

5. ГЕНЕРАТИВНІ МЕРЕЖІ ДЛЯ ПІДВИЩЕННЯ ПРОСТОРОВОГО РОЗРІЗНЕННЯ СУПУТНИКОВИХ ДАНИХ НА ОСНОВІ ТРАНСФОРМЕРІВ З УВАГОЮ (TRANSFORMERS WITH ATTENTION)

Алла Лавренюк, доцент

Кафедра математичного моделювання та аналізу даних
Навчально-науковий Фізико-технічний інститут
Національний технічний університет України «Київський
політехнічний інститут імені Ігоря Сікорського»

alla.lavrenyuk@gmail.com

ВСТУП

Протягом останніх кількох років спостерігається стрімке зростання доступності супутникових даних середнього та високого просторового розрізнення, які можна вільно завантажувати з відкритих джерел, таких як платформи Copernicus чи USGS. Вільний доступ до цих даних відкриває нові перспективи для наукових досліджень у різних галузях — від моніторингу довкілля та змін клімату до сільського господарства і просторового планування. Супутникові зображення надають унікальну можливість для дослідження як локальних, так і глобальних змін, що відбуваються на поверхні Землі. Раніше дослідники часто стикалися з вибором: або працювати з даними високого просторового розрізнення на обмеженій території, або ж використовувати супутникові дані з глобальним охопленням, але низьким просторовим розрізненням. Однак сьогодні це протиріччя значною мірою знімається завдяки новітнім супутниковим платформам, що пропонують як високе просторове розрізнення, так і достатню просторову та спектральну деталізацію.

Однією з найпопулярніших і найбільш використовуваних науковцями платформ є супутникова місія Sentinel-2, яка входить до програми Європейського Союзу Copernicus. Sentinel-2 надає спектральні зображення з просторовим розрізненням до 10 для чотирьох каналів: синього (490 нм), зеленого (560 нм), червоного (665 нм) та ближнього інфрачервоного (842 нм). Ці канали

2.5. Генеративні мережі для підвищення просторового розрізнення...

охоплюють важливі діапазони, які дозволяють аналізувати стан рослинності, водойм та урбанізованих територій. Однак Sentinel-2 також має інші, більш специфічні спектральні канали, зокрема червоний край спектра та короткохвильові інфрачервоні канали з просторовим розрізненням 20 м. Ці канали, хоч і мають менше просторове розрізнення, є важливими для детального аналізу екосистем і здоров'я рослинності. У багатьох дослідженнях ці канали не використовуються через нижче просторове розрізнення, що обмежує точність аналізу, особливо на локальних територіях.

Інші популярні джерела супутникових даних, такі як місії Landsat-8 та Landsat-9, також відіграють важливу роль у моніторингу земної поверхні, особливо завдяки своїй довгій історії спостережень (місія Landsat триває з 1972 року). Дані Landsat мають просторове розрізнення 30 м, що є нижчим показником у порівнянні із Sentinel-2, проте вони охоплюють ширший спектральний діапазон, включаючи термальні інфрачервоні канали, що особливо корисні для аналізу температурних змін поверхні. Важливо відзначити, що спільне використання даних Sentinel-2 та Landsat може значно покращити якість моніторингу, оскільки ці платформи взаємодоповнюють одна одну, зокрема щодо часової частоти знімків і спектрального покриття.

Однак використання супутникових даних з різним просторовим розрізненням викликає певні технічні труднощі. Однією з таких проблем є гармонізація зображень з різним розрізненням для отримання єдиного продукту з високою точністю. Вирішення цієї проблеми стало одним із ключових напрямів розвитку в супутниковій індустрії та наукових дослідженнях, спрямованих на підвищення просторового розрізнення зображень. Збільшення просторового розрізнення зображень, відоме як суперрезолюція, дозволяє перетворювати зображення з низькою просторовою деталізацією на зображення вищого просторового розрізнення. Це дає можливість більш точно вивчати об'єкти на поверхні Землі, включаючи моніторинг змін ландшафтів, виявлення аномалій та інші завдання.

Одним із перспективних підходів до розв'язання задачі суперрезолюції є застосування методів глибокого навчання, зокрема генеративних змагальних мереж (GANs). GANs дозволяють створювати нові зображення на основі вже існуючих даних, і сьогодні активно застосовуються для покращення якості зображень у різних галузях. Сучасні дослідження показали, що GANs є ефективними не лише для суперрезолюції зображень, але й для вирішення широкого спектра задач комп'ютерного зору,

таких як сегментація зображень, відновлення пошкоджених даних та інші.

В даному дослідженні пропонується методологія для підвищення просторового розрізнення супутникових зображень Sentinel-2 та Landsat на основі генеративних мереж. Однією з ключових інновацій цього підходу є використання сучасних нейронних мереж із блоками channel attention та self-attention, що дозволяє краще враховувати взаємозв'язки між різними спектральними каналами та підвищити точність генерованих зображень. Це дозволяє не лише отримати продукти високого просторового розрізнення, але й покращити аналіз рослинності та інших елементів ландшафту.

Застосування таких підходів особливо ефективно для аналізу рослинності, оскільки канали червоного краю спектра, які мають нижче просторове розрізнення, містять важливу інформацію про стан рослин та їх здоров'я. Використання методів суперрезолюції дозволяє «підтягнути» ці канали до рівня розрізнення основних каналів Sentinel-2, що значно покращує результати аналізу рослинності.

Експериментальні результати, отримані для даних Sentinel-2, підтвердили ефективність підходу, що розглядається в даному розділі. Нижче продемонстровано, що застосування глибоких нейронних мереж дозволяє значно підвищити якість зображень і створити продукти, які можуть бути використані для детального аналізу земної поверхні. Крім того, проведені експерименти показали, що цей підхід також успішно підвищує просторове розрізнення супутникових зображень Landsat-8 та Landsat-9, що відкриває нові можливості для використання даних цих місій у широкому спектрі застосувань. Особливо перспективним є використання цього підходу для досліджень в галузі екології та сільського господарства, де точність і деталізація супутникових знімків мають критичне значення.

Таким чином, запропонована методологія дозволяє значно підвищити якість супутникових зображень, що відкриває нові можливості для моніторингу природних процесів, вивчення змін у довкіллі та аналізу змін земної поверхні. Завдяки використанню передових методів глибокого навчання, це робить важливий внесок у розвиток сучасних технологій обробки супутникових даних і відкриває нові перспективи для наукових досліджень у цій сфері.

5.1. ОГЛЯД ЛІТЕРАТУРИ

Існує широкий набір супутникових даних, що дозволяють вирішувати різноманітні завдання у галузі спостереження Землі, таких як класифікація земельного покриву, прогнозування врожайності сільськогосподарських культур, виявлення вирубок лісів та контроль природних пожеж. Супутникові зображення надають безцінну інформацію про стан земної поверхні, яку можна використовувати для моніторингу змін довкілля, планування ресурсів та оцінки ризиків у реальному часі. Широкий спектр супутникових місій, доступних для наукових досліджень, включає платформи різного рівня просторового та спектрального розрізнення, що забезпечує можливість адаптації вибору даних до специфічних потреб завдань.

Залежно від мети дослідження та території, що вивчається, важливим є правильний вибір супутникових даних. Наприклад, для точного моніторингу невеликих цільових областей, таких як окремі поля, лісові ділянки або урбанізовані території, найбільш ефективними є дані високого просторового розрізнення [1]. Водночас для глобальних проєктів або національних програм моніторингу, де охоплюються великі площі, перевагу надають супутниковим зображенням середнього просторового розрізнення [2]–[4]. Зокрема, дані, отримані з платформ Sentinel-1 та Sentinel-2, що входять до програми Copernicus, є одними з найбільш популярних та затребуваних у наукових дослідженнях [5], [6]. Це зумовлено їх високим просторовим та спектральним розрізненням, що дозволяє отримати детальну інформацію про стан земної поверхні на глобальному рівні, з можливістю вивчення екосистем, оцінки змін в аграрних і лісових ресурсах, а також спостереження за урбанізованими територіями.

