УДК 004.932
DOI: 10.18523/2617-3808.2025.8.88-96

Чоловський С. О., Бучко О. А.

АУГМЕНТАЦІЯ ДАНИХ У КОМП'ЮТЕРНОМУ ЗОРІ ІЗ ВИКОРИСТАННЯМ ГЕНЕРАТИВНИХ МОДЕЛЕЙ

У статті представлено огляд сучасних підходів до використання генеративних моделей для аугментації даних у задачах комп'ютерного зору. Показано, що ці моделі здатні генерувати високоякісні зображення та різні типи розмітки, що забезпечує їхню ефективність у широкому спектрі прикладних задач. Важливою умовою є недопущення витоку даних під час застосування переднавчених моделей. Проаналізовано методи оцінювання якості синтетичних даних, зокрема використання метрик візуальної якості та відповідності обумовлення, часто із залученням допоміжних моделей. Окреслено перспективні напрями подальших досліджень, зокрема забезпечення якості генерації без використання допоміжних моделей та розроблення методів вибору зразків для аугментації для найбільш ефективного навчання.

Ключові слова: аугментація, комп'ютерний зір, генеративні моделі, генеративно-змагальні мережі, дифузійні мережі.

Вступ

Аугментацією, або доповненням даних, називають набір прийомів, спрямованих на штучне розширення навчальної вибірки шляхом створення модифікованих версій наявних даних. Модель — нейронну мережу, яка навчається для розв'язання певної задачі з використанням аугментованих даних, — далі називатимемо цільовою моделлю, а саму задачу цільовою задачею відповідно.

Для аугментування даних можуть використовуватись як одиночні операції, так і багатоетапні схеми перетворень. Для пошуку оптимальних значень параметрів аугментації використовують, зокрема, навчання з підкріпленням [2], еволюційні підходи [23]. Стратегії, у яких вибір аугментаційних перетворень залежить від властивостей навчальних даних, відомі як метааугментації (meta-learning augmentations) [24]. Дослідження в цьому напрямі ведуть, зокрема, й українські науковці [3]. Додаткова обчислювальна складність визначається складністю самих перетворень зображень, простором та стратегією пошуку параметрів.

Синтетичні дані застосовуються при розв'язанні різних задач машинного навчання, в тому числі комп'ютерного зору. Метою цього огляду є опис та аналіз проблематики застосування генеративних моделей для аугментації даних у задачах комп'ютерного зору.

Базові аугментації

Для аугментування зображень спершу використовували примітивні операції над зображеннями: афінні трансформації, згорткові трансформації, перетворення в просторі кольорів, лінійне змішування та інші. На рис. 1 наведено приклади таких перетворень. У літературі стратегії аугментації, в яких використовують лише примітивні трансформації або їхні композиції, зазвичай класифікують як класичні методи аугментації. Навіть примітивні перетворення дають можливість відобразити значну частину варіативності вхідних даних. Їх широко використовують на практиці, про що свідчить, наприклад, той факт, що їх інтегровано у популярні бібліотеки машинного навчання. Вважають, що класичні методи аугментації дозволяють отримати невеликий, але гарантований приріст показників цільової моделі.

Варто зазначити, що існує формалізація [12; 13], в рамках якої диференційовані трансформації розглядаються як ще один шар нейронної мережі і для підбору параметрів перетворень використовується градієнтний спуск.

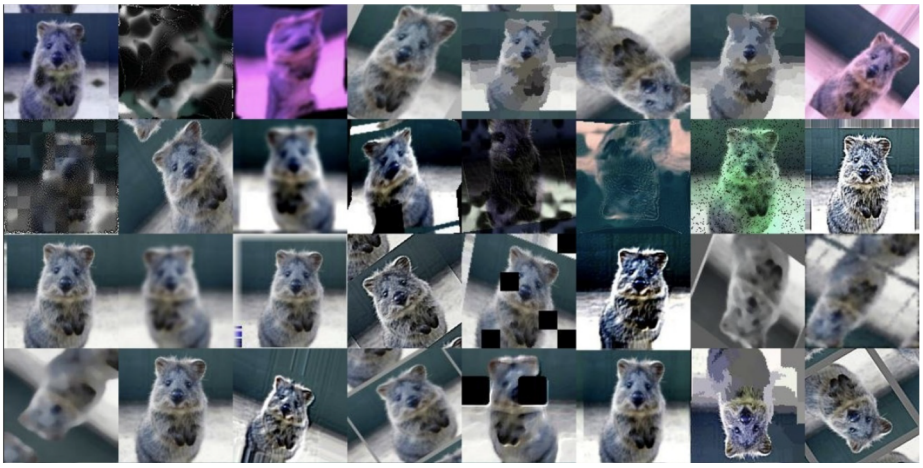

Рис. 1. Приклади базових аугментацій, застосованих до зображення

Застосування генеративних моделей для аугментації даних

Створення синтетичних даних є одним із важливих напрямів використання генеративних моделей, зокрема для аугментації навчальних вибірок. Переважна більшість таких моделей належить до одного з трьох основних класів: варіаційні автоенкодери (VAE), генеративно-змагальні мережі (GAN), дифузійні моделі (diffusion models).

