Available online at www.jpit.az13 (1)
2022

Analysis of generative adversarial networks

Yadigar N. Imamverdiyev ^a, Firangiz I. Musayeva ^b

^aAzerbaijan Technical University, Institute of Artificial Intelligence, H. Javid, ave. 25, AZ1148 Baku, Azerbaijan

^bInstitute of Information Technology, Azerbaijan National Academy of Sciences, B. Vahabzade str., 9A, AZ1141 Baku, Azerbaijan

^a yadigar@iit.science.az; ^b sadiyeva.firangiz@gmail.com

ARTICLE INFO

<http://doi.org/10.25045/jpit.v13.i1.03>

Article history:

Received 2 September 2021

Received in revised form 18 November 2021

Accepted 21 January 2022

Keywords:

Neural networks

Generative models

Generative adversarial networks

Auto encoders

Generator

Discriminator

Generativ rəqib şəbəkələrin analizi

Açar sözlər:

Neyron şəbəkələr

Generativ modellər

Generativ rəqib şəbəkələr

Avtoenkoder

Generator

Diskriminator

Анализ генеративных состязательных сетей

Ключевые слова:

Нейронные сети

Генеративные модели

Генеративные конкурирующие сети

Автокодеры

Генератор

Дискриминатор

ABSTRACT

Recently, a lot of research has been done on the use of generative models in the field of computer vision and image classification. At the same time, effective work has been done with the help of an environment called generative adversarial networks, such as video generation, music generation, image synthesis, text-to-image conversion. Generative adversarial networks are artificial intelligence algorithms designed to solve the problems of generative models. The purpose of the generative model is to study the set of training patterns and their probable distribution. The article discusses generative adversarial networks, their types, problems, and advantages, as well as classification and regression, segmentation of medical images, music generation, best description capabilities, text image conversion, video generation, etc. general information is given. In addition, comparisons were made between the generative adversarial network algorithms analyzed on some criteria.

Son zamanlar kompüterlə görmə və təsvirin təsnifatı sahəsində generativ modellərin istifadəsi ilə bağlı xeyli tədqiqat aparılmışdır. Eyni zamanda, video generasiya, musiqi generasiyası, təsvirin sintezi, mətndən şəkil çevrilməsi kimi generativ rəqib şəbəkələr adlanan mühitin köməyi ilə effektiv işlər görülmüşdür. Generativ rəqib şəbəkələr generativ modellərin problemlərini həll etmək üçün nəzərdə tutulmuş süni intellekt alqoritmləridir. Generativ modelin məqsədi, təlim nümunələri toplusunu və onların ehtimal paylanmasını öyrənməkdir. Məqalədə generativ rəqib şəbəkələr, onların növləri, problemləri və üstünlükləri, həmçinin təsnifat və reqressiya, tibbi təsvirlərin seqmentasiyası, musiqi generasiyası, ən yaxşı təsvir imkanları, mətn təsvirinin çevrilməsi, video generasiya və s. haqqında ümumi məlumat verilir. Bundan əlavə, analiz edilən generativ rəqib şəbəkə alqoritmləri arasında bəzi meyarlar üzrə müqayisələr aparılmışdır.

В последнее время было проведено много исследований по использованию генеративных моделей в области компьютерного зрения и классификации изображений. В то же время была проведена эффективная работа с помощью среды, называемой генеративно-состязательными сетями, такими как генерация видео, генерация музыки, синтез изображений, преобразование текста в изображение. Генеративно-состязательные сети — это алгоритмы искусственного интеллекта, предназначенные для решения задач генеративных моделей. Целью генеративной модели является изучение набора тренировочных паттернов и их вероятного распределения. В статье обсуждаются генеративно-состязательные сети, их виды, проблемы и преимущества, а также классификация и регрессия, сегментация медицинских изображений, генерация музыки, возможности наилучшего описания, преобразование текстового изображения, генерация видео и т. д. Приводится общая информация. Кроме того, были проведены сравнения алгоритмов генеративно-состязательной сети, проанализированных по некоторым критериям.

1. Giriş

Süni intellektin inkişafına dair mövcud yanaşmaların əksəriyyəti əsasən maşın təliminə əsaslanır. Bu günə kimi ən çox istifadə olunan və uğurlu maşın təlimi növü supervizorlu təlimdir. Süni intellekt sahəsində son illərdə çox populyar olan generativ rəqib şəbəkələr (GRŞ, Generative Adversarial Network, GAN) – dərin neyron şəbəkələrinin bir sinfidir. Bu şəbəkələr *Ian Goodfellow* və həmkarları tərəfindən 2014-cü ildə təklif olunub (*Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, ... & Bengio, 2020*) və çox qısa müddətdə müxtəlif tətbiq sahələrində – video, şəkil, musiqi və sintetik verilənlərin generasiyası və s. sahələrdə maraqlı nəticələr təqdim edilmişdir. GRŞ-lər kompüter görmə, təbii dilin emalı, zaman sıralarının sintezi və semantik seqmentasiya kimi müxtəlif sahələrdə tətbiq edilir. GRŞ-lər maşın təlimindəki generativ modellər ailəsinə aiddir. Variasiyalı avtoenkoderlər kimi digər generativ modellərlə müqayisədə GRŞ-lər dəqiq proqnoz edilən sıxlıq funksiyalarını idarə etmək, istənilən nümunələri səmərəli şəkildə generasiya etmək kimi üstünlüklər təklif edirlər (*Warde-Farley, Bengio, 2017*).