Одним із важливих факторів, що обмежує використання супутникових зображень у завданнях з високими вимогами до деталізації, є обмежена кількість каналів високого просторового розрізнення. Зокрема, для побудови карти класифікації земного покриву з просторовим розрізненням 10 м слід використовувати лише канали із просторовим розрізненням 10 м [7], [8]. У випадку з Sentinel-2 це чотири канали: синій, зелений, червоний та ближній інфрачервоний. Проте супутник також має спектральні канали з просторовим розрізненням 20 м, такі як червоний край спектру, які є надзвичайно корисними для аналізу стану рослинності. Втім, через нижче просторове розрізнення ці канали часто залишаються невикористаними в практичних додатках, що

знижує ефективність класифікаційних алгоритмів та моніторингових завдань.

Крім Sentinel-2, іншими поширеними джерелами супутникових даних є місії Landsat-8 та Landsat-9, які мають просторове розрізнення 30 м. Попри нижче просторове розрізнення ці дані все ж можуть суттєво покращити результати моніторингу завдяки ширшому спектральному покриттю. Проте інтеграція даних з різним просторовим розрізненням, таких як Sentinel-2 і Landsat, вимагає виконання процесу, відомого як підвищення розрізнення, або суперрезолюція. Це завдання виникає в багатьох практичних додатках і полягає в тому, щоб отримати зображення з більшим просторовим розрізненням на основі даних з низьким розрізненням [5]. Взагалі задача підвищення розрізнення виникає в багатьох прикладних завданнях для обробки зображень та відео [9]. Методи підвищення просторового розрізнення зображень можуть використовувати кілька зображень з низьким просторовим розрізненням для створення одного зображення з високим просторовим розрізненням або працювати лише з одним вхідним зображенням з низьким просторовим розрізненням. Вирішення проблеми за участю кількох вхідних зображень детально описано в [9]. Нижче розглядається розв'язання цієї проблеми у випадку, коли є лише одне вхідне зображення.

Одними з найпростіших і найвідоміших методів підвищення просторового розрізнення зображень є інтерполяційні підходи. До таких методів належать метод найближчого сусіда, білінійна інтерполяція, бікубічна інтерполяція та сплайни. Кожен з цих методів має свої особливості та використовується в різних контекстах залежно від вимог до обчислювальної потужності та точності. Метод найближчого сусіда є найпростішим і полягає у виборі значення найближчого пікселя при збільшенні зображення. Хоча цей метод виконується швидко, він створює помітні "блокові" артефакти в зображенні, які спотворюють реальну структуру об'єктів. Білінійна інтерполяція використовує лінійне усереднення значень найближчих пікселів для розрахунку нових точок, що забезпечує більш плавний перехід між пікселями, проте може призводити до розмиття деталей зображення. Бікубічна інтерполяція є вдосконаленим підходом, який враховує не лише найближчі, але й більш віддалені пікселі, що дозволяє отримати ще гладкіші результати, проте також не завжди точно відображає контури об'єктів. Сплайни, як інтерполяційний метод, забезпечують ще вищу якість згладжування, проте, як і попередні

2.5. Генеративні мережі для підвищення просторового розрізнення...

методи, не здатні точно відтворити дрібні деталі зображення та контури об'єктів.

Основною проблемою цих методів є те, що вони не враховують зміст зображення, його контекст чи структури об'єктів. Такі методи працюють виключно з інтенсивністю пікселів, не беручи до уваги форми чи межі об'єктів, що призводить до виникнення артефактів і розмитості, особливо на складних зображеннях з великою кількістю деталей. У випадках, коли важливо зберегти чіткість контурів, таких як у супутникових знімках або медичних зображеннях, ці методи є недостатньо ефективними. З розвитком глибоких методів машинного навчання та нейронних мереж з'явилися більш ефективні підходи для підвищення просторового розрізнення зображень. Завдяки здатності глибоких моделей вивчати складні патерни та враховувати контекст у зображенні, ці методи дозволяють значно покращити якість результуючих зображень, зберігаючи чіткі деталі та структури об'єктів. Нейронні мережі, зокрема згорткові нейронні мережі (CNN), стали популярними завдяки своїй здатності розпізнавати текстури, межі та інші елементи на зображеннях, що робить їх незамінними в завданнях покращення просторового розрізнення.

У 2015 р. було запропоновано метод Super-Resolution Convolutional Neural Network (SRCNN) як один із перших глибоких методів для розв'язання цієї задачі [10]. Однак автори цієї роботи не пропонують рішення «від початку до кінця», а спочатку збільшують зображення до бажаного розміру за допомогою стандартного методу бікубічної інтерполяції, а лише потім застосовують запропоновану нейронну мережу для покращення якості зображення. Метод SRCNN був вдосконалений у [11] моделлю Very Deep Super Resolution (VDSR), яка складалася з 20 шарів порівняно з трьома шарами у моделі SRCNN. Крім того, VDSR намагається оцінити залишкове значення між цільовим зображенням та інтерпольованим, а не повністю здійснює перетворення від одного зображення до іншого, як це було запропоновано у випадку SRCNN. У цих обох методів є недолік використання інтерполяції на початку, що призводить до великої кількості параметрів у моделях та складнощів у навчанні для підвищення просторового розрізнення зображення. Для подолання цих проблем у [12] була представлена модель Fast Super-Resolution Convolutional Neural Networks (FSRCNN). Вона приймає на вхід зображення з низьким просторовим розрізненням без будь-якої попередньої обробки.

Інші методи із подібним підходом також були запропоновані в [13] - [15].

Альтернативним методом для підвищення просторового розрізнення зображення є використання генеративних мереж (GAN) [16], що складаються з двох нейронних мереж: генератора та дискримінатора. Такий підхід був вперше запропонований в [17] з Super Resolution Generative Adversarial Networks (SRGAN). Мережа генератора створює зображення вищого просторового розрізнення, тоді як мережа дискримінатора намагається визначити, чи зображення є реальним, чи згенерованим генератором. Навчання моделі GAN завершується, коли генератор навчився будувати зображення високого просторового розрізнення, які не відрізняються від справжніх зображень для дискримінатора. Кілька публікацій [18] - [20] пропонують використання моделей GAN для розв'язання задач збільшення просторового розрізнення зображень і показують, що ці моделі надають найточніші результати та є найефективнішими для розв'язання задачі підвищення просторового розрізнення.

Однак ці методи застосовувалися до звичайних зображень, а не до супутникових багатоспектральних даних. У публікаціях [21] і [22] було запропоновано використовувати моделі GAN для задач обробки даних спостереження Землі, таких як підвищення просторового розрізнення цифрових моделей рельєфу (DEMs) та збільшення просторового розрізнення даних метеорадару, що дозволяє отримувати більш точні прогнози погоди. Тем не менше, немає досліджень щодо підвищення просторового розрізнення каналів Sentinel-2 чи Landsat-8/9. Таким чином, основна мета даного дослідження полягає у заповненні цієї прогалини і розробці методології підвищення просторового розрізнення всіх інших доступних каналів Sentinel-2 з розрізненням 20 м до 10 м [23], а також всіх доступних каналів Landsat із 30 м до 10 м.

Водночас нейронні мережі із механізмом уваги [24], зокрема трансформери, здобули велику популярність при розв'язанні різних задач, таких як обробка зображень, обробка природної мови, системи рекомендацій, генерація зображень [25] і т.д. головним чином через те, що трансформери перевершили інші звичайні моделі глибокого навчання, які переважно базуються на використанні згорткових шарів без уваги [26]. Тому запропоновано метод для підвищення просторового розрізнення всіх доступних каналів Landsat із 30 м до 10 м на основі моделі GAN, яка складається з трансформерів із блоками уваги [27].

5.2. ПОСТАНОВКА ЗАДАЧІ

Метою цього дослідження є розробка універсального підходу та ефективної моделі до підвищення просторового розрізнення супутникових зображень, який буде ефективним для різних супутникових систем і дозволить покращити просторове розрізнення зображень з 20 або навіть 30 м до 10 м. Основний акцент зроблено на використанні генеративно-змагальних мереж (GAN), які здатні навчатися на зображеннях з високим просторовим розрізненням та застосовувати отримані знання для покращення якості зображень з нижчим просторовим розрізненням.

