NEURAL NETWORK-BASED IMAGE GENERATION METHODS IN DIGITAL ART AND GRAPHIC DESIGN

Yalg‘osheva Ozoda Farhod qizi
Master’s Student, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
E-mail: ozoda4449@icloud.com
Scientific Supervisor: Ismailov Kamoliddin Saidaxmadovich
Associate Professor, Department of Television and Media Technologies, Tashkent University of Information Technologies, Doctor of Philosophy (PhD) in Technical Sciences

Abstract

The article presents a systematic analysis of the main neural network-based image generation methods used in digital art and graphic design: generative adversarial networks (GANs), variational autoencoders (VAEs), diffusion models and multimodal transformer-based models. The study employs theoretical comparative analysis, mathematical modelling and a meta-analysis of experimental results reported in the international literature. The models are compared in terms of image quality, diversity, generation speed and controllability using the FID, IS and CLIP Score metrics. The results show that latent diffusion models with classifier-free guidance provide the highest quality and controllability in text-to-image generation, while GANs retain their advantage in tasks that require real-time performance. A hybrid approach based on a cultural consistency loss is proposed to address the problem of accurately reproducing national cultural elements (ornament, embroidery, miniature painting).

Keywords

neural networks, image generation, GAN, VAE, diffusion models, transformer, CLIP, digital art, graphic design, national ornament.

RAQAMLI SAN'AT VA GRAFIK DIZAYNDA NEYRON TARMOQLAR ASOSIDA TASVIR GENERATSIYA QILISH USULLARI

Yalg'osheva Ozoda Farhod qizi

Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti magistranti

E-mail: ozoda4449@icloud.com

Ilmiy rahbar: Ismailov Kamoliddin Saidaxmadovich

TATU "Televizion va media texnologiyalar" kafedrasi dotsenti, s.f.f.d. (PhD)

Annotatsiya. Maqolada raqamli san'at va grafik dizayn sohasida neyron tarmoqlar asosida tasvir generatsiya qilishning asosiy usullari – generativ raqobatlashuvchi tarmoqlar (GAN), variatsion avtokodlovchilar (VAE), diffuziya modellari hamda transformer arxitekturasiga asoslangan multimodal modellar tizimli tahlil qilingan. Tadqiqotda nazariy-qiyosiy tahlil, matematik modellashtirish va xalqaro ilmiy manbalarda e'lon qilingan eksperimental natijalarni meta-tahlil qilish usullaridan foydalanilgan. Modellar tasvir sifati, xilma-xilligi, generatsiya tezligi va boshqaruvchanlik mezonlari bo'yicha FID, IS va CLIP Score metrikalari asosida taqqoslangan. Natijalar shuni ko'rsatadiki, klassifikatorsiz yo'naltirishga ega latent diffuziya modellari matnli tavsif asosida tasvir yaratishda eng yuqori sifat va boshqaruvchanlikni ta'minlaydi, GAN modellari esa real vaqt rejimida ishlash talab etiladigan vazifalarda afzalligini saqlab qolmoqda. Maqolada milliy madaniy elementlarni (naqsh, kashtachilik, miniatyura) aniq qayta tiklash muammosini hal etish uchun madaniy moslik yo'qotish funksiyasiga asoslangan gibrid yondashuv taklif etilgan.

Kalit so'zlar: *neyron tarmoqlar, tasvir generatsiyasi, GAN, VAE, diffuziya modellari, transformer, CLIP, raqamli san'at, grafik dizayn, milliy naqsh.*

МЕТОДЫ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ В ЦИФРОВОМ ИСКУССТВЕ И ГРАФИЧЕСКОМ ДИЗАЙНЕ

Аннотация. В статье проведён системный анализ основных методов генерации изображений на основе нейронных сетей в цифровом искусстве и графическом дизайне: генеративно-состязательных сетей (GAN), вариационных автокодировщиков (VAE), диффузионных моделей и мультимодальных моделей на основе архитектуры трансформер. В исследовании использованы методы теоретико-сравнительного анализа, математического моделирования и метаанализа экспериментальных результатов, опубликованных в международных научных источниках. Модели сопоставлены по критериям качества, разнообразия, скорости генерации и управляемости на основе метрик FID, IS и CLIP Score. Результаты показывают, что латентные диффузионные модели с бесклассификаторным управлением обеспечивают наиболее высокое качество и управляемость при генерации изображений по текстовому описанию, тогда как GAN-модели сохраняют преимущество в задачах, требующих работы в реальном времени. Предложен гибридный подход на основе функции потерь культурной согласованности для точного воспроизведения национальных культурных элементов (орнамент, вышивка, миниатюра).