GRŞ-lər bir-biri ilə rəqabətdə olan iki şəbəkənin təlimi ilə xarakterizə edilir. Bu şəbəkələr ilə bağlı bəzi metaforalar var: biri rəsm əsəri oğrusu və digəri isə onu müəyyən edən ekspert olaraq düşünülür. GRŞ-lərdə generator (G) kimi tanınan rəsm əsəri oğrusu real obrazlar yaratmaq məqsədilə saxta nümunələr generasiya edir, diskriminator (D) kimi tanınan ekspert isə həm saxta nümunələri, həm də real (orijinal) şəkilləri alır və onları bir-birindən ayırmağa çalışır. Hər ikisi eyni zamanda və bir-biriləri ilə rəqabət edərək öyrədilir. Ən əsası isə generatorun real şəkillərə birbaşa çıxışı yoxdur. Diskriminator həm saxta nümunələrə, həm də real şəkillər toplusundan götürülmüş nümunələrə giriş imkanı əldə edir. Diskriminatorun köməyiylə eyni səhv signalı generatorun daha keyfiyyətli və reala daha bənzər saxta nümunə generasiya edə bilməsi üçün istifadə edilir (*Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, ... & Bengio, 2020*).

Bu məqalədə öncə generativ rəqib şəbəkələrin nəzəri əsasları və strukturu təqdim edilmiş, daha sonra onların problemləri, üstünlükləri, öyrənmə alqoritmləri və bir çox sahələrə tətbiqlərinin müqayisəli analizi aparılmışdır.

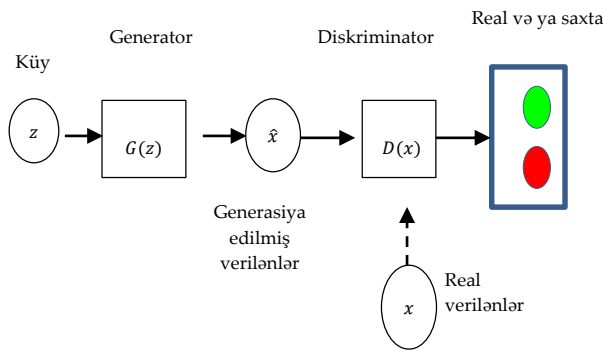
2. GRŞ nəzəriyyəsi

GRŞ-lərin ilk tətbiqindən indiyə kimi şəkil generasiyası və digər sahələrdə tətbiqi üçün bir neçə maraqlı tədqiqatlar yerinə yetirilib. Ümumi olaraq desək, generativ modellər iki kateqoriyaya ayrılır: məhdud Bolsman maşınları (Restricted Boltzmann Machine, RBM), Naive Bayes modeli (NBM) və gizli markov modelindən (Hidden Markov Model, HMM) ibarət maşın təliminə əsaslanan ənənəvi generativ modellər (*Donahue, Kr'ahenb'uhl and Darrell, 2016*). Digəri isə, variasiyalı avtoenkoder (Variational AutoEncoder, VAE), GRŞ-lər və törəmə modellərindən ibarət təlim modelləridir. GRŞ gizli dəyişənlərin köməyiylə məqsəd verilənləri generasiya edən generativ modeldir. Xüsusi olaraq, modeldəki generator və diskriminator arasında oyun təlimi aparılır və real verilənlərin paylaşılmasından ibarət olan hədəf dəyişənləri təsadüfi dəyişənlər tərəfindən generasiya edilir. Ənənəvi maşın təlimi alqoritmləri ilə müqayisədə model funksional və daha çox tətbiq ssenarisinə malikdir. Bunlar, adətən, ImageNet CIFAR kimi böyük verilənlər toplusunda daha yaxşı nəticə göstərirlər. GRŞ-lərin inkişaf prosesi təxminən üç mərhələdə izah edilə bilər. Birinci mərhələ 2014-2015-ci illərdə GRŞ modellərinin yaradılmasından dərin konvolusiyalı GRŞ (Deep Convolutional Generative Adversarial Networks, DCGAN) modelinin inkişafına qədərki dövr, ikinci mərhələ 2015 - 2017-ci illərdə DCGAN modelinin tətbiqindən Vasserşteyn GRŞ (Wasserstein GAN, WGAN) modelinin inkişafına qədərki dövr və üçüncü mərhələ isə 2017 - ci ildən bu günə kimi WGAN modelinin inkişafından sonrakı dövr hesab edilir. DCGAN və WGAN modelləri hər mərhələnin əsasını təşkil edir. Əvvəlki GRŞ-lər ilə müqayisədə DCGAN daha asan tətbiq edilə bilər və rejim pozulması problemini aradan qaldırır. WGAN-ın meydana gəlməsi GRŞ modelləri üçün mühüm əhəmiyyət kəsb etməyə başladı. Bu model ən yüksək keyfiyyətli nümunələri generasiya etmək bacarığına malikdir. Müxtəlif nümunə və ssenarilərin hər birinin müxtəlif tələblərə görə fərqli GRŞ modelləri bir-birinin ardınca ortaya çıxdı. Bundan başqa, tibb, təhlükəsizlik kimi sahələrdə də yaxşı nəticələr əldə edilmişdir. BigGAN ilk olaraq yüksək dəqiqlik və aşağı çeşidli şəkillər generasiya edir. StyleGAN GRŞ

sahəsində tamamilə fərqli inkişaf etmiş və insan üzünün təsvirlərinin generasiya edilməsi istiqamətində çox yaxşı nəticələr əldə etmişdir. Bundan əlavə, bu modelin köməyiylə digər yüksək keyfiyyətli təsvirlərin də generasiyası mümkündür. Üz tanınması üçün AgecGAN kimi alqoritmlərin istifadəsi daha məqsədəuyğundur (Dumoulin, Belghazi, Poole, Mastropietro, Lamb, Arjovsky, & Courville, 2016, Creswell, White, Dumoulin, Arulkumaran, Sengupta, & Bharath, 2018).

