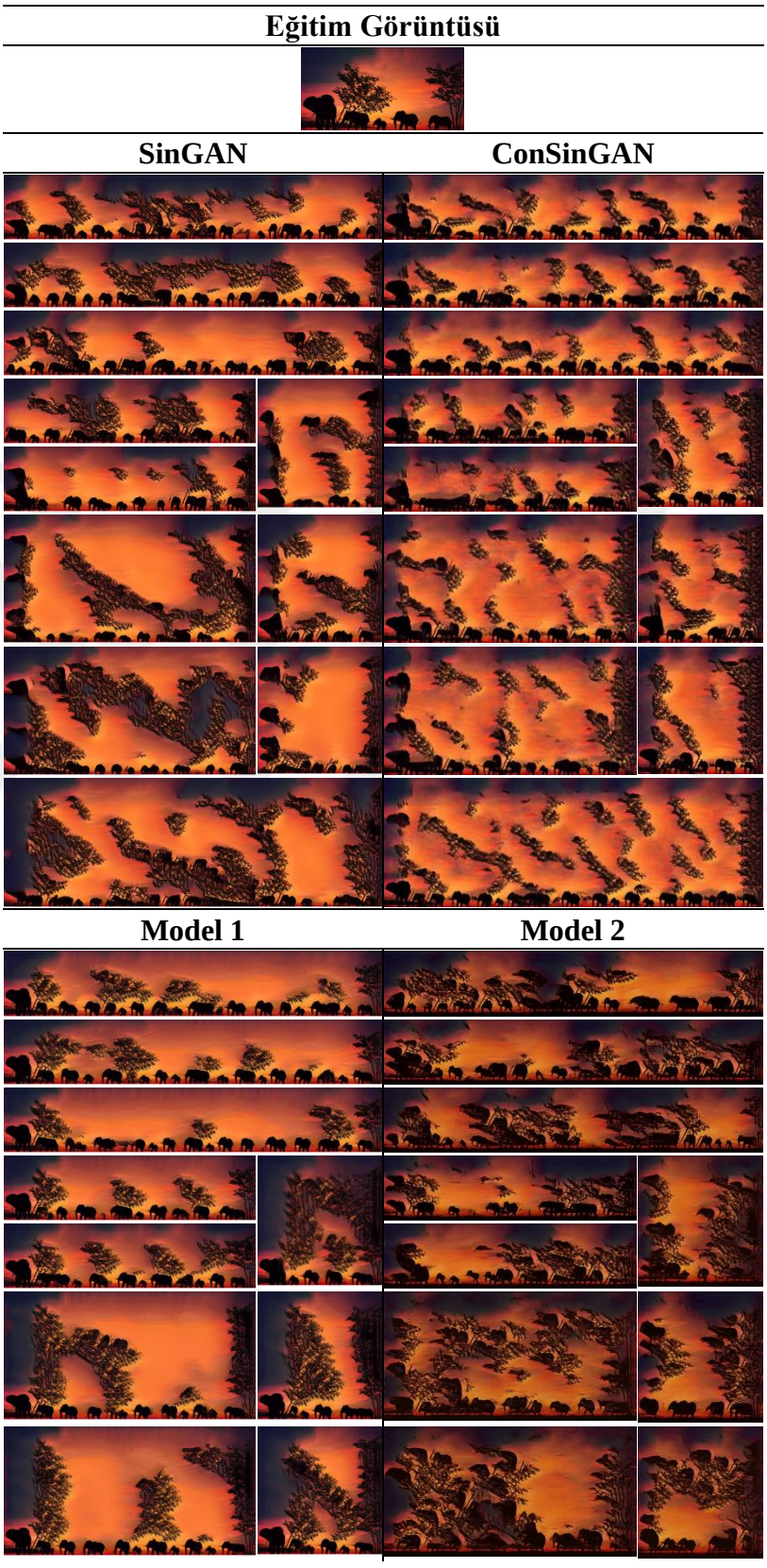


Model 2



Şekil 6.7: Animasyon görüntüleri (2).

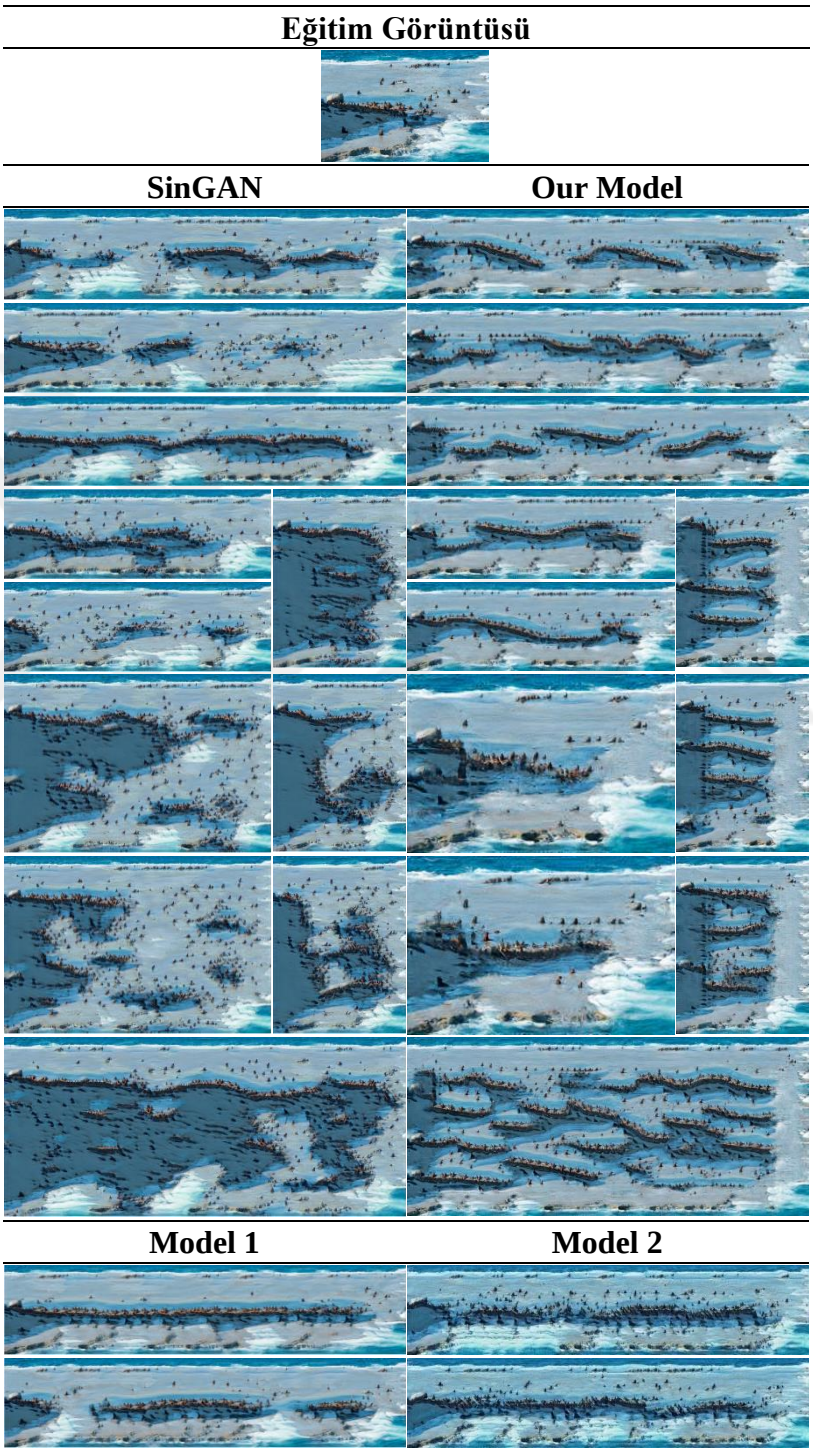
I. Farklı boyutlarda görüntü üretme uygulaması örneği



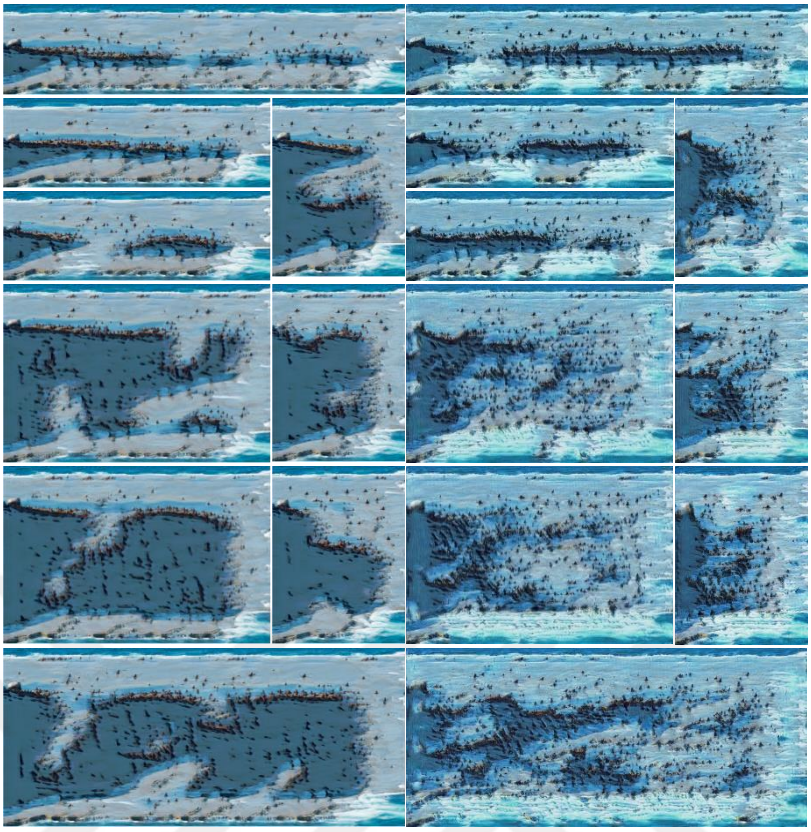
Şekil 6.8: Farklı boyutlarda üretilen görüntüler (1).