Ключевые слова: *нейронные сети, генерация изображений, GAN, VAE, диффузионные модели, трансформер, CLIP, цифровое искусство, графический дизайн, национальный орнамент.*

KIRISH

So‘nggi o‘n yillikda sun’iy intellekt texnologiyalari ijodiy sohalarga jadal kirib kelib, raqamli san’at va grafik dizayn amaliyotini tubdan o‘zgartirmoqda. Agar ilgari kompyuter grafikasi asosan dizayner tomonidan qo‘lda boshqariladigan vositalar

majmuasi sifatida qaralgan bo‘lsa, bugungi kunda neyron tarmoqlar matnli tavsif, eskiz yoki namuna tasvir asosida yangi vizual kontentni mustaqil ravishda yarata oladigan generativ tizimlarga aylandi. Bu jarayon dizaynerning roli, ijodiy mehnat unumdorligi va vizual madaniyatning rivojlanish yo‘nalishlari haqidagi ilmiy savollarni ham kun tartibiga qo‘ymoqda.

Neyron tarmoqlar asosidagi tasvir generatsiyasining zamonaviy bosqichi 2014-yilda I. Gudfellou va hammualliflar tomonidan generativ raqobatlashuvchi tarmoqlar (Generative Adversarial Networks, GAN) taklif etilishi bilan boshlandi [1]. Undan bir oz avval D. Kingma va M. Velling variatsion avtokodlovchi (VAE) modelini ishlab chiqib, ehtimoliy latent fazo g‘oyasini generativ modellashtirishga olib kirdi [2]. 2020-yildan boshlab esa J. Ho va hammualliflarning shovqinni bosqichma-bosqich olib tashlashga asoslangan diffuziya modellari [3] va R. Rombax guruhining latent diffuziya modellari [4] soha yetakchisiga aylandi. Parallel ravishda A. Vasvani va hammualliflar taklif etgan transformer arxitekturasi [5] hamda A. Redford guruhining matn va tasvirni yagona semantik fazoda moslashtiruvchi CLIP modeli [6] matnli tavsif asosida tasvir yaratish (text-to-image) imkoniyatini amaliy darajaga olib chiqdi.

O‘zbekiston Respublikasida sun’iy intellekt texnologiyalarini rivojlantirish davlat siyosatining ustuvor yo‘nalishlaridan biri sifatida belgilangan. Xususan, O‘zbekiston Respublikasi Prezidentining 2021-yil 17-fevraldagi PQ-4996-son qarori [7] va 2024-yil 14-oktabrdagi PQ-358-son qarori bilan tasdiqlangan 2030-yilgacha sun’iy intellekt texnologiyalarini rivojlantirish strategiyasi [8] ushbu sohadagi ilmiy tadqiqotlar va amaliy ishlanmalarni qo‘llab-quvvatlashga qaratilgan. Shu bilan birga, milliy ijodiy industriyada generativ modellardan foydalanish bo‘yicha tizimli ilmiy tadqiqotlar hali yetarli emas.

Mavjud xorijiy tadqiqotlar tahlili ko‘rsatadiki, zamonaviy generativ modellar asosan g‘arb vizual an’analarini aks ettiruvchi katta hajmdagi ma’lumotlar to‘plamlarida

o‘qitilgan. Natijada ular o‘zbek milliy naqshlari, suzana kashtachiligi, Buxoro va Samarqand me’moriy bezaklari yoki Sharq miniatyurasi kabi madaniy elementlarni yaratishda geometrik simmetriyani buzish, uslublarni aralashtirish va madaniy kontekstga mos kelmaydigan detallarni qo‘shish kabi xatoliklarga yo‘l qo‘yadi. Bu holat milliy dizayn amaliyotida generativ modellardan samarali foydalanishni cheklaydi.

Shu munosabat bilan ushbu tadqiqotning **maqsadi** raqamli san’at va grafik dizaynda qo‘llaniladigan neyron tarmoqlar asosidagi tasvir generatsiyasi usullarini tizimli tahlil qilish, ularni miqdoriy va sifat mezonlari bo‘yicha qiyosiy baholash hamda milliy madaniy elementlarni generatsiya qilish uchun gibrid yondashuvni asoslashdan iborat. Maqsadga erishish uchun quyidagi vazifalar belgilandi: generativ modellarning asosiy sinflari matematik asoslarini yoritish; ularning sifat ko‘rsatkichlarini xalqaro benchmark natijalari asosida taqqoslash; dizayn amaliyotidagi qo‘llanish sohalarini aniqlash; milliy madaniy kontekstga moslashtirilgan gibrid model arxitekturasini taklif etish.

