

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені Ігоря СІКОРСЬКОГО»  
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ  
ІНСТИТУТ

Кафедра математичного моделювання та аналізу даних

«До захисту допущено»

Завідувач кафедри

\_\_\_\_\_ Н.М. Куссуль

«\_\_» \_\_\_\_\_ 2022 р.

**Дипломна робота**  
**на здобуття ступеня бакалавра**

зі спеціальності: 113 Прикладна математика  
на тему: «Застосування генеративно-змагальних мереж для  
покращення якості сегментації супутникових знімків»

Виконав: студент 4 курсу, групи ФІ-81  
Шкаліков Олег Володимирович

Керівник: проф., д.т.н., зав. каф. ММАД Куссуль Н.М. \_\_\_\_\_

Консультант: \_\_\_\_\_

Рецензент: проф., д.т.н., проф. ННІПСА Данилов В.Я. \_\_\_\_\_

Засвідчую, що у цій дипломній  
роботі немає запозичень з праць  
інших авторів без відповідних  
посилань.

Студент \_\_\_\_\_

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені Ігоря СІКОРСЬКОГО»  
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ  
ІНСТИТУТ**

**Кафедра математичного моделювання та аналізу даних**

Рівень вищої освіти — перший (бакалаврський)  
Спеціальність (освітня програма) — 113 Прикладна математика,  
ОПП «Математичні методи моделювання, розпізнавання образів та  
безпеки даних»

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ Н.М. Куссуль

«\_\_» \_\_\_\_\_ 2022 р.

**ЗАВДАННЯ  
на дипломну роботу**

Студент: Шкаліков Олег Володимирович

1. Тема роботи: *«Застосування генеративно-змагальних мереж для покращення якості сегментації супутникових знімків»*,

керівник: проф., д.т.н., зав. каф. ММАД Куссуль Н.М.,

затверджені наказом по університету №\_\_ від «\_\_» \_\_\_\_\_ 2022 р.

2. Термін подання студентом роботи: «\_\_» \_\_\_\_\_ 2022 р.

3. Вихідні дані до роботи: композит супутникових знімків Sentinel-2, зроблених з 1 липня по 1 серпня 2021 р. та маска класифікації для Київської області.

4. Зміст роботи: аналіз проблем, що виникають при розв'язку задачі сегментації супутникових знімків за допомогою сучасних нейромережевих архітектур та розробка методів аугментації незбалансованих наборів даних за допомогою змагально-генеративних мереж.

5. Перелік ілюстративного матеріалу: презентація доповіді.

6. Дата видачі завдання: 30 вересня 2021 р.

## Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання	Примітка
1	Узгодження теми роботи із науковим керівником	01.09.2021 - 30.09.2021	Виконано
2	Огляд сучасних методів розв'язання задачі семантичної сегментації	01.10.2021 - 11.11.2021	Виконано
3	Дослідження проблем, які виникають при сегментації супутникових знімків	12.11.2021 - 17.02.2021	Виконано
4	Дослідження архітектур GAN у задачі image-to-image translation	18.12.2021 - 15.02.2022	Виконано
5	Проведення експериментів з генерації аугментованих навчальних вибірок	05.02.2022 - 15.05.2022	Виконано
6	Аналіз впливу аугментації на якість сегментації	16.05.2022 - 31.05.2022	Виконано

Студент

\_\_\_\_\_ Шкаліков О.В.

Керівник

\_\_\_\_\_ Куссуль Н.М.

## РЕФЕРАТ

Кваліфікаційна робота містить: 56 стор., 11 рисунки, 5 таблиць, 17 джерел.

Застосування сучасних нейромережевих архітектур у задачі семантичної сегментації супутникових знімків дозволяє має великий прикладний сенс. Проте незбалансованість класів у навчальних вибірках, що є притаманним саме супутниковим даним, є істотною проблемою, яка знижує їх ефективність.

Дана робота полягає у розробці методів застосуванні доповнення навчальних вибірок, що має на меті підвищити якість семантичної сегментації. Для цього було проаналізовано різні підходи до генеративних моделей, зокрема генеративно-змагальні мережі. Особлива увага була приділена архітектурі Pix2Pix та її вдосконаленням. Також було розроблено і імплементовано алгоритм генерації штучних масок класифікації і, як наслідок, повний процес аугментації тренувального набору.

У ході виконання роботи було проведені експерименти, які довели ефективність запропонованого підходу. Попри незначне погіршення метрик для мажоритарних класів, якість класифікації міноритарних класів значно зросла. Таким чином було отримано метод аугментації наборів даних супутникових знімків, який дозволяє покращити якість семантичної сегментації.

СЕМАНТИЧНА СЕГМЕНТАЦІЯ, UNET, НЕЗБАЛАНСОВАНІСТЬ КЛАСІВ, ГЕНЕРАТИВНО-ЗМАГАЛЬНІ МЕРЕЖІ, PIX2PIX,

## ABSTRACT

The qualifying paper contains: 56 pages, 11 figures, 5 tables, 17 sources.

Using of modern neural network architectures in the problem of semantic segmentation of satellite images has a great practical meaning. However, the imbalance of classes in the training samples, which is inherent in the satellite data, is a significant problem that reduces their effectiveness.

This work is to develop methods for applying the augmentation of training datasets, which aims to improve the quality of semantic segmentation. For this purpose, different approaches to generative models have been analyzed, in particular generative-adversarial networks. In particular, the Pix2Pix architecture and its improvements. An algorithm for generating artificial classification masks and, as a result, a complete process of augmentation of the training set has been also developed and implemented.

In the course of the work, experiments have been conducted that proved the effectiveness of the proposed approach. Despite a slight deterioration in the metrics for majority classes, the quality of the classification of minority classes has increased significantly. Thus, a method of augmentation of satellite image data sets has been obtained, which allows to improve the quality of semantic segmentation.

SEMANTIC SEGMENTATION, UNET, CLASS IMBALANCE, GENERATIVE-ADVERSARIAL NETWORKS, PIX2PIX

## ЗМІСТ

Перелік умовних позначень, скорочень і термінів .....	7
Вступ.....	8
1 Огляд сучасних методів сегментації супутникових знімків .....	10
1.1 Задача семантичної сегментації супутникових знімків .....	10
1.2 Сучасні нейромережеві методи розв’язання задачі семантичної сегментації .....	17
1.3 Проблеми, що виникають при застосуванні нейромережевих підходів.....	21
Висновки до розділу 1.....	22
2 Генеративні моделі у задачах сегментації супутникових знімків .....	24
2.1 Різновиди генеративних моделей.....	24
2.2 Генеративно-змагальні мережі .....	27
2.3 GAN у задачах image-to-image translation .....	32
2.4 Перспективи застосування GAN для генерації штучних супутникових знімків .....	37
Висновки до розділу 2.....	39
3 Аналіз застосування GAN для аугментації навчальних вибірок .....	41
3.1 Попередня обробка навчальних даних .....	41
3.2 Результати сегментації без використання GAN.....	43
3.3 Результати сегментації з використанням аугментованого набору даних .....	45
Висновки до розділу 3.....	50
Висновки .....	52
Перелік посилань .....	54

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

GAN — generative-adversarial network, генеративно-змагальна мережа

IoU — intersection over union, міра Жаккара

w/o  $W$  — функція помилки без використання вагів

$W$  — зважена функція помилки

## ВСТУП

**Актуальність дослідження.** У сучасному світі задачі сегментації супутникових знімків мають широке прикладне значення, і використовуються для побудов мап сільськогосподарських культур [1], детектування незаконної вирубки лісів [2], звалищ, пожеж, тощо. Проте сучасні методи розв'язку цієї задачі здебільшого є алгоритмами навчання з учителем та потребують розмічених навчальних вибірок. Створення даного матеріалу, а саме ground truth масок для навчання потребує значних людських ресурсів, натомість велика кількість людино-годин може не призвести до значного покращення результатів сегментації. До того ж, навіть при використанні такого підходу, точність класифікації міноритарних класів є низькою, це пов'язано з тим, що загальна площа, яку вони займають на супутникових знімках з вибірки мала.

Ідея роботи полягає у тому, щоб застосувати генеративні моделі для створення штучних навчальних вибірок, тим самим збільшити й урізноманітнити навчальні вибірки для моделей сегментації.

**Метою дослідження** є покращення точності семантичної сегментації супутникових знімків. **Задача дослідження** полягає у дослідженні застосування генеративно-змагальних нейронних мереж для аугментації навчальних вибірок. Для розв'язання задачі необхідно вирішити такі **завдання**:

- 1) оглянути сучасні методів розв'язання задачі семантичної сегментації;
- 2) дослідити проблеми, які виникають при сегментації супутникових знімків;
- 3) дослідити архітектур GAN у задачі image-to-image translation та перспективи їх застосування для аугментації навчальних вибірок супутникових знімків;
- 4) провести експерименти з генерації аугментованих навчальних



вибірок;

5) проаналізувати вплив аугментації на якість семантичної сегментації.