1 *2* *3*

Рис. 2. Схематичне представлення можливостей генеративних аугментацій: вісь абсцис — інформація про положення на зображенні, вісь ординат — інформація про вміст зображення, чорним — елементи з вибірки, жовтим — генеральна сукупність, синім — простір генерації штучних зображень. *1* — у порівнянні з класичними аугментаціями (позначені червоним); *2* — моделювання генеральної сукупності за допомогою генеративної моделі; *3* — простір генерації перевищує простір даних

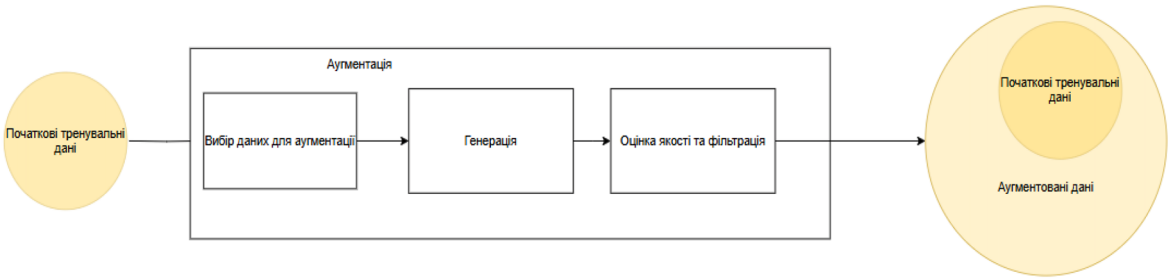

Рис. 3. Типова послідовність етапів аугментації із використанням генеративних моделей

Генеративні моделі дають можливість не лише отримувати більш різноманітні варіанти наявних зображень порівняно з базовими аугментаціями (рис. 2, *1*), а й ефективно моделювати розподіли навчальних даних (рис. 2, *2*). Використання генеративних моделей, попередньо навчених на великих наборах даних (мільярди зображень), дає змогу додати до цільового датасету інформацію про ширший клас можливих зображень (рис. 2, *3*). Однак це потребує накладання додаткових обмежень на згенеровані зображення, щоб уникнути витоку інформації чи спотворення відображення реальних даних. На рис. 3

наведено типову послідовність етапів аугментації із використанням генеративних моделей, яка передбачає такі кроки: формування штучної вибірки (вибір даних для аугментації), генерацію, оцінювання якості та фільтрацію штучно згенерованих даних. Як буде показано нижче, стратегії формування штучних даних можуть варіюватися — від рівномірної вибірки з простору ознак до аугментації конкретних екземплярів навчальних даних на основі оцінок їхньої інформативності.

Використання мереж типу кодувальник — декодувальник для аугментації

Автоенкодер є моделлю, яка кодує дані у вектор простору ознак і декодує їх з нього. Також цю модель називають латентним простором ознак. Ознаки латентного простору кодують як просторові, так і семантичні характеристики зображень.

Визначальною рисою варіаційних автоенкодерів (variational avtoencoder, VAE) є те, що енкодер навчається таким чином, щоб розподіл представлень у латентному просторі наближався до бажаного розподілу, зазвичай нормального. Це дає змогу генерувати нові приклади шляхом вибірки з цього розподілу.

У роботі [25] використано кодувально-декодувальну (encoder-decoder) згорткову нейронну мережу для перенесення стилів, із метою створення варіантів зображень при розв'язанні задач оцінки глибини (monocular depth estimation) і класифікації. Модель, що була попередньо навчена на наборі даних картин відомих художників, використали для створення варіантів фотографій об'єктів і дорожніх ситуацій. Стиль вибирався випадковим чином як вектор у просторі стилів. Проведений експеримент на наборі даних OFFICE, що складається з зображень 31 класу, кожне з яких належить до одного з трьох доменів, показав покращення узагальнюючих властивостей навчених моделей порівняно з класичними методами аугментації. Автори варіювали інтенсивність перенесення стилю та співвідношення штучних і природних зображень.

У роботах [4; 22] згенеровані варіаційним автоенкодером повністю синтетичні зображення апробовано на модельних наборах MNIST, Fashion-MNIST, Omniglot. У роботі [21] при розв'язанні задачі виявлення об'єктів, за допомогою варіаційного автоенкодера, створювали варіанти зображень об'єкта (людини) всередині розміченої рамки (bounding box), але залишали решту зображення незмінною.

Використання генеративно-змагальних мереж для аугментацій

У фреймворці генеративно-змагальних мереж (generative adversarial network, GAN) мережа складається з двох частин — генератора і дискримінатора, які суміщені в процесі навчання, але мають протилежні цільові функції. Генератор намагається створити дані, подібні до справжніх, тоді як дискримінатор прагне відрізнити згенеровані зразки від реальних.

Українські дослідники у своєму дослідженні [7] порівнювали класичні та генеративні аугментації на модельних датасетках MNIST і CIFAR-10. У роботі [19] при розв'язанні задачі класифікації зображень комах порівнюються базові аугментації та підхід із додаванням згенерованих за допомогою генеративно-змагальної мережі зображень. Для якісного аналізу використовується t-sne візуалізація представлень у латентному просторі.

