11. – 501. – Access mode: https://doi.org/10.3389/fnins.2017.00501(date of access: 05.10.2025).
6. Armitage, R. S. Artificial general intelligence and its threat to public health /R. S. Armitage // PLoS/PMC. – 2025. – Access mode: https://pmc.ncbi.nlm.nih.gov/articles/PMC12415933(date of access: 05.10.2025).

DOI 10.70286/ISU-15.10.2025.005

ДИФУЗІЙНІ МОДЕЛІ ЯК СУЧАСНИЙ СТАНДАРТ ВІДНОВЛЕННЯ ВІЗУАЛЬНИХ ДАНИХ

Рогов Андрій Володимирович
кандидат технічних наук, доцент
https://orcid.org/0009-0009-3972-4931
Кафедра інформаційних технологій та математичного моделювання
Харківський національний університет імені В. Н. Каразіна, Україна

У сучасному світі цифрових технологій зображення є одним з найважливіших носіїв інформації. Їхня роль простягається від збереження історичної та культурної спадщини до щоденної комунікації у соціальних мережах, від медичних досліджень до автономних транспортних систем. Візуальні дані не лише відображають реальність, а й стають основою аналітики, моделювання, автоматизованого прийняття рішень та штучного інтелекту. Разом із тим, значна частина цих даних потребує обробки та відновлення. Це можуть бути пошкоджені архівні фотографії, артефакти, що виникають під час стиснення або передавання даних, небажані об’єкти на знімках, дефекти сенсорів камер чи зони, які необхідно приховати з міркувань безпеки та конфіденційності. У сфері цифрової реставрації, відеомонтажу, комп’ютерних ігор та доповненої реальності виникає потреба у технологіях, що здатні реалістично відновлювати або заповнювати відсутні частини зображення відповідно до завдання.

Саме тому одним із найактуальніших завдань сучасного комп’ютерного зору є дорисовування (inpainting) та маскування об’єктів на зображеннях і відео. Розвиток цього напряму безпосередньо пов’язаний із появою нових моделей штучного інтелекту, які дозволяють не просто копіювати текстури, а семантично «розуміти» сцену – передбачати структуру, кольори й контексти об’єктів, яких бракує. Саме поєднання математичних методів відновлення з генеративними можливостями нейронних мереж відкрило шлях до якісно нових підходів у роботі з візуальними даними. Метою дослідження є аналіз еволюції методів відновлення візуальних даних та виявлення переваг дифузійних моделей як сучасного інструменту inpainting.

Ці технології активно застосовуються у сферах цифрової безпеки, кінематографії, AR/VR-середовищах, медичних системах діагностики та реконструкції [5]. Протягом останнього десятиліття відбувся суттєвий перехід

від класичних методів інтерполяції до глибинних генеративних моделей, здатних відновлювати відсутні ділянки сцени на семантичному рівні [7]. Серед таких моделей саме дифузійні підходи (Diffusion Models) за останні три роки стали ключовим стандартом, демонструючи найвищу якість відновлення деталей і стабільність результатів [6, 10].

Перші алгоритми дорисовування базувалися на інтерполяції або розв'язанні рівнянь часткових похідних – методи Telea, Navier-Stokes Inpainting [13, 1]. Вони забезпечували прийнятну якість лише для невеликих або однорідних ділянок зображення. Подальший розвиток відбувся через patch-based методи [3], що використовували текстурне копіювання з контекстних областей. Проте такі підходи залишались локальними й не враховували глобальну структуру сцени.

Поява нейронних мереж привнесла семантичний рівень відновлення. Алгоритми Context Encoders, Partial Convolution, Gated Convolution, DeepFill v2 уперше дозволили генерувати об'єкти, яких не існувало в початкових даних [9, 7]. Водночас GAN-моделі (Generative Adversarial Networks) забезпечили високу реалістичність, але мали нестабільність навчання та можливість появи артефактів на кордонах заміщення [5].

Дифузійні моделі реалізують поступовий процес шумування і відновлення зображення. Під час прямого проходу до даних додається шум, а під час зворотного – нейронна мережа поетапно «очищує» зображення, навчаючись відновлювати структуру сцени з будь-якого рівня спотворення [6]. Цей підхід дозволив відмовитись від змагання двох мереж, як у GAN, і зробив навчання стабільнішим.

Серед найвідоміших архітектур слід визначити DDPM (Denoising Diffusion Probabilistic Models, [6], Improved DDPM [8], та Latent Diffusion Model (LDM) [10], що стала основою Stable Diffusion. Усі вони демонструють значне покращення метрик PSNR, SSIM та FID порівняно з GAN-підходами [4].

Основними перевагами дифузійних моделей вважають наступні:

- висока деталізація та реалістичність результатів;
- стійкість до різних типів шуму й дефектів;
- гнучкість у навчанні на текстових або зображувальних описах;
- універсальність – модель можна донавчати під конкретні сценарії: маскування об'єктів відео, реставрація фото чи створення контенту.