**Об'єктом дослідження** є засоби, які дозволяють збільшити якість семантичної сегментації супутникових знімків.

**Предметом дослідження** є методи аугментації навчальних вибірок за допомогою генеративно-змагальних мереж.

При розв'язанні поставлених завдань використовувались такі *методи дослідження*: методи математичної статистики, математичного аналізу та методи побудови та навчання глибоких нейронних мереж.

**Наукова новизна** отриманих результатів полягає у застосуванні генеративно-змагальних мереж для генерації навчальних прикладів, які можуть збалансувати тренувальні вибірки супутникових знімків і, як наслідок, покращити якість їх семантичної сегментації.

**Практичне значення** результатів полягає у підвищенні якості сегментації міноритарних класів на супутникових знімках.

**Апробація результатів та публікації.** Частина даної роботи була представлена на ESA Living Planet Symposium (23-27 травня 2022 р., м. Бонна, ФРН) та XVIII Науково-практичній конференції студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики» (15-16 червня 2022 р., м. Київ).

# 1 ОГЛЯД СУЧАСНИХ МЕТОДІВ СЕГМЕНТАЦІЇ СУПУТНИКОВИХ ЗНІМКІВ

У першому розділі даної роботи ми розглянемо постановку задачі семантичної сегментації, нейромережеві методи, які застосовують для її розв’язку, а також специфічні проблеми, які виникають при застосуванні цих методів до супутникових знімків. Також мова піде про функції помилки та метрики якості сегментації, які будуть використані у наступних розділах.

## 1.1 Задача семантичної сегментації супутникових знімків

Серед широко спектру задач, які пов’язані з використанням супутникових знімків, особливе місце займає задача семантичної сегментації. Розв’язок даної задачі дозволяє вирішувати такі прикладні задачі, як, наприклад, детектування вирубки лісів [2], класифікацію сільськогосподарських полів [1] і відіграє визначну роль у задачах моніторингу навколишнього середовища та сільського господарства.

### 1.1.1 Постановка задачі сегментації

Як і інші задачі навчання з учителем, задача семантичної сегментації передбачає існування навчальної вибірки, тобто набору  $X$ , який складається з зображень  $I \in (R^C)^{H \times W} = \mathcal{I}$  та міток класів  $k \in K$  для кожного пікселя кожного зображення, де  $R \subset \mathbb{R}$  – множина можливих значень кожного каналу пікселя,  $H, W, C$  – висота, ширина та кількість каналів зображення,  $K = \overline{1, M}$  – множина класів ( $M \in \mathbb{N}$  – кількість класів). Зазвичай множина  $R = \overline{0, \dots, 255}$ , або  $R = [0, 1]$ .

**Означення 1.1.** Нехай ми маємо навчальну вибірку  $X = (I_i, Y_i)_{i=1}^N$ ,

де  $I_i \in \mathcal{I}$ ,  $Y_i \in K^{H \times W}$ . Задача семантичної сегментації полягає у тому, щоб знайти таке відображення  $f : \mathcal{I} \rightarrow K^{H \times W}$ , що деякий функціонал помилки  $L(f, X)$  приймає мінімальне значення.

Тобто задача семантичної сегментації є задачею класифікації, але не цілого зображення, а кожного окремого пікселя.

### 1.1.2 Функції помилки

Важливим питанням є вибір функціоналу помилки, які може суттєво вплинути на результат. Абсолютна більшість використовуваних на практиці функцій помилок мають наступну форму: розраховується певна функція помилки  $l(f, p, k)$  для кожного окремого пікселя  $p \in I$ , після чого застосовується певна операція редукції по усім пікселям та зображенням. Одним з найчастіших виборів операції редукції є звичайне середнє, і у цьому випадку загальний функціонал помилки виглядає наступним чином:

$$L(f, X) = \frac{1}{N} \sum_{n=1}^N \frac{1}{H \cdot W} \sum_{(p,k) \in (I_n, Y_n)} l(f, p, k) \quad (1.1)$$

Однією з найбільш використовуваних функцій помилки у задачах класифікації та сегментації є крос-ентропія.

**Означення 1.2** (Cross-Entropy). Нехай  $p \in I$  – піксель зображення,  $k \in K$  – правильний клас,  $f_i(p)$  – передбачувана ймовірність належності класу  $i \in K$  для пікселя  $p$ , тоді крос-ентропія обчислюється наступним чином:

$$\text{CE}(f, p, k) = - \sum_{i=1}^{|K|} 1\{i = k\} \ln f_i(p) = - \ln f_k(p)$$

Інтуїтивний сенс функції крос-ентропії наступний: найменшого значення дана функція набуває тоді, коли розподіли правильних класів та передбачуваних однакові. Тобто дана функція є певною мірою різниці цих

розподілів. Але у реальних застосуваннях, а особливо при сегментації супутникових знімків, дуже часто розподіл класів далекий від рівномірного, тобто присутня незбалансованість класів. У таких випадках застосування класичної крос-ентропії може призвести [3] до того, що якість класифікації міноритарних класів буде низькою, бо значення помилки буде не великим. Для того, щоб врахувати цю обставину застосовують модифікацію, яка має назву зважена крос-ентропія.

**Означення 1.3** (Weighted Cross-Entropy). Нехай  $p \in I$  – піксель зображення,  $k \in K$  – правильний клас,  $f_i(p)$  – передбачувана ймовірність належності класу  $i \in K$  для пікселя  $p$ ,  $\omega_i \in (0, 1]$  – вага класу  $i \in K$ . Тоді зважена крос-ентропія обчислюється наступним чином:

$$\text{WCE}(f, p, k) = - \sum_{i=1}^{|K|} \omega_i \ln f_i(p) 1\{i = k\} = -\omega_k \ln f_k(p)$$

Тобто ми для кожного з класів задаємо ваги, тобто міру того, наскільки нам важливо врахувати помилку у класифікації кожного пікселя цього класу. У більшості випадків, ваги задаються  $1 - \frac{N_k}{N}$ , де  $N_k$  – кількість пікселів класу  $k$  у вибірці,  $N$  – загальна кількість пікселів.

Іншою розповсюдженою функцією помилки для задач семантичної сегментації, яка застосовна [3] при сильно незбалансованих класах є Focal Loss.

**Означення 1.4** (Focal Loss). Нехай  $p \in I$  – піксель зображення,  $k \in K$  – правильний клас,  $f_i(p)$  – передбачувана ймовірність належності класу  $i \in K$  для пікселя  $p$ ,  $\gamma \in (0, 1)$  – параметр,  $\omega_i \in (0, 1]$  – вага класу  $i \in K$ . Тоді зважена Focal Loss обчислюється наступним чином:

$$\text{FL}(f, p, k) = - \sum_{i=1}^{|K|} \omega_i (1 - f_i(p))^\gamma \ln f_i(p) 1\{i = k\} = -\omega_k (1 - f_k(p))^\gamma \ln f_k(p)$$

Як можна побачити [3], даний вигляд функції помилки дозволяє сфокусуватися на складних прикладах та зменшити вагу помилок тих

прикладів, які вже правильно класифікуються, але ймовірність яких все ще не рівна 1.

Звичайно, що це далеко не повний перелік усіх можливих варіацій функціоналу помилки, проте це найпопулярніші функції, які широко застосовуються при розв'язанні задач семантичної сегментації у тому числі супутникових знімків.

### 1.1.3 Метрики якості

Окрім визначення функціоналів помилки, доцільно згадати й про ті засоби, які дозволяють оцінити наскільки якісна побудована модель семантичної сегментації. Для надання формальних та точних оцінок, які дозволять порівнювати різні моделі з різними функціями помилки, використовуються метрики.

Грунтуючись на тому, що задача семантичної сегментації являє собою задачу класифікацію кожного пікселя зображення, ми будемо розглядати і застосовувати метрики специфічні для класифікації. Для того щоб визначити більшість з цих метрик зручно використовувати матрицю невідповідностей [4].

**Таблиця 1.1** – Матриця невідповідностей

		Справжні				
Передбачувані		$k_1$	$k_2$	$\dots$	$k_m$	
	$k_1$	$n_{11}$	$n_{12}$		$n_{1m}$	$N_{P1}$
	$k_2$	$n_{21}$	$n_{22}$		$n_{2m}$	$N_{P2}$
	$\vdots$			$\ddots$		
	$k_m$	$n_{m1}$	$n_{m2}$		$n_{mm}$	$N_{P3}$
		$N_{T1}$	$N_{T2}$		$N_{Tm}$	$N$

Матриця невідповідностей являє собою квадратну матрицю розмірності  $t \times t$ , де  $t$  - кількість класів. Кожний компонент  $n_{ij}$ , де

$i, j = \overline{1, m}$ , даної матриці дорівнює кількості пікселів, які були віднесені нашої моделлю до класу  $i$ , при тому, що справжня правильна мітка класу  $j$ . Також дуже корисними для визначення метрик є значення сум усіх елементів, або ж тільки по рядкам і стовпцям. Тут і надалі ми будемо використовувати наступні позначення:

$$N_{Tj} = \sum_{l=1}^m n_{lj}, \quad N_{Pi} = \sum_{l=1}^m n_{il}, \quad N = \sum_{i,j=1}^m n_{ij}.$$

Тобто  $N_{Tj}$  дорівнює кількості пікселів, які належать до класу  $j$ ,  $N_{Pi}$  – це кількість пікселів, які були віднесені моделлю до класу  $i$ , і, відповідно,  $N$  – це сумарна кількість пікселів.