В [11] запропоновано підхід для генерації зображень одночасно з масками сегментації. Автори спиралися на твердження, що для досить реалістичних генеративних моделей інформація про семантику окремих пікселів наявна в латентному просторі, а оскільки маски сегментації структурно простіші за відповідні зображення, навчити модель декодувати в маску сегментації можна з відносно невеликої кількості масок. Для генерації зображень використали мережу StyleGAN, із внутрішніх шарів якої будували карту ознак (feature-map) та отримували маски, класифікуючи кожен піксель за допомогою ансамблю багатошарових перцептронів (MLP). Отримані зображення додатково ранжувалися та фільтрувалися за значенням дивергенції Єнсена — Шеннона для оцінок кожного з пікселів — однієї з варіацій дивергенції Кульбака — Лейблера. Якість згенерованих масок оцінювалася шляхом підрахунку усередненого за категоріях індексу Жаккара (mIoU) з ручною анотацією згенерованих зображень. У роботі [1] цей підхід застосовано для аугментації медичних даних. Порівнювались методи аугментації, що використовують мережу StyleGAN, як у оригінальній роботі, та її наступника, StyleGAN-2.

У роботі [8] для отримання зображень рослин за їх масками сегментації використовували архітектуру SPADE, також відому за її реалізацією GAU-GAN. Отримані таким чином штучні зображення рослин вставляли на місце реальних на природні зображення та використовували для навчання моделі сегментації рослин. Аналогічний підхід застосували у пізнішій роботі з цього напряму [5].

Використання дифузійних мереж для аугментацій

Робота дифузійних моделей відбувається в два етапи: спершу до даних поступово додається шум (процес дифузії), унаслідок чого розподіл вхідних даних наближається до нормального, а потім відбувається відновлення даних, доданий шум усувається крок за кроком (знешумлення, denoising). При навчанні тренуються обидва оператори: дифузії та відновлення. Такий підхід дозволяє генерувати високоякісні зображення та інші типи даних.

Значний прогрес у генерації зображень досягнуто з появою генеративних мереж Stable Diffusion [18]. Автори перенесли дифузію у латентний простір автоенкодера, відділивши таким чином процес генерації від операцій переходу з простору пікселів до простору ознак і навпаки. Це дало можливість зменшити обчислювальну складність, покращити якість і забезпечити контрольованість генерації. Для контролю генерації використовуються мультимодальні обумовлення, найпоширеніші формати — це текстове обумовлення (промпт), зображення та псевдо-зображення (маски, карти глибин та подібні).

У роботі [26] використовували генеративну мережу Stable Diffusion в задачі опису зображень (image captioning). Проведено низку експериментів, для дослідження різних аспектів їхнього підходу, вплив на якість цільової моделі оцінювали на різних бенчмарках, зокрема для навчання на малих даних (few-shot learning). Розглянуто три стратегії побудови описів: часткові, де вказуються не всі об'єкти на зображенні, повні та синтетичні, отримані шляхом перефразування повних за допомогою мовної моделі чи з використанням моделі допоміжної моделі опису зображень. Для оцінки відповідності згенерованих зображень текстовим описам обчислювали показники CLIP-score та VIFIDEL, а для візуальної якості обрали показник MUSIQ замість стандартної відстані сприйняття за Фреше (FID). Оцінювали вплив фільтрації штучних зображень за цими показниками. За результатами проведених експериментів автори зазначили, що значення метрик для штучних даних, які близькі до реальних, не гарантують можливості повної заміни реальних даних штучними.

У [16] також використовується Stable Diffusion — генеративна мережа, де основну увагу приділено задачі класифікації при навчанні на невеликій кількості даних. Замість використання текстових обумовлень використовувалася техніка текстової інверсії (textual inversion) для побудови векторного представлення промпту. Окремо розглянуто умови для унеможливлення витоків даних при оцінюванні на стандартних бенчмарках.

При розв'язанні проблеми підрахунку кількості об'єктів запропоновано підхід, що використовує ControlNET для контрольованої генерації [14]. Щоби обійти необхідність забезпечення якості генерації, автори змінювали співвідношення між штучними та природними зображеннями при навчанні.

У роботі [6] представлено ще один підхід аугментації для задачі розпізнавання об'єктів на основі кількох зображень (few-shot detection), з використанням ControlNet. Нижче розглянемо цей підхід детально. До випадково вибраних зображень застосовується випадкове примітивне перетворення (зсув, обрізання або масштабування) та добувається апріорна інформація (prior extraction) для отримання обумовлювального зображення. На основі розмітки будується текстовий опис (промпт) і з побудованих обумовлень генерується штучне зображення. Якість генерації оцінюється за допомогою метрики CLIP-score, яка порівнює відповідність згенерованої ділянки текстовому опису та визначає її рейтинг серед інших згенерованих зображень для того самого класу об'єктів. Фінальний показник розраховується як середнє усіх згенерованих об'єктів, для подальшого навчання використовується лише певна частка найбільш відповідних зображень. Розглядалися різні стратегії отримання апріорної інформації: детекція контурів і сегментація, а також отримання описів типу «зображення {назви класів через кому}». Інші гіперпараметри: частка справжніх зображень, сила фільтрації.

