



Görüntüden Görüntüye Dönüşüm ve Görüntü Sentezleme Yapan Üretici Çekişmeli Ağların İncelenmesi

Sara Altun^{1*}, Muhammed Fatih Talu²

¹ İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Malatya, Türkiye (ORCID: 0000-0003-2877-7105) sara.altun@inonu.edu.tr

² İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Malatya, Türkiye (ORCID: 0000-0003-1166-8404) fatihaltu@gmail.com

(3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications June 11-13, 2021)

(DOI: 10.31590/ejosat.949116)

ATIF/REFERENCE: Sara, A. & Muhammed Fatih, T. (2021). Görüntüden Görüntüye Dönüşüm ve Görüntü Sentezleme Yapan Üretici Çekişmeli Ağların İncelenmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (26), 53-60.

Öz

Görüntüden görüntüye dönüşüm işlemi; görüntü işleme, bilgisayar grafikleri ve bilgisayarla görme alanındaki problem çözümlerinde kullanılmaktadır. Görüntüden görüntüye dönüşüm, belirli bir girdinin bir görsel temsilini başka bir temsille eşleştirmeyi öğrenmeyi gerektirmektedir. Üretici çekişmeli ağlarla (GAN'lar) görüntüden görüntüye dönüşüm yoğun bir şekilde incelenmiştir. Ve incelenen yöntemler çok modlu, süper çözünürlüklü, nesne dönüştürme ile ilgili dönüşüm vb. gibi birçok kısma uygulanmıştır. Ancak, görüntüden görüntüye dönüşüm teknikleri, istikrarsızlık ve çeşitlilik eksikliği gibi bazı sorunlardan muzdariptir. Bu çalışmada, GAN algoritmalarına ve türevlerine dayanan görüntüden görüntüye dönüşüm yöntemlerine kapsamlı bir genel bakış açısı sağlama amaçlanmıştır. Ayrıca, görüntüden görüntüye dönüşüm teknikleri tartışılıp ve analiz edilmektedir. Son olarak gelecekteki araştırılan yöntemler özetlenmekte ve tartışılmaktadır.

Anahtar Kelimeler: Görüntüden görüntüye dönüşüm, üretici çekişmeli ağlar, derin öğrenme

Review of Generative Adversarial Networks for Image-to-Image Translation and Image Synthesis

Abstract

Image to image translation process is used in image processing, computer graphics and computer vision problem solving. Image-to-image translation requires learning to match one visual representation of a given input to another representation. Image-to-image translation has been extensively studied with producer adversarial networks (GANs). And the methods studied are multimodal, super tool, transformation into an object, and so on. It has been applied to many parts such as. However, it suffers from some issues such as image-to-image formatting, instability, and lack of variety. This article is intended to provide a comprehensive overview of the image translation methods according to GAN algorithms and their derivatives. In addition, its image-to-image formation is discussed and analyzed. Finally, future research methods are summarized and discussed.

Keywords: Image-to-image translation, generative adversarial networks, deep learning.

* Sorumlu Yazar: İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Malatya, Türkiye, ORCID: 0000-0003-2877-7105, sara.altun@inonu.edu.tr

1. Giriş

Derin öğrenme algoritmalarının hızlı ilerleyişi ile birçok bilgisayar görme uygulaması için dijital görüntüleri analiz etme ve anlama işlemleri son yıllarda ilgi odağı olmuştur. Bu çalışma GAN algoritmalarını ve varyantlarını kullanarak görüntüden görüntüye dönüşüme kapsamlı bakış sağlar. Üretken çekişmeli ağlara (GANlar) genel giriş yapılı ve GAN değişkenleri, yapıları ve amaç işlevleri gösterilir. Görüntüden görüntüye dönüşüm yaklaşımları, en son teknolojiye sahip algoritmalar, teori, uygulamalar ve açık zorluklar dâhil olmak üzere ayrıntılı olarak tartışılmaktadır. Görüntüden görüntüye dönüşüm yaklaşımları, denetimli ve denetimsiz türler olarak sınıflandırılır. Bu çalışmanın katkıları şu şekilde özetlenebilir: (1)Görüntüden görüntüye dönüşüm yaklaşımları, derinlemesine açıklamalarla denetimli ve denetimsiz türler olarak sınıflandırılmaktadır. (2) Aynı zamanda karşılaştırma veri kümelerini, değerlendirme metriğini ve görüntüden görüntüye dönüşüm uygulamalarını da özetlenmektedir.

Bu çalışmanın yapısını özetleyecek olursak 2.bölümde, hem denetimli hem de denetimsiz görüntüden görüntüye dönüşüm tekniklerini tanıtılmış ve tartışılmıştır. Bölüm 3'te görüntüden görüntüye dönüşüm yapıldıktan sonra kullanılan değerlendirme metrikleri açıklanmıştır. Devamında kullanılan yöntemlerin genel özellikleri tablo ile açıklanmıştır. Bölüm 4'te genel olarak elde edilen sonuçlara ve gelecekteki çalışmalara nasıl bir yön vereceği açıklanıp bitirilmiştir.