Однією з найпоширеніших та популярних метрик є точність (ассурасу), яка є нічим іншим як долею правильних передбачень відносно усього набору даних.

**Означення 1.5** (Ассурасу). Нехай отримана матриця невідповідностей  $\|n_{ij}\|_{i,j=1}^m$ . Тоді метрика точність обчислюється наступним чином:

$$\text{Ассурасу} = \frac{\sum_{i=1}^m n_{ii}}{N}$$

Точність є метрикою, яка є чудово інтерпретованою, проте має і суттєві недоліки. Ці суттєві недоліки безпосередньо пов'язані з незбалансованістю класів, що є притаманним для задач семантичної сегментації супутникових знімків. А саме: при наявності незбалансованості класів дана метрика може приймати високі значення, навіть якщо якість розпізнавання якогось з міноритарних класів дуже низька. Це обумовлено тим, що сам спосіб обрахунку метрики залежить від кількості представників кожного з класів: значення  $n_{ii}$  не може перевищувати реальну кількість пікселів для класу  $i$ , і саме тому внесок міноритарних класів значно менший ніж мажоритарних.

Для того, щоб врахувати недоліки точності та мати змогу аналізувати якість класифікації для кожного класу окремо ми можемо використовувати

метрики User Accuracy та Producer Accuracy.

**Означення 1.6** (User Accuracy). Нехай отримана матриця невідповідностей  $||n_{ij}||_{i,j=1}^m$ . Тоді метрика User Accuracy для класу  $k_i \in K$  обчислюється наступним чином:

$$UA(k_i) = \frac{n_{ii}}{N_{Pi}}$$

**Означення 1.7** (Producer Accuracy). Нехай отримана матриця невідповідностей  $||n_{ij}||_{i,j=1}^m$ . Тоді метрика Producer Accuracy для класу  $k_i \in K$  обчислюється наступним чином:

$$PA(k_i) = \frac{n_{ii}}{N_{Ti}}$$

Як можна побачити, метрика User Accuracy відображає долю пікселів, які дійсно мають клас  $k_i$ , при умові, що модель передбачила саме цей клас. На противагу цьому існує метрика Producer Accuracy, яка відображає долю пікселів класу, які коректно класифіковані. Таким чином існує інструмент аналізу якості класифікації кожного окремого класу, попри це для аналізу загальної якості треба або ж усереднювати значення цих метрик за класами та між собою, або ж використовувати інші підходи.

Однією з метрик, які дозволяють оцінити загальну якість класифікації при умові незбалансованості вибірки є каппа Коена [4].

**Означення 1.8** (Cohen's Каппа). Нехай отримана матриця невідповідностей  $||n_{ij}||_{i,j=1}^m$ . Тоді метрика каппа Коена для класу  $k_i \in K$  обчислюється наступним чином:

$$\kappa = \frac{p_o - p_e}{1 - p_e},$$

де  $p_o$  – частка спостережуваної згоди, що є нічим іншим, як Accuracy.  $p_e$  – очікувана точність, тобто точність, яка відповідає випадковим збігам реальних та передбачуваних міток класів. Вона обраховується,

ґрунтуючись на гіпотезі незалежності реальних міток та передбачуваних, наступним чином:

$$p_e = \sum_{i=1}^m \frac{N_{Pi} \cdot N_{Ti}}{N^2}$$

Таким чином каппа Коена надає нам оцінку, яку можна інтерпретувати як міру того, наскільки передбачення моделі узгоджені зі справжніми мітками класів. Більш того, дана метрика дозволяє порівнювати різні моделі, які були навчені на різних наборах даних, бо використовує саме очікувану точність, що дозволяє прибрати залежність від конкретної вибірки.

Іншим підходом, який притаманний саме задачам семантичної сегментації є обрахунок метрики під назвою індекс Жаккара, або intersection-over-union (IoU). Дана метрика вимірює міру схожості двох множин, а саме відношення їх перетину до об'єднання:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

У застосуванні до задач семантичної сегментації IoU являє собою міру того, як співвідносяться один до одного множини пікселів, які справді мають певний клас з множиною пікселів, яким моделлю було передбачено цей клас.

**Означення 1.9** (IoU). Нехай отримана матриця невідповідностей  $\|n_{ij}\|_{i,j=1}^m$ . Тоді метрика IoU для класу  $k_i \in K$  обчислюється наступним чином:

$$\text{IoU}(k_i) = \frac{n_{ii}}{N_{Pi} + N_{Ti} - n_{ii}}$$

Дана метрика приймає значення у діапазоні  $[0, 1]$ , де значення 1 означає повну відповідність передбачених міток зі справжніми, а 0 – ситуацію, коли не спостерігається жодного пікселя, який був би віднесений до класу і справді б належав цьому класу.

Для отримання ж оцінки для усієї моделі можливо усереднити отримані значення метрики IoU за усіма класами.



## 1.2 Сучасні нейромережеві методи розв'язання задачі семантичної сегментації

До розв'язку задачі семантичної сегментації існує багато підходів. Найбільш ефективним, сучасними та широко відомими з них є методи, які ґрунтуються на штучних нейронних мережах.

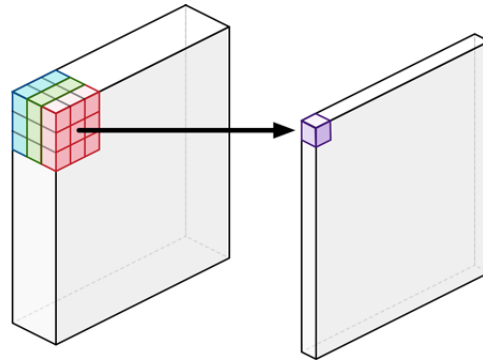
### 1.2.1 Згорткові нейромережі

Серед широко спектру можливих архітектур особливе місце у задачах класифікації, сегментації, детекції та багатьох інших, що пов'язані з зображеннями, займають згорткові нейронні мережі [5]. Даний мереж поєднує три архітектурні принципи, для того, щоб забезпечити інваріантність моделі до зсуву, масштабування та іншим перетворенням зображення, а саме:

- 1) локальні рецептивні поля
- 2) спільні ваги
- 3) просторові підвибірки.

Перші два принципи, а у багатьох сучасних архітектурах і третій, імплементуються за допомогою використання операції згортки. Основна ідея застосування таких мереж полягає у тому, що на кожному етапі операція скалярного множення з ядром застосовується до послідовних маленьких частинок зображення, розмір яких відповідає розміру ядра згортки (рис. 1.1). Це допомагає виокремити специфічні ознаки, які стають ще більш високорівневими на кожному подальшому шарі, бо розмір рецептивного поля збільшується (якщо розмір ядра більше 1). Важливою властивістю згорткових шарів є і інваріантність, щодо зміщення та інших подібних перетворень, тобто локальні ознаки змінять своє положення відповідно до зміщення, але не змінять своє значення. Крім того, на кожному кроці ми можемо застосовувати декілька ядер і

отримувати одразу велику кількість різних ознак. Таким чином послідовно застосовуючи згорткові шари ми можемо віднайти таке представлення даних, яке дозволить ефективно розв'язувати поставлені задачі, у тому числі задачу семантичної сегментації.



**Рисунок 1.1** – Принцип роботи згорткового шару  $3 \times 3$  [6]

При усіх цих перевагах згорткові мережі залишаються ефективними, бо використовують спільні ваги для різних частин зображення. Крім того, ми маємо змогу змінювати крок (stride) згортки, тобто параметр, який визначає на якій відстані у пікселях будуть знаходитись ті частини зображення до яких буде застосована згортка. Для регуляції ж розміру вихідного зображення і ефектів, пов'язаних з роботою операції згортки на краях зображення ми здатні використовувати padding, тобто розширення зображення за допомогою додавання нових пікселів з усіх боків. Значення та кількість цих пікселів залежать від визначених архітектором мережі стратегій. Найбільш використовуваними є варіанта заповнення нулями, або ж віддзеркалення значень з країв зображення.

Просторові ж підвибірки можливо реалізовувати або ж знову таки за допомогою операції згортки з різними параметрами stride, або ж за допомогою pooling. Цей метод зменшення розмірності ставить на меті вирішення проблеми, яка пов'язана з тим, що одні й ті ж самі ознаки можуть знаходитись у різних частинах зображення, що призводить до різних вихідних мап ознак. Сам процес полягає у тому, що ми певним

чином для кожної мапи частини вхідної мапи ознак даємо у відповідність одне значення, яке узагальнює ці ознаки. Найпоширенішими ж стратегіями є наступні:

- 1) average pooling — усереднення усіх вхідних ознак
- 2) max pooling — підрахунок найбільшого значення.

