

Я. О. Ісаєнков¹
О. Б. Мокін¹

САМОВАЛІДОВАНИЙ U-GAN ДЛЯ ТРАНСФОРМАЦІЇ ЦІЛЬОВОГО КЛАСУ В ЗАДАЧАХ СЕГМЕНТАЦІЇ

¹Вінницький національний технічний університет

Розглянуто проблему браку даних для навчання автоматизованих інтелектуальних систем у різних специфічних сферах, таких як медицина, аналіз супутникових знімків, аграрна галузь, екологія, мова. Описано сучасні методи розв'язання цієї проблеми, зокрема аугментація, генеративні змагальні мережі, дифузійні моделі, інпейнтинг. Основну увагу приділено задачі сегментації, де, окрім зображення, необхідно створювати маски для нових об'єктів. Також зазначено проблему суб'єктивного і ручного процесу вибору найкращої епохи під час навчання моделі, і окреслено альтернативи, які можуть допомогти вирішити цю проблему, такі як inception score та frechet inception distance.

Запропоновано покращену модель часткової трансформації цільового класу сегментації, яка включає нові компоненти самовалідації, такі як додаткова функція втрат, що контролює схожість вихідного зображення на вхідне, претреновану модель сегментації і метрику для оцінки якості згенерованих масок з масками сегментації згенерованих зображень. Ці вдосконалення дозволяють системі ефективніше перетворювати фоновий або нульовий клас у цільовий, створювати точніші сегментаційні маски та автоматично вибирати найкращі епохи під час навчання.

Експерименти на наборі даних панорамних знімків зубів показали, що використання цієї технології дозволило збільшити точність сегментації пломб на 0,9 %, підвищивши значення коефіцієнта Жаккара з 90,5 % до 91,4 %. Модель генеративної змагальної мережі тренувалася впродовж 150 епох з автоматичним вибором кращої епохи, якою була 135-та епоха. Якість згенерованих зображень цієї епохи була підтверджено експертною оцінкою. На супутникових знімках кораблів використання моделі показало підвищення точності сегментації з 63,4 % до 65,2 %. Незважаючи на складність даних, модель змогла адекватно трансформувати вхідні дані порожнього моря в об'єкти кораблів. Найкращі результати досягнуто на 82-й епосі, що також збіглося з вибором найкращої епохи експертом, демонструючи важливість автоматичного вибору оптимальної епохи під час навчання для вилучення з цього процесу додаткових суб'єктивних факторів та прискорення підготовки моделей.

Ці результати підтверджують ефективність запропонованого підходу, демонструючи поліпшення метрик в задачах сегментації та автоматизацію процесу вибору моделі. Запропоновані методи та підходи мають потенціал для широкого застосування у різних сферах, сприяючи розвитку нових інтелектуальних систем та підвищенню їхньої точності.

Ключові слова: аугментація, генерування даних, генеративна змагальна мережа, ГЗМ, сегментація, глибоке навчання, U-GAN, U-генератор.

Вступ

Однією з основних проблем створення нових автоматизованих інтелектуальних систем є брак даних для навчання. Найпоширеніша ця проблема у специфічних сферах життя, таких як медицина. Наприклад, для тренування моделі детектування пломб на скані комп'ютерної томографії (КТ) [1], необхідно мати велику кількість зображень, на яких чітко видно ці пломби. Проте, через обмежену кількість доступних медичних даних, створення таких моделей стає викликом.

Аналогічна ситуація спостерігається у сфері аналізу супутникових знімків [2]. Для визначення різних типів об'єктів на супутникових знімках, модель повинна бути навчена на великій кількості фото чи відео матеріалів, де зафіксовані різні типи об'єктів в різних погодних умовах та сезонних варіаціях. Однак, такі дані часто є закритими або поганими за якістю.

Ще одним прикладом є аграрна галузь, де автоматизовані системи використовуються для моніторингу стану посівів. Для того, щоб система могла точно визначати хвороби рослин, дефіцит по-

живних речовин, кількість врожаю необхідно мати велику кількість високоякісних зображень різних культур на різних стадіях їхнього росту. Проте, зібрати такі дані є складним завданням через сезонність та різноманітність кліматичних умов [3].

У сфері екології автоматизовані системи використовуються для моніторингу стану навколишнього середовища, наприклад, для виявлення забруднення води чи повітря [4]. Для цього необхідні дані з різних географічних регіонів і в різних часових інтервалах, що часто є важкодоступним через технічні або фінансові обмеження.

Недостатність даних є значною проблемою у напрямку дослідження мови, особливо коли йдеться про тренування текстових представлень для менш популярних мов, таких як українська [5]. На відміну від англійської, для якої існує величезна кількість текстових даних у різних форматах, ресурси для української мови значно обмежені. Це ускладнює процес створення високоякісних лінгвістичних моделей, здатних адекватно розуміти та обробляти тексти цією мовою.

Таким чином, проблема браку даних для навчання є ключовим викликом у багатьох сферах, що вимагає пошуку нових підходів до збору, обробки та аналізу даних. Цю проблему можна вирішувати за допомогою кількох сучасних методів. Насамперед, метод аугментації [6] дозволяє значно збільшити кількість навчальних даних шляхом створення нових зображень на основі існуючих. Це може включати різні трансформації, такі як обертання, масштабування, зміну яскравості та контрасту, що допомагає моделі навчатися на різноманітніших даних і підвищувати свою стійкість до змін. Іншим підходом є використання генеративних моделей, таких як генеративні змагальні мережі (ГЗМ) [7], чи складніших — дифузійних [8]. Ці моделі здатні створювати нові реалістичні зображення, які можуть бути використані для навчання. Метод інпейнтингу [9] дозволяє заповнювати відсутні, пошкоджені, виділені частини зображень, що також може бути корисним для створення нових навчальних даних.