2.1. GRŞ-nin strukturu

Generativ rəqib şəbəkələrin ideyası olduqca sadədir: iki neyron şəbəkəsi (generator və diskriminator) götürülür və təlim prosesində onlar öz aralarında “yarışirlar” – generator verilənlərin statistik modelini qurur və onun əsasında reallığa bənzər obyektlər yaradır, diskriminator isə onların həqiqi və ya saxta olduğunu müəyyən edir (şəkil 1).



Şəkil 1. Genarativ rəqib şəbəkələrin ümumi sxemi

Generator təsadüfi küydən istifadə edərək real verilənlərin paylanması ilə nümunələr generasiya edir. Diskriminator isə real verilənlər ilə saxta verilənləri bir-birindən ayırır. Bunlara əsaslanaraq, GRŞ modeli üçün iki oyunçu oyunu reallaşdırılır. Oyunun təlimindən modelin ümumiləşdirmə qabiliyyətini yaxşılaşdırmaqla iki şəbəkə arasındakı çəki parametrlərini optimallaşdırmaq üçün istifadə edilir. Nəticədə, generator tərəfindən generasiya edilən saxta nümunələrin paylanması real nümunələrin paylanmasına uyğunlaşdırılır. Modelin tarazlıq vəziyyəti diskriminatorun saxta nümunələri real nümunələrdən fərqləndirə bilmədiyi vəziyyətdir. Yuxarıda göstərilmiş GRŞ arxitekturasında (Şəkil 1.) $G(z)$ generatoru tərəfindən saxta nümunələr generasiya edilir.

Diskriminator iki növ giriş verilənlərini (real və saxta nümunələr) qəbul edir, onların real olub-olmadığını müəyyən edir. GRŞ-lərin itki funksiyası sıfır cəmi oyun mühitində bir-biri ilə rəqabətdə olan şəbəkədən ibarət iki oyunçunun min-max oyununa əsaslanır. Diskriminatorun generatordan çıxan verilənləri real verilənlərdən ayırması və geriye yayma alqoritmi ilə şəbəkə modelinin çəkirlərini optimallaşdırması lazımdır. Diskriminatorun giriş parametrləri x və $\theta(D)$, itki funksiyası isə aşağıdakı kimidir (Yang, Yan, Zhang, Yu, Shi, Mou, & Wang, 2018).

$$V(D, \theta^{(D)}) = -E_{x \sim p_r(x)}[\log D(x)] - E_{z \sim p_g(z)}[\log(1 - D(G(z)))]. \quad (1)$$

Burada, $p_r(x)$ – real verilənlərin, $p_g(z)$ – daxil edilən küyün paylanmasıdır (adətən, təsadüfi Qauss paylanması götürülür), $D(x)$ – diskriminator funksiyası $G(z)$ – generator funksiyasını göstərir. Generatorun itki funksiyası düstur (2) - də göstərilir.

$$V(G, \theta^{(G)}) = -E_{z \sim p_g}[-\log D(G(z))]. \quad (2)$$

“Rəqib” sözü bu iki şəbəkə arasında gedən rəqabəti simvolizə edir, çünki bu şəbəkələrin məqsədləri bir-birinə ziddir və onların münasibətini antaqonist oyun kimi xarakterizə etmək mümkündür. Generator və diskriminator aşağıdakı min-max oyununda bir-birinə qarşı “rəqabət aparırlar”:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_r(x)}[\log D(x)] + \mathbb{E}_{z \sim p_g(z)}[\log(1 - D(G(z)))]. \quad (3)$$

Diskriminator $D(G(z))$ –in ehtimalını 0-a yaxınlaşdırmaq istəyir, yəni $(1 - D(G(z)))$ –i maksimallaşdırmağa cəhd edir. Eyni zamanda, Generator $D(G(z))$ –i 1-ə yaxınlaşdırmağa çalışır ki, diskriminator səhv etsin və generasiya edilmiş nümunəni həqiqi qəbul etsin (düstur (3)). Yəni, generator $(1 - D(G(z)))$ –i minimallaşdırmaq istəyir. Rəqabətli birgə təlim prosesində sistem Neş tarazlığına çata bilər ki, bu zaman şəbəkədə generasiya olunan nümunələri təbii nümunələrdən fərqləndirmək mümkün olmur. Baxılan min-max oyununun ixtiyari G və D funksiyaları fəzasında yeganə həlli vardır. $p_r(x) = p_g(z)$, $\forall x$ təlim verilənlərinin

paylanması bərpa edir və ixtiyari x üçün $D = 1/2$ olur.

GRŞ təlimi zamanı generator və diskriminator növbə ilə (öncə D diskriminatoru, sonra isə G generatoru) öyrədilir. Şəbəkələrdən biri öyrədildikdə, digəri sabit saxlanılır. Nəzəri olaraq $V(G, D)$ -də ən uyğun həll əldə etmək üçün öncə k dəfə diskriminator öyrədilməli və daha sonra generator bir dəfə öyrədilməlidir. Lakin praktikada $k = 1$ götürülür. $V(G, D)$ –nin maksimum qiymət alması üçün onun törəməsindən sonra aşağıdakı $D^*(x)$ alınır.

$$D^*(x) = \frac{p_r(x)}{p_r(x) + p_g(x)}. \quad (4)$$

$p_r(x)$ və $p_g(z)$ ehtimallı paylanmaları arasındakı KL (Kullback-Leibler) fərqliliyi yuxarıdakı məqsəd funksiyasına tətbiq etməklə əldə edilə bilər. Bu da modelin təlimini daha yaxşı izah edə bilir (Lucic, Kurach, Michalski, Gelly, & Bousquet, 2017, Alguliyev, Abdullayeva, & Ojagverdiyeva, 2020).