Використовуючи вищезазначені складові блоки ми можемо будувати архітектур нейронних мереж, які будуть ефективно застосовні до даних у вигляді зображень та будуть відповідати усім необхідним вимогам, які ми ставимо перед мережею.

### 1.2.2 Архітектура UNet

Однією з таких архітектур, що є визнаним лідером у задачі семантичної сегментації, у тому числі, супутникових зображень є архітектура UNet [7].

Вона являє собою повністю згорткову (fully-convolutional) мережу, тобто не містить ніяких інших типів шарів, окрім згортки, транспонованої згортки чи інших методів збільшення розміру та pooling-у (і у деяких випадках, шарів, що нормалізують). Головною ідеєю даної архітектури, як це проілюстровано на рис. 1.2, є додавання додаткових шляхів (skip connections) з шарів, які зменшують розмір простору ознак, до шарів, які збільшують. Зменшуючи розмір простору ознак ми з кожним кроком виділяємо все більш високорівневі ознаки, але значна частина інформації, особливо щодо розташування на зображенні, може бути втрачена. Застосування ж такої архітектури дозволяє нам враховувати ознаки з різних етапів. І крім того, у деяких випадках, подолати явище затухання градієнтів.

Особливої уваги заслуговує і той факт, що кількість каналів ознак для кожного шару, що збільшує, залишається великою, на відміну від класичних згорткових мереж. Це дає змогу більш ефективно враховувати контекстну інформацію.



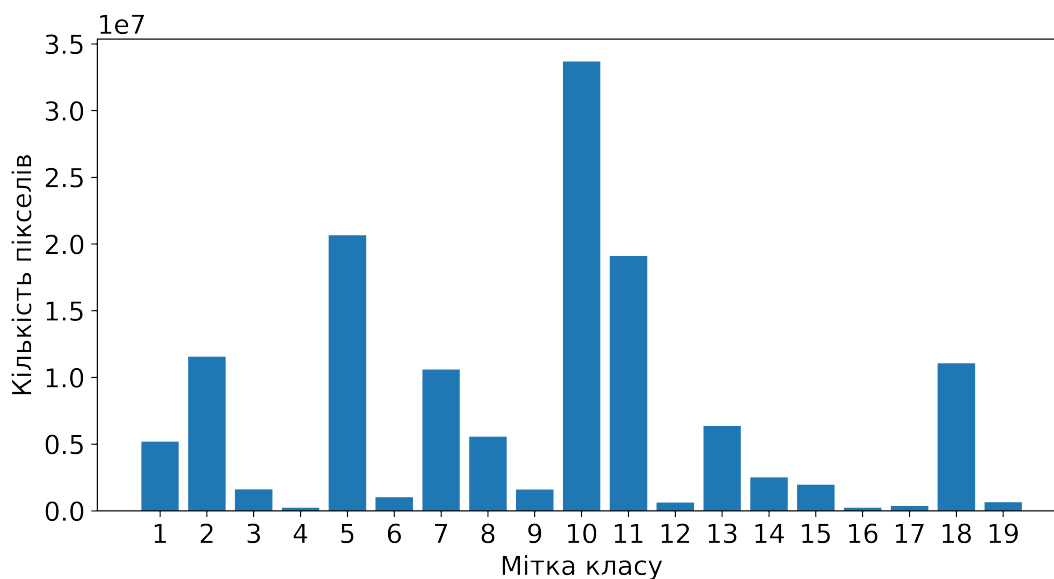
### 1.3 Проблеми, що виникають при застосуванні нейромережевих підходів

Попри використання сучасних нейромережевих архітектур, при розв'язку задач семантичної сегментації супутникових знімків виникають істотні проблеми, які не дозволяють досягти максимальної якості. Ці перепони є наслідком самої структури задачі, а саме того, що семантична сегментація у постановці 1.1 є задачею навчання з учителем, тобто вимагає розміченого набору даних, де кожному супутниковому знімку буде відповідати відповідна ground-truth маска сегментації.

Тож першою істотною проблемою є те, що створення навчальних вибірок, а саме: розмітка супутникових знімків, вимагає значної кількості людських ресурсів. Причому кінцева якість сегментації безпосередньо залежить від розміру тренувального набору. Тобто потреба у покращенні значень метрик вимагає, серед усього іншого, істотного збільшення навчальних прикладів, а отже й створених людьми ground-truth масок. Крім того, у випадку створення масок для супутникових знімків, не завжди людина може правильно класифікувати усі об'єкти ґрунтуючись тільки на зображенні, що може бути пов'язано з порою року, розміром об'єкту, роздільною здатністю знімка, схожістю різних класів, тощо.

Друга істотна проблема пов'язана з дуже сильною незбалансованістю класів, особливо у задачах класифікації сільськогосподарських культур. Як доказ цього, можна навести статистику (рис. 1.3) ground truth маски по кількості пікселів по кожному з типів полів та кількості зображень, отриманих після поділу композиту на частини по  $256 \times 256$  пікселів на яких представлений даний клас для Київської області. Вищезгаданий поділ обумовлений тим, що нейромережеві моделі не можуть ефективно працювати з дуже великими зображеннями, тож ми вимушені різати їх на частини.

І попри застосування функцій помилок, які можуть враховувати



**Рисунок 1.3** – Кількість пікселів для різних класів

сільськогосподарських культур у досліджуваному наборі даних

дисбаланс у класах, досягнути бажаних значень метрик, що описують точність моделей не вдається, що ми побачимо у подальших розділах цієї роботи. Особливо незадовільні результати виникають саме для класів, які мають маленьку кількість прикладів у навчальній вибірці. А доповнити навчальну вибірку достатньою кількістю прикладів для певних класів може бути взагалі неможливо, через особливості сільського господарства у досліджуваному регіоні або самої сільськогосподарської культури. Тобто знайти ділянку земної поверхні, на якій окремий клас буде достатньо представлений, у деяких випадках не є можливим. Останній факт як раз і є специфічним саме для аналізу супутникових знімків і виділяє їх серед усіх інших сфер, де розв'язується задача семантичної сегментації.

## Висновки до розділу 1

У даному розділі ми розглянули постановку задачі семантичної сегментації супутникових знімків, яка є однією з найважливіших задач у цій сфері, бо має багато прикладних застосувань. Були розглянуті різні

вигляди функцій помилки, а саме: крос-ентропія, зважена крос-ентропія та Focal Loss, які можуть суттєво впливати на якість сегментації при застосуванні сучасних методів.

Визначено й засоби оцінки та порівняння різних моделей, тобто метрики, які застосовні до розглянутої задачі. Серед них точність (Accuracy), User Accuracy, Producer Accuracy, каппа Коена та міра Жаккара. Усі з перелічених метрик можливо ефективно обрахувати використовуючи матрицю невідповідностей, і при цьому кожна з них має зрозумілий для кінцевого користувача сенс.

Згорткові нейромережі — це ефективний спосіб розв’язання задач, пов’язаних з обробкою зображень, у тому числі задач семантичної сегментації супутникових знімків. Архітектура ж UNet, яка була розглянута у даному розділі, на даний момент є одним з найпотужніших, визнаних науковою спільнотою підходів до цієї задачі. Вона використовує додаткові шляхи (skip connections), щоб враховувати інформацію з мал ознак різної розмірності, що особливо важливо при аналізі супутникових знімків.

Було виявлено, що значними проблемами при розв’язку задачі семантичної сегментації за допомогою сучасних нейромережових підходів є складність формування великих вибірок, а також, що є специфічним для супутникових даних, сильна незбалансованість класів, яку інколи неможливо змінити, бо додати реальні спостереження, з бажаним розподілом класів, не є можливим.

## 2 ГЕНЕРАТИВНІ МОДЕЛІ У ЗАДАЧАХ СЕГМЕНТАЦІЇ СУПУТНИКОВИХ ЗНІМКІВ

У другому розділі ми розглянемо генеративні моделі, які можуть бути застосовані для генерації штучних супутникових знімків. Особлива увага буде приділена генеративно-змагальним нейронним мережам. Також піде мова про вирішення задачі image-to-image translation за допомогою GAN, а саме про архітектуру Pix2Pix та її вдосконалені версії. Та нарешті буде описано механізм застосування генеративно-змагальних мереж для аугментації навчальних вибірок, що має за мету вирішити проблеми, які були описані у попередньому розділі.