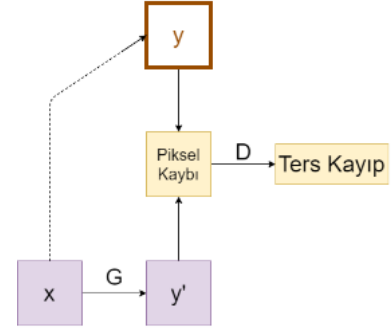
2. Üretici Çekişmeli Ağlarda Görüntüden Görüntüye Dönüşüm

2.1. Görüntüden Görüntüye Dönüşüm Teknikleri

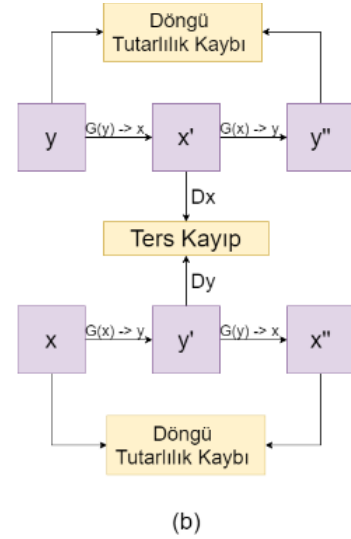
Bu bölümde, görüntüden görüntüye dönüşüm yöntemleri, Şekil 1'de gösterildiği gibi denetimli ve denetimsiz türler olarak sınıflandırılmıştır.

Denetimli dönüşüm yöntemi farklı iki alandaki (X, Y) , (x, y) görüntü çiftini barındırır ve her $x \in X$ görüntüsü ile başka bir $y \in Y$ alanından karşılık gelen bir görüntü için $G: X \rightarrow Y$ dönüşümcüsünü kullanırken bir olasılık dağılımını öğrenir. Denetimli dönüşüm bazı durumlarda yüksek kaliteli görüntü üretmek için kaynak görüntülere göre koşullandırılmış etki alanı görüntülerini veya sınıf etiketlerini kullanır. Denetimli dönüşüm ayrıca aşağıdaki bölümlerde açıklandığı gibi yönlü ve çift yönlü dönüşüm olarak ikiye ayrılır. Bu yöntemlerin nasıl bir ilerleyişe sahip olduğu Şekil 2'de gösterilmektedir.

Denetimsiz dönüşüm yöntemi gösterildiği gibi, iki üretici haritalama işlemi $G: X \rightarrow Y$ ve $F: Y \rightarrow X$ ve iki düşman ayırıcı D_y ve D_x olan simetrik bir yapıdır. Denetimsiz dönüşüm yöntemlerinin nasıl bir ilerleyişe sahip olduğu Şekil 3'de gösterilmektedir.



Şekil 2. Denetimli görüntüden görüntüye dönüşüm yöntemleri (Pix2Pix gibi denetlenen yöntemler)



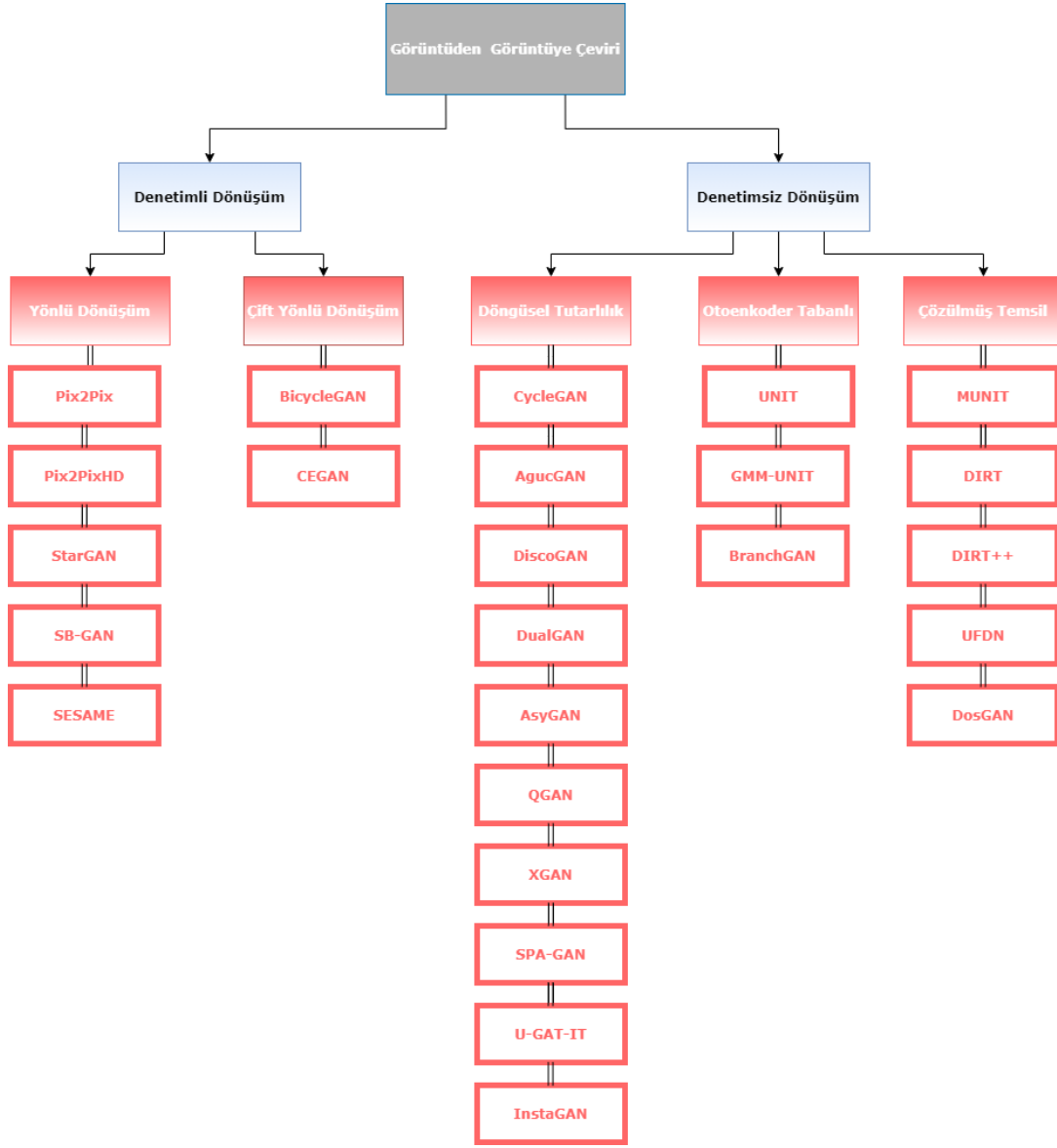
Şekil 3. Denetimsiz görüntüden görüntüye dönüşüm yöntemleri (CycleGAN, DualGAN ve DiscoGAN, InstaGAN gibi denetimsiz yöntemler)