Однак, у задачах сегментації виникає додаткова проблема: для згенерованих даних потрібні маски. Маски визначають, які частини зображення належать до конкретних об'єктів, і без них модель не може правильно навчатися. Створення масок для згенерованих даних може вимагати значних зусиль та ресурсів, що додає складності до процесу підготовки даних. Тому для розв'язання цієї проблеми необхідно розробляти нові методи автоматичного або напівавтоматичного створення масок, що дозволить ефективніше використовувати згенеровані дані [10].

Під час навчання генеративних змагальних мереж стає гострою проблема вибору найкращої епохи навчання. ГЗМ складаються з двох моделей: генератора і критика, які змагаються між собою. Генератор намагається створювати реалістичні зображення, в той час як критик намагається відрізнити справжні зображення від згенерованих. У процесі навчання обидві моделі поліпшуються, і визначення моменту, коли навчання досягло оптимального результату і не почався процес деградації чи перенавчання, є досить складним завданням.

Вибір найкращої епохи є суб'єктивним процесом, оскільки він часто базується на візуальній оцінці якості згенерованих зображень. Цей підхід може призвести до різних результатів залежно від інтерпретації та досвіду дослідника. Однак існують метрики, які можуть допомогти об'єктивно оцінити якість згенерованих даних та визначити оптимальну епоху для зупинки навчання. Однією з найпоширеніших метрик є Inception Score, яка оснований на двох критеріях: вірність (fidelity), що відповідає за якість створених даних, та різноманітність (diversity), що відповідає за те, наскільки різні дані може створювати модель. Ще однією важливою метрикою є Fréchet Inception Distance (FID), яка вимірює відстань між розподілами реальних і згенерованих зображень. Низьке значення FID вказує на те, що згенеровані зображення мають розподіл, близький до розподілу реальних зображень, що свідчить про високу якість генератора. До того ж, існують спеціалізовані метрики для оцінки якості зображень в конкретних доменах. Наприклад, для медичних зображень можуть використовуватися метрики, що оцінюють точність відтворення патологічних ознак [11].

Таким чином, хоча вибір найкращої епохи під час навчання ГЗМ є суб'єктивним процесом, використання об'єктивних метрик може значно полегшити це завдання і забезпечити надійніші результати. Метрики, такі як Inception Score, FID, влучність (precision) та повнота (recall), дозволяють оцінити якість згенерованих зображень та допомагають дослідникам зробити обґрунтований вибір оптимальної епохи для зупинки навчання, але знову виникає проблема: як оцінити якість згенерованих сегментаційних масок, коли йдеться про задачу сегментації?

Часто виникає ситуація, коли є дані нульового класу (фон або відсутність об'єкта) і дуже сильно незбалансовані дані першого класу (цільові об'єкти). Це створює додаткові труднощі у навчанні моделей, оскільки більшість алгоритмів можуть зміщувати прогнози в бік більшого класу. Але з

іншого боку, присутність цих даних може бути використана в моделі для їхньої трансформації в інший клас, як показано авторами у попередній роботі [10]. Це означає, що дані нульового класу можуть бути використані для генерації нових прикладів першого класу, що допоможе збалансувати набір даних і покращити навчання моделі.

У цій роботі запропоновано покращення наявної моделі часткової трансформації цільового класу сегментації [10] за рахунок введення додаткових компонентів самовалідації, що дозволяє системі:

- якісніше перетворювати нульовий клас у цільовий, зменшуючи зміни у контексті для вхідної картинки;
- створювати точніші сегментаційні маски;
- автоматично вибирати кращу епоху під час навчання.

Вдосконалена модель трансформації цільового класу для задач сегментації

У роботі [10] авторами запропоновано модель часткової трансформації цільового класу сегментації, показано на рис. 1.

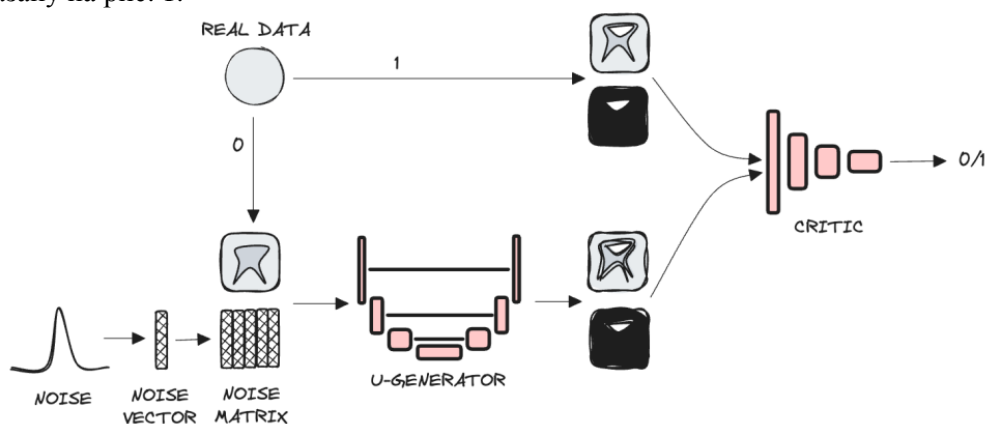


Рис. 1. Схема часткової трансформації цільового класу сегментації

Основні сутності діаграми включають:

- Real data (реальні дані) – доступні розмічені дані фонового (0) та цільового (1) класів.
- Noise (шум), noise vector (вектор шуму), noise matrix (матриця шуму) — алгоритм створення додаткового каналу шуму для того, щоб тими самими вхідними зображеннями можна було створювати різні результати. Спочатку створюється вектор шуму. Далі він дублюється по осі ординат стільки разів, скільки в нього елементів, утворюючи квадратну матрицю, яка і буде додатковим каналом зображення.
- U-generator (U-генератор) — генератор з архітектурою U-GAN [12], що приймає на вхід зображення фонового класу (чорно-біле або кольорове) з додатковим каналом шуму та трансформує його у відповідне зображення цільового класу з маскою сегментації створеного на зображенні об'єкта.

– Critic (критик) — приймає на вхід зображення цільового класу з маскою. Ця пара може бути реальною або згенерованою, і задача критика — розрізнити їх. Цей компонент є ключовим для системи, оскільки він визначає, наскільки добре генератор впорався зі своєю задачею. Використання критика забезпечує зворотний зв'язок, необхідний для вдосконалення генератора.

Попри гарний результат генерування та підвищення точності у задачі сегментації пломб на окремих зубах, у попередній версії підходу була проблема, показана на рис. 2: ліве зображення, верхня щелепа, другий зуб ліворуч (16 — за Міжнародною двоцифровою нумерацією зубів) та перший праворуч (27); праве зображення, перший зуб ліворуч на верхній (17) та нижній (47) щелепі. Зображення, що трансформувалось, могло змінити контрастність, насититись артефактами та шумами. Авторами застосовано проміжне рішення, яке використовувало агрегацію нетрансформованого зображення з трансформованим, використовуючи маску. Але цей додатковий крок ігнорував деякі змінені деталі нового зображення, що не підпадали під маску.

В цій роботі пропонується поліпшити це рішення, шляхом додавання нових сегментаційних та валідаційних компонентів.

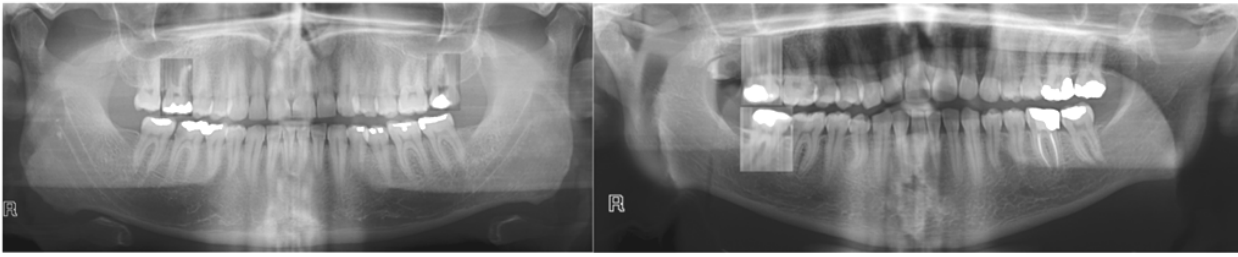


Рис. 2. Приклад недоліків генерування попереднього підходу

Удосконалена архітектура рішення показана на рис. 3, де:

– Additional loss (додаткова втрата) — новий компонент, що відповідає за додатковий штраф (окрім критика) за те, наскільки сильно змінилося зображення. Цей компонент впроваджено для забезпечення більшої стабільності та точності генерації. Використання цього компонента дозволяє зберігати початкові характеристики зображення, мінімізуючи небажані зміни, які можуть виникнути під час трансформації.

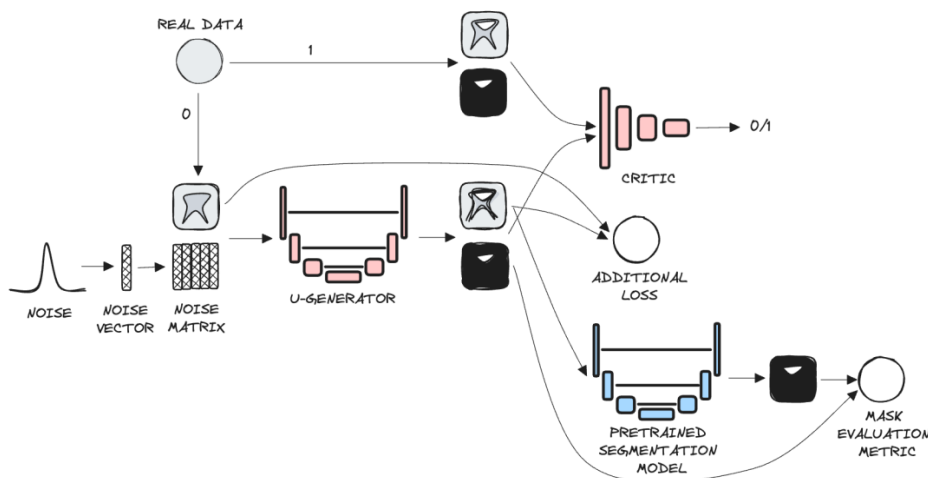


Рис. 3. Схема вдосконаленої трансформації цільового класу сегментації

– Pretrained segmentation model (претренована сегментаційна модель) — модель побудована на доступних даних цільового класу. Приймає на вхід згенероване зображення і намагається зробити маску цільового об'єкта.