Şekil 6.8 (devam): Farklı boyutlarda üretilen görüntüler (1).

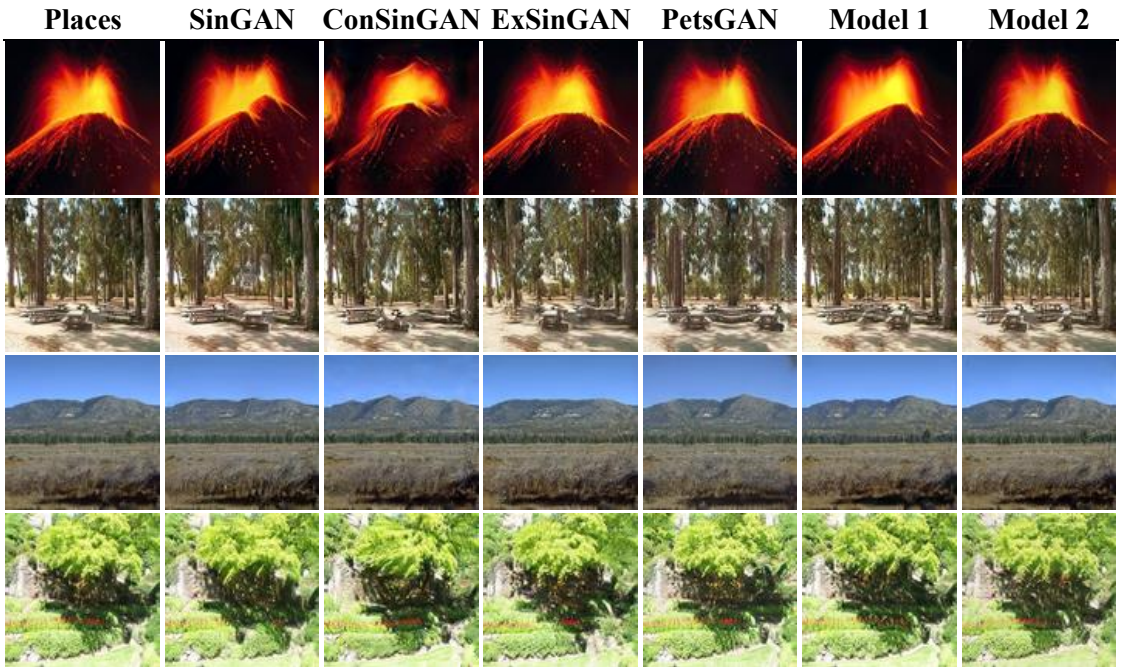


Şekil 6.9: Farklı boyutlarda üretilen görüntüler (2).

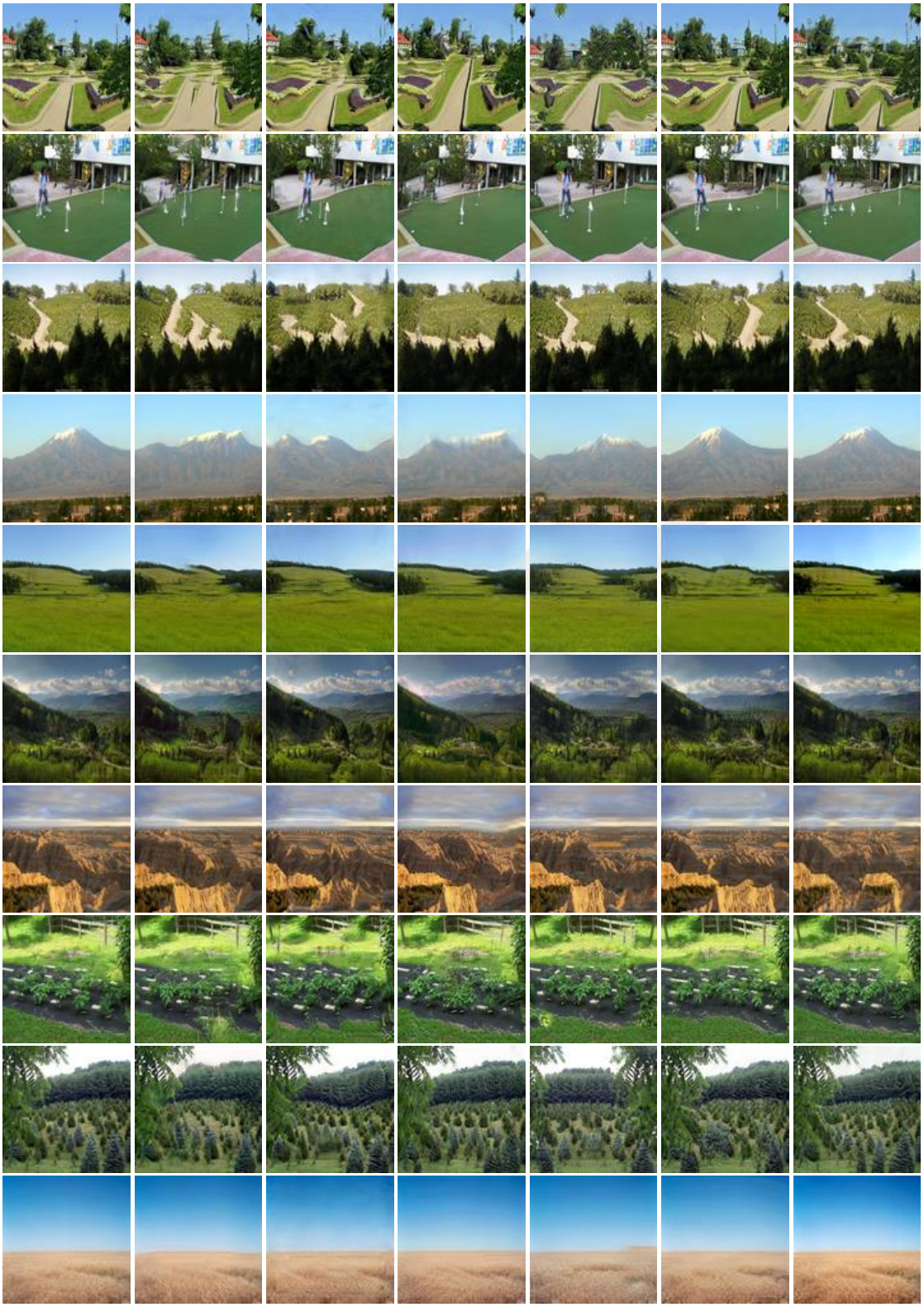


Şekil 6.9 (devam): Farklı boyutlarda üretilen görüntüler (2).

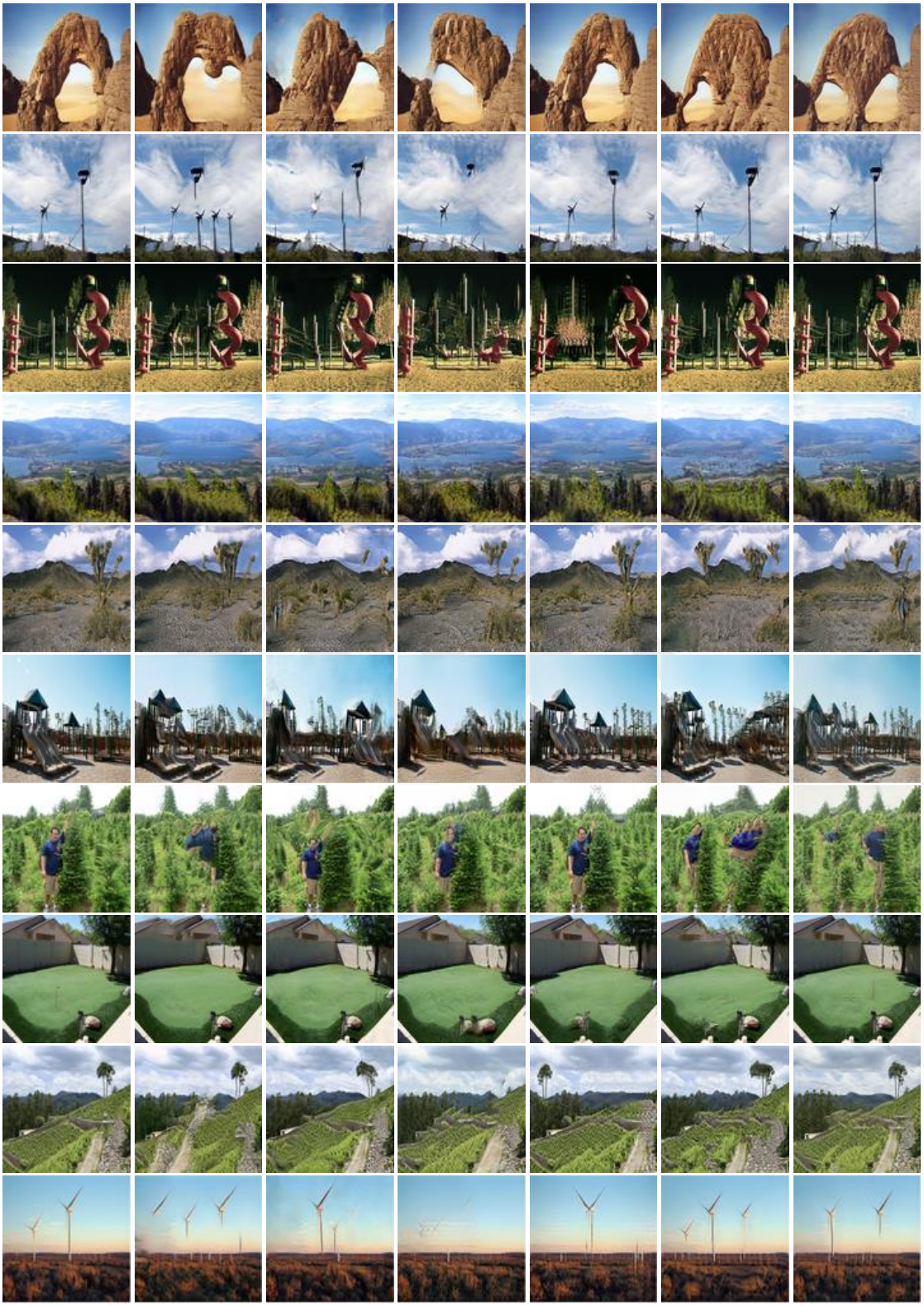
J. Places veri kümesi için modellerin ürettiği görüntüler