2.1.1. Denetimli Dönüşüm

Denetimli dönüşüm yapmak için birçok yöntem önerilmiştir. Bu yöntemler yönlü ve çift yönlü dönüşüm olmak üzere ikiye ayrılır.

a. Yönlü dönüşüm

Pix2Pix (Isola ve ark., 2017) koşullu üretici çekişmeli ağa dayanan denetimli görüntüden görüntüye dönüşüm yaklaşımıdır. Pix2Pix, iki veri kümesi kullanır ve bire bir eşlemeyi öğrenmek için eşleştirilmiş görüntüler gerektirir. Veri kümelerinden biri girdi olarak kullanılırken diğeri koşul girdisi olarak kullanılır. Üretici ağ, her katmana atlama bağlantılarına dayanan "U-Net" tabanlı bir mimari kullanır. Buna karşılık ayırıcı, sınıflandırıcı olarak evrişime dayalı bir "PatchGAN" kullanır. Pix2Pix'in amaç işlevi, L2 yerine L1 normuna sahip cGAN'ı kullanmaktadır. Bu da daha az bulanıklığa yol açar.



Şekil 1. Genel Görüntüden Görüntüye Dönüşüm Yapan Üretici Çekişmeli Ağlar

Pix2Pix gibi cGAN'a dayanan çeşitli görüntüden görüntüye dönüşüm uygulamalarına olanak sağlasa da, üretilen görüntüler yine de düşük çözünürlüklü ve bulanık olabilmektedir. Wang ve arkadaşları (Wang ve ark., 2018) genelden özele bir üretici ağ, üç çok boyutlu ayrıştırıcıya dayalı bir mimari ve güçlü bir çekişmeli öğrenme işlemleri kullanarak çıktı görüntülerinin çözünürlüğünü 2048×1024 'e yükseltmek için Pix2pixHD'yi önermiştir.

Önceki bahsedilen çalışmalar iki alanla (X, Y) sınırlı olduğu bilinmektedir. Bu düşünceden yola çıkılarak, yalnızca tek bir üretici ve bir ayrıştırıcı ağ kullanarak çok alanlı görüntüden görüntüye dönüşüm için birleşik bir GAN olan StarGAN (Choi ve ark., 2018) önerilmiştir. Üretici ağ girdi görüntüsünü (x) çıktı görüntüsüne (y) çevirmek için eğitilir, etki alanı etiket bilgisi c , $G(x, c) \rightarrow y$ üzerinde koşullandırılır. Çoklu veritabanları arasında k alanı $k(k-1)$ arasındaki eşlemeyi öğrenmek için etki alanı etiketlerini kontrol etmek, bilinmeyen etiketleri göz ardı etmek ve belirli bir veri kümesine ait belirli bir etikete odaklanmak için maske vektörü m kullanılır. StarGAN, birden fazla alanı kontrol

etmek için ayrıştırıcı ağın üstüne bir yardımcı sınıflandırıcı ekleyerek ve üretici ağa döngü tutarlılığı uygulayarak birden çok veritabanını aynı anda eğitmek için tek bir üretici ağ ve bir ayrıştırıcı ağ gerektirir (Wu ve ark., 2017).

SB-GAN (Azadi ve ark., 2019) rastgele görüntüden başlayarak, bir bölütleme düzeni sentezler ve bölütleme sentezi ağını gerçekçi görünen bölütleme düzenlerine yönlendirmek için ayrıştırıcı ağ kullanır. Oluşturulan düzen daha sonra son görüntünün sentezlenmesi için bir koşullu görüntü sentez ağına girdi olarak sağlanır. Koşullu görüntü sentez ağını gerçek bölütleme düzenleriyle eşleştirilmiş gerçekçi görüntülere yönlendirmek için ikinci bir ayrıştırıcı ağı kullanılır. Son olarak, koşullu görüntü sentez ağını temel gerçeğe uyan görüntüler oluşturmaya yönlendirmek için üçüncü bir koşulsuz ayrıştırıcı ağ kullanılır.