– Mask evaluation metric (метрика оцінки маски) – метрика для порівняння маски, згенерованої генератором, з маскою, яку формує претренована сегментаційна модель. У випадку, якщо маска генератора відповідає згенерованому об'єкту і об'єкт настільки точний, що претренована модель змогла визначити його маску, значення метрики буде високим.

Проаналізуємо ці нові компоненти детальніше.

Додаткова втрата

Цей компонент додано, щоб подолати проблему надмірної трансформації зображення фонового класу у разі перетворення у цільовий клас. Підхід додає штраф за те, наскільки зображення відрізняється від вхідного після проходження генератора. Це дозволяє зберігати вхідне зображення та скорочує час (кількість епох), необхідний генератору для створення реалістичних зображень.

Гіперпараметри цього компонента:

– Функція втрат — може бути середньоквадратичним відхиленням (MSE), кореневим середньоквадратичним відхиленням (RMSE), абсолютним відхиленням (MAE) тощо.

– Коефіцієнт впливу — коефіцієнт, що контролює вплив цієї функції на загальну втрату. Може бути більше одиниці (основна втрата) або менше одиниці (додаткова втрата).

– Функція зміни коефіцієнта впливу — сила впливу може контролюватись певною математичною моделлю і зменшуватись в залежності від номера поточної епохи. До прикладу, спочатку визначати більший штраф за відхилення від вхідного зображення, щоб генератор зрозумів концепт, а згодом зменшувати його, дозволяючи моделі більше вчитися від критика, який «знає», як саме має

виглядати об'єкт цільового класу. Функція може мати лінійний характер, вигляд сигмоїди, бути константною.

Загальну функцію втрат можна записати у такому вигляді:

$$\text{TotalLoss} = \text{CriticLoss} + X \cdot \text{SimilarLoss}, \quad (1)$$

де CriticLoss — штраф основної моделі критика, SimilarLoss — додаткова функція втрат, X — коефіцієнт впливу додаткового компонента.

Претренована сегментаційна модель

Додаткова модель, що тренується для задачі сегментації цільового класу. Мета цієї моделі — на даних, створених генератором, знайти маску сегментації. На рис. 4 показано приклад роботи цієї моделі на даних, створених під час ранніх епох навчання. Перший ряд показує зображення і згенеровану маску, другий — те саме зображення і маску сегментації, яку створила ця модель. Через наявні на зображеннях артефактів модель не може якісно сформулювати маску, що буде впливати на результати метрики оцінки маски.

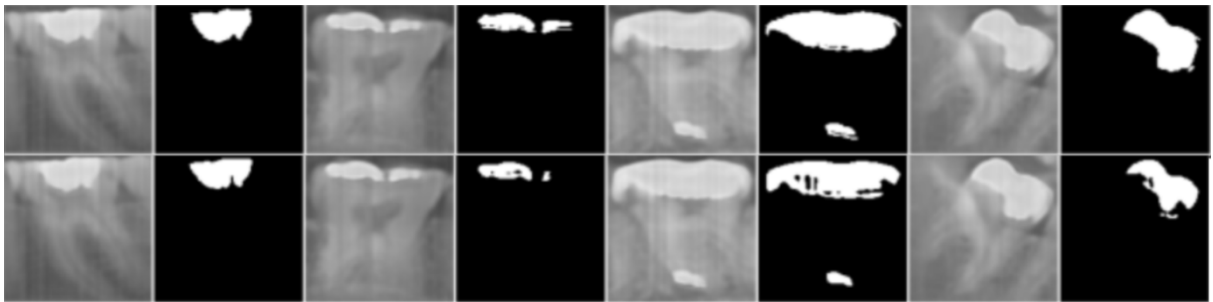


Рис. 4. Приклад генерування зображення та маски (перший ряд) та результат від сегментаційної моделі (другий ряд) на етапі навчання ГЗМ

Метрика оцінки маски

Цей компонент порівнює згенеровану маску з маскою, створеною моделлю сегментації. Чим більше вони схожі, тим більше має бути значення метрики. Однією з таких метрик може бути коефіцієнт Жаккара (Jaccard index або IoU). Він використовується для кількісного визначення просторового перекриття між прогнозованою та справжньою масками. Значення варіюється від 0 (без перекриття) до 1 (повне перекриття)

$$\text{IoU} = \frac{A \cap B}{A \cup B}, \quad (2)$$

де A — згенерована сегментаційна маска, B — передбачена сегментаційна рамка.

На рис. 4 (верхній графік) показано зміни цієї метрики під час навчання моделі. Чітко видно, коли модель почала генерувати хоч трохи схожі на реальні результати, тобто близько 60-ї епохи (сегментаційна модель почала знаходити маску, і ця маска почала збігатися зі згенерованою). Експерименти з моделлю, описані у подальших розділах, показали, що найкращі за метрикою епохи переважно перетинаються з вибраними авторами вручну. Для цього вибрано 10 перших за метрикою епох і також вручну вибрано 10 найкращих результатів серед 150-ти епох. В середньому результат спільних номерів епох між визначеними моделлю і авторами серед цих 10 дорівнює 70 % (тобто 7 з 10 номерів збіглися), використовуючи різні значення для випадкового стану (random seed).