2.2. GRŞ-lərin problemləri

GRŞ-lərin bir sıra problemləri var. Ən əsası, bu şəbəkələrin öyrədilməsi olduqca çətin və qeyri-stabil prosesdir, çünki həm onların hər birini öyrətmək, həm də onların tanıma dəqiqliyinin balansını təmin etmək lazımdır.

- 1) Rejimin pozulması (mode collapse – girişin istənilən qiymətlərində generator eyni bir obyekt verir) problemi GRŞ-nin effektivliyinin qiymətləndirilməsinə ümumi qəbul edilmiş yanaşmaların olmamasıdır. Generatorun bir-birinə bənzər nümunələr generasiyası zamanı qismən pozulma (partial collapse), ən pis halda isə yalnız bir nümunənin generasiyası baş verirsə, bu, tam pozulma (complete collapse) adlanır. Diskriminator tərəfindən real və saxta toplu generasiya olunan nümunələrin paylanması tarazlaşdırmaq üçün praktiki hiylələr və ya müxtəlif ehtimal paylama rejimlərini əhatə edən bir neçə GRŞ istifadə etməklə generatorun çeşidliliyini artırmaq olar (Bang, & Shim, 2018, July, Che, Li, Jacob, Bengio, & Li, 2016, Metz, Poole, Pfau, & Sohl-Dickstein, 2016).

- 2) Təlimin qeyri-sabitliyi (saddle points). GRŞ-də Hessian itki funksiyası sonsuzdur. Buna görə optimal həll yolu olaraq minimum nöqtədən çox, yəhərvari nöqtəsi tapmaqdır. Dərin təlimdə əksər optimizatorlar yalnız itki funksiyasının birinci dərəcədən törəməsindən asılıdır. GRŞ-lər üçün yəhər nöqtəsinə yaxınlaşmaq yaxşı başlanğıc hesab edilir (Arjovsky, & Bottou, 2017).
- 3) Generativ modellərin qiymətləndirilməsi. Generativ modellər tərəfindən generasiya edilmiş nümunələrin həqiqiliyini necə ölçmək olar? Ehtimal qiymətləndirilməsindən istifadə edilməlidir? və s. kimi suallar yalnız GRŞ-lər üçün deyil, ümumi olaraq ehtimal modelləri üçündür. (Theis, Oord, & Bethge, 2015)-də GRŞ-ləri fərqli metrikalardan istifadə etməklə qiymətləndirilməsi aparılır. Generasiya olunan nümunələrin keyfiyyəti haqqında müxtəlif nəticələr əldə etməyin yolları nəzərə çatdırılır.

Bu problemlərin həlli istiqamətində müxtəlif praktiki və nəzəri həllərin təklif edilməsi aktualdır. GRŞ-nin əsas problemlərindən biri də p_r və p_g ehtimal paylanmaları arasındakı məsafəni ölçmək və minimuma endirməkdir. Generator sabit saxlandıqda, diskriminator təlimin çarpaz entropiyası minimuma endirilir. Yuxarıda qeyd edilən GRŞ-nin strukturunda (4) və (3) düsturlarından aşağıdakı düsturu almaq olar.

$$V(D, \theta^{(G)}) = KL(p_r || \frac{p_r + p_g}{2}) + KL(p_g || \frac{p_r + p_g}{2}) - 2 \log 2. \quad (5)$$

Beləliklə, GRŞ-nin əsas problemi, hazırda, məsafə funksiyasını seçməkdir.

Generativ model verilənlərin paylanması approksimasiya etmək üçün qurulur. Tutaq ki, $x \sim p_{data}(x)$ və bu paylanmadan olan sonlu nümunə toplusu $X = \{x | x \sim p_r(x)\}, |X| = n$ verilib. Elə model qurmaq lazımdır ki, $p_{model}(x, \theta) \sim p_r(x)$ olsun. Generativ modelindən mövcud olanlara analoji olan yeni verilənlər generasiya etmək üçün istifadə etmək olar. Məsələn, generativ model vasitəsilə obyektlərin və sinflərin birgə paylanması qurmaq, sonra isə Bayes düsturundan istifadə edərək obyektin sinfə aid olması ehtimalını tapmaq olar.

3. Generativ rəqib şəbəkələrin əsas modelləri

Tədqiqatçılar GRŞ-nin arxitekturasında istifadə olunan neyron şəbəkə növünə görə müxtəlif formaları fərqləndirirlər:

A. Taməlaqəli GRŞ (Vanilla GAN) – təklif edilmiş ilk GRŞ arxitekturasıdır. Həm diskriminator, həm də generator çoxlaylı irəliyaşılma şəbəkəsidir (məsələn, çoxlaylı perseptron). Bu arxitektura MNIST (Modified National Institute of Standards and Technology database), CIFAR (Canadian Institute for Advanced Research) və Toronto verilənlər toplusu üçün tətbiq edilir. Praktikada generatoru optimallaşdırmaq üçün yox olan qradiyent problemi yarana bilər və G-nin öyrədilməsi üçün $D(G(z))$ ifadəsinin maksimum qiyməti tapılmalıdır.