MATERIALLAR VA METODLAR

Tadqiqot nazariy-qiyosiy tahlil, matematik modellashtirish hamda xalqaro nufuzli konferensiyalar (CVPR, ICCV, NeurIPS, ICLR, ICML) materiallarida e’lon qilingan eksperimental natijalarni meta-tahlil qilish usullariga asoslandi. Qiyosiy tahlil uchun modellar to‘rtta asosiy sinfga ajratildi: generativ raqobatlashuvchi tarmoqlar, variatsion avtokodlovchilar, diffuziya modellari va transformer asosidagi avtoregressiv hamda multimodal modellar. Solishtirma natijalar yagona standartlashtirilgan protokollar – ImageNet 256×256 ma’lumotlar to‘plamida sinfga shartlangan generatsiya va MS-COCO ma’lumotlar to‘plamida matn asosida zero-shot generatsiya bo‘yicha olindi, bu esa turli mualliflar natijalarini o‘zaro taqqoslash imkonini beradi.

Generativ raqobatlashuvchi tarmoqlar.

GAN ikki neyron tarmoq – generator *G* va diskriminator *D* o‘rtasidagi raqobatli o‘yinga asoslanadi [1]. Generator tasodifiy shovqin vektori *z* dan real ma’lumotlarga o‘xshash tasvir yaratishga, diskriminator esa real va sintetik tasvirlarni farqlashga intiladi. O‘qitish jarayoni quyidagi minimaks masalasini yechishdan iborat:

$$\min_G \max_D V(D, G) = E_{x\sim pdata}[\log D(x)] + E_{z\sim pz}[\log(1 - D(G(z)))] \quad (1)$$

bu yerda p_{data} – real tasvirlar taqsimoti, p_z – latent shovqin taqsimoti. GAN oilasining dizayn uchun eng muhim vakillari – StyleGAN seriyasi bo‘lib, unda latent vektor xaritalash tarmog‘i orqali oraliq *W* fazosiga o‘tkaziladi va adaptiv normallashtirish yordamida turli masshtab darajalarida uslubni boshqarish imkoniyati yaratiladi [9]. StyleGAN3 modelida esa signalni qayta ishlash nazariyasiga asoslangan holda aliasing effekti bartaraf etilib, tasvir detallarining koordinatalarga “yopishib qolishi” muammosi hal qilindi [10]. Pix2Pix [11] va CycleGAN [12] kabi shartli modellar eskizni tayyor tasvirga, bir uslubni boshqasiga o‘tkazish vazifalarida keng qo‘llaniladi. GAN larning asosiy kamchiliklari – o‘qitishning beqarorligi va rejim qulashi (mode collapse), ya’ni generatorning faqat cheklangan turdagi tasvirlarni ishlab chiqarishi hisoblanadi.

Variatsion avtokodlovchilar.

VAE kodlovchi $q_\varphi(z|x)$ va dekodlovchi $p_\theta(x|z)$ tarmoqlaridan tashkil topgan bo‘lib, ma’lumotlar log-ehtimolligining quyi chegarasi (ELBO) ni maksimallashtirish orqali o‘qitiladi [2]:

$$L(\theta, \varphi; x) = E_{q\varphi(z|x)}[\log p_\theta(x|z)] - D_{KL}(q_\varphi(z|x) \parallel p(z)) \quad (2)$$

Birinchi had tasvirni qayta tiklash aniqligini, ikkinchi had esa Kulbak–Leybler divergensiyasi orqali latent taqsimotning standart normal taqsimotga yaqinligini ta’minlaydi. VAE uzluksiz va silliq latent fazo hosil qilishi tufayli uslublar orasida interpolyatsiya qilish va dizayn variantlarini boshqariladigan tarzda o‘zgartirish uchun

qulay. Biroq sof VAE modellari yaratgan tasvirlar odatda xiralashgan bo‘ladi. Shu sababli zamonaviy tizimlarda VAE asosan mustaqil generator sifatida emas, balki tasvirni siqilgan latent fazoga o‘tkazuvchi komponent sifatida, xususan, latent diffuziya modellarining avtokodlovchisi sifatida qo‘llaniladi [4], [13].

Diffuziya modellari.

Diffuziya modellari ikki jarayonga asoslanadi. To‘g‘ri (forward) jarayonda real tasvir x_0 ga T qadam davomida Gauss shovqini bosqichma-bosqich qo‘shiladi [3]:

$$q(x_t \mid x_{t-1}) = N(x_t;\ \sqrt{(1-\beta_t)}\, x_{t-1},\ \beta_t I) \quad (3)$$

bu yerda β_t – shovqin jadvali parametri. Teskari jarayonda neyron tarmoq ε_θ har bir qadamda qo‘shilgan shovqinni bashorat qilishni o‘rganadi va soddalashtirilgan yo‘qotish funksiyasi quyidagicha ifodalanadi:

$$L_{simple} = E_{t,\, x0,\, \varepsilon}[\|\varepsilon - \varepsilon_\theta(x_t,\ t)\|^2] \quad (4)$$