SESAME yöntemi (Ntavelis ve ark., 2020) üretici ağı, anlamsal maske tarafından belirlenen pikselleri oluşturmayı amaçlar. Sınıf içi üretime rehberlik etmek için kod çözücündeki alıcı alanı ve SPADE katmanlarını artırmak için genişletilmiş evrişimlere sahip bir kodlayıcı-kod çözücü mimarisine sahiptir. SESAME ayrıştırıcı ağı, yaygın olarak kullanılan

PatchGAN'ın aksine RGB görüntüsünü ve semantiğini bağımsız olarak işler. Son evrişimli katmandan önce iki akış, DRGB ve DSem birleştirilir. SESAME'de, Toplam Küresel Havuzlama işlemiyle çıktı yamalarının sayısına eşit uzamsal boyutların 2B matrisine indirgenir. Her yoldaki DRGB özellik vektörü DSem tarafından ölçeklendirilir ve ürüne bir kalıntı eklenir.

b. Çift yönlü dönüşüm

BicycleGAN (Zhu ve ark., 2017), eşleştirilmiş eğitim görüntüleri gerektiren çok modellenli bir alanlar arası dönüşüm yöntemidir. Gerçekçi ve çeşitli çıktılar oluşturmak için koşullu değişken otomatik kodlayıcıyı (cVAE) koşullu gizli regresör (cLR) ile birleştirir. Bu model, üretici ağ için bir U-Net ve ayırıcı için iki PatchGAN kullanan geleneksel LSGAN'dır. BicycleGAN, eğitim sürecini stabilize etmek için çapraz entropi kaybı yerine en küçük kareler kayıp fonksiyonunu kullanır.

Bu yöntemlere ek olarak, gerçekçi ve çeşitli görüntüler oluşturmak için koşullu üretme modellerine dayanan çok modellenli haritalamayı öğrenmeye yönelik yeni bir görüntüden görüntüye dönüşüm yaklaşımı olarak CEGAN (Xiong ve ark., 2019) önerilmiştir. CEGAN, gizli alan ile gerçek görüntü alanı arasında sıkı bağlantılar kurarak olası çoklu sonuç modlarının dağılımını yakalar. Model, üretici G, ayırıcı D ve kodlayıcı E'den oluşur. Bu modelde ayırıcı ağ, diğer önerilen GAN yöntemlerinden farklı olarak, gerçekçi görünümü görüntüler üretme yaparken fazlalık ve gürültünün etkisini azaltmak için gerçek görüntü alanı yerine gizli uzaydaki gerçek ve sahte örnekleri ayırt etmede kullanır.

2.1.2. Denetimsiz Görüntüden Görüntüye Dönüşüm

Eşleştirilmiş görüntüler olmadan iki veya daha fazla etki alanı arasındaki eşlemeyi öğrenmeyi amaçlamaktadır ve son zamanlarda görüntüden görüntüye dönüşümde çapraz eşlemeyi öğrenme yeteneği nedeniyle yoğun bir şekilde araştırılmıştır. Denetimsiz dönüşüm yapmak ve sınırlı eğitim örnekleri sorununu hafifletmek için birçok yöntem önerilmiştir. Bu yöntemler üç gruba ayrılır: döngüsel tutarlılık tabanlı, otomatik kodlayıcı tabanlı ve çözülmüş gösterim kullanan yöntemler.

a. Döngü tutarlılığı ile denetimsiz dönüşüm

Eşleştirilmemiş görüntüden görüntüye dönüştürme sorunlarının üstesinden gelmek için döngüsel kayıplar, üç tür kayıp kullanarak girdi ve çevrilen görüntü arasındaki temel öznitelikleri korumak için kullanılır: ters kayıp, döngüsel tutarlılık kaybı ve yeniden yapılandırma kaybı. Yeniden yapılandırma kaybı, dönüşümü, iki eşleme işlevinden oluşan kimlik eşlemesine yakın olacak şekilde düzenlemek için kullanılır. Burada birincisi, giriş etki alanından hedef etki alanına ileri döngü eşlemesini ve ikinci işlevi, yani bir geriye doğru işlev eşlemesini öğrenir, ileriye doğru haritalamanın tersini öğrenir (Tripathy ve ark., 2018).

Her iki işlev de, Şekil 2'te gösterildiği gibi, iki kayıp işlevi kullanılırken, yani ters kayıplar ve döngüsel tutarlılık kayıpları kullanılarak aynı anda eğitilir. Zhu ve arkadaşları (Zhu ve ark., 2017) iki üretici ağ ve iki ayırıcı kullanarak iki farklı alandan eşleştirilmemiş görüntülerin bir koleksiyonu arasındaki eşleştirmeyi öğrenen denetimsiz bir görüntüden

görüntüye dönüşüm yöntemi önermektedir. CycleGAN iki haritalama işlevi $G: X \rightarrow Y$ ve $F: Y \rightarrow X$ ve iki çekişmeli ayırıcı D_y ve D_x olan simetrik bir yapıdır. Üretici G , girdiyi X alanından Y 'ye eşler, üretici F , Y 'den X 'e girişi eşler. Ayırıcı $D(y)$, y 'yi $G(x)$ 'den ayırt ederken, ayırıcı $D(x)$, x 'i $F(y)$ 'den ayırır. CycleGAN'ın amaç işlevi iki terim içerir: rakip kayıplar ve döngüsel tutarlılık kayıpları. Hedef görüntülerin oluşturduğu görüntülerdeki çekişmeli kayıplar, aşağıdakilerin dağılımını eşleştirmek için kullanılır.

CycleGAN yöntemi; nesne dönüştürme, koleksiyon stili aktarımı, sezon aktarımı, fotoğraf geliştirme vb. dahil olmak üzere birçok görüntüden görüntüye dönüşüm görevine başarıyla uygulanmıştır. Aynı zamanda CycleGAN yalnızca bire bir eşleştirmeyi öğrenebilir. Yakın zamanda önerilen artırılmış CycleGAN (AgucGAN) (Almahairi ve ark., 2018), CycleGAN'a benzer bir fikir ortaya koyar. Ancak her bir alana yardımcı gizli kodlar ekleyerek denetimsiz ortamda iki alan arasında çoktan çoğa eşleştirme öğrenir.