До базових недоліків дифузійних моделей, що суттєво обмежують їхнє застосування, зазвичай відносять велику обчислювальну складність і тривалість генерації. Цей аспект є критичним, оскільки кожна генерація високоякісного зображення або відео вимагає багаторазового проходу через модель із послідовними кроками демо-шумлення (denoising), що створює значне навантаження на апаратне забезпечення, зокрема на GPU (графічні процесори) з великим об'ємом пам'яті. Це безпосередньо впливає на вартість експлуатації та швидкість відгуку системи, роблячи їх менш придатними для сценаріїв, де потрібна генерація в реальному часі або в умовах обмежених ресурсів.

Саме ця проблематика стимулює активний розвиток оптимізованих варіантів [12]. Дослідники зосереджуються на зменшенні кількості необхідних кроків демо-шумлення (наприклад, за допомогою методів DDIM або LoRA), на

оптимізації архітектури самої моделі для зменшення її розміру та на впровадженні ефективніших механізмів уваги. Мета полягає у досягненні балансу між якістю результату та обчислювальною ефективністю.

З метою ширшого розуміння позиції дифузійних моделей у сфері обробки зображень, проведено аналіз методів відновлення зображень, що базуються на різних підходах, із визначенням переваг та недоліків кожного (табл. 1).

Таблиця 1 – Порівняльний аналіз методів відновлення

Підхід	Клас методів	Якість	Переваги	Недоліки
PDE-inpainting	Класичний	Низька	Простота, швидкість	Немає семантики
Patch-based	Текстурний	Середня	Збереження текстур	Не враховує структуру
GAN-моделі	Генеративний	Висока	Семантична узгодженість	Артефакти, нестабільність
Diffusion Models	Новітній	Дуже висока	Реалістичність, стабільність	Висока складність

Джерело: узагальнено автором на підставі [3, 4, 6, 7, 13]

Як видно, дифузійні моделі демонструють найкращий компроміс між якістю, узгодженістю та стабільністю, поступово витісняючи GAN у завданнях inpainting і object-removal [10, 12]. Отримані результати порівняння підтверджують поступовий розвиток методів inpainting у напрямі підвищення семантичної узгодженості реконструкції.

Еволюцію підходів до відновлення зображень представлено на рисунку 1.

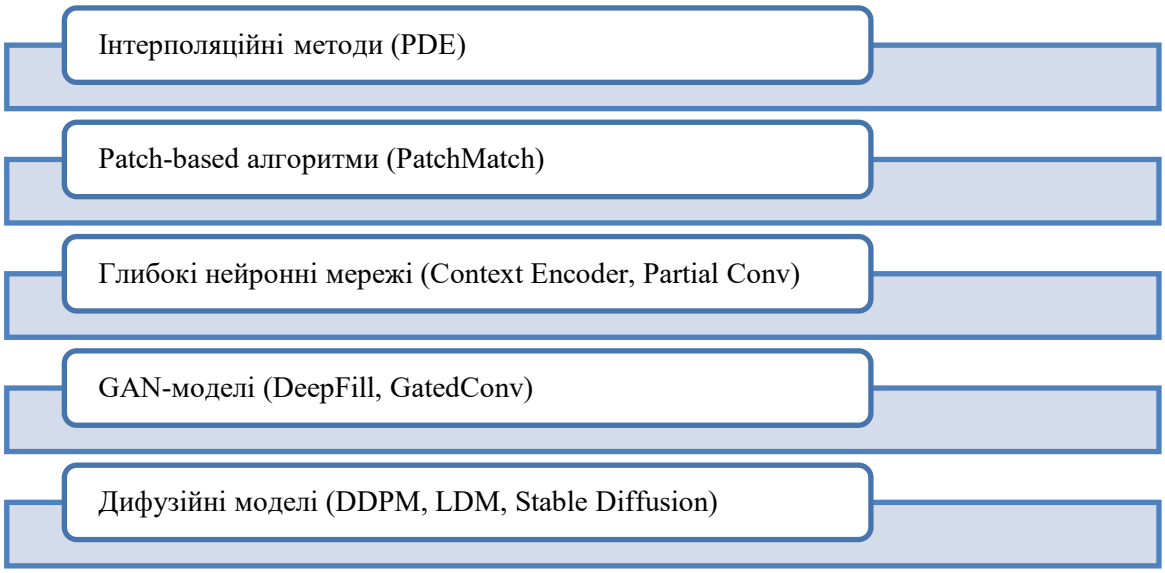

Рисунок 1 – Еволюція методів inpainting у напрямі підвищення якості відновлення зображень

Вона демонструє поступовий перехід від класичних методів інтерполяції та диференціальних рівнянь до генеративних підходів, здатних відтворювати відсутні фрагменти сцени на семантичному рівні. Якщо ранні алгоритми працювали лише з локальними текстурами, то сучасні глибинні та дифузійні моделі забезпечують узгодженість глобальної структури, високу деталізацію та стійкість до шуму.