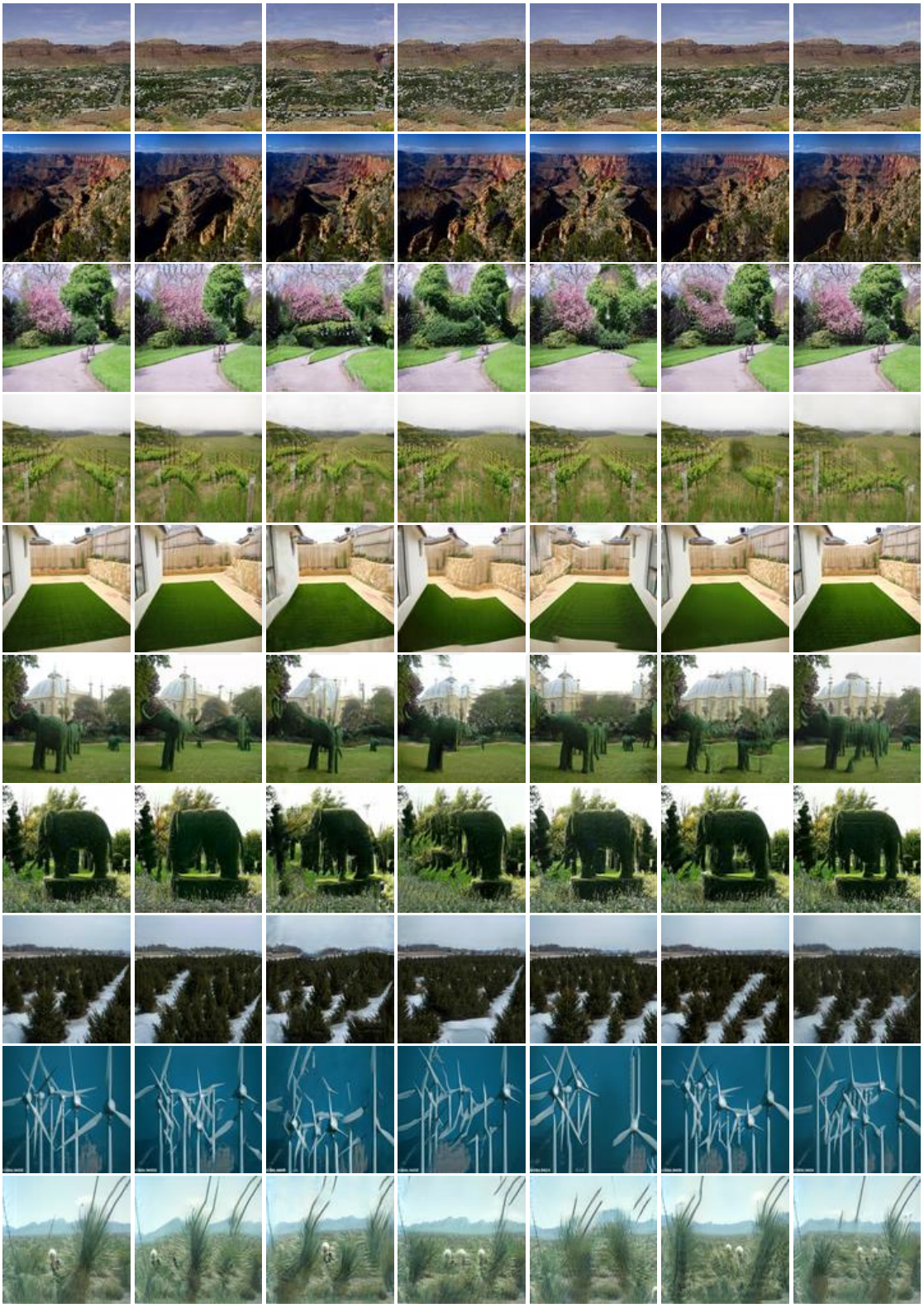
Şekil 6.10: Tüm modellerin ürettiği görüntüler (Places).



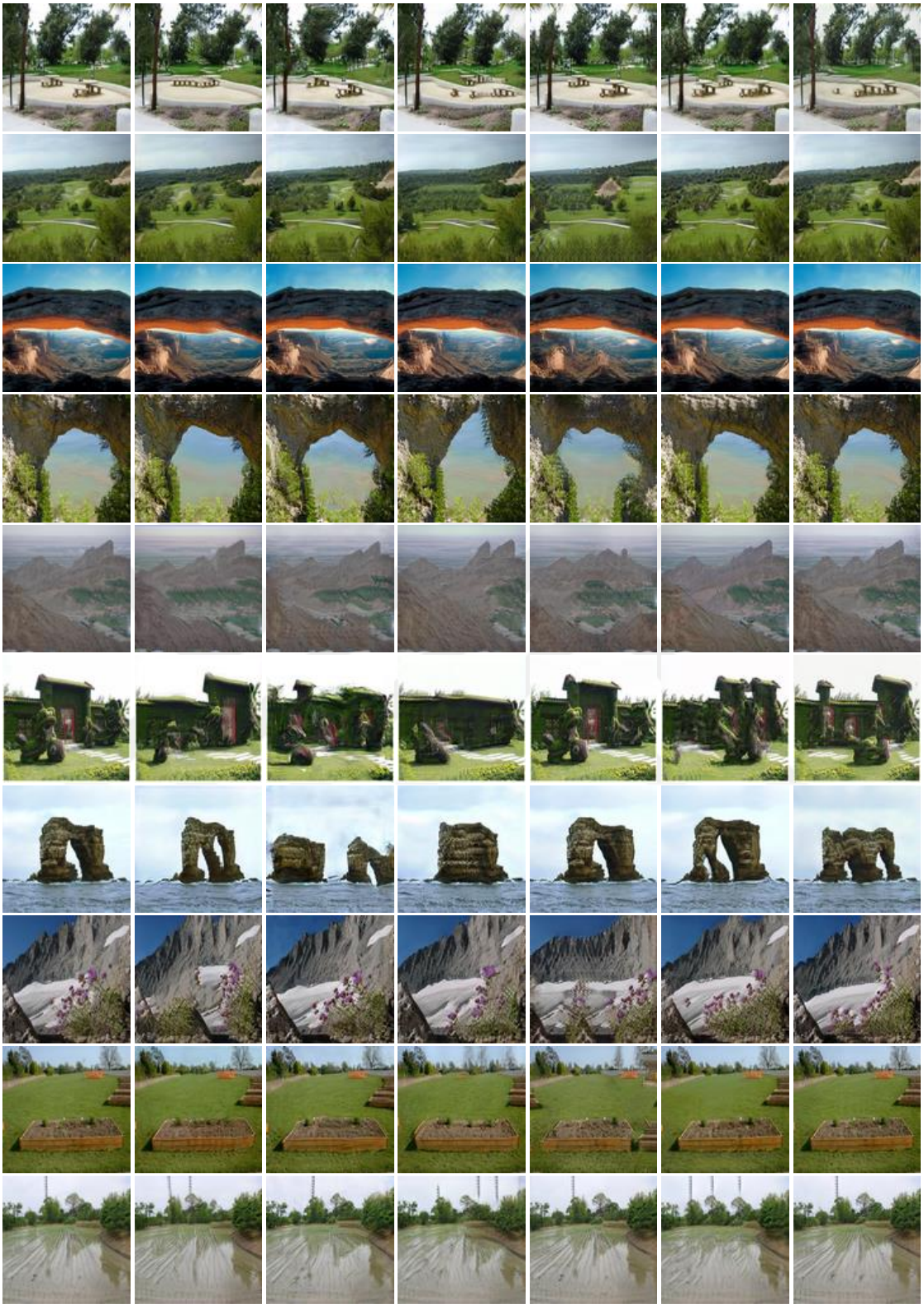
Şekil 6.10 (devam): Tüm modellerin ürettiği görüntüler (Places).