У роботах [9; 10; 20] по-різному розвинуто ідею аналізу карт збудження шарів уваги (attention maps) для забезпечення одночасної генерації як штучних зображень, так і масок сегментації. Детальне порівняння наведено в табл. 1.

Таблиця 1. **Порівняння підходів одночасної генерації зображень і масок**

Робота	Основний фокус у роботі	Метод отримання масок	Формування текстових описів	Забезпечення якості генерації	Порівняння з іншими методами аугментації
DatasetDM [9]	Запропоновано метод генерації і його апробацію на few-shot learning задачах	Побудова маски сегментації за допомогою окремо навченого декодера	Перефразування інформації з розмітки за допомогою GPT-4	Окремо не розглядається	Виконується порівняння з класичними аугментаціями як частина ablation study
Dataset Diffusion [10]	Дослідження запропонованого методу на широкому наборі задач	Уточнення карти збудження за допомогою шарів уваги	Порівнюються стратегії побудови на основі розмітки та перефразування за допомогою мовних моделей	Якісний аналіз згенерованих даних. Розрахунок IoU зі вручну еталонними масками сегментації	Порівняння з іншим подібними методами аугментації та вичерпний ablation study
Mosaic Fusion [20]	Відтворення схеми аугментації, представленої у більш ранній роботі з використанням дифузійної моделі замість побудови колажів. Апробація запропонованого методу на широкому наборі задач	Бінаризація карт збудження за рівнем відсічення	На основі розмітки у форматі «a photo of {назва класу}»	Окремо не розглядається, виконується фільтрація масок шляхом аналізу компонент зв'язності. Отримані зображення реалістичні в деталях, але мозаїчні	Наявне порівняння з більш ранніми методами аугментації, порівнюються різні ваги однієї генеративної моделі

У роботі [17] використовується модель FreeStyleNet для генерації зображень за допомогою природних масок сегментації. Автори відмовляються від кількісного оцінювання якості згенерованих зображень загалом, натомість пропонують при навчанні ігнорувати лише «складні» пікселі на синтетичних зображеннях. Для обчислення складності окремих пікселів пропонується усереднювати функцію втрат, обчислену для допоміжної сегментаційної моделі. На основі цієї ж метрики, обчисленої для реальних зображень, пропонується генерувати з більшою ймовірністю варіанти зображень зі складнішими масками.

Обговорення

Попри доведену ефективність і гнучкість, використання генеративних моделей для аугментації має принципові обмеження. Розглянемо їх у порядку значущості: ризик витоку даних, необхідність забезпечення якості генерації та визначення меж застосовності конкретних методів. Найважливішим є забезпечення безпеки даних, оскільки витік даних ставить під сумнів як результати теоретичних досліджень, так і практичні застосування. Забезпечення якості генерації є менш критичним, оскільки навіть у випадку неідеальних результатів розглянуті методи можуть бути результативними на практиці у випадкових умовах (у форматі ad hoc).

Одним із важливих аспектів використання генеративних моделей для аугментації є забезпечення безпеки даних. Це пов'язано з тим, що інформація, яка міститься в оригінальних наборах даних, може потрапляти у згенеровані дані. Наприклад, розглянуті нами дифузійні генеративні моделі, навчені на варіантах датасету LAION, що складається з мільярдів зображень, зібраних із відкритих джерел. Існує ризик того, що частину цих зображень буде відтворено при генерації нових варіантів. Якщо згенероване зображення буде надмірно схожим на тестові дані з оцінювального набору даних, це може поставити під сумнів надійність результатів і призвести до витоку даних, що негативно позначиться на достовірності експериментів. Таким чином, є ризик компрометації експерименту. Схематично описаний витік даних зображено на рис. 4.

Окрім загального ризику витоку даних, важливо також забезпечити коректність використання генеративних моделей у рамках конкретної задачі. Деякі з розглянутих робіт застосовували аугментацію у контексті навчання на малих вибірках (few-shot learning). Якщо обмежитися лише вказанням текстового опису «зображення {назва класу}», то після генерації в навчальну вибірку потраплять не тільки варіанти об'єктів, які є у навчальній вибірці, а й абсолютно нові зображення об'єктів цього класу, що ставить під сумнів достовірність результатів експерименту.

Рис. 4. Схема витоку даних

Забезпечення якості генерації даних є ключовим аспектом при використанні генеративних моделей для аугментації зображень. Попри значно вищу якість отриманих зображень порівняно з попередниками, в усіх перелічених вище роботах, що використовують дифузійні моделі, автори так чи інакше вказують на обмеження якості, особливо при генерації складних сцен із великою кількістю об'єктів. Незважаючи на це, деякі дослідження обмежуються тільки опосередкованим оцінюванням якості, порівнюючи показники моделей, які навчалися на повністю штучних і повністю природних даних, що не дає точного уявлення про реальну ефективність застосування генеративних моделей для аугментації.