Generatsiya jarayonida model sof shovqindan boshlab uni ketma-ket tozalash orqali yangi tasvir hosil qiladi. Latent diffuziya modellarida (Stable Diffusion) bu jarayon piksel fazosida emas, balki VAE yordamida siqilgan latent fazoda amalga oshiriladi, bu esa hisoblash xarajatlarini bir necha barobar kamaytiradi [4]. Matnli shart c ni hisobga olish uchun klassifikatorsiz yo‘naltirish (classifier-free guidance) usuli qo‘llaniladi [14]:

$$\hat{\varepsilon}_\theta(x_t,\ c) = \varepsilon_\theta(x_t,\ \varnothing) + w \cdot (\varepsilon_\theta(x_t,\ c) - \varepsilon_\theta(x_t,\ \varnothing)) \quad (5)$$

bu yerda w – yo‘naltirish kuchi koeffitsiyenti bo‘lib, uning qiymati oshgan sari tasvirning matnga mosligi kuchayadi, ammo xilma-xillik kamayadi. Dizayn amaliyotida diffuziya modellarining imkoniyatlari ControlNet [15] (chiziqli eskiz, chuqurlik xaritasi, poza orqali fazoviy boshqaruv), DreamBooth [16] (kam sonli namunalar asosida muayyan obyekt yoki uslubni o‘rgatish) va LoRA [17] (past rangli adapterlar orqali modelni resurs tejamkor moslashtirish) usullari bilan sezilarli darajada kengaytirildi.

Transformer va multimodal modellar.

Transformer arxitekturasining asosini diqqat (attention) mexanizmi tashkil etadi [5]:

$$Attention(Q, K, V) = softmax(QK^T / \sqrt{d_k})\, V \quad (6)$$

bu yerda Q, K, V – so‘rov, kalit va qiymat matritsalari, d_k – kalit vektorlari o‘lchami. Tasvir generatsiyasida transformerlar uch yo‘nalishda qo‘llaniladi: tasvirni diskret tokenlar ketma-ketligi sifatida avtoregressiv modellashtirish (VQGAN va DALL·E) [18]; diffuziya modellarida U-Net o‘rniga asosiy tarmoq sifatida foydalanish (Diffusion Transformer, DiT) [19]; matnli tavsifni kodlash va kross-diqqat mexanizmi orqali generatsiya jarayoniga uzatish. CLIP modeli 400 million “tasvir–matn” juftligida kontrastiv o‘qitilgan bo‘lib, matn va tasvirni umumiy semantik fazoga joylashtiradi [6]. Aynan shu xususiyat matnli so‘rov (prompt) orqali tasvir yaratish va generatsiya qilingan tasvirning matnga mosligini baholash imkonini beradi.

Baholash metrikalari.

Generativ modellarni miqdoriy baholash uchun uchta asosiy metrikadan foydalanildi. Frechet Inception Distance (FID) real va generatsiya qilingan tasvirlarning Inception-v3 tarmog‘i xususiyatlar fazosidagi taqsimotlari orasidagi masofani o‘lchaydi [20]:

$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \quad (7)$$

bu yerda μ va Σ – mos ravishda real (r) va generatsiya qilingan (g) tasvirlar xususiyatlarining o‘rta qiymati va kovariatsiya matritsasi. FID qiymati qanchalik past bo‘lsa, tasvirlar sifati va xilma-xilligi shunchalik yuqori hisoblanadi. Inception Score (IS) tasvirlarning aniq sinflanuvchanligi va sinflar bo‘yicha xilma-xilligini baholaydi [21], uning yuqori qiymati yaxshi natijani bildiradi. CLIP Score esa generatsiya qilingan tasvir va matnli tavsif orasidagi semantik moslikni CLIP embeddinglari orasidagi kosinus o‘xshashlik orqali o‘lchaydi [22].

Taklif etilayotgan gibrid yondashuv.

Milliy madaniy elementlarni generatsiya qilishdagi muammolarni hal etish uchun latent diffuziya modeli, madaniy CLIP kodlovchisi va StyleGAN3 arxitekturasiga asoslangan diskriminatorni birlashtiruvchi gibrid yondashuv taklif etiladi. Madaniy CLIP kodlovchisi E_c o‘zbek milliy naqshlari, kashtachilik namunalari, me’moriy bezaklar va miniatyura asarlaridan iborat maxsus ma’lumotlar to‘plamida CLIP modelini qo‘shimcha o‘qitish (fine-tuning) orqali hosil qilinadi. Generatsiya qilingan tasvir $\hat{x}$ ning madaniy namunaga mosligini ta’minlash uchun madaniy moslik yo‘qotish funksiyasi (Cultural Consistency Loss) kiritiladi:

$$L_{CCL} = 1 - cos(E_c(\hat{x}), E_c(x_{ref})) \qquad (8)$$

bu yerda x_{ref} – madaniy element namunasi. Umumiy yo‘qotish funksiyasi diffuziya yo‘qotishi, madaniy moslik hadi va diskriminator beradigan raqobatli hadning vaznli yig‘indisi sifatida ifodalanadi:

$$L_{total} = L_{simple} + \lambda_c L_{CCL} + \lambda_a L_{adv} \qquad (9)$$

bu yerda λ_c va λ_a – mos hadlarning ta’sir darajasini belgilovchi giperparametrlar. StyleGAN3 asosidagi diskriminator aliasingdan xoli arxitekturasi tufayli naqshlarning mayda geometrik detallari va simmetriyasini nozik farqlay oladi, bu esa takrorlanuvchi ornamental tuzilmalarning buzilishini kamaytiradi.

NATIJALAR

Standart ImageNet 256×256 protokoli bo‘yicha sinfga shartlangan generatsiyada turli sinfdagi modellarning xalqaro manbalarda e’lon qilingan ko‘rsatkichlari 1-jadvalda umumlashtirildi.

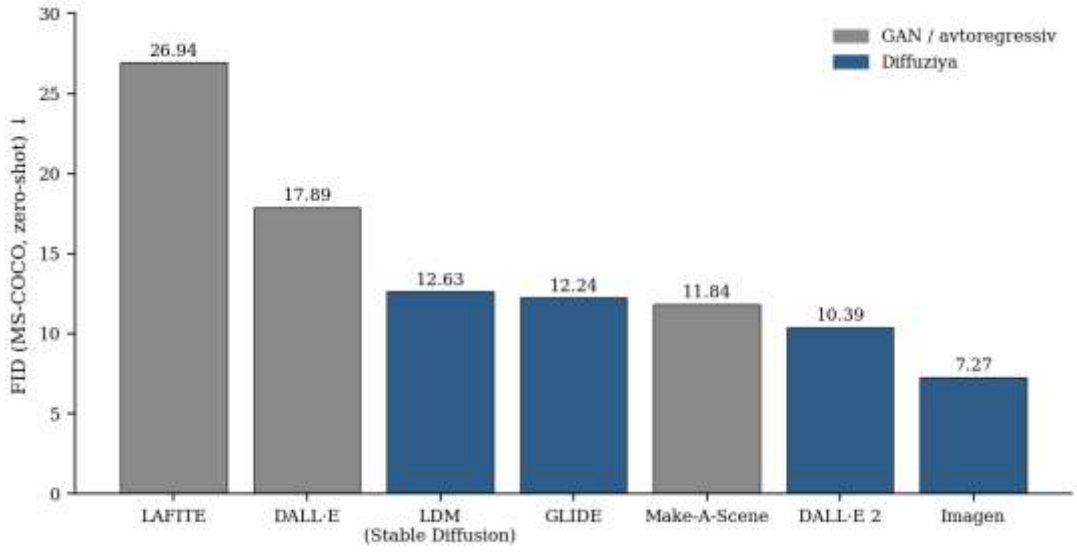
1-jadval

ImageNet 256×256 ma’lumotlar to‘plamida sinfga shartlangan generatsiya natijalari

Model	Model sinfi	Manba	FID ↓	IS ↑
BigGAN-deep	GAN	Brock va boshq. [23]	6,95	171,4
ADM-G	Diffuziya	Dhariwal, Nichol [24]	4,59	186,7
LDM-4-G	Latent diffuziya	Rombach va boshq. [4]	3,60	247,7
StyleGAN-XL	GAN	Sauer va boshq. [25]	2,30	265,1
DiT-XL/2-G	Diffuziya transformer	Peebles, Xie [19]	2,27	278,2

1-jadval ma'lumotlaridan ko'rinib turibdiki, 2019–2023-yillar oralig'ida FID ko'rsatkichi 6,95 dan 2,27 gacha, ya'ni taxminan uch barobar yaxshilangan. Diffuziya modellari klassik GAN modellarini sifat bo'yicha ortda qoldirgan, ammo StyleGAN-XL natijasi masshtablashtirilgan GAN arxitekturalari ham diffuziya modellari bilan raqobatlasha olishini ko'rsatadi. Eng yaxshi natija diffuziya jarayoni va transformer arxitekturasini birlashtirgan DiT modeliga tegishli bo'lib, bu arxitekturalarning konvergensiyasi tendensiyasini tasdiqlaydi.

Dizayn amaliyoti uchun muhimroq bo'lgan matnli tavsif asosida tasvir yaratish vazifasida MS-COCO ma'lumotlar to'plamida zero-shot rejimida olingan FID ko'rsatkichlari 1-rasmda keltirilgan.