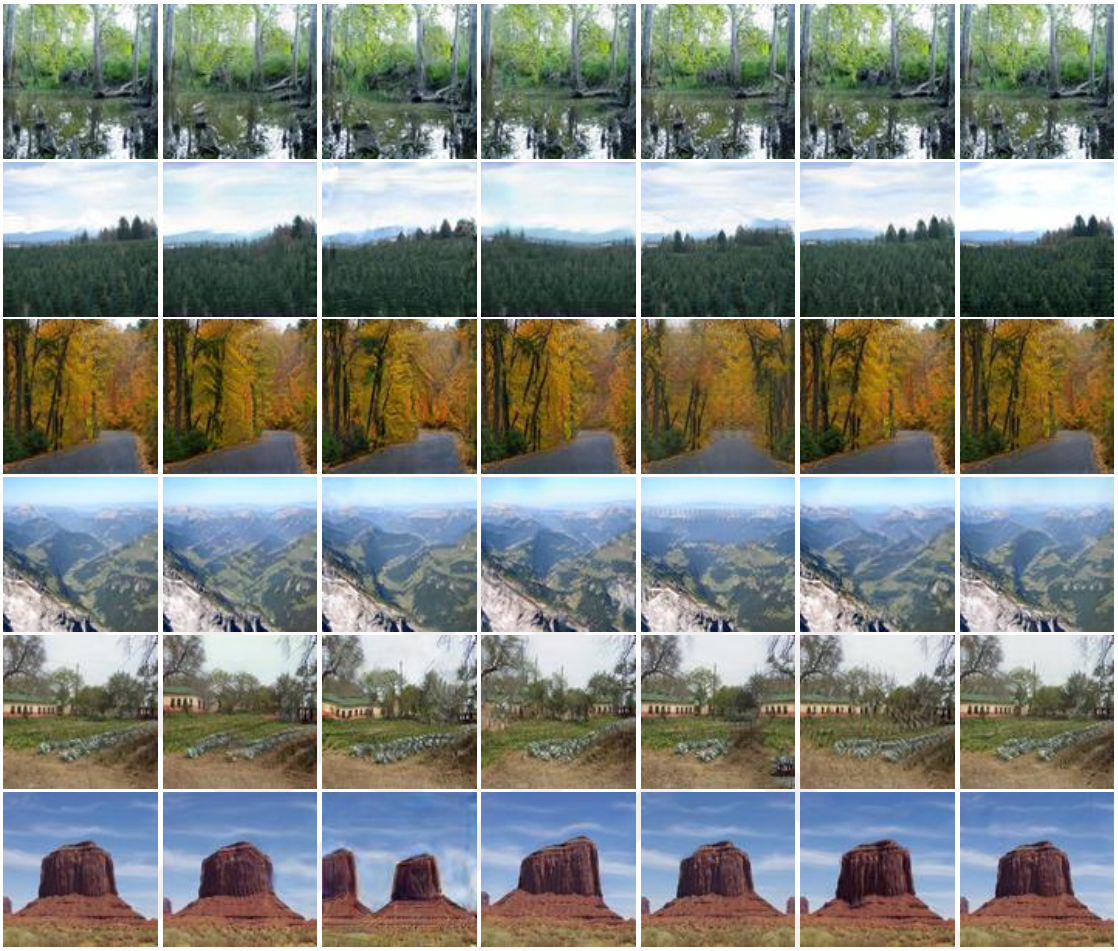
Şekil 6.10 (devam): Tüm modellerin ürettiği görüntüler (Places).



Şekil 6.10 (devam): Tüm modellerin ürettiği görüntüler (Places).

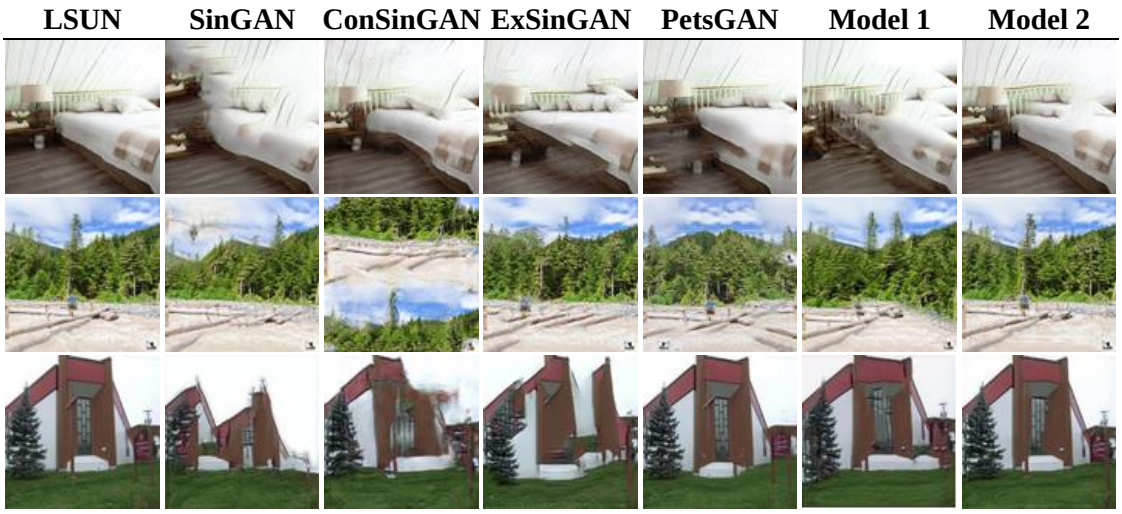


Şekil 6.10 (devam): Tüm modellerin ürettiği görüntüler (Places).

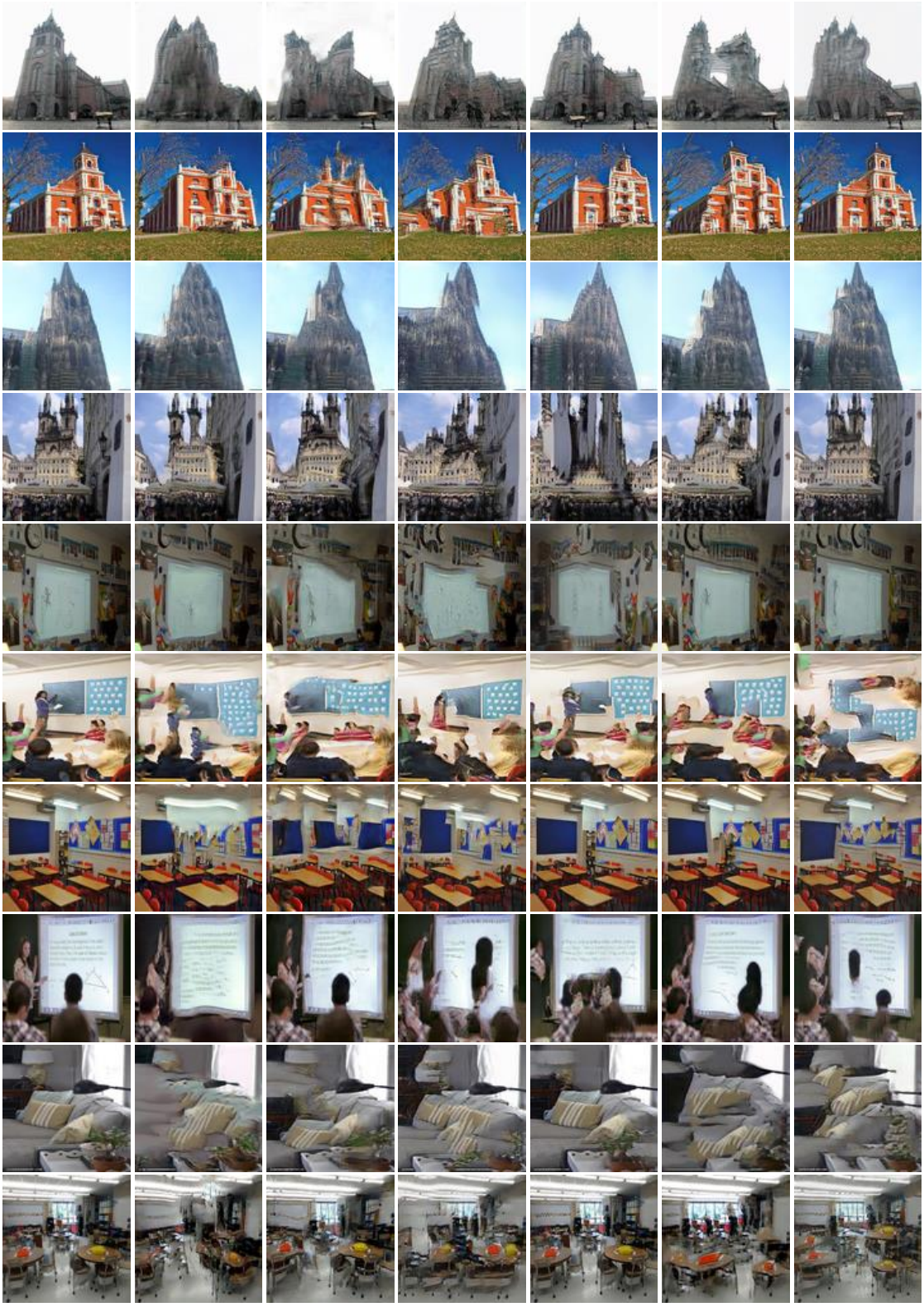


Şekil 6.10 (devam): Tüm modellerin ürettiği görüntüler (Places).

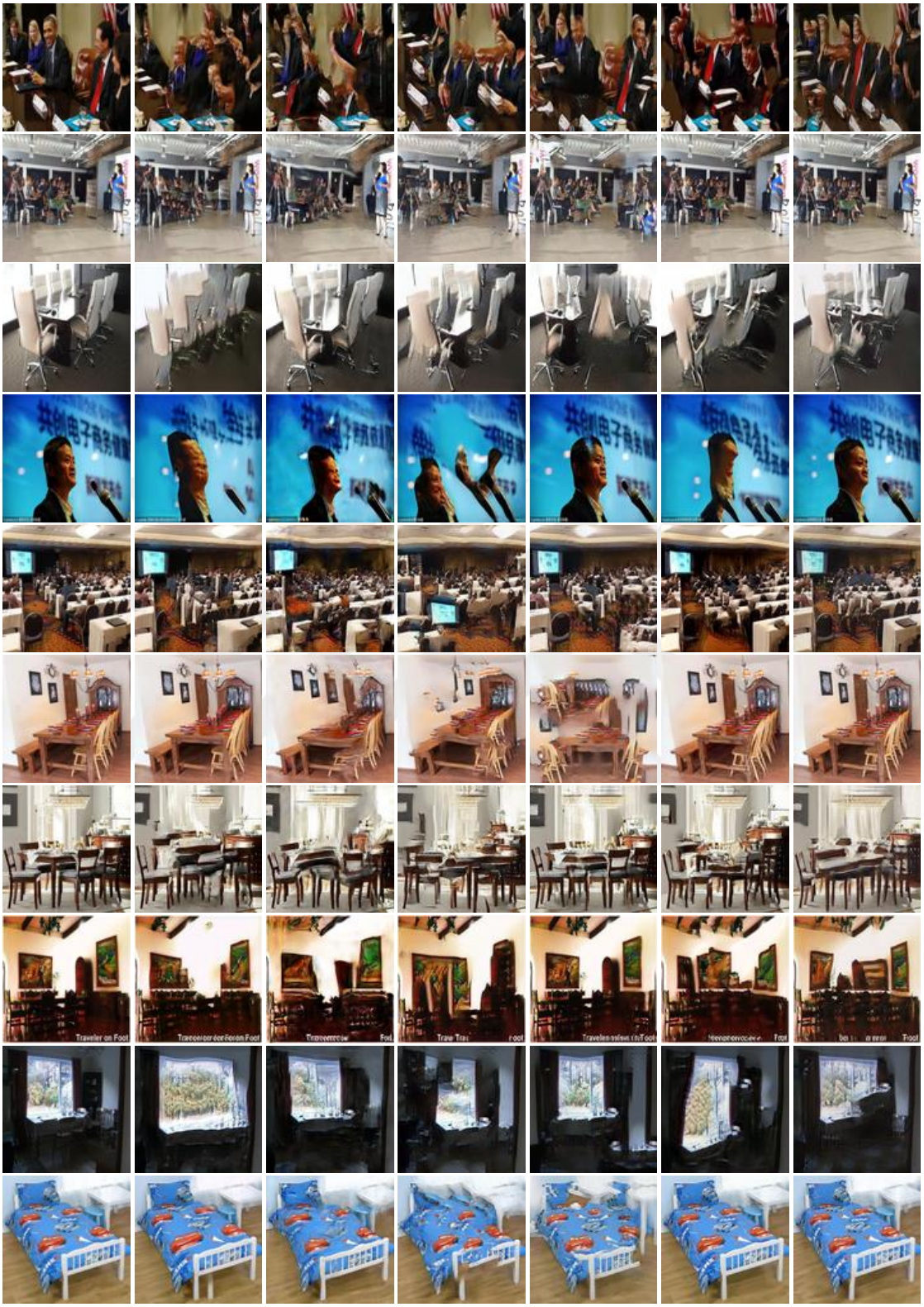
K. LSUN veri kümesi için modellerin ürettiği görüntüler