Для порівняння на нижньому графіку рис. 5 показано зміну значень функцій втрат генератора та критика в залежності від номера епохи. Вибрати кращу епоху за цими показниками неможливо, оскільки вони не корелюють з реальною якістю генерування і немає критерію, яке значення є кращим на думку моделі.

Для перевірки ефективності розробленого підходу проведено кілька експериментів на різних наборах даних. Перший експеримент базувався на наборі даних панорамних знімків щелепи [1]. Перша частина цього експерименту заснована на попередній роботі авторів [10], де вдалося досягти підвищення точності сегментації, а також вдосконалити фінальну модель сегментації пломб на зображенні одного зуба розміром 128×128 пікселів.

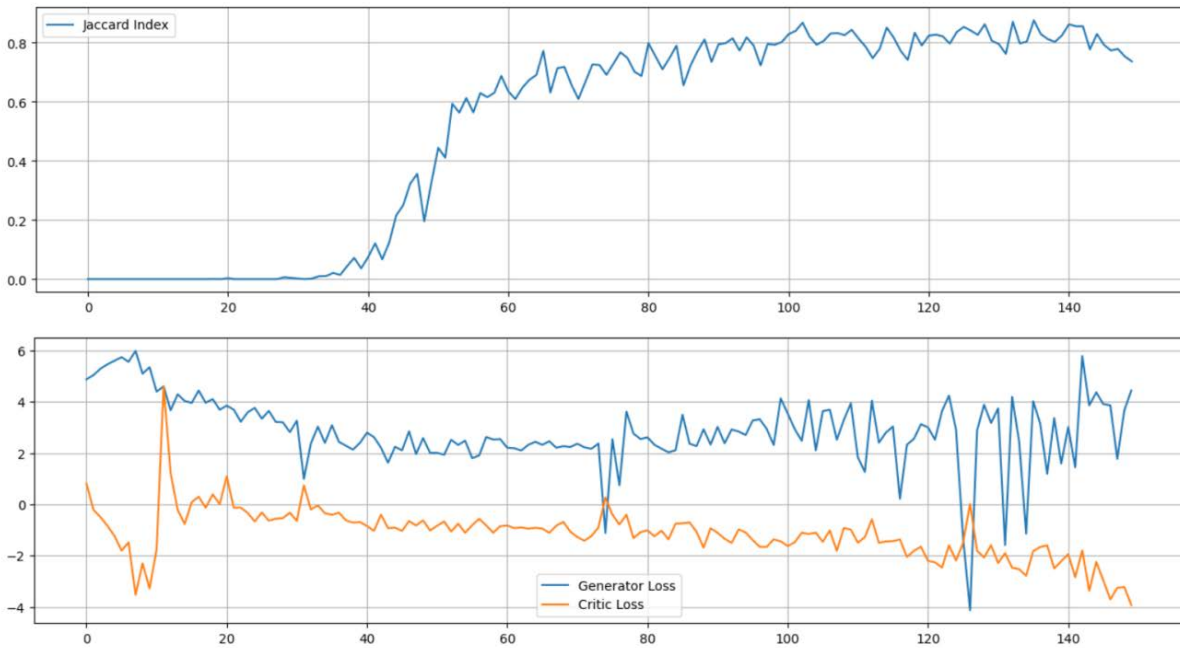


Рис. 5. Графіки зміни значень коефіцієнта Жаккара (Jaccard index), функції втрат генератора (Generator Loss) та критика (Critic Loss) під час навчання ГЗМ.
Аугментація плombs на панорамних стоматологічних рентгенівських знімків

Метою цього експерименту була побудова аналогічної моделі, але за використання додаткових, описаних вище, компонентів. Це дозволить автоматично визначати кращу епоху, оптимізуючи час, витрачений на вибір моделі, а також уникнути кроку постобробки згенерованого зображення. Для цього натреновано модель сегментації на 128 зображеннях та виконано тестування ще на 20 зображеннях. Результати тестування показали високу точність моделі, яка становила 94,2 % за коефіцієнтом Жаккара. Цю модель використано як оцінювальну для перевірки якості згенерованих даних.

Далі вже впродовж 150 епох натреновано модель ГЗМ. Найкраща епоха вибиралася на основі метрики коефіцієнта Жаккара. Для підвищення якості навчання додатково використовувалася функція втрат, заснована на квадратичній помилці. Початковий коефіцієнт для цієї функції встановлено на рівні 10 з лінійним зменшенням до 1 впродовж 150 епох. Найкращою епохою, згідно з обраною метрикою, визначено 135-ту епоху.

На рис. 5 показано графік цього експерименту, де можна спостерігати зміну метрик генерування впродовж навчання. Приклад згенерованих зображень подано на рис. 6. У першому ряду показані вхідні зображення, які передані у генератор, та шум, який дозволяє кожен раз створювати нові варіації. Другий ряд демонструє результати генерації з доданою пломбою та її маскою. Помітно, що зображення, за винятком місця пломби, піддаються мінімальним змінам.