1-rasm. Matn asosida tasvir generatsiyasi modellarining MS-COCO to‘plamidagi zero-shot FID ko‘rsatkichlari (manbalar: [4], [26], [27])

1-rasmdan ko‘rinadiki, diffuziya asosidagi modellar (GLIDE, LDM, DALL·E 2, Imagen) avtoregressiv va GAN asosidagi modellarga (DALL·E, LAFITE) nisbatan sezilarli darajada past FID qiymatiga ega. Imagen modelining 7,27 lik natijasi katta hajmdagi til modelini matn kodlovchisi sifatida qo‘llash tasvir sifatiga kuchli ta’sir ko‘rsatishini isbotlaydi [27]. Stable Diffusion (LDM) esa eng yuqori natija bermasa-da, ochiq kodli ekanligi va oddiy grafik protsessorlarda ishlashi tufayli dizayn amaliyotida eng keng tarqalgan modelga aylangan.

Modellarning sifat ko‘rsatkichlaridan tashqari, dizayn jarayoni uchun ahamiyatli bo‘lgan amaliy xususiyatlari ham tahlil qilindi. Olingan natijalar 2-jadvalda umumlashtirildi.

2-jadval

Generativ model sinflarining dizayn amaliyoti nuqtayi nazaridan qiyosiy tavsifi

Mezon	**GAN**	**VAE**	**Diffuziya**	**Transformer (avtoregressiv)**
Tasvir sifati	Yuqori	O‘rta (xiralashish)	Juda yuqori	Yuqori
Xilma-xillik	Past–o‘rta (rejim qulashi)	Yuqori	Juda yuqori	Yuqori
Generatsiya tezligi	Juda yuqori (1 qadam)	Juda yuqori	Past–o‘rta (20–50 qadam)	Past (tokenma-token)
Matn orqali boshqaruv	Cheklangan	Cheklangan	Juda yuqori	Yuqori

Mezon	GAN	VAE	Diffuziya	Transformer (avtoregressiv)
O‘qitish barqarorligi	Past	Yuqori	Yuqori	Yuqori
Asosiy qo‘llanish	Uslub ko‘chirish, yuz sintezi, real vaqt	Latent fazo, interpolyatsiya	Konsept-art, illustratsiya, reklama	Multimodal generatsiya

Grafik dizayn amaliyotida qo‘llanilishini tahlil qilish asosida generativ modellarning quyidagi asosiy funksional yo‘nalishlari aniqlandi. Birinchidan, konsept-art va g‘oya generatsiyasi bosqichida diffuziya modellari dizaynerga bir necha daqiqa ichida o‘nlab vizual variantlarni ko‘rib chiqish imkonini beradi, bu esa dastlabki izlanish bosqichini sezilarli qisqartiradi. Ikkinchidan, uslub ko‘chirish (style transfer) vazifalarida konvolyutsion neyron tarmoqlarga asoslangan klassik usul [28] hamda CycleGAN [12] va DreamBooth [16] kabi zamonaviy yondashuvlar muayyan badiiy uslubni yangi kontentga o‘tkazishni ta’minlaydi. Uchinchidan, ControlNet [15] yordamida kompozitsiyani eskiz yoki kontur orqali qat’iy belgilash imkoni paydo bo‘lib, bu generativ modelni tasodifiy tasvir generatoridan dizaynerning niyatini aniq bajaruvchi vositaga aylantiradi. To‘rtinchidan, tasvirning yo‘qolgan qismlarini to‘ldirish (inpainting), chegaralarini kengaytirish (outpainting) va sifatini oshirish (super-resolution) funksiyalari tayyor maketlar ustida ishlashni soddalashtiradi.

Milliy madaniy elementlarni generatsiya qilish nuqtayi nazaridan umumiy maqsadli modellarning cheklovlari aniq namoyon bo‘ladi. Ular “o‘zbek naqshi” yoki “suzana” kabi so‘rovlarga javoban umumiy sharqona bezakni eslatuvchi, biroq girih geometriyasining qat’iy qoidalari, islimiy naqshlarning o‘ziga xos ritmi va mintaqaviy maktablarning farqlarini aks ettirmaydigan tasvirlar yaratadi. Taklif etilgan gibrid

yondashuvda madaniy CLIP kodlovchisi aynan shu semantik farqlarni o‘rganishga, StyleGAN3 asosidagi diskriminator esa geometrik aniqlik va simmetriyani nazorat qilishga yo‘naltirilgan.