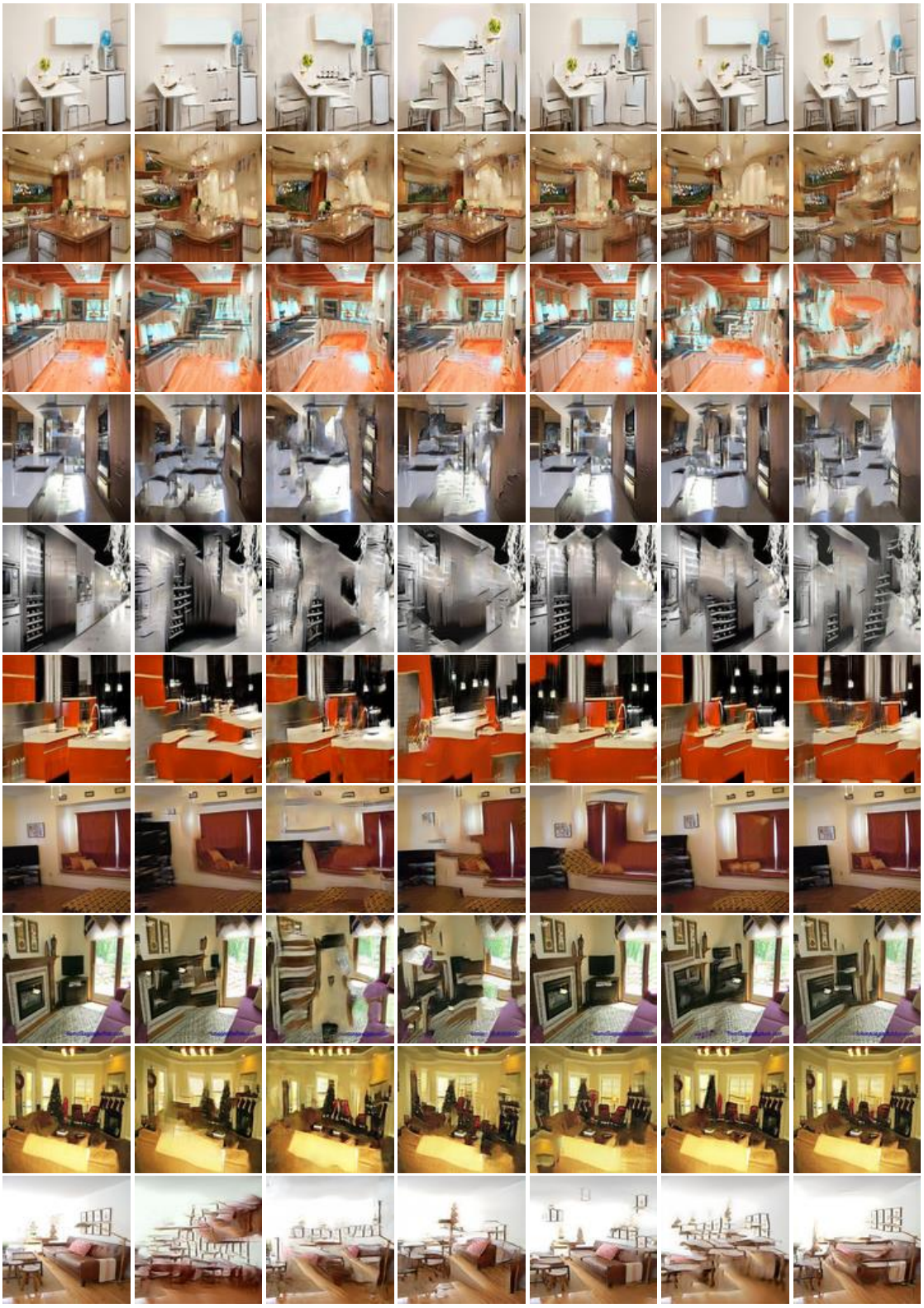
Şekil 6.11: Tüm modellerin ürettiği görüntüler (LSUN)



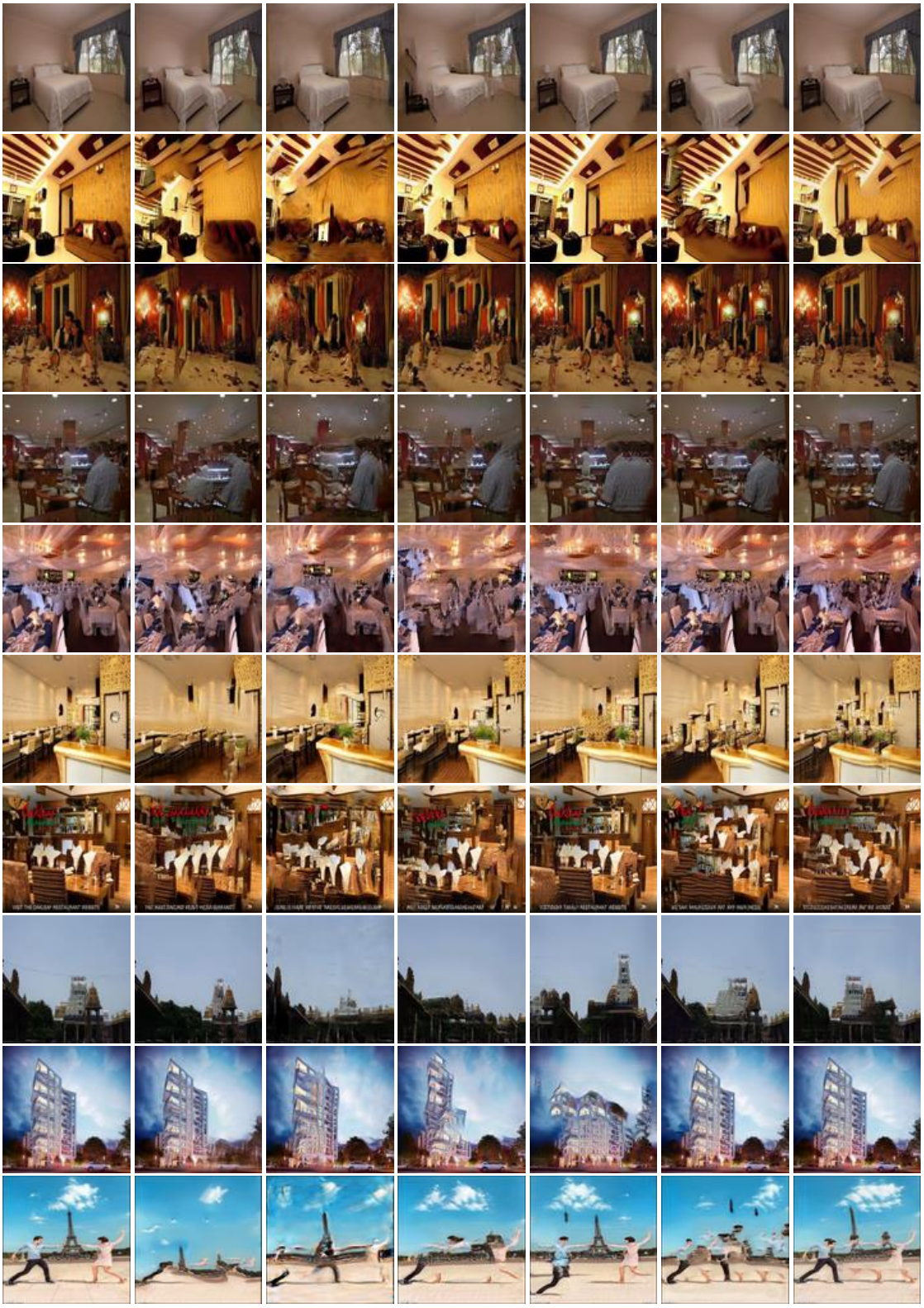
Şekil 6.11 (devam): Tüm modellerin ürettiği görüntüler (LSUN)