Поширеною мірою подібності наборів зображень є відстань сприйняття за Фреше (Frechet Inception Distance FID-score), вона вважається золотим стандартом у багатьох задачах генерації зображень і застосовується в розглянутих вище роботах, утім у [17; 26] вказується обмеження використання цієї метрики при аугментації. У розглянутих вище роботах також використовуються специфічні показники, характерні для різних задач. Наприклад, обчислення CLIP-score як показника відповідності між зображенням та його текстовим описом надає певну інтерпретованість результатів генерації. Цей підхід є зручним не тільки для задачі опису зображень, а й для широкого кола інших застосувань, оскільки дає можливість оцінити, наскільки добре згенероване зображення відповідає заданому текстовому опису, що є важливим для задач, де текстова інформація відіграє роль у контексті генерації.

Набувають популярності експерименти з підбором найкращих текстових описів (промптів), зокрема використання мовних моделей для перефразування. Такий підхід дає змогу покращити точність і релевантність згенерованих зображень. У задачах сегментації та детекції часто використовуються показники, як-от індекс Жаккара (intersection over union, IoU, mIoU), для оцінювання якості штучно згенерованих даних. Для обчислення цих метрик або здійснюється ручна розмітка штучного зображення, або використовується додаткова модель сегментації, що дає змогу порівняти згенеровані зображення з реальними.

Подальші дослідження в галузі аугментації даних із використанням генеративних моделей мають кілька перспективних напрямів. Одним з основних є застосування цих методів у сферах з обмеженим обсягом даних, таких як медичні зображення. В розглянутих роботах дослідження виконувалось на великих універсальних наборах даних (COCO, ADE, PASCAL VOC та ін.). З огляду на можливу відсутність якісних допоміжних моделей у спеціалізованих доменах, доцільно обмежити їх використання для оцінювання якості генерації. Як альтернативу оцінці за допомогою допоміжних моделей пропонується використовувати метрику на основі відстані сприйняття за Фреше для порівняння згенерованих і анотованих вручну (еталонних) масок сегментації.

Із подальшим розвитком генеративних моделей нинішні обмеження якості генерації, імовірно, буде подолано, а сучасні архітектури, зокрема дифузійні та генеративно-змагальні моделі, можуть бути замінені більш ефективними підходами. Водночас стратегії вибору, до яких саме екземплярів вибірки доцільно застосовувати аугментацію, значною мірою не залежать від конкретної архітектури генеративної моделі, а отже, можуть досліджуватися окремо. Перспективним напрямом є розвиток методів ранжування даних за ступенем невизначеності [11] або значенням функції втрат [17], а також їхнє поєднання з підходами, подібними до тих, що були запропоновані у [15], де авторами виконувався аналіз представлень «складних» і «простих» для класифікації екземплярів у просторі ознак.

Висновки

У статті розглянуто сучасні підходи до використання генеративних моделей для аугментації даних у задачах комп'ютерного зору. Такі моделі демонструють значно більшу варіативність доповнення даних порівняно з класичними методами та забезпечують вищу гнучкість при розв'язанні спеціалізованих задач. Напрям активно розвивається, поєднуючи методи з різних підгалузей комп'ютерного зору та машинного навчання.

Показано розвиток підходів до використання генеративних моделей від ранніх моделей на модельних наборах, до комплексних рішень, що дозволяють одночасне отримання зображень і відповідної розмітки завдяки механізмам обумовлення. Істотне розширення можливостей досягнуто з появою дифузійних моделей, які забезпечують високу якість та контрольованість синтезу даних.

Проаналізовано сучасні підходи до використання генеративних моделей для аугментації даних, виділено проблематику та окреслено перспективні напрями подальших досліджень. Зокрема, запропоновано метрику оцінювання якості генерації для сегментації, яку планується перевірити в наступних дослідженнях.