B. Dərin konvolyusiya GRŞ – çoxlaylı konvolyusiya şəbəkələri istifadə edilir, şəkillərin sintezi məsələlərinə tətbiq edilir. Bu mühit iki şəbəkədən ibarətdir: biri şəbəkə generatoru adlanan konvolyusiya neyron şəbəkə (Convolutional Neural Network, CNN), digəri diskriminator adlanan de-CNN olaraq işləyir. Nöqsanı – modelin öyrədilməsi prosesinin xeyli uzun olmasıdır. CNN arxitekturasına daxil olan məhdudiyətlər aşağıdakılardır (Salimans, Goodfellow, Zaremba, Cheung, Radford, & Chen, 2016).

- Bütün səviyyəli birləşdirmə (pooling) təbəqələrini addımlama konvolyusiyaları ilə silir.
- G və D üçün Bolsman maşınları istifadə edilir.
- Generator və diskriminator şəbəkələrində uyğun olaraq ReLU (Rectified Linear Unit) və leaky-ReLU istifadə edilir.

C. Şərti GRŞ (Conditional GAN, CGAN) – generatoru giriş olaraq c (sinif nişanı, mətn və ya şəkil) və z gizli küy vektorunu alır. Beləliklə, $G(z/c)$ reala bənzər nümunələrin generasiyasıdır. Generator və diskriminator konkret sinfi göstərməklə seçmə toplusunun paylaşılmasını generasiya edə bilərlər. Şərti GRŞ-lərin itki funksiyası aşağıda göstərilmişdir.

$$L_{CGAN} = -E_{x \sim p_r(x)} \left[\log D \left(\frac{x}{c} \right) \right] - E_{z \sim p_z(z)} \left[\log \left(1 - D(G(z/c)) \right) \right]. \quad (6)$$

Burada, $p_z(z)$ küy və c şərti verilənlər generatorun, x real və c şərti verilənləri isə diskriminatorun girişləridir (Mirza, & Osindero, 2014, Isola, Zhu, Zhou, & Efros, 2017, Odena, Olah, & Shlens, 2017, July).

D. Dövri GRŞ (CycleGAN) – təsvirlərin generasiyası üçün ən qabaqcıl GRŞ-dir. Bu şəbəkələr domenlər arasında keçid üçün cüt verilənlər topluları tələb etmirlər, çünki belə verilənləri almaq çox çətindir. Buna baxmayaraq, dövri şəbəkələr iki müxtəlif X və Y domenlərindən olan verilənlərlə öyrədilməlidirlər (məsələn, X – atlar, Y – zebirlər). Bir domendən digərinə keçidi məhdudlaşdırmaq üçün “dövrün ardıcıl itirilməsi” mexanizmi istifadə edilir (Yeh, Chen, Yian Lim, Schwing, Hasegawa-Johnson, & Do, 2017).

E. Vasserşteyn GRŞ itki funksiyası Vasserşteyn məsafəsi daxil edilməklə dəyişdirilir. Nəticədə Vasserşteyn şəbəkəsinin itki funksiyası təsvirin keyfiyyəti ilə əlaqədardır. Bundan başqa, öyrənmənin dayanıqlılığı yaxşılaşır və arxitekturdan asılı olmur (Arjovsky, Chintala, & Bottou, 2017, July).

F. Təkmilləşən GRŞ (Progressive GAN, ProGAN) – təkmilləşdirilmiş Vasserşteyn şəbəkəsi əsasında yaradılıb, öyrədilmə zamanı yeni laylar tədricən əlavə olunur. Bu layların hər biri həm diskriminator, həm də generator üçün təsvirlərin çözümlünü artırır. Yüksək görüntülü təsvirlərin yaradılması böyük problemdir. Təsvir nə qədər böyükdürsə, şəbəkənin səhv etməsi o qədər asandır, çünki o daha mürəkkəb və incə detalları yaratmağı öyrənməlidir. ProGAN şəbəkəsində əvvəlcə kiçik miqyaslı laylar öyrənilir, sonra model diqqəti irimiqyaslı strukturların təmizlənməsinə cəmləşdirir (Karras, Aila, Laine, & Lehtinen, 2017).

G. StyleGAN – ProGAN-nın təsvirin generasiyası zamanı spesifik xüsusiyyətlərinin pozulması problemini həll etmək üçün üslub əsaslı generativ model olaraq tətbiq edilir. StyleGAN generator şəbəkəsinin arxitekturasını yenidən qurur, tərtib edilmiş keyfiyyətə ziyan vuran üslublara kiçik miqyasda dəyişikliklər etməklə təsvir sintezini düzgün idarə etməyə imkan verir. StyleGAN çox yüksək çözümlü (high resolution) təsvirlər generasiya etməyə qadir GRŞ-dir (məsələn, 1024×1024 ölçülü). İdeyası laylar

steki qurmaqdan ibarətdir. Başlanğıc laylar aşağı ölçülü təsvirlər generasiya edir (2*2-dən başlayaraq) və sonrakı laylar tədricən ölçünü artırır (Karras, Laine, & Aila, 2019).

GRŞ-nin video (MocoGAN, Pose-GAN, VGAN), təsvir (CycleGAN, DiscoGAN, PAN, Pix2pix), musiqi (C-RNN-GAN, ORGAN, SeqGAN) generasiyası, mətnin təsvirə çevrilməsi (StackGAN, TAC-GAN) və s. üçün müxtəlif arxitekturları təklif edilmişdir (Chen, Duan, Houthoof, Schulman, Sutskever, & Abbeel, 2016, December, Wang, Z., She, & Ward, 2021, Hitawala, 2018).

Bir çox müxtəlif generativ modellər var və əslində, GRŞ ilk generativ model deyil. GRŞ-nin əsas alternativləri variasiya avtoenkoderi, avtoregressiv modellər, axın modelləri və hibrid modellərdir. Onların hər birinin GRŞ ilə müqayisədə oxşar və fərqli, üstün və çatışmayan cəhətləri vardır (Nowozin, Cseke, & Tomioka, 2016, December).