### 2.1 Різновиди генеративних моделей

Для того, щоб наблизитись до розв'язку проблеми генерації штучних зображень, розглянемо наступну задачу: нехай існує набір даних  $(x_1, x_2, \dots) \in X$ , які мають невідомий розподіл  $\pi$ . Ми намагаємось знайти його оцінку

$$p^* \in \arg \min_p \mathcal{D}(\pi||p),$$

де  $\mathcal{D}$  - деяка міра близькості розподілів, яка задовольняє умови

$$\begin{cases} \mathcal{D}(\pi||p) \geq 0, & \pi \neq p \\ \mathcal{D}(\pi||p) = 0, & \pi = p \end{cases}$$

Наприклад дивергенція Йонсена-Шеннона:

$$JSD(\pi||p) = \frac{1}{2} \int_{\mathbb{R}} \left[ \pi(x) \ln \frac{2\pi(x)}{p(x) + \pi(x)} + p(x) \ln \frac{2p(x)}{p(x) + \pi(x)} \right] dx \quad (2.1)$$

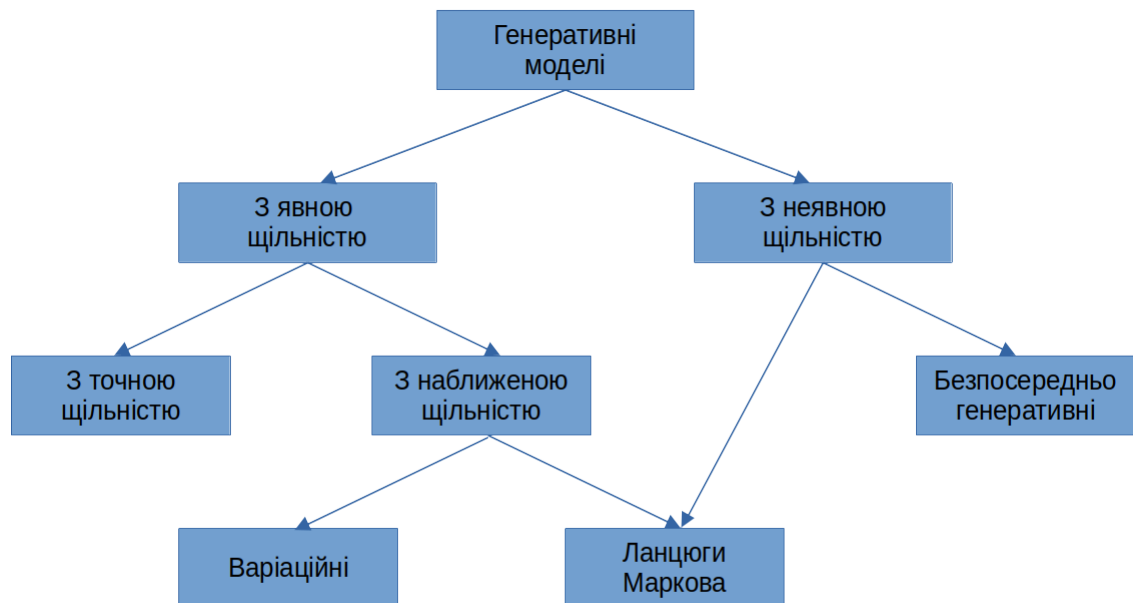
Якщо ми зможемо знайти дану оцінку розподілу для реальних



супутникових знімків, то ми зможемо генерувати випадкові величини з даного розподілу, які і будуть являти собою штучні знімки. Звичайно, що пошук по усім можливим розподілам є дуже складною задачею, тому ми зосередимось на параметричних моделях, тобто:

$$p(x|\theta) \in \arg \min_{\theta} \mathcal{D}(\pi||p_{\theta}),$$

Існує багато методів розв'язку подібних задач, які можна поділити на декілька великих груп [8], що зображено на рисунку 2.1.



**Рисунок 2.1** – Таксономія генеративних моделей

Моделі з явною щільністю дозволяють обраховувати значення щільності. Максимізація правдоподібності для даних моделей є простою, бо являє собою просту підстановку щільності, що описує модель, у вираз для правдоподібності. Найскладніше для даної групи підібрати таку модель, яка зможе описувати усю складність реальних даних і при цьому дозволить обраховувати щільність розподілу. Тому виділяють дві

стратегії:

1) моделі, які гарантують можливість точного обрахунку (tractability);

2) моделі, які ґрунтуються на обрахунку наближених значень.

Моделі, які гарантують можливість точного обрахунку щільності, такі як FVBN (fully visible belief networks), нормалізуючи потоки, мають перевагу у тому, що дозволяють проводити оптимізацію безпосередньо функції правдоподібності, проте це вимагає від них жорстких обмежень. До того ж більшість подібних моделей виконують обрахунки послідовно, тобто для генерації необхідно виконати декілька послідовних кроків, які не можуть бути розпаралелені, що значно збільшує час необхідний для генерації.

Для уникнення проблем, пов'язаних з жорсткими обмеженнями на модель, які необхідні для точного обрахунку щільності, використовуються моделі, які дозволяють отримувати наближені значення щільності. Вони у свою чергу поділяються на дві групи: з детерміністичною апроксимацією, здебільшого це автоенкодера, та зі стохастичною, як наприклад ланцюги Маркова, машини Больцмана.

Звичайно, що і у даних підходів є суттєві недоліки. Варіаційні автоенкодера використовують у своїй роботі нижню границю (ELBO), яка при виборі "слабкого" апостеріорного або ж апіорного може значно відрізнятись від справжньої правдоподібності, що призведе до того, що розподіл генератора та реальний розподіл будуть значно відрізнятись. На практиці варіаційні моделі все ж здатні гарно апроксимувати правдоподібність, проте якість згенерованих зображень є низькою. Якщо ж казати про недоліки методів побудованих на основі ланцюгів Маркова, то слід зазначити їх низьку швидкодію, що обумовлена як необхідністю виконувати операції покроково, так і тим, що збіжність дуже повільна та не має жодного методу, який би міг вказати чи вже наявна збіжність чи потребується продовжити оптимізацію.

Концептуально іншим підходом є застосування моделей, які не

оперують щільністю або її наближеннями як такими, а навчаються лише ґрунтуючись на тих прикладах, які надає генератор. Тобто використовують не сам розподіл як такий, а певні спостереження з нього. Деякі з цих моделей знову ж таки використовують ланцюги Маркова для генерування прикладів, здебільшого це генеративні стохастичні мережі. Але вони мають притаманні усім ланцюгам Маркова недоліки, які були описані вище. І на сам кінець існують моделі, які здатні безпосередньо генерувати зображення за один крок. Найбільш відомими й поширеними з них є генеративно-змагальні мережі, які ми розглянемо більш детально.

## 2.2 Генеративно-змагальні мережі

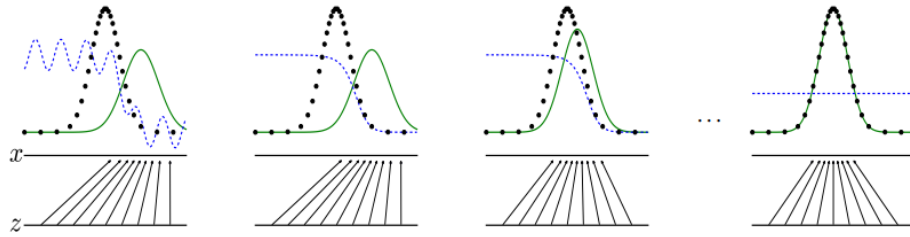
Ідея оригінальної GAN [9] полягає у тому, що ми маємо дві сутності: генератор — диференційовну функцію  $G : Q \times \Theta_G \rightarrow X$ , яка на основі елемента з множини  $Q$  (так званого латентного простору) та параметрів  $\theta_g \in \Theta_G$  генерує елемент з множини  $X$ ; та дискримінатор — диференційовну функцію  $D : X \times \Theta_D \rightarrow \{0, 1\}$ , яка намагається на визначити чи є елемент штучно згенерованим, чи належить розподілу навчальних даних. Для пошуку функції генератора та дискримінатора застосовується міні-максний функціонал якості:

$$\min_G \max_D [\mathbb{E}_{x \sim \pi} \ln D(x) + \mathbb{E}_{q \sim p_Q} \ln(1 - D(G(q)))] , \quad (2.2)$$

Тобто відбувається змагання між генератором, який прагне генерувати найбільш схожі на реальні зображення, щоб ввести в оману дискримінатор, який навпаки прагне якомога краще відрізнити штучно згенеровані та реальні приклади.

Але для того, щоб знайти відповідні оптимальні значення параметрів дискримінатора та генератора необхідний алгоритм навчання, який з кожним кроком буде все ближче і ближче наближати розподіл генератора до реального, аж поки дискримінатор не втратить будь-яку

можливість відрізняти реальні та штучно згенеровані зображення. Дану ідею зображено на рисунку 2.2.



**Рисунок 2.2** – Ілюстрація ідеї GAN [9]. Зеленим позначено розподіл генератора, синім — дискримінатора, чорними точками — реальних даних

Як це притаманно штучним нейронним мережам, алгоритм оптимізації ґрунтується на обрахунку і використанні градієнтів цільової функції помилки від параметрів. Але, через те, що наша модель не є класичною нейромережею, необхідний окремий підхід до навчання GAN.

**Алгоритм 2.1** (Алгоритм навчання GAN). Параметри:  $k$  - кількість ітерацій навчання дискримінатора за 1 епоху,  $N$  - кількість епох навчання,  $m$  - розмірність батчу.

Вхідні дані: навчальна вибірка  $X$  з розподілом  $\pi$ .