Benzer şekilde, DiscoGAN (Kim ve ark., 2017) ve DualGAN (Yi ve ark., 2017), döngüsel tutarlılığa dayalı eşleştirilmemiş görüntüden görüntüye dönüşüm sorununu çözmek için aynı anda önerilmiştir. DiscoGAN, yeniden yapılanma kayıplarını kullanarak farklı alanlar arasındaki ilişkileri keşfetmeyi öğrenirken; DualGAN, Wasserstein GAN kullanarak ikili öğrenmeye dayanmaktadır.

Son zamanlarda döngüsel tutarlılık oluşturulurken yardımcı değişken (aux) ekleyerek asimetrik alanlar arasında eşleşmemiş görüntüden görüntüye dönüşüm modellemek için asimetrik çerçeve kullanan AsymGAN (Li ve ark., 2019) önerilmiştir. Aux, verimsiz ve verimli alanlar arasındaki ekstra bilgileri öğrenmek için kullanılır. Farklı alanlar ve dengesiz bilgi arasındaki karmaşıklığı içeren zorluklar bu değişken kullanılarak çözümler ve daha iyi dengelenir. AsymGAN, farklı aux değişkenlerini kullanarak çeşitli ve daha yüksek kaliteli çıktılar üretebilmektedir. CycleGAN ve denetimsiz piksel düzeyinde görüntüden görüntüye dönüşüm odaklanan ilgili yaklaşımların aksine, yüksek kaliteli ve çeşitli görüntüler ürettiği kanıtlanmıştır.

Chen ve ark. (2019) denetimsiz görüntüden görüntüye dönüşüm için kalite kaybına dayanan birleşik bir kalite duyarlı GAN (QGAN) çerçevesi önermektedir. QGAN tasarımı, kaliteye duyarlı kaybın - klasik kalite değerlendirme kaybı ve rakip kayba ek olarak uyarlanabilir algısal kalite farkındalık kaybı olmak üzere iki ayrıntılı uygulamasını içerir.

XGAN (Royer ve ark., 2020), semantik tutarlılık kaybına dayalı olarak denetimsiz görüntüden görüntüye dönüşüm için önerilen ikili bir düşman otomatik kodlayıcıdır. XGAN, piksel seviyesindeki bilgilerden ziyade ortak özellik seviyesindeki bilgileri öğrenmek için her iki alanın paylaşılan özellik temsili yakalar. Etki alanları arasında görüntünün anlamsal içeriğini korumak için her iki alanın dönüşümünde de anlamsal tutarlılık kaybı kullanılır.

SPA-GAN (Emami ve ark., 2020), denetimsiz görüntüden görüntüye dönüşüm işlevleri için etkili bir uzamsal dikkat mekanizması sunmaktadır. SPA-GAN, ayırıcıdaki uzamsal dikkat haritalarını hesaplar ve doğrudan üretici ağı geri besler. Bu, üretici ağı görüntüden görüntüye dönüşümde en çok ayırt edici alanlara odaklanmaya zorlar. SPA-GAN ayrıca, gerçek

ve sahte görüntülerin özellik haritalarının farkı olarak tanımlanan özellik haritası kaybı terimi sunmaktadır ve üretici ağları daha gerçekçi çıktı görüntülerine yol açan görüntü dönüşümsü sırasında alana özgü özellikleri korumaya teşvik eder. SPA-GAN'da, giriş görüntülerini sınıflandırmanın yanı sıra ayrıştırıcı, üretici ağla beslenen ve en ayırt edici nesne parçalarına odaklanmasına yardımcı olan uzamsal dikkat haritaları da üretir. Özellik eşleme kaybı, görüntü dönüşümünde alana özgü özellikleri korumak için kullanılır.

U-GAT-IT (Kim ve ark., 2019) sabit bir ağ mimarisi ve hiper parametre ile çeşitli veri setlerinde görsel olarak daha güzel sonuçlar üreten dikkat modülü ve AdaLIN (Uyarlanabilir Katman-Örnek Normalleştirme) ile denetimsiz görüntüden görüntüye çeviri yöntemidir. AdaLIN işlevi, veri setlerine bağlı olarak öğrenilen parametrelerle şekil ve dokudaki değişiklik miktarını esnek bir şekilde kontrol etmek için dikkat yönlendirmeli modeline yardımcı olur. Uyarlanabilir Katman Örneği Normalizasyonunun (AdaLIN), farklı miktarda geometri ve stil değişiklikleri içeren çeşitli veri kümelerini çevirmek için gerekli olduğu bulunmuştur.

Mo ve ark. (2018), temelini CycleGAN'dan alan InstaGAN'ı önermişlerdir. InstaGAN, önceki yöntemlerden farklı olarak birden fazla hedef görüntüye ulaşabilmeyi sağlayan görüntüden görüntüye dönüşüm yöntemi önermişlerdir. Örneğin moda görüntülerinde kottan eteğe çevirme gibi görüntüden görüntüye dönüşümde başarı göstermiştir. Her ağ, hem bir görüntüyü hem de bir dizi örnek maskesini kodlamak için tasarlanmıştır. G permütasyon eşdeğeridir ve D set sırasına göre permütasyon değişmezdir. Özellikleri elde etmek için, değişmezlik için tüm set öğelerinin özelliklerini toplar ve sonra onu eşdeğerlik için kimlik eşlemesiyle birleştirilir. Karşılaştırma yaparken CycleGAN + Seg denilen maske ile artırılmış görüntüyü dönüştüren ek bir giriş kanalı ile bir CycleGAN modelini eğitmektedir.

b. Otomatik kodlayıcı tabanlı modellerle denetimsiz dönüşüm

Otomatik kodlayıcı, hem bir kodlayıcıdan hem de giriş görüntülerini gizli gösterime dönüştüren bir kod çözücüdür oluşur. Bu sıkıştırılmış vektör, yüksek kaliteli görüntüler oluşturmak için üretici ağ beslenir.