MUHOKAMA

Olingan natijalar generativ modellashtirish sohasida sifat va boshqaruvchanlik o‘rtasidagi muvozanat masalasini yaqqol ko‘rsatadi. Diffuziya modellarining ustunligi ularning ehtimoliy asoslangan barqaror o‘qitish jarayoni, rejim qulashiga moyil emasligi va matnli shartni klassifikatorsiz yo‘naltirish orqali moslashuvchan tarzda hisobga olishi bilan izohlanadi. Bu xulosa P. Dhariwal va A. Nichol [24] hamda R. Rombax va hammualliflarning [4] tadqiqotlari natijalariga mos keladi. Shu bilan birga, diffuziya modellarining asosiy kamchiligi – ko‘p qadamli generatsiya tufayli sekinligi – real vaqt rejimidagi interaktiv dizayn vositalarida hali ham muhim cheklov bo‘lib qolmoqda, bu sohada bir qadamli generatsiya qiluvchi GAN lar o‘z ahamiyatini yo‘qotmagan.

Tadqiqot natijalari arxitekturalar konvergensiyasi tendensiyasini ham tasdiqlaydi. Zamonaviy yetakchi tizimlar alohida model sinfiga emas, balki bir necha yondashuvning kombinatsiyasiga tayanadi: VAE latent fazo hosil qiladi, diffuziya jarayoni generatsiyani amalga oshiradi, transformer esa asosiy tarmoq va matn kodlovchisi vazifasini bajaradi. Masalan, SDXL modeli [13] kengaytirilgan U-Net, ikkita matn kodlovchisi va aniqlashtiruvchi modelni birlashtiradi. Taklif etilgan gibrid yondashuv ham shu mantiqqa asoslanib, har bir komponentning kuchli tomonini madaniy aniqlik vazifasiga yo‘naltiradi.

Shuni ta’kidlash lozimki, FID va IS metrikalari ImageNet sinflarida o‘qitilgan Inception tarmog‘iga tayanadi va madaniy-estetik jihatlarni to‘liq baholay olmaydi. Madaniy elementlarning haqqoniyligi, uslubiy tozaligi va an’anaviy kompozitsion qoidalarga muvofiqligini baholash uchun mutaxassislar ishtirokidagi sifat baholash hamda maxsus madaniy metrikalarni ishlab chiqish zarur. Bundan tashqari, generativ

modellardan foydalanish mualliflik huquqi, o‘qitish ma’lumotlarining kelib chiqishi va an’anaviy hunarmandlar merosini himoya qilish bilan bog‘liq axloqiy-huquqiy masalalarni ham keltirib chiqaradi. Mazkur jihatlar keyingi tadqiqotlarda alohida o‘rganilishi lozim.

Tadqiqotning cheklovi shundan iboratki, qiyosiy tahlil turli mualliflar tomonidan standart protokollar asosida e’lon qilingan natijalarga tayanadi. Taklif etilgan gibrid modelning milliy madaniy elementlar ma’lumotlar to‘plamidagi eksperimental verifikatsiyasi dissertatsiya tadqiqotining keyingi bosqichida amalga oshiriladi.

XULOSA

O‘tkazilgan tadqiqot natijasida raqamli san’at va grafik dizaynda qo‘llaniladigan neyron tarmoqlar asosidagi tasvir generatsiyasi usullari tizimli tahlil qilindi va quyidagi xulosalarga kelindi.

Birinchidan, generativ modellarning to‘rtta asosiy sinfi – GAN, VAE, diffuziya va transformer asosidagi modellar – turli matematik tamoyillarga asoslanadi va dizayn amaliyotida bir-birini to‘ldiruvchi vazifalarni bajaradi. Ikkinchidan, xalqaro benchmark natijalari tahlili diffuziya modellarining tasvir sifati va matn orqali boshqaruvchanlik bo‘yicha yetakchiligini tasdiqladi: ImageNet 256×256 protokolida FID ko‘rsatkichi 2019–2023-yillarda 6,95 dan 2,27 gacha yaxshilangan, MS-COCO zero-shot generatsiyasida esa diffuziya modellari 7,27–12,63 oralig‘ida natija ko‘rsatgan. Uchinchidan, GAN modellari yuqori generatsiya tezligi tufayli real vaqt rejimidagi dizayn vazifalarida o‘z ahamiyatini saqlab qolmoqda. To‘rtinchidan, umumiy maqsadli modellar milliy madaniy elementlarni aniq qayta tiklashda jiddiy cheklovlarga ega bo‘lib, bu muammoni hal etish uchun madaniy CLIP kodlovchisi, madaniy moslik yo‘qotish funksiyasi va StyleGAN3 asosidagi diskriminatorni birlashtiruvchi gibrid yondashuv taklif etildi.

Tadqiqot natijalaridan dizayn studiyalari, ijodiy industriya korxonalari va oliy ta’lim muassasalarida “Kompyuter grafikasi va dizayn” yo‘nalishi bo‘yicha o‘quv jarayonida foydalanish mumkin. Keyingi tadqiqotlar milliy madaniy elementlarning maxsus ma’lumotlar to‘plamini shakllantirish, taklif etilgan modelni eksperimental verifikatsiya qilish va madaniy haqqoniylikni baholovchi metrikalarni ishlab chiqishga yo‘naltiriladi.