Список літератури

1. Application of DatasetGAN in medical imaging: preliminary studies [Electronic resource] / Zong Fan [et al.] // Medical Imaging 2022: Image Processing. — [S. l.], 2022. — Pp. 452–458. — Mode of access: https://doi.org/10.1117/12.2611191 (date of access: 24.06.2025). — Title from screen.
2. AutoAugment: Learning Augmentation Strategies From Data [Electronic resource] / Ekin D. Cubuk [et al.] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — [S. l.], 2019. — Pp. 113–123. — Mode of access: https://doi.org/10.1109/CVPR.2019.00020 (date of access: 24.06.2025). — Title from screen.
3. Baran I. Safe Augmentation: Learning Task-Specific Transformations from Data [Electronic resource] / Irynei Baran, Orest Kupyn, Arseny Kravchenko. — Mode of access: https://doi.org/10.48550/arXiv.1907.12896 (date of access: 24.06.2025). — Title from screen.
4. Chadebec C. Data Augmentation with Variational Autoencoders and Manifold Sampling [Electronic resource] / Clйment Chadebec, Stйphanie Allassonniиre // Deep Generative Models, and Data Augmentation, Labelling, and Imperfections: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings. — Berlin, Heidelberg, 2021. — Pp. 184–192. — Mode of access: https://doi.org/10.1007/978-3-030-88210-5_17 (date of access: 24.06.2025). — Title from screen.
5. Cho S.-B. Enhanced classification performance through GauGAN-based data augmentation for tomato leaf images [Electronic resource] / Seung-Beom Cho, Yu Cheng, Sanghun Sul // IET Image Processing. — 2024. — Vol. 18, no. 14. — Pp. 4887–4897. — Mode of access: https://doi.org/10.1049/ipr2.13069 (date of access: 24.06.2025). — Title from screen.
6. Data Augmentation for Object Detection via Controllable Diffusion Models [Electronic resource] / Haoyang Fang [et al.] // 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). — [S. l.], 2024. — Pp. 1246–1255. — Mode of access: https://doi.org/10.1109/WACV57701.2024.00129 (date of access: 24.06.2025). — Title from screen.
7. Data Augmentation Method Using Generative Adversarial Networks [Electronic resource] / Oleksandr Chaikovskyi [et al.] // Technical Sciences and Technologies. — 2021. — Pp. 83–91. — Mode of access: https://doi.org/10.25140/2411-5363-2021-2(24)-83-91. — Title from screen.
8. Data Augmentation Using GANs for Crop/Weed Segmentation in Precision Farming [Electronic resource] / Mulham Fawakherji [et al.] // 2020 IEEE Conference on Control Technology and Applications (CCTA). — [S. l.], 2020. — Pp. 279–284. — Mode of access: https://doi.org/10.1109/CCTA41146.2020.9206297 (date of access: 24.06.2025). — Title from screen.
9. Dataset Diffusion: Diffusion-based Synthetic Data Generation for Pixel-Level Semantic Segmentation [Electronic resource] / Quang Nguyen [et al.] // Advances in Neural Information Processing Systems. — 2023. — Vol. 36. — Pp. 76872–76892. — Mode of access: https://proceedings.neurips.cc/paper_files/paper/2023/hash/f2957e48240c1d90e62b303574871b47-Abstract-Conference.html (date of access: 24.06.2025). — Title from screen.
10. DatasetDM: synthesizing data with perception annotations using diffusion models / Weijia Wu [et al.] // Proceedings of the 37th International Conference on Neural Information Processing Systems. — Red Hook, [NY], [USA], 2023. — Pp. 54683–54695.
11. DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort [Electronic resource] / Yuxuan Zhang [et al.] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — [S. l.], 2021. — Pp. 10140–10150. — Mode of access: https://doi.org/10.1109/CVPR46437.2021.01001 (date of access: 24.06.2025). — Title from screen.
12. Differentiable Automatic Data Augmentation [Electronic resource] / Yonggang Li [et al.] // Computer Vision — ECCV 2020 / ed. by A. Vedaldi [et al.]. — Cham, 2020. — Pp. 580–595. — Mode of access: https://doi.org/10.1007/978-3-030-58542-6_35. — Title from screen.
13. Differentiable Data Augmentation with Kornia [Electronic resource] / Jian Shi [et al.]. — Mode of access: https://doi.org/10.48550/arXiv.2011.09832 (date of access: 02.02.2024). — Title from screen.
14. Diffusion-based data augmentation for object counting problems [Electronic resource] / Zhen Wang [et al.] // ICASSP 2025 — 2025 IEEE international conference on acoustics, speech and signal processing (ICASSP). — [S. l.], 2025. — Pp. 1–5. — Mode of access: https://doi.org/10.1109/ICASSP49660.2025.10888449 (date of access: 24.06.2025). — Title from screen.
15. Distilling Model Failures as Directions in Latent Space [Electronic resource] / Saachi Jain [et al.]. — Mode of access: https://doi.org/10.48550/arXiv.2206.14754 (date of access: 23.06.2025). — Title from screen.
16. Effective Data Augmentation With Diffusion Models [Electronic resource] / Brandon Trabucco [et al.]. — Mode of access: https://doi.org/10.48550/arXiv.2302.07944 (date of access: 24.06.2025). — Title from screen.
17. FreeMask: synthetic images with dense annotations make stronger segmentation models / Lihe Yang [et al.] // Proceedings of the 37th International Conference on Neural Information Processing Systems. — Red Hook, [NY], [USA], 2023. — Pp. 18659–18675.