UNIT (Liu ve ark., 2017), denetimsiz görüntüden görüntüye dönüşüm için CoGAN ve VAE'lerin birleşim çerçevesidir. Paylaşılan gizli alan varsayımına dayanmaktadır. Ağırlık paylaşımı kısıtlaması, paylaşılan gizli alanı iki alanda karşılık gelen görüntüleri oluşturmaya zorlamak için kullanılır. UNIT çerçevesi iki kodlayıcı, iki üretici ve iki ayrıştırıcıdan oluşur. Bununla birlikte, UNIT iyi performans göstermesi için iki alanın benzer modellere sahip olmasını gerektirir. Birleştirilmiş üretici rakip ağ (CoGAN)(Liu ve ark., 2016) yönteminin yazarları, çok alanlı görüntülerin denetimsiz bir şekilde ortak dağıtımını öğrenmek için bir çift GAN kullanmayı önermiştir. CoGAN, gerekli alan adlarını eşleştirmek için aynı yapıya sahip GAN çiftlerini kullanır. Her bir GAN çifti, ilk birkaç katmanın ve son birkaç katmanın ağırlıklarını paylaşmaya zorlanır, bu da CoGAN'ın marjinal alan dağılımlarından ayrı olarak alınan örneklerden çoklu alanların ortak dağılımlarını öğrenmesini sağlar.

Daha yakın zamanlarda, Liu ve ark. (2020), temsil içeriğini etki alanı özneliklerinden ayıran denetimsiz görüntüden görüntüye dönüşüm yöntemi önermiştir. Gizli alan özneliği bir Gauss karışım modeli (GMM) ile modellendiğinden GMM-UNIT modeli olarak adlandırılmıştır. Bu modelde, karışımdaki her Gauss bileşeni bir alan ile ilişkilendirilmiştir. GMM-UNIT'in iki temel avantajı vardır. Birincisi, çok modellenli ve çok alanlı dönüşüme izin verir. İkincisi, alanlar arasında enterpolasyona ve görünmeyen alanlara ekstrapolasyona ve dönüşüme izin verir. DIRT ++ 'dan farklı olarak GMM-UNIT, sıfır veya az vuruşlu görüntüler oluşturmaya izin veren sürekli bir etki alanı kodlaması önerisini gerektirmektedir.

Zhou ve ark. (2019), iki alan için tek kodlayıcı-çift kod çözücü mimarisine dayanan denetimsiz bir karşılıklı görüntüden görüntüye dönüşüm modeli BranchGAN'ı önerdi. BranchGAN, aynı kodlayıcıyla iki alanın paylaşılan dağıtımını kullanarak bir görüntüyü bir etki alanından başka bir etki alanına aktarır. Modeli, iki görüntü alanının ortak dağılımını öğrenmek üzere eğitmek için yeniden yapılandırma kaybı, kodlama kaybı ve rakip kaybı kullanır.

c. Çözülmüş temsil ile denetimsiz dönüşüm

Son zamanlarda denetimsiz görüntüden görüntüye dönüşüm yaklaşımları, bir içerik kodlayıcı ve bir stil kodlayıcı (Guo ve ark.,2019) ile veri varyasyonlarının faktörlerini modelleyen çözülmüş gösterimi öğrenir. Huang ve ark. (2018), görüntülerin gizli uzayının ikiye ayrıştırılabileceğini varsayan, denetimsiz çok modlu bir görüntüden görüntüye dönüşüm (MUNIT) çerçevesi önerdi. MUNIT modelinde bir içerik alanı ve bir stil alanı mevcuttur. MUNIT, iki otomatik kodlayıcıdan oluşur ve her bir otomatik kodlayıcının gizli kodu, içerik kodu ve stil koduyla çarpanlara ayrılmıştır. İçerik kodu, dönüşüm sırasında korunması gereken temel uzamsal yapıyı kodlarken; stil kodu, giriş görüntüsünde yer almayan yapının oluşturulmasını temsil eder. Görüntüyü hedef alana dönüştürmek, çeşitli ve çok modlu çıktılar üretmek için içerik kodu, hedef stil uzayındaki stil uzayından örneklenen farklı bir stil koduyla yeniden birleştirilir.

Model, her alan için bir tane olmak üzere iki otomatik kodlayıcıdan (sırasıyla kırmızı ve mavi oklarla gösterilir) oluşur. Her bir otomatik kodlayıcının gizli kodu, bir içerik kodu c ve bir stil kodundan oluşur. Modeli, hem görüntüleri hem de gizli kodları yeniden yapılandıran iki yönlü yeniden yapılandırma hedeflerinin (kesikli çizgiler) yanı sıra, dönüştürülen görüntülerin hedef alandaki gerçek görüntülerden ayırt edilemez olmasını sağlayan rakip hedeflerle (noktalı çizgiler) eğitirler. Karşılıklı kayıp ve çift yönlü yeniden yapılandırma kaybı terimlerinin ağırlıklı toplamı olan nihai hedefi optimize etmek için kodlayıcıları, kod çözücülerini ve ayırıcıları birlikte eğitirler.