Задача полягає в адаптації існуючих моделей суперрезолуції до різних типів супутникових зображень, враховуючи їх особливості, такі як спектральні характеристики та наявність шумів. Також важливо забезпечити, щоб запропонований метод був універсальним, тобто міг застосовуватися до різних супутникових даних, таких як Sentinel-2 і Landsat, що мають різне просторове розрізнення.

5.3. РОЗВ'ЯЗАННЯ ЗАДАЧІ

В даному розділі для розв'язання задачі підвищення розрізнення зображень запропоновано використовувати модель GAN, зокрема, для перетворення каналів Sentinel-2 з розрізненням 20 м в зображення з розрізнення 10 м. Схема запропонованої методології показана на рис. 1. Запропонована модель складається з двох нейронних мереж — генератора та дискримінатора. В якості базової моделі була використана ідея SRGAN [17], яка використовувалася для розв'язання задач підвищення розрізнення звичайних зображень, яка була адаптована для роботи з супутниковими даними.

На першому етапі модель GAN навчалася перетворювати на канали червоного, зеленого, синього та NIR з розрізненням 20 м в трансформовані канали з розрізненням 10 м. Після цього навчена модель GAN могла б використовуватися для перетворення інших каналів з просторовим розрізненням 20 м на 10 м. На відміну від традиційних зображень супутникового зображення повинно бути розділено на квадрати.

Розмір вхідного зображення дорівнює $128 \times 128 \times 1$ пікселів, розмір вихідного зображення дорівнює $256 \times 256 \times 1$ пікселів.

Враховуючи, що краї зображення розділяють об'єкти на частини, пропонується генерувати квадрати з перекриттям у 1/4 розміру зображення з усіх сторін. В результаті відсікається центральна частина отриманого зображення ($256 \times 256 \times 1$ пікселів) розміром $128 \times 128 \times 1$ пікселів, за винятком квадратів, які були взяті по краям супутникового зображення.

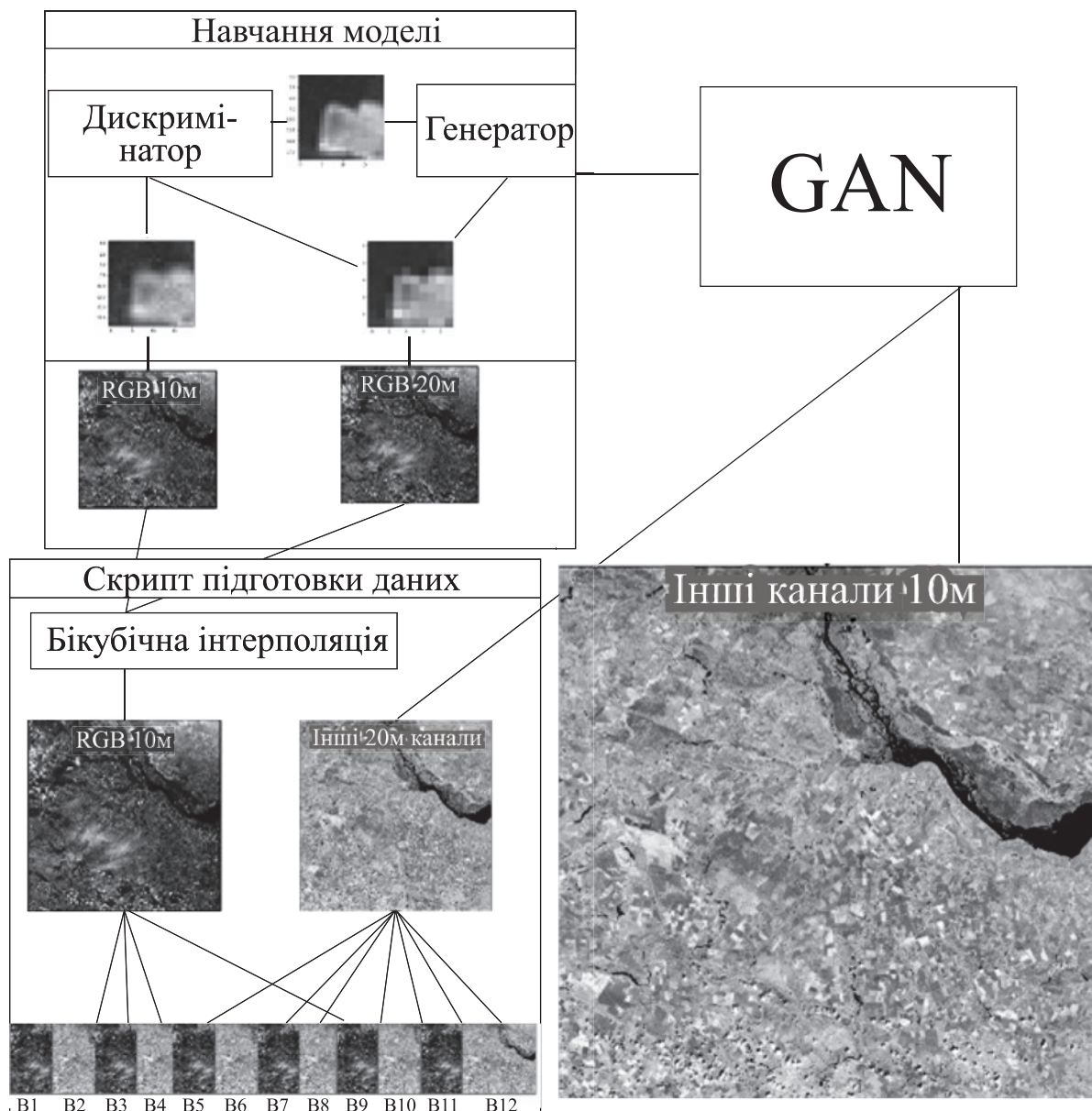


Рис. 1. Схема розв'язання задачі підвищення розрізнення зображень

Дискримінатор — це звичайна нейронна мережа для бінарної класифікації (реальне зображення 10 м чи згенероване з 20 м) з сигмоїдальним останнім шаром. Таким чином, в якості базового

2.5. Генеративні мережі для підвищення просторового розрізнення...

дискримінатора була обрана мережа ResNet, однак для цієї мети можна використовувати будь-яку іншу базову модель. ResNet легко замінити більш сучасними моделями, такими як Efficientnet, Normalizer Free Network (NFnet) або трансформерами.

В даному випадку генератор складається з 16 ідентичних блоків, з'єднаних зі зворотними з'єднаннями за допомогою операції поелементного множення. Кожен блок має два згорткових шари з ядром 3×3 і 64 ознаками, за якими слідує шар нормалізації та функція активації. Для обох наших моделей ми використовували функцію активації Leaky ReLU. Після цього є специфічний блок для підвищення просторового розрізнення карти ознак. Враховуючи, що потрібно отримати точні значення пікселів без розмиття, такі підходи як інтерполяція чи, наприклад, шар для підвищення дискретизації із зворотною згорткою, не є підходящими. Для цього було обрано ефективний субпіксельний згортковий шар, який спеціально розроблений для завдань підвищення розрізнення зображень.