FOYDALANILGAN ADABIYOTLAR

1. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets // Advances in Neural Information Processing Systems. – 2014. – Vol. 27. – P. 2672–2680.

2. Kingma D. P., Welling M. Auto-encoding variational Bayes // International Conference on Learning Representations (ICLR). – 2014. – arXiv:1312.6114.

3. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 6840–6851.

4. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-resolution image synthesis with latent diffusion models // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2022. – P. 10684–10695.

5. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998–6008.

6. Radford A., Kim J. W., Hallacy C. et al. Learning transferable visual models from natural language supervision // Proceedings of the 38th International Conference on Machine Learning (ICML), PMLR 139. – 2021. – P. 8748–8763.

7. O‘zbekiston Respublikasi Prezidentining 2021-yil 17-fevraldagi “Sun’iy intellekt texnologiyalarini jadal joriy etish uchun shart-sharoitlar yaratish chora-tadbirlari to‘g‘risida”gi PQ-4996-son qarori // Qonunchilik ma’lumotlari milliy bazasi (lex.uz).

8. O‘zbekiston Respublikasi Prezidentining 2024-yil 14-oktabrdagi “2030-yilgacha sun’iy intellekt texnologiyalarini rivojlantirish strategiyasini tasdiqlash to‘g‘risida”gi PQ-358-son qarori // Qonunchilik ma’lumotlari milliy bazasi (lex.uz).

9. Karras T., Laine S., Aila T. A style-based generator architecture for generative adversarial networks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 4401–4410.

10. Karras T., Aittala M., Laine S., Härkönen E., Hellsten J., Lehtinen J., Aila T. Alias-free generative adversarial networks // Advances in Neural Information Processing Systems. – 2021. – Vol. 34. – P. 852–863.

11. Isola P., Zhu J.-Y., Zhou T., Efros A. A. Image-to-image translation with conditional adversarial networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 1125–1134.

12. Zhu J.-Y., Park T., Isola P., Efros A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks // Proceedings of the IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2223–2232.

13. Podell D., English Z., Lacey K., Blattmann A., Dockhorn T., Müller J., Penna J., Rombach R. SDXL: Improving latent diffusion models for high-resolution image synthesis // International Conference on Learning Representations (ICLR). – 2024.

14. Ho J., Salimans T. Classifier-free diffusion guidance // arXiv preprint arXiv:2207.12598. – 2022.

15. Zhang L., Rao A., Agrawala M. Adding conditional control to text-to-image diffusion models // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – 2023. – P. 3836–3847.

16. Ruiz N., Li Y., Jampani V., Pritch Y., Rubinstein M., Aberman K. DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2023. – P. 22500–22510.

17. Hu E. J., Shen Y., Wallis P., Allen-Zhu Z., Li Y., Wang S., Wang L., Chen W. LoRA: Low-rank adaptation of large language models // International Conference on Learning Representations (ICLR). – 2022.

18. Esser P., Rombach R., Ommer B. Taming transformers for high-resolution image synthesis // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2021. – P. 12873–12883.

19. Peebles W., Xie S. Scalable diffusion models with transformers // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – 2023. – P. 4195–4205.

20. Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium // Advances in Neural Information Processing Systems. – 2017. – Vol. 30.

21. Salimans T., Goodfellow I., Zaremba W., Cheung V., Radford A., Chen X. Improved techniques for training GANs // Advances in Neural Information Processing Systems. – 2016. – Vol. 29.

22. Hessel J., Holtzman A., Forbes M., Le Bras R., Choi Y. CLIPScore: A reference-free evaluation metric for image captioning // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2021. – P. 7514–7528.

23. Brock A., Donahue J., Simonyan K. Large scale GAN training for high fidelity natural image synthesis // International Conference on Learning Representations (ICLR). – 2019.

24. Dhariwal P., Nichol A. Diffusion models beat GANs on image synthesis // Advances in Neural Information Processing Systems. – 2021. – Vol. 34. – P. 8780–8794.

25. Sauer A., Schwarz K., Geiger A. StyleGAN-XL: Scaling StyleGAN to large diverse datasets // ACM SIGGRAPH 2022 Conference Proceedings. – 2022.

26. Ramesh A., Dhariwal P., Nichol A., Chu C., Chen M. Hierarchical text-conditional image generation with CLIP latents // arXiv preprint arXiv:2204.06125. – 2022.

27. Saharia C., Chan W., Saxena S. et al. Photorealistic text-to-image diffusion models with deep language understanding // Advances in Neural Information Processing Systems. – 2022. – Vol. 35. – P. 36479–36494.

28. Gatys L. A., Ecker A. S., Bethge M. Image style transfer using convolutional neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 2414–2423.