Одним із ключових викликів при застосуванні дифузійних моделей у відеообробці є забезпечення часової узгодженості кадрів, тобто стабільності

об’єктів і фону під час послідовного відновлення зображень. Використання класичних 2D-дифузійних моделей, спочатку розроблених для статичних зображень, у відеопослідовностях призводить до ефектів «миготіння», розсинхронізації текстур або дрібних зміщень деталей між сусідніми кадрами. Це пов’язано з тим, що кожен кадр обробляється незалежно, без урахування інформації про попередні або наступні зображення сцени. Для усунення мерехтіння також застосовують cycle-consistent training і optical-flow-guided warping як додаткові обмеження до temporal attention, що підвищує стабільність текстур між кадрами [11, 2].

Сучасні дослідження пропонують моделі [11, 2], які поєднують просторову та часову обробку даних, забезпечуючи збереження цілісності руху та стабільності відтворюваних об’єктів. Такі підходи реалізуються через додавання механізмів тривимірної згортки (3D convolution), рекурентних зв’язків між кадрами або умовного навчання на послідовностях відео. Наприклад, у моделях Make-A-Video (Meta AI) та Stable Video Diffusion застосовується принцип temporal attention, який дозволяє дифузійній мережі аналізувати контекст декількох кадрів одночасно, зберігаючи безперервність руху.

Таке поєднання просторово-часової обробки створює основу для автоматизованого маскування об’єктів у відео, коли модель не лише відновлює зображення у межах одного кадру, а й підтримує узгодженість текстур і контурів між ними. У результаті отримується більш плавна, реалістична та послідовна відеопослідовність без видимих артефактів, що робить ці методи перспективними для практичних систем відеоредагування, цифрової реставрації й доповненої реальності.

Таким чином, поява глибинних нейронних мереж, зокрема архітектур на основі CNN, GAN і механізмів уваги, докорінно змінила якість дорисовування, надавши змогу створювати правдоподібні та структурно узгоджені результати. Дифузійні моделі, що представляють сучасний етап розвитку, поєднали найкращі риси попередніх підходів і забезпечили високу деталізацію, здатність працювати з великими ділянками та можливість інтерактивного керування через текстові підказки. Проведений огляд довів, що дорисовування є динамічною та перспективною галуззю, яка поступово перетворилася з локальної техніки відновлення пікселів на універсальний інструмент генеративної обробки зображень. Це робить його особливо актуальним у прикладних завданнях, пов’язаних із маскуванням об’єктів, де від алгоритму вимагається не лише висока якість результату, але й максимальна непомітність втручання для стороннього спостерігача. З урахуванням цього у практиці доцільно поєднувати дифузійну модель із механізмами temporal attention та оптичним потоком для узгодженості маскування у відео в реальному часі. Отже, дифузійні моделі можна розглядати як універсальний підхід до генеративної обробки візуальних даних, що поєднує точність математичного моделювання та креативність нейромережевих генераторів.

Список використаних джерел

1. Image inpainting / M. Bertalmio [et al.]. In SIGGRAPH. 2000. P. 417–424.

2. Blattmann A., Rombach R., Dockhorn T., Ommer B. Stable video diffusion: Learning coherent video generation. arXiv preprint arXiv:2304.08818. 2023. Режим доступу: https://arxiv.org/abs/2304.08818.
3. Criminisi A., Pérez P., Toyama K. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing. 2004. Vol. 13, no. 9. P. 1200–1212.
4. Dhariwal P., Nichol A. Diffusion models beat GANs on image synthesis. Advances in Neural Information Processing Systems. 2021. Vol. 34. P. 8780–8794.
5. Goodfellow I., Bengio Y., Courville A. Deep learning. MIT Press. 2020. 776 p.
6. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 6840–6851.
7. Liu G. [et al.]. Image inpainting for irregular holes using partial convolutions. Proceedings of the European Conference on Computer Vision (ECCV). 2018. P. 85–100.
8. Nichol A. Q., Dhariwal P. Improved denoising diffusion probabilistic models. arXiv preprint arXiv:2102.09672. 2021. Режим доступу: https://arxiv.org/abs/2102.09672.
9. Pathak D. [et al.]. Context encoders: Feature learning by inpainting. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 2536–2544.
10. Rombach R. [et al.]. High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 10684–10695.
11. Singer U. [et al.]. Make-A-Video: Text-to-video generation without text-video data. arXiv preprint arXiv:2209.14792. 2022. Режим доступу: https://arxiv.org/abs/2209.14792.
12. Song J., Meng C., Ermon S. Denoising diffusion implicit models. Proceedings of the International Conference on Learning Representations (ICLR). 2021. Режим доступу: https://arxiv.org/abs/2010.02502.
13. Telea A. An image inpainting technique based on the fast marching method. Journal of Graphics Tools. 2004. Vol. 9, no. 1. P. 25–36.

STEAM VƏ ERKƏN TƏHSIL: UŞAQLARDA MARAĞIN FORMALAŞDIRILMASI

Zeynalova Səbinə Cəfər
Zeynalova Sevil Cəfər
b/m
Azərbaycan Dövlət Pedaqoji Universitetinin Ağcabədi filialı

Açar sözlər: Erkən təhsıl, STEAM təhsili, STEAM təhsil proqramı, XXI əsr bacarıqları, STEAM yanaşması

Xülasə: Məqalədə uşaqların erkən yaş dövründən STEAM təhsilinə başlanılmasının əhəmiyyətindən bəhs olunur.