Для спільного навчання дискримінатора та генератора загальна функція втрат складається з двох частин. Перша - стандартна бінарна перехресна ентропія для дискримінатора. Друга - втрата для генератора. У звичайному GAN втрата генератора обчислюється на основі виходів дискримінатора для згенерованих даних. Для завдання підвищення розрізнення зображень основною метою генератора є не тільки вивчення того, як генерувати зображення, схожі на реальні, але й зберігати значення пікселів та форми об'єктів. Таким чином, ми додали втрату вмісту, яка обчислюється на картах ознак моделі Resnet, як в [17], що є більш інваріантною до змін у піксельному просторі, на відміну від втрати MSE, що зазвичай використовується [13]. Отже, якщо X - це вхідне зображення, Y_{pred} - згенероване зображення, (Y_{pred}) - вихід дискримінатора для згенерованого зображення, Y - оригінальне зображення високого просторового розрізнення, втрата GAN може бути оцінена за формулою (5) на основі втрати дискримінатора ($loss1$) та втрати генератора ($loss2$), де α та β - це коефіцієнти ваги.

$$Y_{pred} = generator(X), \quad (1)$$

$$\hat{Y}_{pred} = discriminator(Y_{pred}), \quad (2)$$

$$loss1 = \frac{1}{n} \sum_{i=1}^n Y_{pred_i} \log \hat{Y}_{pred_i} + (1 - Y_{pred_i}) \log (1 - \log \hat{Y}_{pred_i}), \quad (3)$$

$$loss2 = MSE(Y - Y_{pred}), \quad (4)$$

$$loss = \alpha * loss1 + \beta * loss2. \quad (5)$$

Для розв'язання задачі підвищення просторового розрізнення супутникових зображень Landsat з 30 м до 10 м пропонується використовувати дещо іншу модель GAN. Дана модель також складається з двох нейронних мереж: генератора та дискримінатора. Обидві моделі є трансформерами з блоками уваги. Наш підхід базується на моделі Hybrid Attention Transformer (HAT) [28], яка є сучасною та ефективною при розв'язанні задач підвищення розрізнення для звичайних зображень.

Спочатку модель GAN навчається перетворювати трансформовані канали червоного, зеленого та синього кольору з просторовим розрізненням 30 м на оригінальне розрізнення в 10 м. Після цього навчена модель може бути використана для перетворення каналів Landsat з просторовим розрізненням 30 м на розрізнення 10 м. На відміну від традиційних зображень супутникові зображення потрібно розділити на менші квадрати для обробки. Розмір вхідного зображення становить $128 \times 128 \times 1$ пікселів, тоді як розмір вихідного зображення - $384 \times 384 \times 1$ пікселів. Щоб уникнути відсікання об'єктів на краях зображення, генеруються квадрати з перекриттям з усіх сторін, рівним $1/4$ розміру зображення, так само як і у випадку з даними Sentinel-2.

Мережа дискримінатора - це нейронна мережа, яка зазвичай використовується для бінарної класифікації з останнім сигмоїдальним шаром. Найбільш поширеними виборами для дискримінатора є архітектура ResNet, але можна використовувати будь-яку іншу основу, таку як EfficientNet чи Normalizer Free Network. Проте, враховуючи, що трансформери є більш ефективними порівняно з загальнозживаними згортковими мережами, ми зосередились саме на них. Трансформери, такі як ViT, DEiT чи ViLo, мають квадратичну залежність від розміру вхідного зображення, а SWIN [26] має лінійну залежність, тому саме ця архітектура була обрана в якості дискримінатора для розв'язанні поставленої задачі.

Для генератора ми використовуємо мережу трансформера з блоками уваги. SWIN трансформер вже продемонстрував відмінні

2.5. Генеративні мережі для підвищення просторового розрізнення...

результати у завданнях підвищення розрізнення зображень, але, як заявлено в [28], механізм зсуву вікна (основна ідея SWIN) є неефективним для побудови міжвіконного з'єднання. Таким чином, ми покращили його, додавши додаткові блоки уваги, зокрема, увагу перехресного каналу, як у моделі NAT. Загальна архітектура мережі генератора складається з трьох частин: виділення поверхневих ознак, виділення глибоких ознак та відновлення зображення (рис. 2).

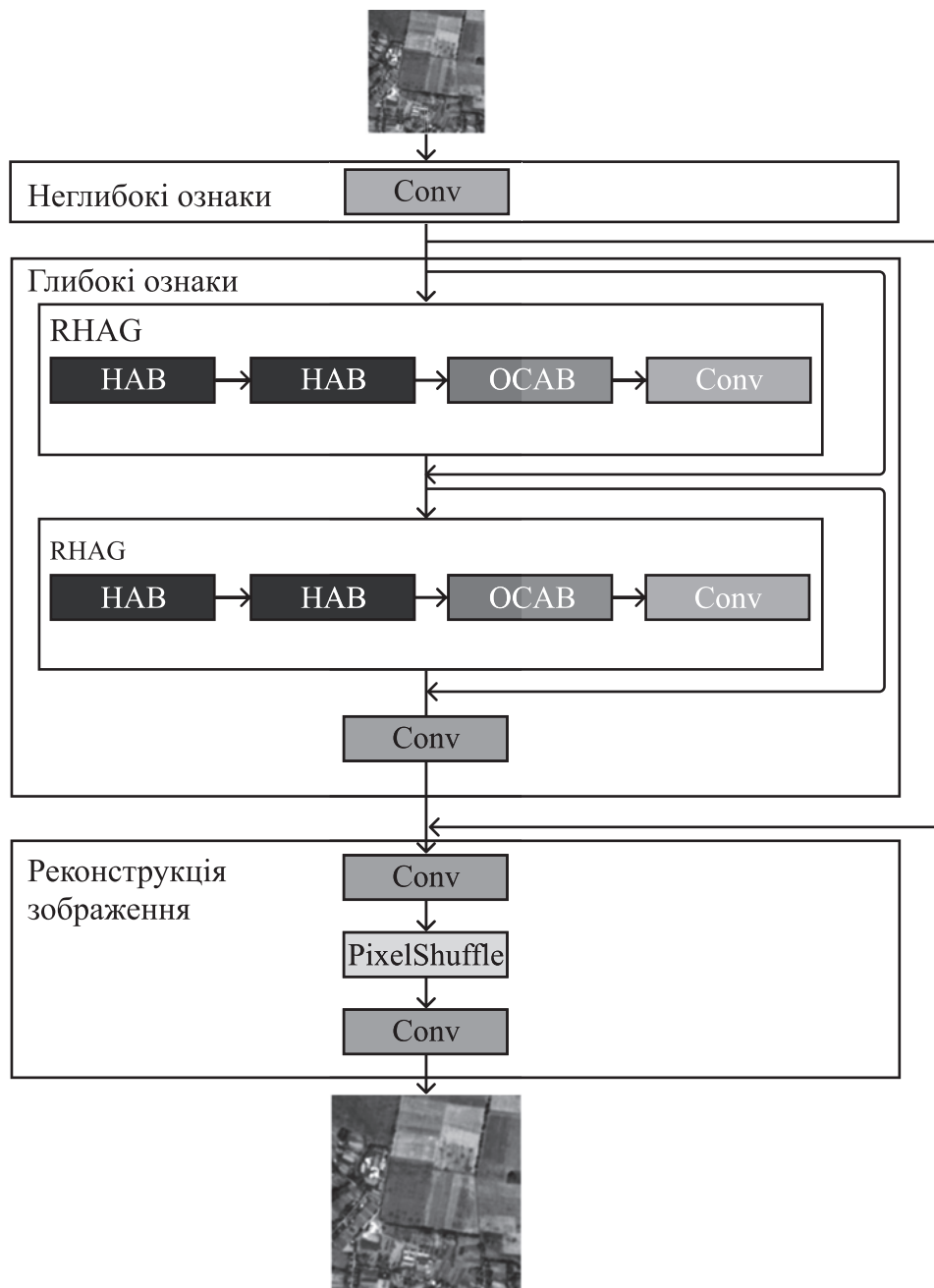


Рис. 2. Загальна архітектура моделі генератора

Для заданого зображення низького просторового розрізнення спочатку використовується один згортковий шар для виділення поверхневих ознак. Вхід перетворюється з простору низької розмірності в простір високої розмірності, створюючи високорозмірне вбудовування (high-dimensional embedding). Глибокі ознаки виділяються за допомогою N послідовних груп гібридної уваги залишкового типу (RHAG) та згорткового шару 3×3 . Після виділення глибоких ознак використовується глобальне залишкове з'єднання для поєднання поверхневих і глибоких ознак. Схожа до SRGAN техніка pixel-shuffle [17] використовується для збільшення просторового розрізнення об'єднаних ознак. В результаті отримуємо зображення більшого просторового розрізнення.