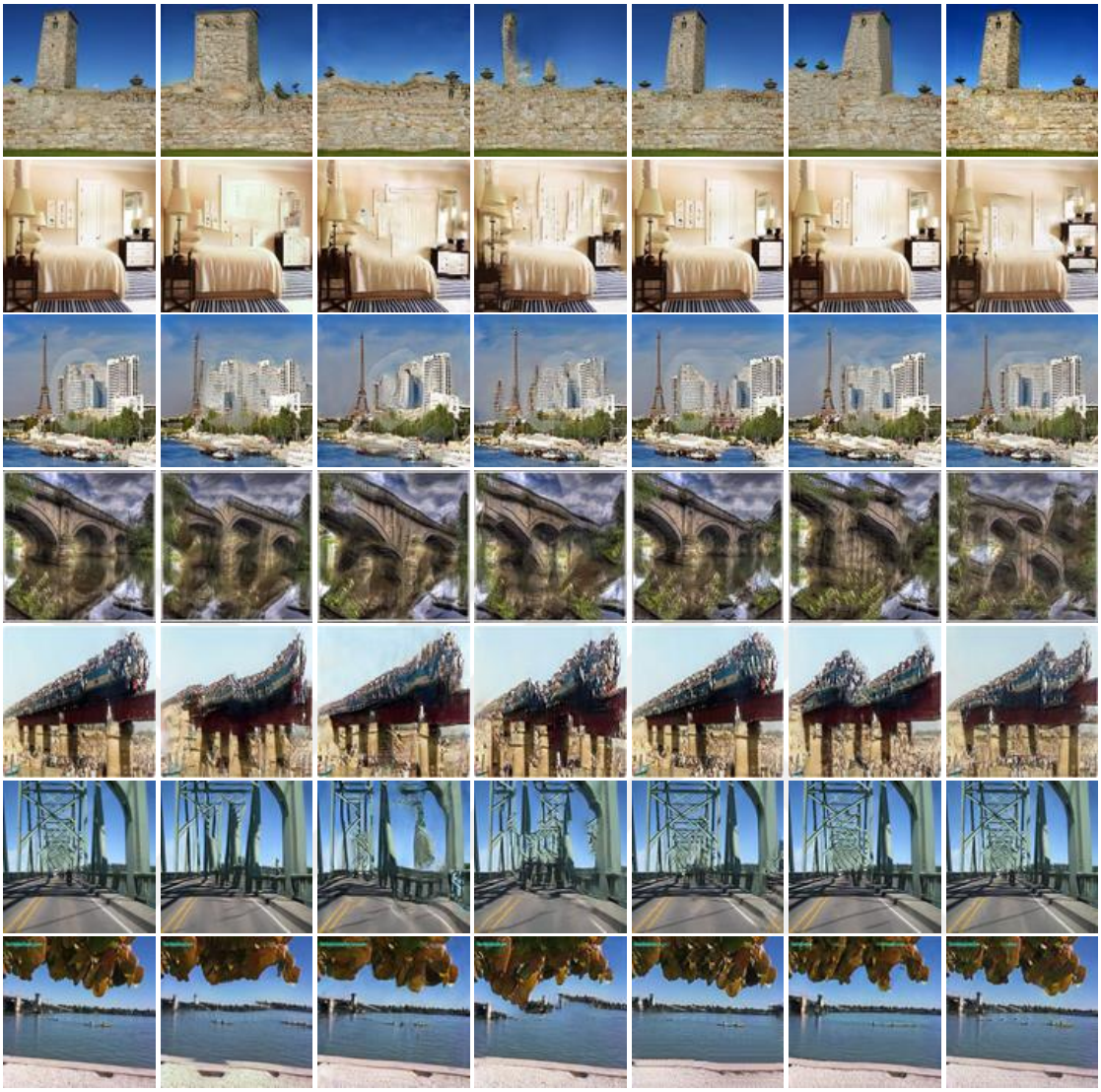
Şekil 6.11 (devam): Tüm modellerin ürettiği görüntüler (LSUN)



Şekil 6.11 (devam): Tüm modellerin ürettiği görüntüler (LSUN)

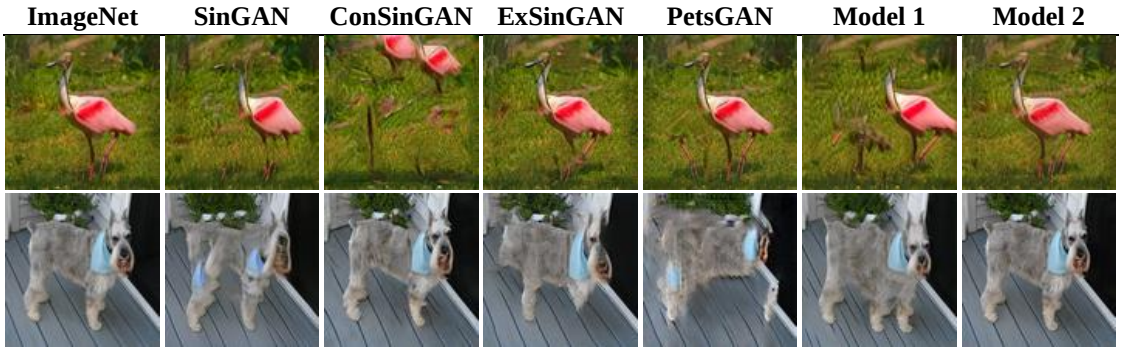


Şekil 6.11 (devam): Tüm modellerin ürettiği görüntüler (LSUN)

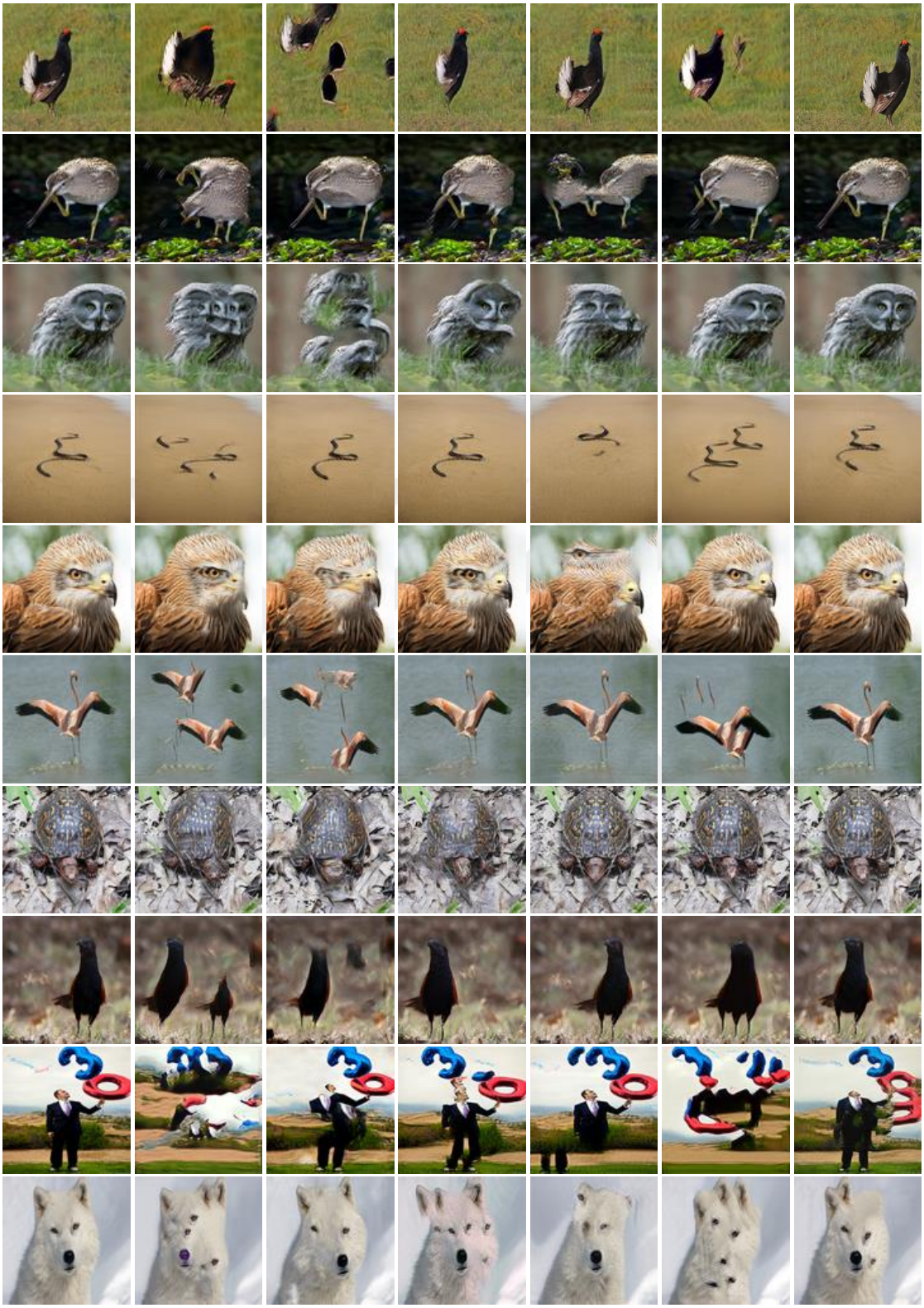


Şekil 6.11 (devam): Tüm modellerin ürettiği görüntüler (LSUN)