18. High-Resolution Image Synthesis with Latent Diffusion Models [Electronic resource] / Robin Rombach [et al.] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — [S. l.], 2022. — Pp. 10674–10685. — Mode of access: https://doi.org/10.1109/CVPR52688.2022.01042 (date of access: 24.06.2025). — Title from screen.
19. Lu C.-Y. Generative Adversarial Network Based Image Augmentation for Insect Pest Classification Enhancement [Electronic resource] / Chen-Yi Lu, Dan Jeric Arcega Rustia, Ta-Te Lin // IFAC-PapersOnLine. — 2019. — Vol. 52, no. 30. — Pp. 1–5. — Mode of access: https://doi.org/10.1016/j.ifacol.2019.12.406 (date of access: 23.06.2025). — Title from screen.
20. MosaicFusion: Diffusion Models as Data Augmenters for Large Vocabulary Instance Segmentation [Electronic resource] / Jiahao Xie [et al.] // Int. J. Comput. Vision. — 2024. — Vol. 133, no. 4. — Pp. 1456–1475. — Mode of access: https://doi.org/10.1007/s11263-024-02223-3 (date of access: 24.06.2025). — Title from screen.
21. Nikolov I. A. Variational Autoencoders for Pedestrian Synthetic Data Augmentation of Existing Datasets: 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP 2024 [Electronic resource] / Ivan Adriyanov Nikolov // Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. — 2024. — Vol. 2. — Pp. 829–836. — Mode of access: https://doi.org/10.5220/0012570700003660 (date of access: 23.06.2025). — Title from screen.
22. Norouzi S. Exemplar VAE: Linking Generative Models, Nearest Neighbor Retrieval, and Data Augmentation [Electronic resource] / Sajad Norouzi, David J. Fleet, Mohammad Norouzi. — Mode of access: https://doi.org/10.48550/arXiv.2004.04795 (date of access: 23.06.2025). — Title from screen.
23. Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules [Electronic resource] / Daniel Ho [et al.]. — Mode of access: https://doi.org/10.48550/arXiv.1905.05393 (date of access: 16.05.2025). — Title from screen.
24. Shorten C. A survey on Image Data Augmentation for Deep Learning [Electronic resource] / Connor Shorten, Taghi M. Khoshgoftaar // Journal of Big Data. — 2019. — Vol. 6, no. 1. — Pp. 60. — Mode of access: https://doi.org/10.1186/s40537-019-0197-0 (date of access: 02.02.2024). — Title from screen.
25. Style Augmentation: Data Augmentation via Style Randomization [Electronic resource] / Philip T. Jackson [et al.]. — Mode of access: https://doi.org/10.48550/arXiv.1809.05375 (date of access: 23.06.2025). — Title from screen.
26. Xiao C. Multimodal Data Augmentation for Image Captioning using Diffusion Models [Electronic resource] / Changrong Xiao, Sean Xin Xu, Kunpeng Zhang // Proceedings of the 1st Workshop on Large Generative Models Meet Multimodal Applications. — [S. l.], 2023. — Pp. 23–33. — Mode of access: https://doi.org/10.1145/3607827.3616839 (date of access: 02.02.2024). — Title from screen.

References

Baran, I., Kupyn, O., & Kravchenko, A. (2019). *Safe Augmentation: Learning Task-Specific Transformations from Data* (arXiv:1907.12896). arXiv. https://doi.org/10.48550/arXiv.1907.12896.

Chadebec, C., & Allassonnière, S. (2021). Data Augmentation with Variational Autoencoders and Manifold Sampling. *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings*, 184–192. https://doi.org/10.1007/978-3-030-88210-5_17.

Chaikovskyi, O., Volokyta, A., Kyrianov, A., & Loutskii, H. (2021). Data Augmentation Method Using Generative Adversarial Networks. *Technical Sciences and Technologies*, 83–91. https://doi.org/10.25140/2411-5363-2021-2(24)-83-91.

Cho, S.-B., Cheng, Y., & Sul, S. (2024). Enhanced classification performance through GauGAN-based data augmentation for tomato leaf images. *IET Image Processing*, *18* (14), 4887–4897. https://doi.org/10.1049/ipr2.13069.

Cubuk, E. D., Zoph, B., Mané, D., Vasudevan, V., & Le, Q. V. (2019). AutoAugment: Learning Augmentation Strategies From Data. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 113–123. https://doi.org/10.1109/CVPR.2019.00020.

Fan, Z., Kelkar, V., Anastasio, M. A., & Li, H. (2022). Application of DatasetGAN in medical imaging: Preliminary studies. *Medical Imaging 2022: Image Processing*, *12032*, 452–458. https://doi.org/10.1117/12.2611191.

Fang, H., Han, B., Zhang, S., Zhou, S., Hu, C., & Ye, W.-M. (2024). Data Augmentation for Object Detection via Controllable Diffusion Models. *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1246–1255. https://doi.org/10.1109/WACV57701.2024.00129.

Fawakherji, M., Potena, C., Prevedello, I., Pretto, A., Bloisi, D. D., & Nardi, D. (2020). Data Augmentation Using GANs for Crop/Weed Segmentation in Precision Farming. *2020 IEEE Conference on Control Technology and Applications (CCTA)*, 279–284. https://doi.org/10.1109/CCTA41146.2020.9206297.

Ho, D., Liang, E., Stoica, I., Abbeel, P., & Chen, X. (2019, May 14). *Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules* (Issue arXiv:1905.05393). arXiv. https://doi.org/10.48550/arXiv.1905.05393.

Jackson, P. T., Atapour-Abarghouei, A., Bonner, S., Breckon, T., & Obara, B. (2019, April 12). *Style Augmentation: Data Augmentation via Style Randomization* (Issue arXiv:1809.05375). arXiv. https://doi.org/10.48550/arXiv.1809.05375.

Jain, S., Lawrence, H., Moitra, A., & Madry, A. (2022, December 2). *Distilling Model Failures as Directions in Latent Space* (Issue arXiv:2206.14754). arXiv. https://doi.org/10.48550/arXiv.2206.14754.

Li, Y., Hu, G., Wang, Y., Hospedales, T., Robertson, N. M., & Yang, Y. (2020). Differentiable Automatic Data Augmentation. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision — ECCV 2020* (pp. 580–595). Springer International Publishing. https://doi.org/10.1007/978-3-030-58542-6_35.