Основним інноваційним блоком генератора є RHAG, який складається з M послідовних гібридних блоків уваги (HAG), блока уваги перехресного перекриття (OCAB) та згорткового шару 3×3 . Блок HUB інтегрує блок згортки на основі каналної уваги в стандартний блок трансформера для подальшого покращення можливостей представлення моделі. Якщо бути більш точним, блок каналної уваги додається до стандартного блоку SWIN трансформера після першого рівня нормалізації та працює паралельно з віконним модулем самоуваги з кількома головами.

5.4. АНАЛІЗ РЕЗУЛЬТАТІВ

Для перевірки ефективності запропонованого підходу було проведено декілька експериментів, які більш детально описані нижче.

Експеримент 1. Підвищення просторового розрізнення всіх доступних каналів Sentinel-2 з просторовим розрізненням 20 м до 10 м.

Для проведення експерименту була завантажена смуга Sentinel-2 (35UQR), що охоплює територію Київської області України за 1 червня 2021 року. Ця смуга була обрана через свою репрезентативність, оскільки вона містить різноманітні типи земельного покриття, такі як сільськогосподарські угіддя, водойми, лісові масиви, штучні об'єкти (міста і дороги), а також необроблювані території.

Така різноманітність ландшафтів дозволяє охопити всі основні категорії земельного покриття, що є важливим для перевірки ефективності запропонованої методології. Крім того,

2.5. Генеративні мережі для підвищення просторового розрізнення...

знімок мав хмарність на рівні 11%, що є прийнятним для аналізу супутникових даних, оскільки хмарність незначно впливає на якість інформації, особливо в спектральних діапазонах, які використовуються для навчання моделі.

Попередня обробка даних здійснювалася за допомогою програмного забезпечення Sen2Core, що є стандартним інструментом для обробки зображень Sentinel-2. Процес обробки включав атмосферну та радіометричну корекцію, що дозволяє враховувати вплив атмосфери на відбивання світла та коригувати похибки, спричинені різницею в умовах освітлення або наявністю аерозолів в атмосфері. Ці кроки є важливими для покращення точності зображень та забезпечення їх придатності для подальшого аналізу, оскільки вони забезпечують реалістичне відображення спектральних характеристик об'єктів на землі.

У процесі навчання моделі використовувалися зелений, синій та ближній інфрачервоний (NIR) канали з просторовим розрізненням 10 м, оскільки вони мають високий рівень деталізації та є важливими для класифікації рослинного покриву, водних об'єктів і штучних структур. Червоний канал, також з просторовим розрізненням 10 м, був використаний для тестування результатів моделі. Для проведення експерименту було створено зображення з нижчим просторовим розрізненням 20 м, трансформували вихідні дані за допомогою кубічної інтерполяції. Це дозволило сформувати навчальний набір, де кожний канал мав свою низькороздільну версію, на якій тренувалася модель, а також використовувалася для подальшого порівняння з оригінальними високороздільними даними.

Навчальні дані були розділені на невеликі квадрати розміром 256×256 пікселів для каналів з просторовим розрізненням 10 м та 128×128 для каналів з розрізненням 20 м. Це дозволило збільшити кількість зразків для тренування моделі та полегшити її навчання на різних ділянках зображення.

Модель була проініціалізована випадковими ваговими коефіцієнтами для всіх шарів. Вхідний шар містив 64 згорткових шари з розміром ядра 9 і активаційною функцією PReLU. Потім сигнал проходив через 16 послідовних блоків залишкового типу з 64 згортковими шарами з розміром ядра 9 та активаційною функцією PReLU, а також 64 згортковими шарами з тим самим розміром ядра та нормалізацією. Після цього слідує ще один блок із 64 згортковими шарами з розміром ядра 3 та нормалізацією. Останнім кроком є передача інформації через транспонований фільтр для збільшення розрізнення та вихідний згортковий фільтр,

який генерує результуюче зображення. Для оцінки втрат генератора використовуються вагові коефіцієнти: 1 для MSE генератора та 0.1 для бінарної крос-ентропії дискримінатора.

Після 10 епох навчання з 10 ітераціями в кожній та розміром порції 32, модель досягла MSE, рівного 2075, для Sentinel-2 нормалізованих пікселів із значеннями від 0 до 10 000. Тестування проводилося розрахунком MSE між згенерованим зображенням GAN з 20 м трансформованого червоного каналу та оригінальним 10 м червоним каналом. На рис. 3 продемонстровано вихід моделі GAN для вхідного зразка.

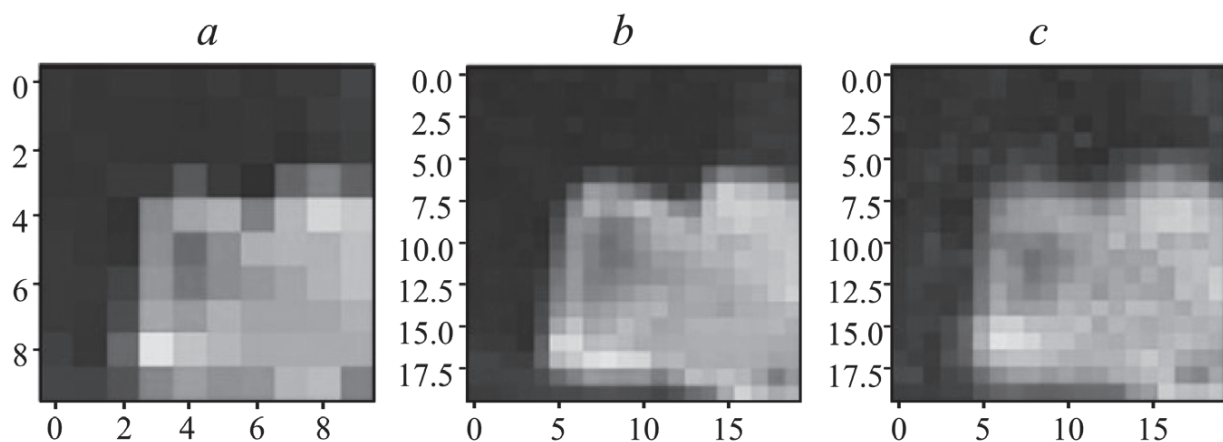


Рис. 3. Підвищення просторового розрізнення на одному зразку: а – трансформоване зображення 20 м, б – оригінальне зображення 10 м, с – вихід GAN

Зазвичай при використанні транспонованих фільтрів результат обирається в центрі вхідного зразка (коли обчислення значень транспонованих значень виконується максимальну кількість разів), а не на кутах, де фільтрацію проводили тільки 1 або кілька разів. Отже, у результуючому зображенні береться лише центральна частина вихідного зразка розміром 128×128 та ковзне вікно для моделювання з розміром кроку 64 пікселі. На рис. 4 продемонстровано результат для повного зображення.

Таким чином, запропоновано методологію для збільшення просторового розрізнення супутникових зображень на основі моделі GAN. Ця модель була успішно навчена на червоних, зелених, синіх та NIR каналах зображень супутників Sentinel-2 і дозволяє збільшити просторове розрізнення всіх інших доступних каналів з розрізненням 20 м до 10 м. Такі канали із збільшеним просторовим розрізненням можуть бути використані для розв'язання різних прикладних задач, особливо крайні червоні

2.5. Генеративні мережі для підвищення просторового розрізнення...

канали для більш точного аналізу рослинності, наприклад, для прогнозу врожаю [29] або визначення глибини по фото [30]–[32]. Обмеження запропонованої методології на основі моделі GAN полягає в можливості збільшення просторового розрізнення із масштабним коефіцієнтом лише 2 або 4. Якщо необхідно збільшити просторове розрізнення із масштабним коефіцієнтом більше 4, потрібно використовувати іншу додаткову інформацію високого просторового розрізнення, наприклад, NDVI, карту земного покриву і т. д.