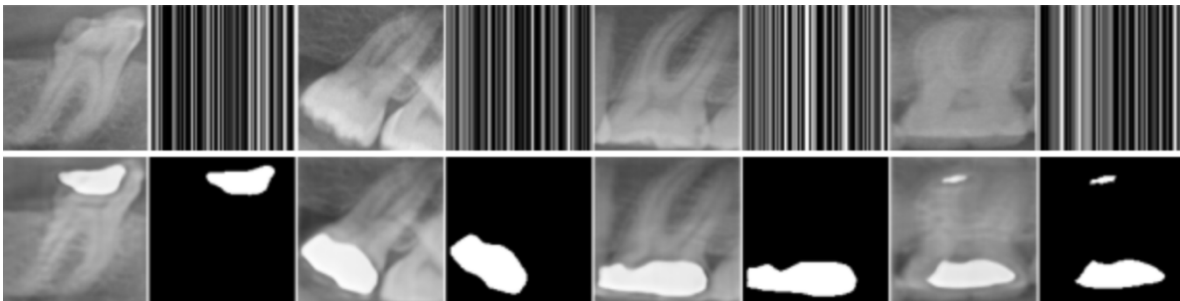


Рис. 6. Результати генерації плombs ГЗМ на кращій епосі

Завдяки роботі генератора базовий набір для тренування валідаційної сегментаційної моделі автоматично розширено на 64 зображення. Валідаційна модель, навчена вже на розширеному наборі, досягла значення метрики 96,2 %, що на 2 % вище ніж модель, навчена без доданих даних. Це підтверджує ефективність використання генеративної моделі для підвищення якості сегментаційної моделі за рахунок автоматичного розширення навчального набору. Подібне поліпшення

(96,1 %) спостерігалось і в попередньому дослідженні авторів, але проміжні етапи допомогли автоматизувати процес тренування і вибору моделі.

Друга частина експерименту полягала у сегментації пломб на повному панорамному знімку зубів. Для цього базовий тренувальний набір даних, який містив 92 картинки, аугментовано та доповнено створеною моделлю, тобто варіаціями кожного наявного зображення за допомогою інструменту для часткової аугментації на основі генеративних моделей. Приклади таких варіацій показано на рис. 7. Варто зазначити, що визначити, які пломби не існували спочатку, є досить складним завданням навіть для людського ока, що підтверджує ефективність запропонованого підходу.

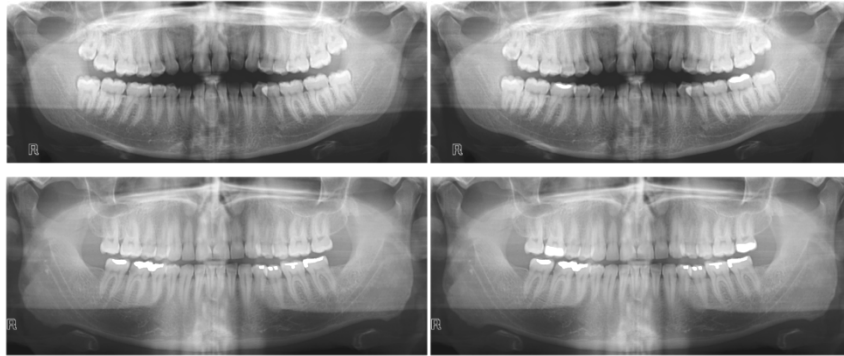


Рис. 7. Приклади аугментування реальних даних КТ-знімків розробленою ГЗМ.
Ліворуч — реальні дані, праворуч — доповненні згенерованими пломбами

Після створення додаткових даних взято стартову модель зі статті про оцінювання якості використання передоброблення панорамних стоматологічних рентгенівських знімків [12]. Її якість на звичайних даних становила 90,5 за коефіцієнтом Жаккара. Під час навчання використовувалися ті самі параметри, але у момент вибору зображення в набір для ітерації обиралося (з ймовірністю 50 %) або реальне зображення, або трансформоване відповідно до описаного вище підходу. Якість такої натренованої моделі становила 91,4, що також свідчить про ефективність рішення.

Аугментація кораблів на супутникових знімках

Дані для цього експерименту взято з конкурсу на платформі Kaggle, де учасники мали за мету досягти найкращого коефіцієнта Жаккара для задачі сегментації кораблів на супутникових знімках [14]. Використання супутникових і дронів знімків стає дедалі важливішим у різних галузях, наприклад, сільськогосподарській, оскільки такі дані дозволяють здійснювати точний моніторинг і аналіз ситуації щодо врожаю або епідемій рослин в реальному часі. Цей експеримент мав на меті не лише перевірити ефективність розробленого підходу в задачі сегментації на медичних знімках, але й оцінити його можливості в інших сферах. Використання супутникових знімків дозволяє протестувати модель у складніших і варіативних умовах, що може відкрити нові перспективи для покращення існуючих рішень у сфері аналізу зображень.

Для проведення експерименту використовувалися ті ж моделі та гіперпараметри, що й у попередньому прикладі панорамних знімків зубів. Відмінністю було те, що на вхід і вихід генератора подавалися кольорові зображення, що збільшувало варіативність і складність даних. Претренована модель оцінки мала точність 81 %. Найкращою епохою генератора вибрано 82-гу епоху, результати якої показані на рис. 8.