4. GRŞ-lərin təlim alqoritmlərinin tətbiqi

GRŞ modelləri bəzi verilənlər üzərində öyrəndikdən sonra real görünən nümunələr yaratmaqda olduqca effektiv generativ modeldir. Bu üstünlüklər, GRŞ-lərin kompüter görmə (Computer Vision, CV) və süni intellektin (Artificial Intelligence, AI) müxtəlif sahələrində tətbiq olunmasına gətirib çıxarır. Burada təsvir, səs və video kimi müxtəlif sahələrdə müxtəlif GRŞ tətbiqlərini müzakirə edək.

4.1. Klassifikasiya və regressiya

Son illərdə konvolyusiyalı neyron şəbəkələri (Convolutional Neural Network, CNN) ilə öyrədilən təlim kompüter görmə tətbiqlərində geniş istifadə olunur və CNN-lərlə supervizorsuz öyrənməyə daha az diqqət yetirilir. (Radford, Metz, & Chintala, 2015)-də DCGAN adlanan CNN sinfi təqdim edilir, onun müəyyən arxitektur məhdudiyətləri var və onların supervizorsuz öyrənmə üçün güclü bir namizəd olduqları nəzərə çatdırılır. Müxtəlif təsvir verilənlər toplusu üzərində öyrədilən dərin konvolyusiya rəqib şəbəkələr cütünün həm generatorunda, həm də diskriminatorunda obyektin kiçik hissələrindən böyük səhnələrə qədər təsvirlərin iyerarxiyası öyrədilir. Əlavə olaraq,

öyrənilmiş əlamətlər yeni tapşırıqlar üçün istifadə edilir.

4.2. Tibbi təsvirlərin seqmentasiyası

(Xue, Xu, Zhang, Long, & Huang, 2018)-də klassik GRŞ-lərdən qaynaqlanan, tibbi təsvirlərin seqmentləşdirilməsi üçün SegAN adlı yeni rəqib neyron şəbəkəsi təklif edilir. Şekli seqmentləşdirmək üçün sıx, piksel səviyyəli sinifləndirmə tələb olunduğu üçün klassik GRŞ diskriminatorunun tək real və ya saxta çıxışı şəbəkələrlə sabit və yetərli dərəcədə əks əlaqə yaratmaqda effektiv ola bilər. Bunun əvəzinə seqmentləşdirmə sinif nişanı asılılıqları yaratmaq üçün generator kimi tam konvolyusiyalı neyron şəbəkəsindən istifadə edilir, diskriminator və generator həm global, həm də yerli əlamətləri öyrənməyə məcbur etmək üçün çox miqyaslı L1 itki funksiyası ilə yeni rəqib şəbəkəsi təklif olunur. SegAN mühitində generator və diskriminator şəbəkələri min-max oyununda alternativ bir şəkildə öyrədilir: Diskriminator bir cüt təsviri giriş olaraq götürür, (orijinal təsviri - proqnozlaşdırılan sinif nişanları asılılığı, orijinal təsviri - əsas həqiqi sinif nişanları asılılığı) və sonra çoxmiqyaslı itki funksiyasını maksimum dərəcədə artıraraq öyrədilir; seqmentor çoxmiqyaslı funksiyasını minimuma endirmək məqsədilə yalnız diskriminatorun keçdiyi qradientlərlə hazırlanır. Belə bir SegAN mühitinin seqmentləşdirmə məsələsi üçün daha effektiv və dayanıqlı olduğu göstərilir və bu, ən müasir U-net seqmentləşdirmə metodundan daha yaxşı nəticə verir.

4.3. Musiqi generasiyası

Hərəkət tanıma ilə bağlı mövcud müasir yanaşmalar daha çox rekurrent neyron şəbəkələrə (Recurrent Neural Networks, RNN) əsaslanır. (Mao, Li, Xie, Lau, Wang, & Paul Smolley, 2017)-də hərəkətin klassifikasiyası və təyini üçün CNN əsaslı bir mühit təklif edilir. Sadə 7 laylı şəbəkə ilə NTU RGB + D verilənlər bazasının validasiya dəstində 89,3 % dəqiqlik əldə edilir. Kəsilməmiş videolarda hərəkətin təyin edilməsi məqsədilə eyni şəbəkə daxilində klassifikasiya edilmiş müvəqqəti seqment nümunələrini çıxarmaq üçün şəbəkə inkişaf etdirilir. Nəticədə, PKU-MMD verilənlər bazasında baza səviyyəsini böyük fərqlə üstələyərək 93,7 % ilə xəritə əldə edilir.

4.4. Ən yaxşı görüntü imkanı

(Ledig, Theis, Huszár, Caballero, Cunningham, Acosta, ... & Shi, 2017)-də ən yaxşı görüntü imkanı (Super Resolution, SR) üçün generativ rəqib şəbəkəsi olan SRGAN təklif edilir. Buna nail olmaq üçün qarşılıqlı itki və məzmun itkisindən ibarət olan qavrayış itkisi funksiyası təklif edilir.

Generativ rəqib şəbəkələr güclü generativ modellərdir, lakin təlimin qeyri-sabitlik problemi var. Son zamanlarda təklif olunan Vasserşteyn GRŞ, GRŞ-lərin sabit təliminə doğru irəliləyir, lakin hələ də yalnız zəif nümunələr generasiya edə bilir. (Gulrajani, Ahmed, Arjovsky, Dumoulin, & Courville, 2017) -də təklif edilən metod standart WGAN-dan üstündür və çox sayda GRŞ arxitekturunun praktik olaraq heç bir hiperparametr olmadan ardıcıl şəkildə öyrədilməsinə imkan verir, bura 101 laydan ibarət ResNets modeli daxildir.