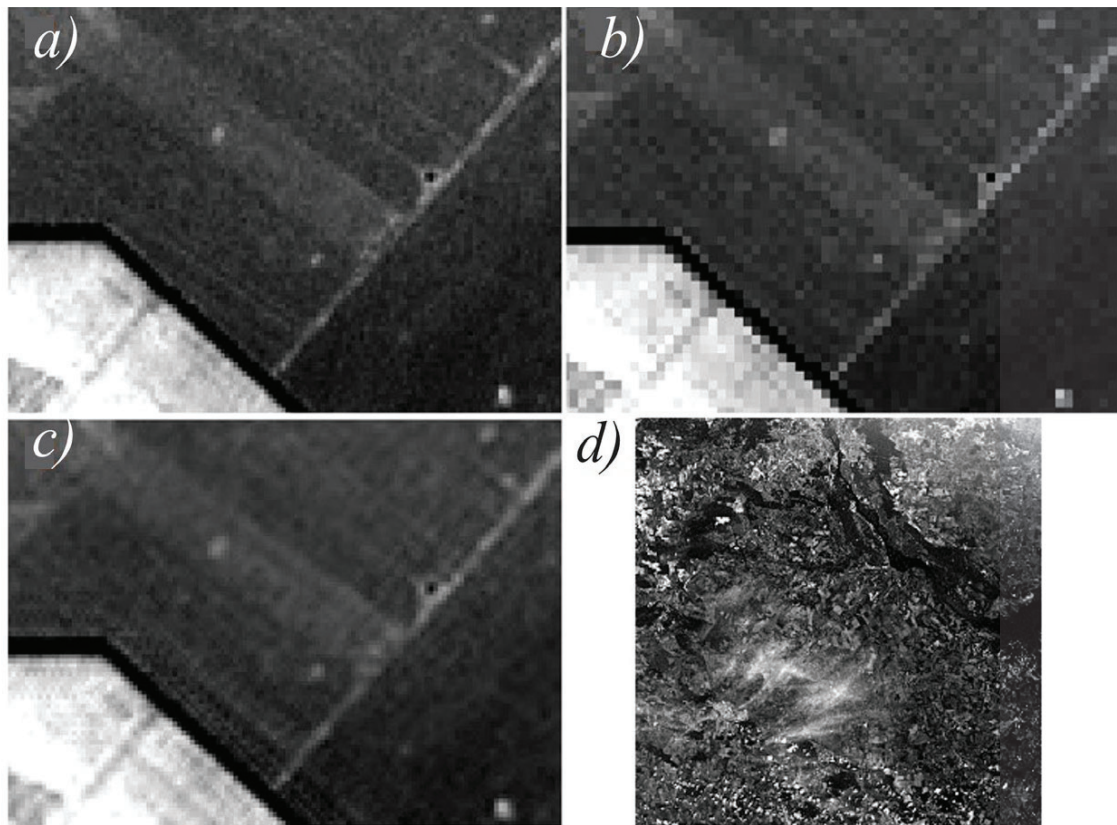


Рис. 4. Результат застосування методу підвищення просторового розрізнення GAN на зображенні: а – оригінальне зображення червоного каналу 10 м, б – трансформоване до 20 м, с – результат моделі, д – результат моделі для повного зображення

Експеримент 2. Підвищення просторового розрізнення всіх доступних каналів Landsat із 30 м до 10 м.

Для проведення експерименту було використано дані супутників Sentinel-2 та Landsat-9 за період з 1 червня 2022 р. до 30 червня 2022 р., експеримент проводився на частині Київської області України. Для забезпечення чітких спостережень для цього

конкретного періоду використані безхмарні композити даних Sentinel-2 та Landsat-9. Набір даних охоплював повне представлення основних типів земельного покриву, включаючи орні землі, водойми, штучні об'єкти, ліси та необроблювані землі. Для підготовки даних використовувалася платформа Google Earth Engine. Червоний, зелений та синій канали використовувалися як для навчання, так і для тестування. Для створення зображення з просторовим розрізненням 30 м із початкових каналів з просторовим розрізненням 10 м було застосовано кубічну інтерполяцію для трансформації зображення. Згодом навчальні дані були розділені на невеликі фрагменти розміром 384×384 пікселів для каналів з розрізненням 10 м і 128×128 пікселів для каналів з розрізненням 30 м.

Попередні дослідження показали, що попереднє навчання є важливим для розв'язання багатьох прикладних задач. Тому що моделі на основі трансформерів потребують більшої кількості даних та ітерацій навчання для отримання базової інформації. Тому моделі генератора та дискримінатора були ініціалізовані з використанням ваг, попередньо отриманих при навчання на наборі даних ImageNet. Для експерименту була встановлено 6 RNAGs та NABs, кількість каналів усієї мережі була 180.

Для оцінки якості отриманої моделі було обрано три стратегії. Перша стратегія полягала в оцінці на основі даних Sentinel-2. Для цього вимірювалася середня похибка (MAE) для спектральних каналів згенерованого зображення (на основі зменшеного зображення Sentinel-2) та оригінальних спектральних каналів із просторовим розрізненням 10 м.

Друга стратегія використовує зменшені зображення Sentinel-2 до 30 м та дані Landsat до 90 м. У цьому випадку модель навчається створювати 30-метровий багатоспектральний канал Landsat на основі 90-метрового каналу. Третя стратегія полягала в аналізі кореляції між згенерованими багатоспектральними каналами та оригінальними даними Sentinel-2. Побудована модель досягла кращих результатів за всіма 3 стратегіями порівняно із класичним підходом GAN, піксельно-орієнтованим підходом глибокої нейронної мережі та методом білінійної регресії, які традиційно використовуються для збільшення розрізнення супутникових даних. Метрика MAE для червоного каналу склала 0,0844, для зеленого каналу - 0,0437, для синього каналу - 0,0507, для всіх каналів - 0,05965. На рис. 5 наведено приклади оригінальних даних Sentinel-2 та Landsat-9 на двох рівнях

2.5. Генеративні мережі для підвищення просторового розрізнення...

масштабування, а також результати підвищення розрізнення, отримані з Landsat-9.

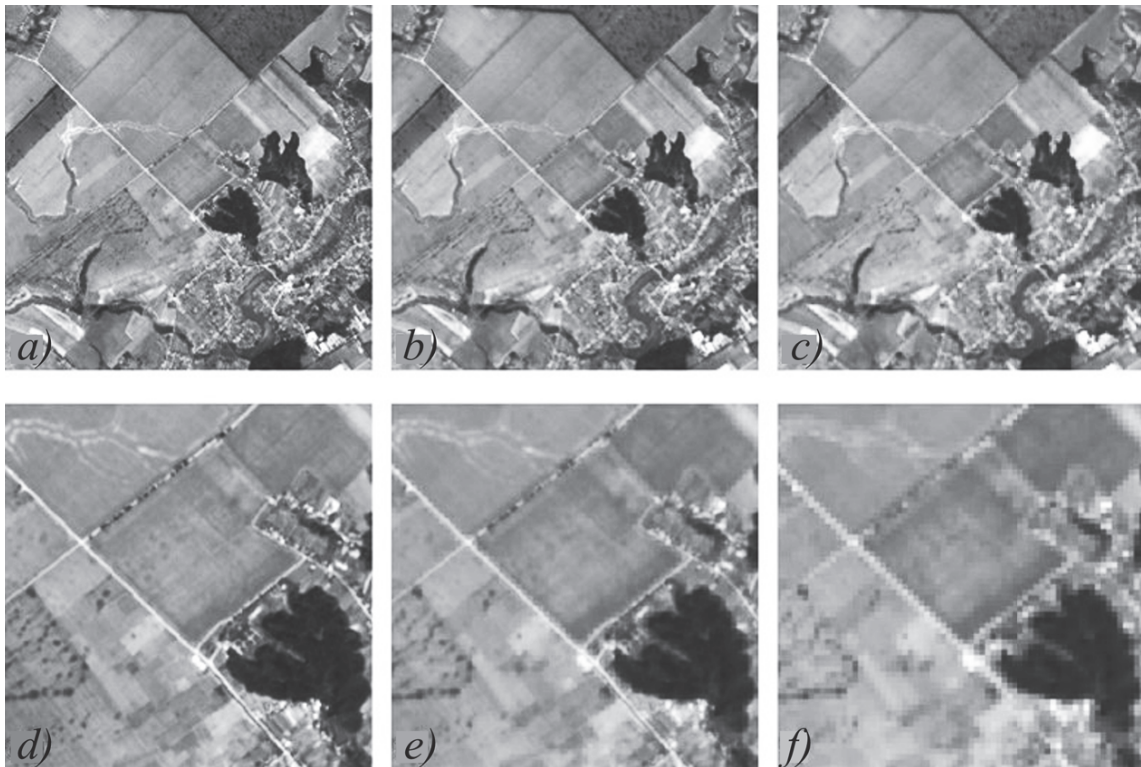


Рис. 5. Результати збільшення розрізнення GAN на зображенні на основі запропонованого піходу: а) - оригінальне RGB-зображення Sentinel-2 (10 м), б) - результат моделі, с) - оригінальне RGB-зображення Landsat-9 (30 м), д) - збільшене оригінальне RGB-зображення Sentinel-2 (10 м), е) - збільшений результат моделі, ф) - збільшене оригінальне RGB-зображення Landsat-9 (30 м)