L. ImageNet veri kümesi için modellerin ürettiği görüntüler



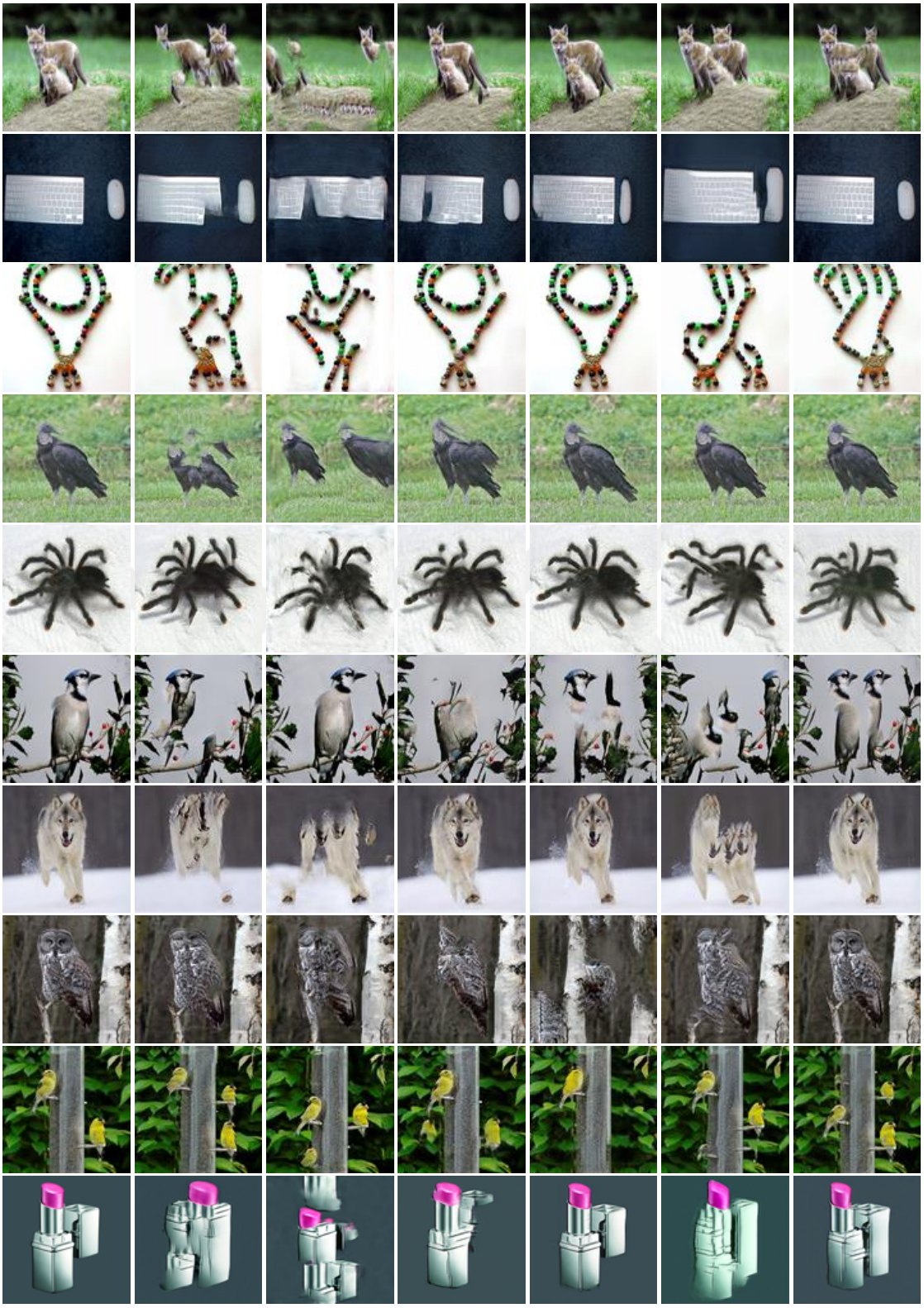
Şekil 6.12: Tüm modellerin ürettiği görüntüler (ImageNet)



Şekil 6.12 (devam): Tüm modellerin ürettiği görüntüler (ImageNet)



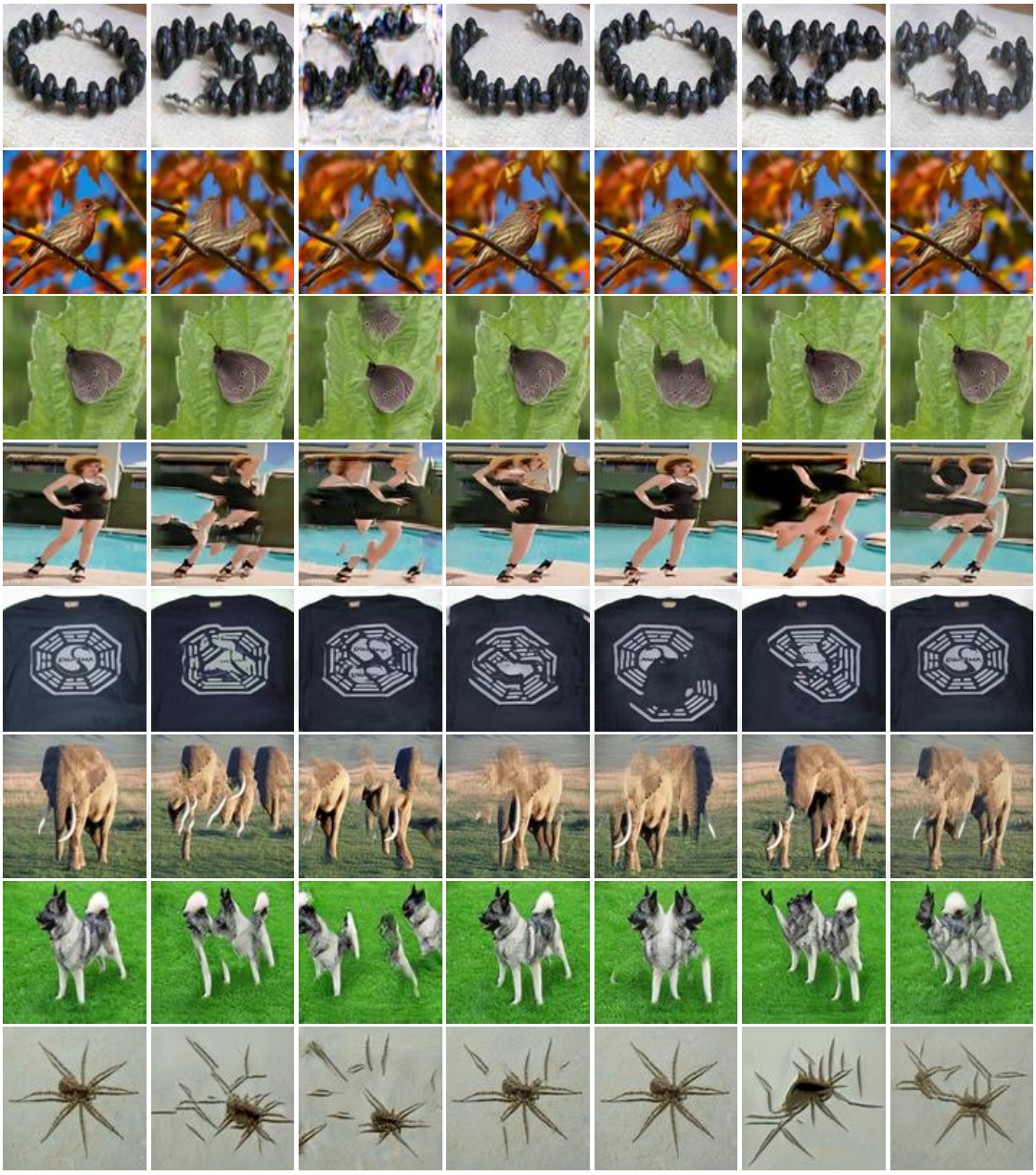
Şekil 6.12 (devam): Tüm modellerin ürettiği görüntüler (ImageNet)



Şekil 6.12 (devam): Tüm modellerin ürettiği görüntüler (ImageNet)



Şekil 6.12 (devam): Tüm modellerin ürettiği görüntüler (ImageNet)



Şekil 6.12 (devam): Tüm modellerin ürettiği görüntüler (ImageNet)

M. Karşılaştırma yapılan modellerin kod linkleri

a. SinGAN [28]: https://github.com/tamarott/SinGAN

b. ConSinGAN [32]: https://github.com/tohinz/ConSinGAN

c. ExSinGAN [35]: https://github.com/zhangzc21/ExSinGAN

d. PetsGAN [67]: https://github.com/zhangzc21/PetsGAN

N. Model 1 ablasyon çalışması parametre değerleri ve eğitim süreleri

Tablo 6.1: Model 1'in farklı konfigürasyonları parametre sayısı ve eğitim süresi

Model 1:	Places		LSUN		ImageNet	
	Parametre Sayısı	**Eğitim Süresi**	**Parametre Sayısı**	**Eğitim Süresi**	**Parametre Sayısı**	**Eğitim Süresi**
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü'de ortalama havuzlama var) + RaLSGAN-GP	~1.251.000	~30 dk.	~1.251.000	~30 dk.	~1.251.000	~30 dk.
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama var) + RaLSGAN-GP	~1.251.000	~30 dk.	~1.251.000	~30 dk.	~1.251.000	~30 dk.
Öz-dikkat + Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama yok) + RaLSGAN-GP	~1.251.000	~30 dk.	~1.251.000	~30 dk.	~1.251.000	~30 dk.
Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama yok) + RaLSGAN-GP	~1.450.000	~32 dk.	~1.450.000	~32 dk.	~1.450.000	~32 dk.
Öz-dikkat + Yoğun bağlantılı blok (üç yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama var) + RaLSGAN-GP	~1.450.000	~32 dk.	~1.450.000	~32 dk.	~1.450.000	~32 dk.
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama yok) + RaLSGAN-GP	~2.100.000	~35 dk.	~2.100.000	~35 dk.	~2.100.000	~35 dk.
Öz-dikkat + Yoğun bağlantılı blok (dört yoğun bağlantı ve bir geçiş katmanlı iki yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama var) + RaLSGAN-GP	~2.100.000	~35 dk.	~2.100.000	~35 dk.	~2.100.000	~35 dk.
Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama var) + RaLSGAN-GP	~1.250.000	~25 dk.	~1.250.000	~25 dk.	~1.250.000	~26 dk.
Yoğun bağlantılı blok (beş yoğun bağlantı ve bir geçiş katmanlı bir yoğun bağlantılı blok, Ü ve A'da ortalama havuzlama yok) + RaLSGAN-GP	~1.250.000	~25 dk.	~1.250.000	~25 dk.	~1.250.000	~26 dk.