Lee ve ark. (2018) , eşleştirilmiş eğitim görüntüleri olmadan çeşitli çıktılar üretmek için çözülmüş bir temsil çerçevesine dayanan çeşitli görüntüden görüntüye dönüşüm (DIRT) yöntemini önerdi. DIRT, gizli alanı paylaşılan bir içerik alanına ve alana özgü öznelik alanlarına ayırır. Ağırlık paylaşımı ve içerik ayırıcı stratejisi, yeni bir çapraz döngü tutarlılık kaybı ve içerik rekabeti kaybı uygulayarak içeriğin ve özneliklerin temsillerini çözmek için kullanılır. DIRT yöntemi sonrasında iki adım uygulayarak DRIT++ 'ya

genişletilmiştir. Bu iki adım: Örnek çeşitliliğini iyileştirmeye yardımcı olan mod çökme problemini hafifletmek için mod arayan düzenlilik terimi dâhil etmek ve ikincisi, çok alanlı görüntüden görüntüye dönüşmek için iki alanlı modeli dönüştürme sorunlarını genelleştirmek (Lee ve ark., 2020) .

Liu ve ark. (2018), birden fazla etki alanındaki verilerden etki alanı değişmeyen gösterimi öğrenen ve sürekli etki alanları arası görüntü dönüşümü ve manipülasyonu gerçekleştirebilen Birleşik Özellik Çözme Ağı (UFDN) adlı kompakt bir model önerdi. Kodlayıcı E, üretici G, piksel alanında Dx bir ayrıştırıcı ve özellik alanı Dv'de bir ayrıştırıcıdan oluşan Birleşik Özellik Çözme Ağına (UFDN)'dır.

Lin ve ark. (2019) etki alanı denetimli GAN (DosGAN) adı verilen denetimsiz görüntüden görüntüye dönüşüm çerçevesi önermiştir. Alana özgü bir özellik çıkarıcı, alandan bağımsız bir özellik çıkarıcı ve daha iyi alana özgü özellikleri ayıklamak ve görüntüleri dönüştürmek için bir görüntü üreticisinden oluşur. Alan bilgilerini, her alanın koşulsuz veya koşullu görüntüden görüntüye dönüşüm için belirli özelliklerini yakalamak üzere açık denetim olarak ele alır.

3. Araştırma Sonuçları ve Tartışma

Bilgisayarla görme uygulamalarında görüntü sentezleme yapılırken GAN'lar büyük başarı elde ettiği görülmüştür (Isola ve ark., 2017; Wang ve ark., 2018; Choi ve ark., 2018; Almahairi ve ark., 2018; Zhu ve ark., 2017; Liu ve ark., 2017

). Bu çalışmada, yapıya dayalı olarak görüntüden görüntüye dönüşüm ile ilgili GAN yöntemlerine kapsamlı bir genel bakış sunmaktadır. Hem denetimli hem de denetimsiz güncel görüntüden görüntüye dönüşüm teknikleri incelenir ve sınıflandırılır. Yakın gelecekte, kuantum üretici çekişmeli ağlar görüntü oluşturmayla ilgili karmaşık sorunların üstesinden gelmek için görüntüden görüntüye dönüşüm yöntemi olarak araştırılacak ve uygulanacaktır. GAN'lar son zamanlarda alanlar arası görüntüden görüntüye dönüşümde çok ilerleme kaydetmiştir. Denetimsiz öğrenme algoritmaları denetimli yaklaşımlarla karşılaştırılabilir performansla sahip olabilir. Eşleştirilmiş eğitim verilerinin edinilmesi pahalı olduğundan denetimsiz öğrenme yöntemine eğilim daha yüksek olacaktır. Açıklanan modelleri tek ve çok modlu algoritmalara da ayırabiliriz. UNIT, CycleGAN ve cGAN tek, CE-GAN, MUNIT, SPA-GAN, DualGAN ve BicycleGAN ise çok modlu algoritmalara aittir. Çok modlu yaklaşımlar, modelin çökmesi sorununu hafifletebilir ve aynı anda hem çeşitli hem de gerçekçi görüntüler üretebilir. Bu nedenle, çok modlu ve denetimsiz yaklaşımlar, gelecekte görüntüden görüntüye çevirinin ana akımı olabilir. Tablo 1'de üretici çekişmeli ağlarda görüntüden görüntüye dönüşüm için güncel yöntemlerin karşılaştırılmasıyla ilgili inceleme yapıp bilgi verilmiştir.

Tablo 1. Görüntüden görüntüye dönüşüm için güncel yöntemlerin karşılaştırılması.

Yöntem	Çift Yönlü Dönüşüm	Paylaşımlı Yapı	Birleşik Yapı	Çok-Alanlı Dönüşüm	Çok-Modlu Dönüşüm	Eşleşmemiş Görüntü
Pix2Pix ((Isola ve ark., 2017)	-	-	-	-	-	-
BicycleGAN ((Zhu ve ark., 2017)	✓	✓	-	-	✓	-
StarGAN (Choi ve ark., 2018)	✓	-	✓	✓	-	✓
CycleGAN (Zhu ve ark., 2017)	✓	-	-	-	-	✓
UNIT (Liu ve ark., 2017)	✓	✓	-	-	-	✓
GMM-UNIT (Liu ve ark., 2020)	✓	-	✓	✓	✓	✓
MUNIT(Guo ve ark.,2019)	✓	✓	-	-	✓	✓
DIRT++ (Lee ve ark., 2020)	✓	✓	-	✓	✓	✓
UFDN (Liu ve ark., 2018)	✓	✓	✓	✓	-	✓

4.Sonuç

Görüntü sentezlemek için ya da görüntüden görüntüye dönüşümde bilgisayar görmesi alanında GAN'lar son yıllarda büyük başarı elde etmişlerdir. Bu çalışmada, yapıya dayalı olarak görüntüden görüntüye dönüşüm ile ilgili GAN değişkenlerine kapsamlı bir genel bakış sunulmaktadır. Hem denetimli hem de denetimsiz son teknoloji görüntüden

görüntüye dönüşüm teknikleri incelenip ve sınıflandırılmaktadır. Gelecekte, görüntü oluşturmayla ilgili karmaşık sorunların üstesinden gelmek için görüntüden görüntüye dönüşüm için kuantum üretken karşıtlık ağı daha da araştırılacak ve uygulanacaktır.