- 1) Доки кількість епох навчання не досягла  $N$ 
  - а) Повторювати  $k$  разів
    - i. Згенерувати  $\{z_1, \dots, z_m\}$
    - ii. Обрати  $\{x_1, \dots, x_m\} \sim \pi(x)$
    - iii. Оновити параметри дискримінатора використовуючи градієнт

$$\nabla_{\theta_D} \left( \frac{1}{m} \sum_{i=1}^m [\ln D(x_i) + \ln(1 - D(G(z_i)))] \right)$$

- б) Згенерувати нові  $\{z'_1, \dots, z'_m\}$

в) Оновити параметри генератора використовуючи градієнт

$$\nabla_{\theta_G} \left( \frac{1}{m} \sum_{i=1}^m \ln(1 - D(G(z'_i))) \right)$$

У якості оптимізатора, який використовує вже обраховані градієнти можна застосовувати довільний алгоритм, у більшості випадків це або звичайний градієнтний спуск, або більш досконалі методи, як наприклад Adam [10].

Окремим питанням, яке заслуговує на увагу, є питання про те, а чи дійсно у результаті розв'язку цієї задачі ми отримаємо генератор, розподіл якого дорівнює розподілу реальних даних. Для того, щоб переконатися у цьому доведемо дві важливі теореми.

**Теорема 2.1** (Гудфеллоу [9]). *Для фіксованого генератора  $G$  з розподілом  $p_G(x)$  оптимальним дискримінатором  $D^*(x)$  є:*

$$D^*(x) = \frac{\pi(x)}{\pi(x) + p_G(x)}.$$

**Доведення.** Запишемо цільову функцію:

$$\min_G \max_D [\mathbb{E}_{x \sim \pi} \ln D(x) + \mathbb{E}_{q \sim p_Q} \ln(1 - D(G(q)))]$$

Генератор, за припущенням є фіксованим, тож ми можемо перейти від математичного сподівання по латентному просторі до математичного сподівання розподілу генератора  $p_G(x)$ .

$$\min_G \max_D [\mathbb{E}_{x \sim \pi} \ln D(x) + \mathbb{E}_{x \sim G(x)} \ln(1 - D(x))]$$

Наступним кроком розпишемо математичні сподівання та властивостями визначених інтегралів:

$$\min_G \int_X \max_D (\pi(x) \ln D(x) + p_G(x) \ln(1 - D(x))) dx$$

Розглянемо умову екстремум для виразу під знаком максимуму і отримаємо:

$$\begin{aligned} \frac{\partial}{\partial D(x)} (\pi(x) \ln D(x) + p_G(x) \ln(1 - D(x))) = \\ = \frac{\pi(x)}{D(x)} - \frac{p_G(x)}{1 - D(x)} = 0 \end{aligned}$$

І виразивши  $D(x)$  остаточно отримаємо:

$$D^*(x) = \frac{\pi(x)}{\pi(x) + p_G(x)}$$

□

**Теорема 2.2** (Гудфелоу [9]). *Глобальний мінімум критерію (2.2) досягається тоді і тільки тоді, коли розподіли генератора  $p_G$  та реальних даних однакові. При цьому значення критерію дорівнює  $-\ln 4$ .*

**Доведення.** Підставимо оптимальний дискримінатор у критерій (2.2).

$$\min_G \left[ \mathbb{E}_{x \sim \pi} \ln \frac{\pi(x)}{\pi(x) + p_G(x)} + \mathbb{E}_{q \sim p_G} \ln \frac{p_G(x)}{\pi(x) + p_G(x)} \right]$$

Тепер помножимо і поділимо кожен дріб на 2 та використаємо властивість логарифмів та математичного сподівання від константи.

$$\min_G \left[ \mathbb{E}_{x \sim \pi} \ln \frac{2\pi(x)}{\pi(x) + p_G(x)} + \mathbb{E}_{q \sim p_G} \ln \frac{2p_G(x)}{\pi(x) + p_G(x)} \right] - \ln 4$$

Вираз у дужках є нічим іншим як дивергенцією Йонсена-Шеннона (2.1). Вона приймає мінімальне значення 0 тоді і тільки тоді, коли розподіли дорівнюють один одному. І, відповідно, значення критерію у випадку рівності 0 мінімуму буде дорівнювати  $-\ln 4$ . □

Проте наявний алгоритм навчання майже завжди не використовується на практиці, що пов'язано з тим, що градієнти параметрів генератора згасають [11, 9]. Це є наслідком самої будови функції помилки, тому її заміняють на наступний аналог:

$$-\ln D(G(z)) \quad (2.3)$$

Притаманні процесу навчання генеративно-змагальних мереж і інші проблеми, як наприклад нестабільність навчання, яка полягає у тому, що у деяких випадках алгоритм не може знайти оптимальні дискримінатор та генератор. Деякі причини цього детально описані у роботі [11], серед них нестабільність градієнтів функції помилки у вигляді (2.3). Іншою можливою причиною може бути те, що як зображено на рис. 2.2, з кожною ітерацією генератор покращується, що робить зворотній зв'язок від дискримінатора все менш правильним, і на сам кінець зовсім випадковим. Тобто генератор починає вчитися на небажаному зворотному зв'язку від дискримінатора.

Одним зі шляхів подолання даних проблем є додавання шуму до входу дискримінатора, застосування спектральних нормалізацій [12] або ж регуляризацій [13].

Іншою істотною проблемою є так званий *mode collapse*. Вона полягає у тому, що вихідні зображення дуже мало відрізняються один від одного та модель не в змозі генерувати різноманітні приклади. Це відбувається через те, що генератор завжди намагається ввести в оману дискримінатор, і для цього генерує зображення, які з найбільшою ймовірністю будуть розпізнані дискримінатором як справжні. Але якщо виявилось так, що дискримінатор потрапив у локальний максимум, з якого не може вибратись, то генератор надмірно пристосовується саме до цього дискримінатора, через що і генерує невелике розмаїття зображень.

Зміна дискримінатора та критерію оптимізації спроможні подолати дану проблему. Одним з таких підходів є Wasserstein GAN [14]. Який змінює дискримінатор, який приймає обмежені значення на критика, виходи якого можуть бути довільні. До того ж сама функція помилки була змінена на таку, що відображає відстань між розподілами. Серед усього іншого, даний підхід спроможний допомогти і у випадку нестабільного навчання.

Попри усі згадані недоліки, якість зображень та швидкодія генерації штучних прикладів для генеративно-змагальних мереж є високою, що і робить GAN визнаним спільнотою лідером у задачах генерації штучних зображень.

Підсумовуючи, генеративно-змагальні мережі мають наступні переваги над іншими моделями:

1) Можливість паралельної генерації, що забезпечує швидкість роботи навченої моделі.

2) На генератора накладається мало обмежень, що дає можливість вибору широко класу архітектур. Це є перевагою порівняно з машинами Больцмана, які допускають лише певні класи ймовірнісних розподілів, а також щодо потоків, що нормалізують, які вимагають, щоб генератор мав зворотнє відображення, а розмірність латентного простіру дорівнювала розмірності даних.

3) Не використовуються ланцюги Маркова з усіма їх недоліками, як у машинах Больцмана та генеративних стохастичних мережах.

4) Генеративна мережа, оптимізація якої була збіжною, точно наближає розподіл реальних даних, а не використовує нижні чи верхні оцінки, як це у випадку з автоенкодерами.

5) Якість зображень, які генеруються за допомогою GAN, здебільшого вища за інші моделі.

Саме ці моменти і обумовлюють вибір саме генеративно-змагальних мереж для генерації штучних супутникових знімків, і, як наслідок, аугментації навчальних вибірок.

## 2.3 GAN у задачах image-to-image translation

Для того, щоб мати вплив на згенеровані класичними генеративно-змагальними мережами зображення, ми маємо досліджувати структуру латентного простору, і навіть якщо нам це, використовуючи певні методи, вдасться, у нас все ще не буде повного контролю за кожним



пікселем зображення, бо розмірність латентного простору майже завжди менша за розмірність зображення. Таким чином точно контролювати розподіл класів у згенерованих зображеннях не є можливим.

Наша ж задача полягає саме у тому, щоб генерувати штучні супутникові знімки і при цьому мати знання про клас кожного зі згенерованих пікселів. Саме тому нам потрібно звернути увагу на ті методи, які дозволяють контролювати кожний з вихідних пікселів.

Однією з задач, яка влаштовує наші вимоги є задача image-to-image translation, яка полягає у тому, що ми повинні перетворити зображення з одного домену у інший, при цьому кількість каналів вхідного та вихідного зображення не повинна бути однаковою. Таким чином подавши маски, на яких будуть зазначені класи кожного з пікселів, ми зможемо генерувати штучні супутникові знімки, з потрібним нам відношенням класів.