4.5. Mətnin təsvirə çevrilməsi

Mətn-təsvirlərindən yüksək keyfiyyətli təsvirlərin sintezi kompüter görməsində çətin bir problemdir və bir çox tətbiq tələb edir. Mövcud mətn-təsvir yanaşmalarının yaratdığı nümunələr verilmiş təsvirlərin mənasını təxminən əks etdirə bilər, lakin lazımi detalları və canlı obyekt hissələrini ehtiva etmir. (Zhang, Xu, Li, Zhang, Wang, Huang, & Metaxas, 2018)-də mətn-təsvirə əsaslanaraq 256x256 piksel ölçüdə real görüntülər yaratmaq üçün toplanmış GRŞ (StackGAN) təklif edilir. Stage-I GAN verilən mətn təsvirinə əsaslanaraq obyektin ibtidai formasını və rənglərini cızır və Stage-I aşağı keyfiyyətli təsvirlər verir. Stage-I nəticələrdəki qüsurları aradan qaldıra və fərqli detallar əlavə edə bilər. Sintez edilmiş təsvirlərin müxtəlifliyini yaxşılaşdırmaq və şərti GRŞ təlimini sabitləşdirmək üçün yeni şərtlərin artırılması (Conditioning Augmentation) üsulu təqdim edilir.

4.6. Video generasiyası

Videodakı siqnallar məzmun və hərəkət olmaqla iki hissəyə bölünə bilər. Məzmun videodakı hansı obyektlərin olduğunu göstərdiyi halda, hərəkət onun dinamikasını müəyyənləşdirir. Bunu əsas götürərək, video generasiyası üçün hərəkət (motion) və məzmun (content) hissələrdən ibarət GRŞ (MoCoGAN)

mühiti təklif edilir (Tulyakov, Liu, Yang, & Kautz, 2018). Təklif olunan mühit təsadüfi vektorları videomühitlərə uyğunlaşdıraraq yeni bir video generasiya edir. Hər bir təsadüfi vektor bir məzmun hissəsindən və hərəkət hissəsindən ibarətdir. Məzmun hissəsi sabit saxlanılarkən, hərəkət hissəsi stoxastik bir proses olaraq həyata keçirilir. Ən yeni yanaşmalarla keyfiyyət və kəmiyyət müqayisəsi apararaq bir neçə verilənlər toplusu üzərində geniş eksperimental nəticələr təklif olunan mühitin effektivliyi qiymətləndirir. Bundan əlavə, MoCoGAN-ın eyni məzmununda, lakin fərqli hərəkəti olan videoları, eyni zamanda fərqli məzmunu və eyni hərəkəti olan videoları generasiya etməyə imkan verdiyini görmək mümkündür.

4.7. Üz əlamətlərinin çevrilməsi

Age-cGAN metodundan istifadə etməklə insan üzünün təsvirinin sintetik yaşlanması reallaşdırılmışdır. Üzün əlamətlərini dəyişmək üçün GRŞ-lərin istifadə olunduğu işlərin əksinə burada şəxsin orijinal yaşlı üzünün əlamətləri qorunub saxlanır. Bu səbəbdən, GAN-ın gizli vektorlarının kimliyin qorumasının (Identity-Preserving) optimallaşdırmasına yeni yanaşma təklif edilir. Əldə edilən yaşlı və cavan üzlərin ən müasir üz tanıma və yaş qiymətləndirmə yolları ilə obyektiv qiymətləndirilməsi təklif olunan metodun yüksək potensialını nümayiş etdirir (Antipov, Baccouche, & Dugelay, 2017, September, Tran, Yin, & Liu, 2017, Fedus, Rosca, Lakshminarayanan, Dai, Mohamed, & Goodfellow, 2017).

5. Nəticə

Generativ rəqib şəbəkələr son illərdə meydana çıxmış və çox sürətlə inkişaf edirlər. GRŞ-lər oyun nəzəriyyəsinə əsaslanan generativ modellərin bir növüdür. GRŞ-lər real verilənlərin, xüsusilə də təsvirlərin generasiyasında praktik olaraq uğur qazanmışdır. Generativ modellərin güclü sinfi olaraq, GRŞ-lər nümunə verilənlərin paylanması aydın şəkildə proqnoz edə bilmirlər, lakin real nümunələrlə eyni paylanmaya uyğun gələn yeni nümunələr generasiya etməyi öyrənirlər. Eyni zamanda bu şəbəkələr müxtəlif təsvirlərin klassifikasiyası, proqnozlaşdırma, üz tanınması kimi effektiv işlərə tətbiq edilmişdir. Məqalədə generativ rəqib şəbəkələrin bəzi ən vacib növləri

qısa analiz edilmiş və müqayisələr aparılmışdır. Bu işdə klassifikasiya və regressiya, tibbi şəkillərin seqmentasiyası, musiqi generasiyası, ən yaxşı görüntü imkanı, mətnin şəklə çevrilməsi, video generasiyası üçün GRŞ-lərin analizi aparılmışdır. Əhəmiyyətli olan odur ki, GRŞ-lər təlimdə daha çox dövrlərlə daha yüksək keyfiyyətli şəkillər istehsal etməyə meyllidirlər.