Kaynakça

- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
- Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8798-8807).
- Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).
- Wu, X., Xu, K., & Hall, P. (2017). A survey of image synthesis and editing with generative adversarial networks. *Tsinghua Science and Technology*, 22(6), 660-674.
- Azadi, S., Tschannen, M., Tzeng, E., Gelly, S., Darrell, T., & Lucic, M. (2019). Semantic bottleneck scene generation. arXiv preprint arXiv:1911.11357.
- Ntavelis, E., Romero, A., Kastanis, I., Van Gool, L., & Timofte, R. (2020, August). Sesame: Semantic editing of scenes by adding, manipulating or erasing objects. In European Conference on Computer Vision (pp. 394-411). Springer, Cham.
- Zhu, J. Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., & Shechtman, E. (2017). Toward multimodal image-to-image translation. arXiv preprint arXiv:1711.11586.
- Xiong, F., Wang, Q., & Gao, Q. (2019). Consistent embedded GAN for image-to-image translation. *IEEE Access*, 7, 126651-126661.
- Tripathy, S., Kannala, J., & Rahtu, E. (2018, December). Learning image-to-image translation using paired and unpaired training samples. In Asian Conference on Computer Vision (pp. 51-66). Springer, Cham.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
- Almahairi, A., Rajeshwar, S., Sordoni, A., Bachman, P., & Courville, A. (2018, July). Augmented cyclegan: Learning many-to-many mappings from unpaired data. In International Conference on Machine Learning (pp. 195-204). PMLR.
- Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017, July). Learning to discover cross-domain relations with generative adversarial networks. In International Conference on Machine Learning (pp. 1857-1865). PMLR.
- Yi, Z., Zhang, H., Tan, P., & Gong, M. (2017). Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE international conference on computer vision (pp. 2849-2857).
- Li, Y., Tang, S., Zhang, R., Zhang, Y., Li, J., & Yan, S. (2019). Asymmetric GAN for unpaired image-to-image translation. *IEEE Transactions on Image Processing*, 28(12), 5881-5896.
- Chen, L., Wu, L., Hu, Z., & Wang, M. (2019). Quality-aware unpaired image-to-image translation. *IEEE Transactions on Multimedia*, 21(10), 2664-2674.
- Royer, A., Bousmalis, K., Gouws, S., Bertsch, F., Mosseri, I., Cole, F., & Murphy, K. (2020). Xgan: Unsupervised image-to-image translation for many-to-many mappings. In Domain Adaptation for Visual Understanding (pp. 33-49). Springer, Cham.
- Emami, H., Aliabadi, M. M., Dong, M., & Chinnam, R. B. (2020). Spa-gan: Spatial attention gan for image-to-image translation. *IEEE Transactions on Multimedia*, 23, 391-401.
- Kim, J., Kim, M., Kang, H., & Lee, K. (2019). U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. arXiv preprint arXiv:1907.10830.
- Mo, S., Cho, M., & Shin, J. (2018). Instagan: Instance-aware image-to-image translation. arXiv preprint arXiv:1812.10889.
- Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. arXiv preprint arXiv:1703.00848.
- Liu, M. Y., & Tuzel, O. (2016). Coupled generative adversarial networks. arXiv preprint arXiv:1606.07536.
- Liu, Y., De Nadai, M., Yao, J., Sebe, N., Lepri, B., & Alameda-Pineda, X. (2020). GMM-UNIT: Unsupervised Multi-Domain and Multi-Modal Image-to-Image Translation via Attribute Gaussian Mixture Modeling. arXiv preprint arXiv:2003.06788.
- Zhou, Y. F., Jiang, R. H., Wu, X., He, J. Y., Weng, S., & Peng, Q. (2019). Branchgan: Unsupervised mutual image-to-image transfer with a single encoder and dual decoders. *IEEE Transactions on Multimedia*, 21(12), 3136-3149.
- Guo, W., Wang, J., & Wang, S. (2019). Deep multimodal representation learning: A survey. *IEEE Access*, 7, 63373-63394.
- Huang, X., Liu, M. Y., Belongie, S., & Kautz, J. (2018). Multimodal unsupervised image-to-image translation.

In Proceedings of the European conference on computer vision (ECCV) (pp. 172-189).

Lee, H. Y., Tseng, H. Y., Huang, J. B., Singh, M., & Yang, M. H. (2018). Diverse image-to-image translation via disentangled representations. In Proceedings of the European conference on computer vision (ECCV) (pp. 35-51).

Lee, H.-Y.; Tseng, H.-Y.; Mao, Q.; Huang, J.-B.; Lu, Y.-D.; Singh, M.; Yang, M.-H. Drit++: Diverse image-to-image translation via disentangled representations. *Int. J. Comput. Vis.* 2020, 1–16.

Liu, A. H., Liu, Y. C., Yeh, Y. Y., & Wang, Y. C. F. (2018). A unified feature disentangler for multi-domain image translation and manipulation. arXiv preprint arXiv:1809.01361.

Lin, J., Chen, Z., Xia, Y., Liu, S., Qin, T., & Luo, J. (2019). Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation. *IEEE transactions on pattern analysis and machine intelligence.*