Набір даних супутникової місії Landsat є одним із найбільших і важливих наукових джерел інформації, яке може бути використано для моніторингу земель, а також аналізу змін земного покриття на багаторічних рядах даних. Однак багато досліджень підтверджують, що просторове розрізнення має важливе значення, тому багато дослідників віддають перевагу даним Sentinel-2. Водночас, найкращі результати моніторингу земель можна досягти лише за умови комбінування даних з різних джерел. Основною перешкодою в гармонізації наборів даних Sentinel і Landsat є їх просторове розрізнення. Це може бути досягнуто з використанням запропонованого нового методу підвищення просторового розрізнення супутникових даних на основі GAN зі згортковими

шарами та блоками уваги, а також з блоками RHAG, які дозволяють вивчати залежності між багатоспектральними та текстурними характеристиками та точно генерувати зображення з просторовим розрізненням 10 м на основі даних Landsat дозволяє отримати.

ВИСНОВКИ

Запропонована методологія показала високу ефективність у підвищенні просторового розрізнення супутникових зображень Sentinel-2 і Landsat-9, що підтверджується результатами проведених експериментів. Використані GAN-моделі забезпечують генерацію зображень з високою точністю, що дозволяє значно покращити деталізацію та якість відображення різних типів земельного покриття.

Одним із ключових аспектів запропонованого підходу є можливість підвищення просторового розрізнення каналів з 20 м до 10 м у Sentinel-2 та з 30 м до 10 м у Landsat-9. Такий підхід є особливо корисним у випадках, коли висока деталізація є критично важливою, наприклад, для моніторингу стану сільськогосподарських угідь, аналізу лісового покриття або вивчення змін у міських середовищах.

З іншого боку, слід зазначити, що в процесі генерації зображень міських районів виникають певні труднощі. Міські ландшафти відрізняються високим рівнем деталізації, складними структурами та значною неоднорідністю, що може створювати проблеми при застосуванні GAN-моделей. Незважаючи на те, що наш підхід дозволяє досягти покращення якості зображень для природних ландшафтів, для міських середовищ потрібні додаткові дослідження та подальша модифікація методики.

Запропонований підхід до підвищення просторового розрізнення на основі GAN значно перевершує традиційні методи, такі як білінійна регресія та піксельно-орієнтовані підходи глибокого навчання. Використання генеративних моделей забезпечує більш точну передачу деталей і текстур зображень, що є критично важливим для аналізу супутникових даних.

Особливо варто відзначити переваги використання блоків уваги (Attention Blocks) та адаптивних блоків залишкової мережі (Residual High Attention Blocks, RHAGs), які дозволяють покращити здатність моделі до навчання на складних зображеннях. Ці блоки забезпечують краще розпізнавання

2.5. Генеративні мережі для підвищення просторового розрізнення...

ключових характеристик зображень, таких як структура ландшафту, різниця в спектральних властивостях різних типів покриву та інші текстурні особливості.

Незважаючи на переваги, запропонований підхід має деякі обмеження. Як було зазначено раніше, труднощі виникають при обробці зображень міського середовища, що може бути пов'язано з недостатньою складністю моделі для адекватного врахування всієї різноманітності текстур та деталей у цих районах. Відповідно, подальші дослідження повинні бути спрямовані на вдосконалення моделі шляхом додавання нових блоків, збільшення кількості шарів або залучення додаткових даних.

Окремої уваги заслуговує аналіз ефективності підходу для різних типів земельного покриву. Наші експерименти охоплювали кілька основних типів ландшафтів, включаючи сільськогосподарські угіддя, водойми, ліси, необроблювані землі та штучні об'єкти. Зображення кожного типу ландшафту мають свої особливості, що впливають на результати генерації.

Для природних ландшафтів, таких як ліси та сільськогосподарські угіддя, модель показала відмінні результати. Це пов'язано з відносно однорідною текстурою та меншою кількістю складних структур у порівнянні з міськими районами. Наприклад, для лісових масивів модель змогла точно відтворити текстуру дерев, зберігаючи високу деталізацію та коректно передаючи зміни у спектральних характеристиках різних ділянок.

Складнішим виявилось завдання генерації зображень водойм. Це пояснюється специфічністю водних поверхонь, які можуть сильно варіюватися залежно від наявності рослинності, глибини, ступеню прозорості води та інших характеристик. У деяких випадках модель не справлялась з точним відтворенням дрібних деталей, наприклад, берегових ліній або окремих елементів, що знаходяться на поверхні води. Проте, загалом, результати генерації для водойм були достатньо точними для використання у практичних завданнях, таких як моніторинг водних ресурсів.

Щодо необроблюваних земель, розроблена модель також показала високу ефективність, що, ймовірно, пов'язано з відносно низькою складністю текстур цього типу ландшафту. Необроблювані землі часто мають менш виражену структуру, що полегшує завдання генерації. Крім того, для цього типу ландшафту модель виявила високу точність у відтворенні спектральних характеристик, що може бути корисним для аналізу змін в умовах посухи або інших природних явищ.

Гармонізація даних з різних супутникових місій є важливим аспектом для забезпечення більш точного аналізу та моніторингу змін земного покриву. Наш підхід дозволяє інтегрувати дані з Sentinel-2 і Landsat-9, що створює умови для комбінованого використання інформації з різних джерел. Це особливо важливо для довгострокового моніторингу, де необхідна висока сумісність даних для проведення ретроспективного аналізу.

Запропонований метод підвищення просторового розрізнення Landsat-9 до рівня Sentinel-2 дозволяє зменшити розбіжності між даними, що полегшує їх інтеграцію та спільне використання в різних дослідницьких задачах. Наприклад, покращена гармонізація даних може бути використана для моніторингу змін у землекористуванні, оцінки впливу кліматичних змін або аналізу екологічних процесів на глобальному рівні.

Однак, як вже було зазначено вище, гармонізація даних супутників з різним просторовим розрізненням потребує врахування додаткових факторів. Наприклад, для досягнення ще кращої сумісності даних може бути корисним використання додаткових показників, таких як NDVI або інші індекси рослинності. Це дозволить підвищити точність аналізу, особливо в умовах, коли різні супутники мають відмінності не лише в просторовому розрізненні, але й у спектральних характеристиках.

Запропонований метод підвищення просторового розрізнення супутникових даних на основі GAN демонструє вагомі переваги в порівнянні з традиційними методами. Результати наших експериментів свідчать, що використання генеративно-змагальних мереж у поєднанні зі згортковими шарами та блоками уваги дозволяє ефективно підвищувати просторове розрізнення супутникових зображень, зокрема, для каналів з розрізненням здатністю 20 м у Sentinel-2 та 30 м у Landsat-9.

Значущим аспектом результатів другого експерименту є підтвердження ефективності комбінування даних різних супутникових місій. Гармонізація даних Sentinel і Landsat, яка досягається завдяки підвищенню просторового розрізнення, відкриває нові можливості для інтеграції даних із різних джерел та забезпечення більш точного аналізу змін на поверхні Землі. Це особливо важливо для задач довгострокового моніторингу, де необхідна максимальна сумісність даних різних супутників.