Ədəbiyyat

- Alguliyev, R. M., Abdullayeva, F. J., & Ojagverdiyeva, S. S. (2020). Protecting children on the internet using deep generative adversarial networks. *International Journal of Computational Systems Engineering*, 6(2), 84-90. <https://ieeexplore.ieee.org/document/771073>
- Antipov, G., Baccouche, M., & Dugelay, J. L. (2017, September). Face aging with conditional generative adversarial networks. In 2017 IEEE international conference on image processing (ICIP) (pp. 2089-2093). IEEE.
- Arjovsky, M., & Bottou, L. (2017). Towards principled methods for training generative adversarial networks. https://openreview.net/pdf?id=Hk4_qw5xe Accessed 11 December 2021.
- Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In *International conference on machine learning* (pp. 214-223). PMLR.
- Bang, D., & Shim, H. (2018, July). Improved training of generative adversarial networks using representative features. In *International conference on machine learning* (pp. 433-442). PMLR.
- Che, T., Li, Y., Jacob, A. P., Bengio, Y., & Li, W. (2016). Mode regularized generative adversarial networks. <https://arxiv.org/abs/1612.02136> Accessed 11 December 2021.
- Chen, X., Duan, Y., Houthoof, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016, December). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (pp. 2180-2188).
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: an overview. *IEEE Signal Process Mag* 35 (1): 53–65. <https://arxiv.org/abs/1710.07035> Accessed 11 December 2021.
- Donahue, J., Krähenbühl, P., & Darrell, T. (2016). Adversarial feature learning. <https://arxiv.org/abs/1605.09782> Accessed 11 December 2021.
- Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., & Courville, A. (2016). Adversarially learned inference. <https://arxiv.org/abs/1606.00704> Accessed 11 December 2021.
- Fedus, W., Rosca, M., Lakshminarayanan, B., Dai, A. M., Mohamed, S., & Goodfellow, I. (2017). Many paths to equilibrium: GANs do not need to decrease a divergence at every step. <https://arxiv.org/abs/1710.08446> Accessed 11 December 2021.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139144. <https://doi.org/10.1145/3422622>
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of wasserstein gans. <https://arxiv.org/abs/1704.00028> Accessed 11 December 2021.
- Hitawala, S. (2018). Comparative study on generative adversarial networks. <https://arxiv.org/abs/1801.04271> Accessed 11 December 2021.
- Xue, Y., Xu, T., Zhang, H., Long, L. R., & Huang, X. (2018). Segan: Adversarial network with multi-scale l1 loss for medical image segmentation. *Neuroinformatics*, 16(3), 383-392. <https://link.springer.com/article/10.1007%2Fs12021-018-9377-x>
- 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).
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. <https://arxiv.org/abs/1710.10196> Accessed 11 December 2021.
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4401-4410).
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).
- Lucic, M., Kurach, K., Michalski, M., Gelly, S., & Bousquet, O. (2017). Are gans created equal? a large-scale study. <https://arxiv.org/abs/1711.10337v4> Accessed 11 December 2021.
- Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Paul Smolley, S. (2017). Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2794-2802).
- Metz, L., Poole, B., Pfau, D., & Sohl-Dickstein, J. (2016). Unrolled generative adversarial networks. <https://arxiv.org/abs/1611.02163> Accessed 11 December 2021.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. <https://arxiv.org/abs/1411.1784> Accessed 11 December 2021.
- Nowozin, S., Cseke, B., & Tomioka, R. (2016, December). f-gan: Training generative neural samplers using variational divergence minimization. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (pp. 271-279).
- Odena, A., Olah, C., & Shlens, J. (2017, July). Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning* (pp. 2642-2651). PMLR.
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. <https://arxiv.org/pdf/1511.06434.pdf%5D> Accessed 11 December 2021.

- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2234-2242.
<https://proceedings.mlr.press/v80/bang18a/bang18a.pdf>
- Theis, L., Oord, A. V. D., & Bethge, M. (2015). A note on the evaluation of generative models.
<https://arxiv.org/pdf/1511.01844.pdf> Accessed 11 December 2021.
- Tran, L., Yin, X., & Liu, X. (2017). Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1415-1424).
- Tulyakov, S., Liu, M. Y., Yang, X., & Kautz, J. (2018). Mocogan: Decomposing motion and content for video generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1526-1535).
- Wang, Z., She, Q., & Ward, T. E. (2021). Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR)*, 54(2), 1-38.
<https://arxiv.org/pdf/1906.01529.pdf>
- Warde-Farley, D., & Bengio, Y. (2016). Improving generative adversarial networks with denoising feature matching.
<https://openreview.net/forum?id=S1X7nhx1>
- Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., ... & Wang, G. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6), 1348-1357.
<https://arxiv.org/pdf/1708.00961.pdf>
- Yeh, R. A., Chen, C., Yian Lim, T., Schwing, A. G., Hasegawa-Johnson, M., & Do, M. N. (2017). Semantic image inpainting with deep generative models. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5485-5493).
- Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2018). Stackgan++: Realistic image synthesis with stacked generative adversarial networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(8), 1947-1962.
<https://arxiv.org/abs/1710.10916>

Yadigar N. İmamverdiyev^a, Firəngiz İ. Musayeva^b

^aAzərbaycan Texniki Universiteti, Süni İntellekt İnstitutu
Azərbaycan, Bakı ş., AZ1148, H.Cavid Prospekti, 25.

^bAMEA İnformasiya Texnologiyaları İnstitutu.
Azərbaycan, Bakı ş., AZ1141, B.Vahabzadə küç., 9A.

Ядигар Н. Имамвердиев^a, Фирангиз И. Мусаева^b

^aАзербайджанский Технический Университет, Институт Искусственного Интеллекта
Азербайджан, г. Баку, AZ1148, пр. Г.Джавида, 25.

^bИнститут Информационных Технологий НАН Азербайджана.
Азербайджан, г. Баку, AZ1148, ул. Б.Вахабзаде, 9А.

 0000-0002-3710-1046