Наївні підходи до розв'язку даної задачі, які полягають у навчанні звичайної згорткової нейронної мережі з мінімізацією евклідової відстані, призведуть до того, що значення пікселів кожного з класів будуть усереднені. І, відповідно, ми отримаємо згенеровані зображення дуже низької якості. Тож слід звернути увагу на GAN, які дозволяють задати кінцеву ціль у вигляді «зробити згенеровані зображення такими, що складно відрізнити від справжніх».

Існує цілий клас генеративно-змагальних мереж, які дозволяють вирішувати дану задачу. Це так звані умовні (conditional) GAN. Однією з провідних архітектур, саме у задачі image-to-image translation, є Pix2Pix [15].

Першим істотною відмінністю архітектури Pix2Pix від класичних GAN є функція помилки, а саме: для вхідного зображення  $x$  та бажаного вихідного зображення  $y$  вона виглядає наступним чином:

$$\mathbb{E}_{x,y} \log D(x,y) + \mathbb{E}_x \log(1 - D(x, G(x))) + \lambda_{L_1} \mathbb{E}_{y,x} \|y - G(x)\|_1 \quad (2.4)$$

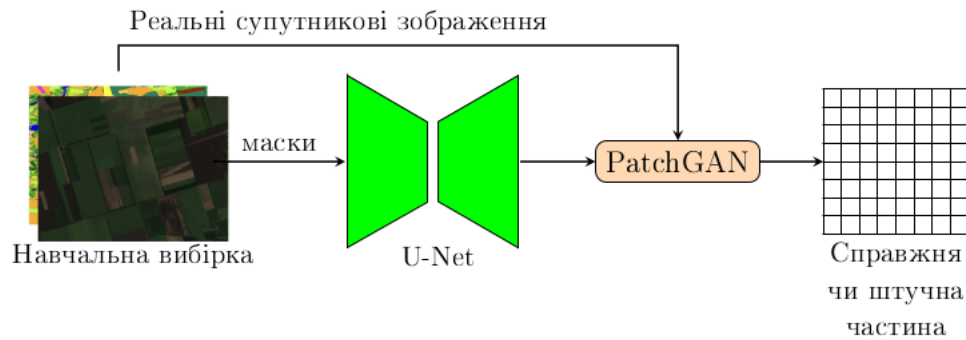
Тобто функція помилки являє собою поєднання класичної помилки для генеративно-змагальних мереж, з тією зміною, що дискримінатор приймає два аргументи, і  $L_1$  помилки. Перша частина як і у всіх інших GAN відповідає за те, щоб згенеровані зображення виглядали подібними до реальних, а друга частина, намагається досягти попиксельної однаковості між справжніми та штучними знімками.  $L_1$  помилка була обрана авторами [15], бо їх дослідження показали, що використання  $L_2$  помилки призводить до більшого рівня розмитості на вихідних зображеннях.

Наступною відмінністю архітектури Pix2Pix є використання так званого PatchGAN у якості дискримінатора. Він аналізує справжність не усього зображення, а лише окремих його частин, і потім усереднює отримані результати. Це дозволяє ефективно моделювати вихідне зображення як випадкове поле Маркова, при умові, що припускається незалежність між пікселями, які потрапили у різні частини. Це можна розуміти [15] як текстурну помилку або помилку стилю.

При чому слід звернути увагу на те, що дискримінатор приймає вхідне та вихідне зображення, тобто має змогу порівняти а чи відповідає згенероване зображення тому входу, який був поданий на генератор.

Генератор повинен генерувати вихідне зображення такого ж розміру який мало вхідне, тож цілком закономірно очікувати, що архітектура генератора повинна бути більш-менш симетричною. Однією з таких, добре відомих архітектур, є вже розглянута нами архітектура UNet. Вона добре зарекомендувала себе у задачах семантичної сегментації, які дещо схожі на задачу image-to-image translation. До того ж використання особливості UNet дозволяють враховувати як високорівневі, так і низькорівневі ознаки. Саме ґрунтуючись на цих принципах, автори [15] і запропонували використати варіацію архітектури UNet у якості генератора.

Таким чином повністю архітектуру Pix2Pix можливо подати за допомогою наступної діаграми (рис. 2.3).



**Рисунок 2.3** – Архітектура Pix2Pix у застосуванні до супутникових знімків

Проте результати роботи запропонованої архітектури інколи бажають покращення, особливо коли мова йде про генерацію деталізованих зображень або зображень великої розмірності. Проте Pix2Pix був і залишається потужною базою, на якій ґрунтуються більш досконалі методи. Одним з таких є архітектура Pix2PixHD [16].

У Pix2PixHD було запропоновано багато вдосконалень класичного Pix2Pix, серед яких були значні зміни генератора, проте ці зміни пов'язані з бажанням генерувати саме зображення більшого розміру. До того ж даний генератор вимагає більших обчислювальних ресурсів. Ми ж зосередимось на тих нововведеннях, які мають на меті покращення, у тому числі, якості зображень.

Першим нововведенням є застосування декількох дискримінаторів. Кожний з цих дискримінаторів має архітектуру PatchGAN та застосовується при різних розмірах зображень. Для отримання піраміди зображень з різними розмірами використовується операція pooling, яка зменшує у 2 рази. Таким чином різні дискримінатори мають різні рецептивні поля, що дозволяє їм разом враховувати різні рівні освідомлення зображення та змінювати генератор так, щоб він генерував глобально узгоджені результати. Результуюча оцінка ж такого складеного дискримінатора являє собою суму оцінок кожної зі складових частин.

Наступні покращення пов'язані з використанням додаткових

функцій помилок. Перша з них являє собою так звану помилку відповідності ознак. Даний критерій дозволяє стабілізувати процес навчання за рахунок того, що ми спонукаємо генератор продукувати зображення, статистики яких будуть наближеними до реальних при різних масштабах. А саме: дана функція помилки обраховується за допомогою результатів внутрішніх шарів усіх дискримінаторів, і призводить до того, що внутрішні представлення дискримінаторів для штучних та справжніх зображень повинні бути наближені один до одного. Позначивши як  $D_l$  дискримінатор  $l$  та, відповідно,  $D_l^{(i)}$  – вихід  $i$ -го шару  $l$ -го дискримінатора, дану похибку можна подати у вигляді:

$$L_{FM}(G, D_k) = \mathbb{E}_{x,y} \sum_{i=1}^T \frac{1}{N_i} \|D_k^i(x, y) - D_k^i(x, G(x))\|_1, \quad (2.5)$$

де  $T$  – загальна кількість шарів дискримінатора,  $N_i$  – кількість елементів на шарі  $i$ . Для обрахунку ж внеску у загальну помилку, значення сумуються за усіма дискримінаторами та множаться на константу  $\lambda_{FM}$ .

Друга функція помилки, яка була представлена і застосована у [16], є помилкою, яка основана на тому, щоб порівнювати внутрішні представлення не тільки дискримінатора, який постійно перебуває у процесі навчання, а і іншою згортковою навченою мережі. А саме, її можна представити наступним чином:

$$L_{VGG} = \sum_{i=1}^T \frac{1}{N_i} \|F_k^i(y) - F_k^i(G(x))\|_1, \quad (2.6)$$

де  $F_k^i$  –  $i$ -ий шар навченої згорткової нейромережі, у більшості випадків використовується VGG,  $T$  – загальна кількість шарів цієї мережі,  $N_i$  – кількість елементів на шарі  $i$ .

Ідея її застосування даної помилки ґрунтується на тому, що високорівневі внутрішні ознаки навченої глибокої згорткової нейромережі для згенерованих та справжніх зображень не повинні сильно відрізнятися.

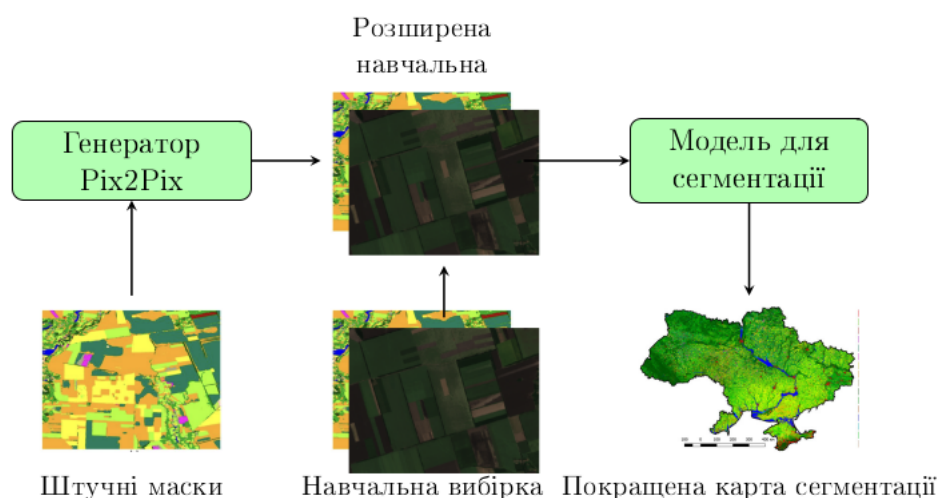
Таким чином використавши ці покращення ми зможемо отримати більш якісні згенеровані зображення, проте питання а чи допоможуть нам аугментовані вибірки отримані таким видозміненим генератором досягти більшої точності семантичної сегментації залишається відкритим і буде досліджено у наступних розділах даної роботи.