Lu, C.-Y., Arcega Rustia, D. J., & Lin, T.-T. (2019). Generative Adversarial Network Based Image Augmentation for Insect Pest Classification Enhancement. *IFAC-PapersOnLine*, *52* (30), 1–5. https://doi.org/10.1016/j.ifacol.2019.12.406.

Nguyen, Q., Vu, T., Tran, A., & Nguyen, K. (2023). Dataset Diffusion: Diffusion-based Synthetic Data Generation for Pixel-Level Semantic Segmentation. *Advances in Neural Information Processing Systems*, *36*, 76872–76892.

Nikolov, I. A. (2024). Variational Autoencoders for Pedestrian Synthetic Data Augmentation of Existing Datasets: 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP 2024. *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, *2*, 829–836. https://doi.org/10.5220/0012570700003660.

Norouzi, S., Fleet, D. J., & Norouzi, M. (2020, November 24). *Exemplar VAE: Linking Generative Models, Nearest Neighbor Retrieval, and Data Augmentation* (Issue arXiv:2004.04795). arXiv. https://doi.org/10.48550/arXiv.2004.04795.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042.

Shi, J., Riba, E., Mishkin, D., Moreno, F., & Nicolaou, A. (2020, November 19). *Differentiable Data Augmentation with Kornia* (Issue arXiv:2011.09832). arXiv. https://doi.org/10.48550/arXiv.2011.09832.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6* (1), 60. https://doi.org/10.1186/s40537-019-0197-0.

Trabucco, B., Doherty, K., Gurinas, M., & Salakhutdinov, R. (2025, June 10). *Effective Data Augmentation With Diffusion Models* (Issue arXiv:2302.07944). arXiv. https://doi.org/10.48550/arXiv.2302.07944.

Wang, Z., Li, Y., Wan, J., & Vasconcelos, N. (2025). Diffusion-based Data Augmentation for Object Counting Problems. *ICASSP 2025 — 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. https://doi.org/10.1109/ICASSP49660.2025.10888449.

Wu, W., Zhao, Y., Chen, H., Gu, Y., Zhao, R., He, Y., Zhou, H., Shou, M. Z., & Shen, C. (2023). DatasetDM: Synthesizing data with perception annotations using diffusion models. *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 54683–54695.

Xiao, C., Xu, S. X., & Zhang, K. (2023). Multimodal Data Augmentation for Image Captioning using Diffusion Models. *Proceedings of the 1st Workshop on Large Generative Models Meet Multimodal Applications*, 23–33. https://doi.org/10.1145/3607827.3616839.

Xie, J., Li, W., Li, X., Liu, Z., Ong, Y. S., & Loy, C. C. (2024). MosaicFusion: Diffusion Models as Data Augmenters for Large Vocabulary Instance Segmentation. *Int. J. Comput. Vision*, *133* (4), 1456–1475. https://doi.org/10.1007/s11263-024-02223-3.

Yang, L., Xu, X., Kang, B., Shi, Y., & Zhao, H. (2023). FreeMask: Synthetic images with dense annotations make stronger segmentation models. *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 18659–18675.

Zhang, Y., Ling, H., Gao, J., Yin, K., Lafleche, J.-F., Barriuso, A., Torralba, A., & Fidler, S. (2021). *DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort*. 10140–10150. https://doi.org/10.1109/CVPR46437.2021.01001.

S. Cholovskyi, O. Buchko

DATA AUGMENTATION IN COMPUTER VISION USING GENERATIVE MODELS

Modern generative models provide high quality of generation, with their usage considered alongside classic data augmentation techniques. The paper provides a review of existing approaches for data augmentation with generative models in computer vision tasks. Reviewed pipelines utilize enhanced conditioning mechanisms of modern generative models to produce both images and labels for various tasks including image captioning, classification, object detection and segmentation.

Data leak prevention is crucial when large pretrained generative models are used for augmentation. Models like Stable Diffusion were trained on billions of publicly available images, which might also be present in popular datasets. A potential methodological weakness in the application to few-shot learning tasks was identified: images generated based on textual prompts are sampled from all possible images of a certain class, but not only from a few given training examples. Controllable sampling should be introduced to prevent possible data leaks.

Despite the high overall quality of generated images, novel diffusional models are still error-prone in the generation of complex scenes. To mitigate this various quality assessment procedures for generated images are used. These methods include visual naturalness evaluation with Fréchet Inception Distance (FID), prompt correspondence control via CLIP-score, intersection-over-union (IoU) with ground truth or predictions of auxiliary segmentation models for segmentation masks.

Further utilization of generative augmentations in data-scarce domains such as medical imaging is needed. To achieve this, it is preferable to eliminate auxiliary predictors from generation quality assessment. It is proposed to compare generated and natural segmentation masks with the FID-score.

Another area for further research is data augmentation sampling strategies that are less dependent on specific generative pipelines and therefore can be considered separately. Targeted augmentation of hard to predict examples is more effective than uniform sampling. It is planned to improve sampling strategies from reviewed papers in future research.

Keywords: augmentation, computer vision, generative models, generative-adversarial networks, diffusional networks.

Матеріал надійшов 27.06.2025