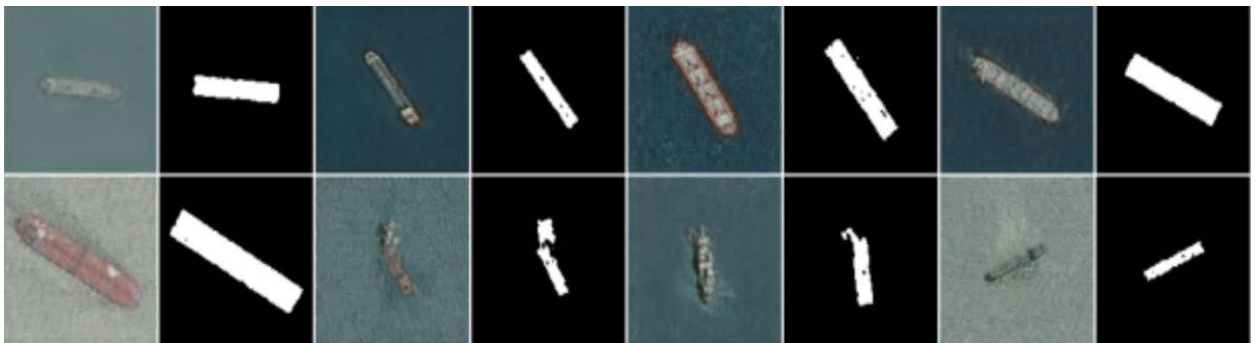


Рис. 8. Результати генерації кораблів ГЗМ на кращій епосі

Як видно з рис. 8, з цією задачею генератор справляється менш якісно через складність даних. Часто згенеровані зображення є розмитими, кораблі не мають чітких контурів, виглядають викривленими або «розламаними», як другий корабель у другому ряду. Маски також часто виходять викривленими, з «дірками» (третій корабель у першому ряду) або з зайвими частинами (третій корабель у другому ряду). Але, незважаючи на неточність генерування, додаткові дані допомогли підвищити якість оцінювальної моделі з 81 % до 84 %.

На рис. 9 показано, як виглядають повноцінні супутникові знімки розміром 768×768 з догенерованими на них кораблями. Перевірено, як запропонований підхід справляється з цією задачею. Як базову модель вибрано U-net34 з публічної роботи цього конкурсу [15]. Взято частину набору даних у розмірі 128 зображень, на які в деяких місцях додано додатково 1–2 цільові кораблі.



Рис. 9. Приклади аугментування супутникових знімків розробленою ГЗМ: зображення 1 та 3 — реальні дані (знімок з Google Maps); 2 та 4 — доповненні згенерованими кораблями

Тренування відбувалося аналогічно до описаного у попередньому розділі. За відсутності додаткових даних точність моделі становила 63,4 %, а з доданими даними — зросла до 65,2 %.

Висновок

У роботі продовжено дослідження проблеми браку даних для навчання автоматизованих інтелектуальних систем у специфічних сферах, таких як медицина, аналіз супутникових знімків, аграрна галузь та екологія. Для розв'язання цієї проблеми запропоновано використання сучасних методів, включаючи аугментацію, генеративні змагальні мережі, інпейнтинг. Проте, у задачах сегментації виникають додаткові виклики, зокрема створення масок для згенерованих даних і вибір найкращої епохи навчання.

Запропоновано вдосконалену модель часткової трансформації цільового класу сегментації, що включає додаткові компоненти самовалідації: додаткову втрату, претреновану сегментаційну модель, метрику оцінки маски. Це дозволяє системі якісніше перетворювати нульовий клас у цільовий, створювати точніші сегментаційні маски та автоматично вибирати кращу епоху в процесі навчання.

Експерименти на різних наборах даних, таких як панорамні знімки щелепи та супутникові знімки кораблів, показали, що запропонований підхід підвищує якість сегментаційних моделей та автоматизує процес вибору фінальної моделі.

Таким чином, використання генеративних моделей для розширення навчальних наборів даних є ефективним рішенням, що дозволяє підвищувати точність моделей машинного та глибокого навчання.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] H. Abdi, S. Kasaei, and M. Mehdizadeh, "Automatic segmentation of mandible in panoramic x-ray," *Journal Med. Imaging (Bellingham)*, vol. 2, no. 4, 044003, 2015. [Electronic resource]. Available: https://www.academia.edu/36038975/PreProcessing_of_Dental_X-Ray_Images_Using_Adaptive_Histogram_Equalization_Method . Accessed on: May 30, 2024.
- [2] A. Samat, et al., "Mapping Blue and Red Color-Coated Steel Sheet Roof Buildings over China Using Sentinel-2A/B MSIL2A Images," *Remote Sens.*, no. 14, 230 p., 2022. <https://doi.org/doi.org/10.3390/rs14010230> .
- [3] C. Zhang, A. Marzougui, and S. Sankaran, "High-resolution satellite imagery applications in crop phenotyping: An overview," *Computers and Electronics in Agriculture*, vol. 175, pp. 105584, 2020. ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2020.105584> .