В цілому проведені дослідження показують, що використання генеративно-змагальних мереж для підвищення просторового розрізнення супутникових зображень є перспективним напрямком, що дозволяє значно покращити якість даних.

2.5. Генеративні мережі для підвищення просторового розрізнення...

Незважаючи на очевидні переваги запропонованого методу, існують деякі обмеження, які необхідно враховувати. Одним із основних обмежень є здатність моделі підвищувати просторове розрізнення лише з масштабним коефіцієнтом до 4. При спробах досягти ще більш високого просторового розрізнення (наприклад, з 30 м до 5 м) модель може не мати достатньої кількості інформації для точного відтворення дрібних деталей, що особливо помітно в складних міських умовах.

Ще одне важливе обмеження стосується впливу спектральної неоднорідності в міських районах. Міські території мають складні структури, що включають різні типи будівель, дороги, зелені насадження, водні об'єкти тощо. Для більш точного відтворення таких середовищ може знадобитися вдосконалення моделі, наприклад, через додавання додаткових шарів або використання інших підходів до обробки зображень, таких як мультиспектральний аналіз.

Таким чином, подальші дослідження можуть бути спрямовані на кілька ключових напрямків.

- **Вдосконалення моделей для міських умов.** Необхідно розробити більш складні архітектури, які могли б краще відтворювати різноманітність текстур і деталей у міських районах. Це може включати використання гібридних підходів, що комбінують генеративні моделі з іншими типами нейронних мереж, такими як рекурентні мережі або мережі зі згортковими блоками.

- **Розширення масштабних можливостей.** Варто дослідити можливість підвищення просторового розрізнення зображень з більшим масштабним коефіцієнтом, що може вимагати залучення додаткових даних або використання інших моделей. Це дозволить розширити область застосування методу, зробивши його ще більш універсальним.

- **Інтеграція з іншими показниками.** Включення додаткових спектральних індексів або карт земельного покриття може підвищити точність генерації зображень, особливо в умовах складних ландшафтів або там, де спектральна неоднорідність є суттєвим фактором.

У цій праці було запропоновано та випробувано новий підхід до підвищення просторового розрізнення супутникових зображень на основі генеративно-змагальних мереж (GAN). До основних висновків слід віднести наступні положення.

1. **Висока ефективність методу.** Використання GAN для підвищення просторового розрізнення супутникових зображень Sentinel-2 і Landsat-9 показало значне покращення якості зображень, що дозволяє отримувати детальніші дані для аналізу різних типів земельного покриття.
2. **Обмеження в міських умовах.** Незважаючи на успіхи в природних ландшафтах, модель стикається з труднощами при генерації зображень міських територій. Це підкреслює необхідність подальших досліджень і вдосконалення архітектури для точного відтворення складних структур і текстур.
3. **Гармонізація даних різних супутників.** Запропонований підхід сприяє ефективній інтеграції даних з різних супутників, що дозволяє використовувати їх спільно для більш точного і всебічного аналізу змін на земній поверхні.

Таким чином, запропонований метод має потенціал для значного покращення якості супутникових зображень і відкриває нові можливості для їх використання у дослідженнях та практичних застосуваннях, пов'язаних з моніторингом земного покриття, оцінкою впливу кліматичних змін та іншими важливими завданнями. Подальші дослідження в цьому напрямку можуть ще більше розширити можливості цього підходу.

ПЕРЕЛІК ПОСИЛАНЬ

1. Kislov, Dmitry E., et al. "Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images." *Remote Sensing in Ecology and Conservation*, vol. 7, no. 3, pp. 355-368, 2021.
2. Hansen, M. C., et al. "High-resolution global maps of 21st-century forest cover change." *Science*, 342(6160), 850-853, 2013.
3. Waldner F., Schucknecht A., Lesiv M. et. al., "Conflation of expert and crowd reference data to validate global binary thematic maps," *Remote sensing of environment*, vol. 221, pp. 235-246, 2019.
4. Shelestov A., Lavreniuk M., Kussul N. et. al., "Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping," *Frontiers in Earth Science*, vol. 5, 2017.

2.5. Генеративні мережі для підвищення просторового розрізнення...

5. d'Andrimont, Raphael, et al. "From parcel to continental scale-A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations," arXiv preprint arXiv:2105.09261, 2021.
6. Yi Zhiwei, Li Jia, and Qiting Chen, "Crop classification using multi-temporal Sentinel-2 data in the Shiyang River Basin of China," *Remote Sensing*, 12.24, pp 1-21, 2020.
7. Shelestov A., Lavreniuk M., Vasiliev V., et. al., "Cloud Approach to Automated Crop Classification Using Sentinel-1 Imagery," *IEEE Transactions on Big Data*, 6(3), pp. 572-582, 2019.
8. Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778-782.
9. Yang, Jianchao, and Thomas Huang. "Image super-resolution: Historical overview and future challenges." *Super-resolution imaging*, pp. 20-34, 2010.
10. Dong, Chao, et al. "Image super-resolution using deep convolutional networks." *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015): 295-307.
11. Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
12. Dong, Chao, Chen Change Loy, and Xiaoou Tang. "Accelerating the super-resolution convolutional neural network." *European conference on computer vision*. Springer, Cham, 2016.
13. Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
14. Lim, Bee, et al. "Enhanced deep residual networks for single image super-resolution." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017.
15. Tai, Ying, Jian Yang, and Xiaoming Liu. "Image super-resolution via deep recursive residual network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

16. Ian, Goodfellow, et al. "Generative adversarial nets." In *Advances in neural information processing systems*. (2014): 2672-2680.

17. Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

18. Yuan, Yuan, et al. "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.

19. Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." *Proceedings of the European conference on computer vision (ECCV) workshops*. 2018.

20. Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." *European conference on computer vision*. Springer, Cham, 2016.

21. Demiray, Bekir Z., Muhammed Sit, and Ibrahim Demir. "D-SRGAN: DEM super-resolution with generative adversarial networks." *SN Computer Science* 2.1 (2021): 1-11.

22. Chen, Hongguang, et al. "Generative adversarial networks capabilities for super-resolution reconstruction of weather radar echo images." *Atmosphere* 10.9 (2019): 555.

23. Lavreniuk, Mykola, et al. "Super Resolution Approach for the Satellite Data Based on the Generative Adversarial Networks." *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 1095-1098.

24. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*, 30 (2017).

25. Zhao, Long, et al. "Improved transformer for high-resolution GANs." *Advances in Neural Information Processing Systems*, 34, pp. 18367-18380, 2021.

26. Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

27. Lavreniuk, M., Shumilo, L., & Lavreniuk, A. "Generative Adversarial Networks for the Satellite Data Super Resolution Based on

2.5. Генеративні мережі для підвищення просторового розрізнення...

the Transformers with Attention,” IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2023, pp. 6294-6297.

28. Chen, Xiangyu, et al. "Activating More Pixels in Image Super-Resolution Transformer." arXiv preprint arXiv:2205.04437, 2022.

29. Kolotii, A., Kussul, N., Shelestov, A., Skakun, S., Yailymov, B., Basarab, R., ..., Ostapenko, V. "Comparison of biophysical and satellite predictors for wheat yield forecasting in Ukraine." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 40(7W3), 2015, pp. 39-44.

30. Lavreniuk, M., Bhat, S. F., Müller, M., & Wonka, P. (2023). EVP: Enhanced Visual Perception using Inverse Multi-Attentive Feature Refinement and Regularized Image-Text Alignment. arXiv preprint arXiv:2312.08548.

31. Spencer, J., Tosi, F., Poggi, M., Arora, R. S., Russell, C., Hadfield, S., ... & Elder, J. H. (2024). The third monocular depth estimation challenge. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1-14).

32. Lavreniuk, M. (2024). SPIdepth: Strengthened Pose Information for Self-supervised Monocular Depth Estimation. arXiv preprint arXiv:2404.12501.