Tablo 6.1 (devam): Model 1'in farklı konfigürasyonları parametre sayısı ve eğitim süresi

Öz-dikkat + RaLSGAN-GP	~1.035.000	~30 dk.	~1.035.000	~30 dk.	~1.035.000	~30 dk.
Öz-dikkat + WGAN-GP	~1.035.000	~30 dk.	~1.035.000	~30 dk.	~1.035.000	~30 dk.

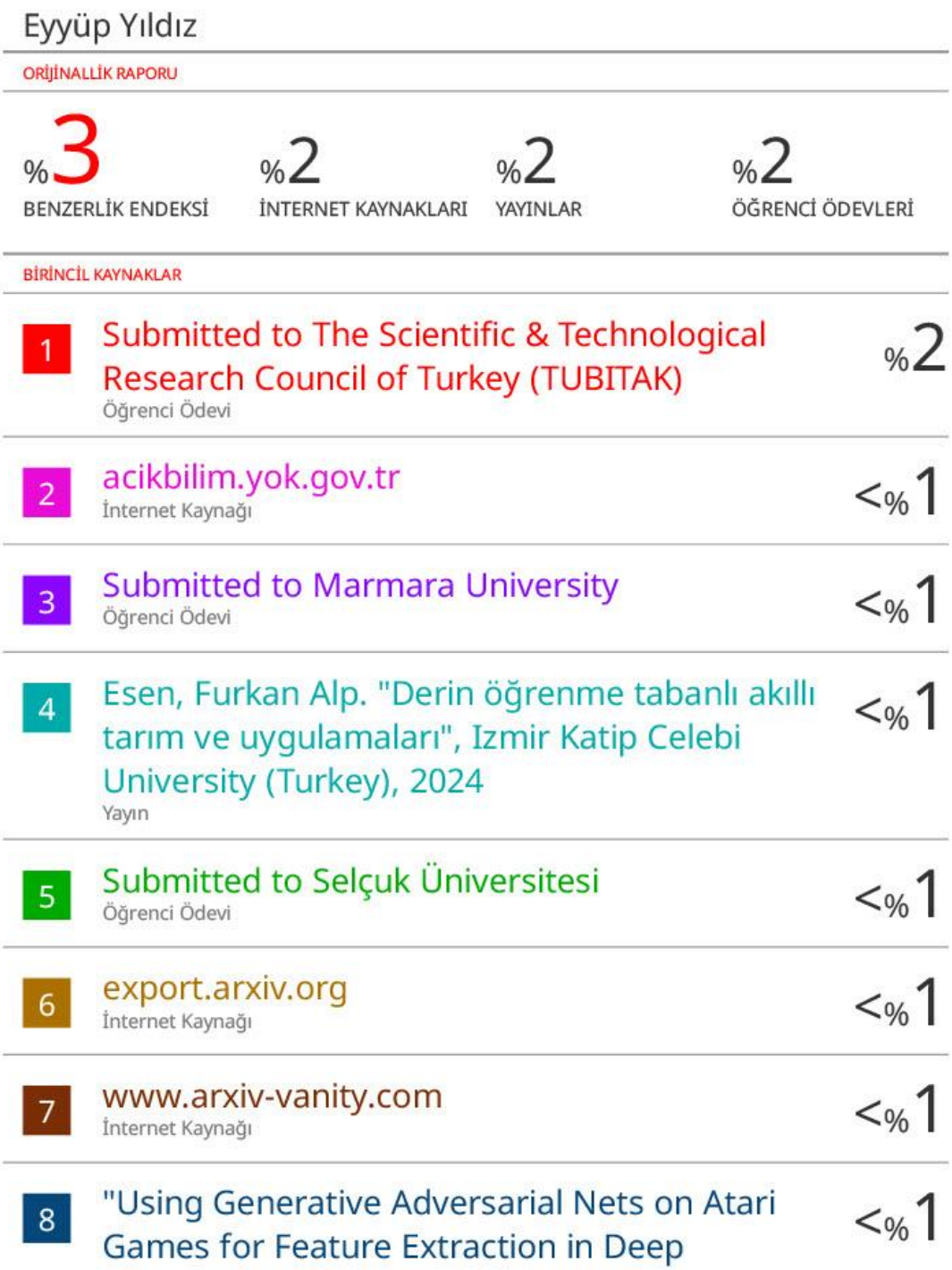
Tablo 6.2: ResNet tabanlı farklı konfigürasyonların parametre sayısı ve eğitim süresi

Model 1:	**Places**		**LSUN**		**ImageNet**	
	Parametre Sayısı	**Eğitim Süresi**	**Parametre Sayısı**	**Eğitim Süresi**	**Parametre Sayısı**	**Eğitim Süresi**
Öz-dikkat + ResNet + RaLSGAN-GP	~1.720.000	~35 dk.	~1.720.000	~35 dk.	~1.720.000	~37 dk.
Öz-dikkat + ResNet + WGAN-GP	~1.720.000	~35 dk.	~1.720.000	~35 dk.	~1.720.000	~37 dk.
ResNet + RaLSGAN-GP	~1.719.000	~25 dk.	~1.719.000	~25 dk.	~1.719.000	~26 dk.
ResNet + WGAN-GP	~1.719.000	~25 dk.	~1.719.000	~25 dk.	~1.719.000	~26 dk.

Tablo 6.3: GoogLeNet tabanlı farklı konfigürasyonların parametre sayısı ve eğitim süresi

Model 1:	**Places**		**LSUN**		**ImageNet**	
	Parametre Sayısı	**Eğitim Süresi**	**Parametre Sayısı**	**Eğitim Süresi**	**Parametre Sayısı**	**Eğitim Süresi**
Öz-dikkat + GoogLeNet + RaLSGAN-GP	~1.505.000	~55 dk.	~1.505.000	~55 dk.	~1.505.000	~55 dk.
Öz-dikkat + GoogLeNet + WGAN-GP	~1.505.000	~55 dk.	~1.505.000	~55 dk.	~1.505.000	~55 dk.
GoogLeNet + RaLSGAN-GP	~1.504.000	~48 dk.	~1.504.000	~48 dk.	~1.504.000	~48 dk.
GoogLeNet + WGAN-GP	~1.504.000	~48 dk.	~1.504.000	~48 dk.	~1.504.000	~48 dk.

İNTİHAL RAPORU İLK SAYFASI



Eyyüp Yıldız

ORİJİNALLİK RAPORU

%3	%2	%2	%2
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	%2
2	acikbilim.yok.gov.tr İnternet Kaynağı	<%1
3	Submitted to Marmara University Öğrenci Ödevi	<%1
4	Esen, Furkan Alp. "Derin öğrenme tabanlı akıllı tarım ve uygulamaları", Izmir Katip Celebi University (Turkey), 2024 Yayın	<%1
5	Submitted to Selçuk Üniversitesi Öğrenci Ödevi	<%1
6	export.arxiv.org İnternet Kaynağı	<%1
7	www.arxiv-vanity.com İnternet Kaynağı	<%1
8	"Using Generative Adversarial Nets on Atari Games for Feature Extraction in Deep	<%1

	Reinforcement Learning", 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020 Yayın	
9	Camila Fernandez Achutti. "Tree bark framework: competences and mindset rearrangements for Digital and Technology Literacy in times of exponential rate of changes", Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA), 2017 Yayın	<%1
10	Ilhan, Ceyda. "Sayisal Analiz Yontemleriyle Mimari Dokuyu Okumak: Bursa - Hisar Bolgesi", Bursa Uludag University (Turkey), 2021 Yayın	<%1
11	toad.halileksi.net İnternet Kaynağı	<%1
12	web.archive.org İnternet Kaynağı	<%1
13	Vadim Sushko, Dan Zhang, Juergen Gall, Anna Khoreva. "One-Shot Synthesis of Images and Segmentation Masks", 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023 Yayın	<%1

14	Yilmaz, Ibrahim. "Bogazkoy Barajinda cevreye Ve Insan sagligina zararli bazi Organik Kirleticilerin Lc-Ms/Ms Ile Tayini Ve cevresel Kalite standartlarina gore degerlendirilmesi", Marmara Universitesi (Turkey), 2020 Yayın	<%1
15	nek.istanbul.edu.tr:4444 İnternet Kaynağı	<%1
16	polen.itu.edu.tr İnternet Kaynağı	<%1
17	tez.sdu.edu.tr İnternet Kaynağı	<%1
18	uu245-211.uludag.edu.tr İnternet Kaynağı	<%1
19	Zilong Zheng, Jianwen Xie, Ping Li. "Patchwise Generative ConvNet: Training Energy-Based Models from a Single Natural Image for Internal Learning", 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021 Yayın	<%1
20	arxiv.org İnternet Kaynağı	<%1
21	ebin.pub İnternet Kaynağı	<%1

22	Guillermo Iglesias, Edgar Talavera, Alberto Díaz-Álvarez. "A survey on GANs for computer vision: Recent research, analysis and taxonomy", Computer Science Review, 2023 Yayın	<%1
23	Sutharsan Mahendren, Chamira U.S. Edussooriya, Ranga Rodrigo. "Diverse single image generation with controllable global structure", Neurocomputing, 2023 Yayın	<%1

Alıntıları çıkart Kapat
Bibliyografyayı Çıkart üzerinde
Eşleşmeleri çıkar Kapat

ETİK KURUL İZİN YAZISI

Uyarı: Canlı denekler üzerinde yapılan tüm araştırmalar için Etik Kurul Belgesi alınması zorunludur.

☐ **Etik Kurul izni gerekmektedir.**

☒ **Etik Kurul izni gerekmemektedir.**

Eyyüp YILDIZ
(İmza)

KURUM İZNİ YAZILARI

Uyarı: Canlı ve cansız deneklerle yapılan tüm çalışmalar için kurum izin belgelerinin eklenmesi zorunludur. Gizlilik ve mahremiyet içeren durumlarda kurum adı kapatılmalıdır.

☐ **Kurum izni gerekmektedir.**

☒ **Kurum izni gerekmemektedir.**

Eyyüp YILDIZ
(İmza)