- [4] В. Мокін, К. Бондалетов, Є. Крижановський, і В. Караваєв, «Метод аугментації текстів про стан масивів вод на основі інтелектуальної прив'язки до багатозв'язних геоінформаційних систем іменованих сутностей», *Вісник Вінницького політехнічного інституту*, № 3, с. 55-65, 2023. <https://doi.org/10.31649/1997-9266-2023-168-3-55-65> .
- [5] O. Bisikalo, O. Kovtun, and V. Kovtun, “Neural Network Concept of Ukrainian-Language Text Embedding,” *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)*, Wrocław, Poland, 2023, pp. 566-569, <https://doi.org/10.1109/ACIT58437.2023.10275511> .
- [6] C. Shorten, and T. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J Big Data*, no. 6, pp.60, 2019. <https://doi.org/10.1186/s40537-019-0197-0> .
- [7] Я. Ісаєнков, і О. Мокін, «Аналіз генеративних моделей глибокого навчання та особливостей їх реалізації на прикладі WGAN», *Вісник Вінницького політехнічного інституту*, № 1, с. 82-94, 2022. <https://doi.org/10.31649/1997-9266-2022-160-1-82-94>.
- [8] M. Chen, S. Mei, J. Fan, and M. Wang, “An overview of diffusion models: Applications, guided generation, statistical rates and optimization,” *arXiv preprint*, arXiv: 2404.07771, April 2024. [Electronic resource]. Available: <https://arxiv.org/abs/2404.07771>. Accessed on: May 30, 2024.
- [9] L. Zhao, and R. Zhao, “Image Inpainting Research Based on Deep Learning,” *International Journal of Advanced Network, Monitoring and Controls*. no. 5. pp. 23-30, 2020. <https://doi.org/10.21307/ijanmc-2020-013> .
- [10] Я. Ісаєнков, і О. Мокін, «Трансформація цільового класу для задачі сегментації з використанням U-GAN», *Вісник Вінницького політехнічного інституту*, № 1, с. 81-87, 2024. <https://doi.org/10.31649/1997-9266-2024-172-1-81-87> .
- [11] K. Falahkheirkhah, et al., “Deepfake Histologic Images for Enhancing Digital Pathology,” *Laboratory Investigation*, vol. 103, Issue 1, pp. 100006, 2023. ISSN 0023-6837. <https://doi.org/10.1016/j.labinv.2022.100006> .
- [12] R. Gulakala, B. Markert, and M. Stoffel, “Generative adversarial network based data augmentation for CNN based detection of Covid-19,” *Sci Rep*, vol. 12, Article no. 19186, 2022. <https://doi.org/10.1038/s41598-022-23692-x> .
- [13] О. В. Коменчук, і О. Б. Мокін, «Аналіз методів передоброблення панорамних стоматологічних рентгенівських знімків для задач сегментації зображень», *Вісник Вінницького політехнічного інституту*, № 5, с. 41-49, 2023. <https://doi.org/10.31649/1997-9266-2023-170-5-41-49> .
- [14] W. Reade, and J. Faudi, “Airbus Ship Detection Challenge,” *Kaggle: Your Machine Learning and Data Science Community*. [Electronic resource]. Available: <https://www.kaggle.com/competitions/airbus-ship-detection> . Accesed: May 30, 2024.
- [15] M. Shugaev, “Unet34 submission TTA (0.699 new public LB),” *Kaggle: Your Machine Learning and Data Science Community*. [Electronic resource]. Available: <https://www.kaggle.com/code/iafoss/unet34-submission-tta-0-699-new-public-lb> . Accesed: May 30, 2024.

Рекомендована кафедрою системного аналізу та інформаційних технологій ВНТУ

Стаття надійшла до редакції 10.06.2024

Ісаєнков Ярослав Олександрович — аспірант кафедри системного аналізу та інформаційних технологій, e-mail: yisaienkov@gmail.com ;

Мокін Олександр Борисович — д-р техн. наук, професор, професор кафедри системного аналізу та інформаційних технологій, e-mail: abmokin@gmail.com .

Вінницький національний технічний університет, Вінниця

Ya. O. Isaienkov¹
O. B. Mokin¹

Self-validated U-GAN for Target Class Transformation in Segmentation Tasks

¹Vinnitsia National Technical University

The paper addresses the problem of data scarcity for training automated intelligent systems in various specific fields such as medicine, satellite image analysis, agriculture, ecology, and language. It describes modern methods for solving this problem, including augmentation, generative adversarial networks, diffusion models, and inpainting. The focus is on the task of segmentation, where it is necessary to create masks for new objects in addition to the image. The subjective and manual process of selecting the best epoch during model training

is also noted, and alternatives that can help solve this problem, such as inception score and frechet inception distance, are outlined.

An improved model of partial transformation of the target class of segmentation is proposed, which includes new self-validation components, such as an additional loss function that controls the similarity of the output image to the input one, a pretrained segmentation model, and a metric for assessing the quality of the generated masks with segmentation masks of generated images. These improvements allow the system to more effectively transform the background or zero class into the target one, create more accurate segmentation masks, and automatically select the best epochs during training.

Experiments on a dataset of panoramic tooth images showed that the use of this technology allowed increasing the accuracy of filling segmentation by 0.9 %, raising the Jaccard coefficient value from 90.5 % to 91.4 %. The generative adversarial network model was trained for 150 epochs with automatic selection of the best epoch, which was the 135th epoch, and the quality of the generated images of this epoch was confirmed by expert evaluation. On satellite images of ships, the use of the model showed an improvement in segmentation accuracy from 63.4% to 65.2 %. Despite the complexity of the data, the model was able to adequately transform the input data of the empty sea into ship objects. The best results were achieved at the 82nd epoch, which also coincided with the expert's choice of the best epoch, demonstrating the importance of automatic selection of the optimal epoch during training to eliminate additional subjective factors from this process and speed up model preparation.

These results confirm the effectiveness of the proposed approach, showing metrics improvements and better automation of the basic approach. The proposed methods and approaches have the potential for wide application in various fields, contributing to the development of new intelligent systems and increasing their accuracy.

Keywords: augmentation, data generation, generative adversarial network, GAN, segmentation, deep learning, U-GAN, U-generator.

Isaienkov Yaroslav — Post-Graduate Student of the Chair of System Analysis and Information Technologies, e-mail: yisaienkov@gmail.com ;

Mokin Oleksandr — Dr. Sc. (Eng.), Professor, Professor of the Chair of System Analysis and Information Technologies, e-mail: abmokin@gmail.com