## **2.4 Перспективи застосування GAN для генерації штучних супутникових знімків**

Маючи ефективні нейромережеві архітектури для розв'язку задачі image-to-image translation ми в змозі перейти до генерації аугментованих навчальних вибірок. Ідея даного процесу полягає у тому, що ми на основі штучних масок з мітками класів, будемо генерувати штучні супутникові знімки. Ці штучні маски у свою чергу будуть такими, що допоможуть збільшити сумарну кількість пікселів для потрібних міноритарних класів і таким чином збалансувати навчальну вибірку. Це у кінцевому результаті очікувано має дати кращі показники якості сегментації, при застосуванні доповнених навчальних матеріалів, отриманих за цією процедурою. Схематично даний процес можна зобразити як на рисунку 2.4.

Архітектура Pix2Pix і її варіації є потужним підґрунтям для реалізація даного процесу, бо дозволяє генерувати якісні зображення, і при цьому надає нам змогу контролювати клас кожного пікселя, без чого неможливо створення штучних вибірок для задачі семантичної сегментації. Окремої уваги заслуговує той факт, що хоч ми і розглядаємо методи покращення згенерованих зображень, все ж таки вирішальним для нас є саме покращення якості семантичної сегментації, особливо щодо міноритарних класів. Тому гонитва за тим, щоб штучно згенеровані за допомогою генеративно-змагальних мереж зображення не було можливо відрізнити від справжніх недоцільна, а звертати увагу треба перш за все на метрики класифікації, що будуть отримані у результаті навчання

моделей семантичної сегментації на наборах даних доповнених за допомогою GAN.



**Рисунок 2.4** – Процес аугментації наборів даних супутникових знімків за допомогою Pix2Pix

Наступним важливим питанням є засоби створення штучних масок, які дозволять збалансувати набір даних. Для вирішення цієї задачі теж можна використовувати багато підходів, у тому числі і генеративно-змагальні мережі, проте простішим і, тим не менш, логічно обґрунтованим є підхід щодо зміни міток класів на вже існуючих масках, які відповідають справжнім супутниковим знімкам. Це дозволить бути впевненими у тому, що просторове розташування класів завжди буде відповідати шаблонам з реального світу.

Для того, щоб змінювати існуючі маски пропонується наступна стратегія: ми будемо змінювати лише декілька класів на кожній ітерації, що дозволить наблизити характеристики штучних масок до реальних. Причому змінювати пропонується класи, які для даного знімку є такими, що мають найбільшу кількість пікселів. Таким чином можемо привести алгоритм генерації штучних масок.

**Алгоритм 2.2** (Алгоритм генерації масок).

Параметри:  $l$  – кількість класів, які будуть змінені Вхідні дані: навчальна вибірка, бажана кількість пікселів для кожного з класів

1) Визначити які класи не досягли достатньої кількості пікселів та додати їх у чергу

2) Поки залишились елементи у черзі

а) Дістати з черги  $l$  міток класів  $k_1, \dots, k_l$ . (Якщо на даній ітерації у черзі менше за  $l$ , то встановити  $l$  таким, що дорівнює довжині черзі).

б) Для наступного елемента у наборі даних, отримати маску, в якій усі мітки  $l$  класів, які займає найбільше пікселів замінити на  $k_1, \dots, k_l$  відповідно

в) Зберегти отриману маску

г) Додати кількість пікселів за кожним класом згенерованої маски до сумарної кількості пікселів за класами. Якщо для якогось з класів  $k_1, \dots, k_l$  ще не досягнута бажана кількість пікселів, то додати такі мітки послідовно у чергу.

Як можна побачити даний алгоритм приймає бажану кількість класів, а не завжди робить розподіл класів наближеним до рівномірного. Це пов'язано з тим, що, по-перше, нас може цікавити підвищення якості семантичної сегментації не усіх класів, а лише певних. По-друге, намагання отримати рівномірний розподіл може призвести до того, що частка згенерованих прикладів буде значно переважати частку справжніх, що може дуже негативно впливати на результати, бо недоліки, у штучно згенерованих за допомогою генеративно-змагальних мереж зображень, все ж таки присутні.

## Висновки до розділу 2

У другому розділі ми зосередились на огляді генеративних моделей та можливостей їх застосування для аугментації навчальних вибірок супутникових знімків.

Перш за все було розглянуто постановку задачі генерування зображень у математичній постановці та різні класи генеративних моделей. Ґрунтуючись на недоліках і перевагах кожного з цих класів, було визначено, що найбільшу перспективу для вирішення задачі генерації супутникових зображень мають генеративно-змагальні мережі.

Тож було докладно проаналізовано властивості збіжності цих мереж, наведено алгоритм навчання та описано основні проблеми, які виникають з GAN та шляхи їх вирішення.

Через те, що задача аугментації вибірок у нашій постановці вимагає контролю за міткою класу кожного з пікселів згенерованого зображення, ми звернулись до генеративно-змагальних мереж, які дозволяють розв'язувати задачу image-to-image translation. Визначними нейромережевими архітектурами у цій галузі є Pix2Pix та нащадки, тож було детально проаналізовано їх застосування до генерації супутникових знімків.

На сам кінець, був запропонований процес аугментації навчальних вибірок супутникових знімків, який полягає у тому, що ми, за описаним алгоритмом, спершу генеруємо маски, які будуть коригувати незбалансованість класів, а після цього, ґрунтуючись на даних масках, генерувати штучні супутникові знімки за допомогою архітектури Pix2Pix та її варіацій.



## **3 АНАЛІЗ ЗАСТОСУВАННЯ GAN ДЛЯ АУГМЕНТАЦІЇ НАВЧАЛЬНИХ ВИБІРОК**

У даному розділі ми застосуємо розглянуті нами у попередніх частинах моделі та методи на практиці, а саме навчимо варіацію моделі архітектури UNet з різними функціями помилки. Обрахуємо метрики і визначимо чи дійсно проблема незбалансованості класів істотно впливає на якість семантичної сегментації супутникових знімків. І на сам кінець за допомогою генеративно-змагальних мереж згенеруємо штучні приклади, які додамо до реальних даних та будемо проводити навчання вже на цій аугментованій вибірці. І визначимо чи дійсно подібний спосіб дозволяє покращити якість семантичної сегментації.

Вихідний код, який був створено для усіх наведених у даному розділі експериментів доступний на GitHub <sup>1</sup>.

### **3.1 Попередня обробка навчальних даних**

Для перевірки того, чи дійсно запропонований підхід дозволить збільшити якість семантичної сегментації супутникових знімків проведемо експерименти з класифікації на 19 класів композиту Sentinel-2A [17] для Київської області взятих з 1 липня по 1 серпня 2021 року. Класи представляють собою здебільшого сільськогосподарські культури, тому розв'язання задачі сегментації саме на таких даних має широкий прикладний потенціал.

Ми використовували саме композит супутникових знімків, бо територія Київської області надто велика, щоб вона могла повністю розміститися на одному супутниковому знімку Sentinel-2A. Саме тому був взятий часовий ряд червоного, зеленого, синього та ближнього

---

<sup>1</sup><https://github.com/ShkalikovOleh/SatelliteGAN>

інфрачервоного каналів за липень 2021 р., який потім був об'єднаний. У тих частинах, що перетиналися були, були обрані медіанні значення пікселів, також знімки були очищені від детектованих хмар та тіней від них.

Після чого ці дана колекція знімків була об'єднана у один великий композит для Київської області та до цього композиту була додана окремим каналом маска з мітками класів, яка була отримана у результаті досліджень науковцями інституту космічних досліджень Національної академії наук України.

Сучасні нейромережеві архітектури не дозволяють нам працювати з зображення дуже великого розміру, тому ми вимушені розрізати отриманий композит на квадратні частини розміром  $256 \times 256$  пікселів. Такий вибір обґрунтований тим, що саме такий розмір зображення генерувався авторами класичного Pix2Pix [15].

Попри те, що ми брали дані за місяць, все одно залишились території, які не були покриті супутником за спостережуваний період. До того ж, контури Київської області не є ідеальною геометричною фігурою, тож на деяких знімках, на яких показані кордони області відсутні значення пікселів або масок, які відповідають іншим областям. Через ці два фактори у деяких отриманих зображеннях присутні пікселі, у яких відсутнє значення. Такі приклади були відфільтровані та не використовувалися при навчанні нейронних мереж.

На сам кінець, через те, що використовувалися знімки з попередньою обробкою та корекцією класу ТОВА, то область значень кожного з каналів була не класичною  $[0, 1]$ , а мала більше максимальне значення. У свою чергу більшість нейромережових архітектур працює зі значеннями у проміжку  $[-1, 1]$ . Тож був застосований аналог min-max нормалізації, а саме кожне значення пікселю кожного каналу було поділено на максимальне значення для даного каналу даного зображення, а після цього класична стандартизація з математичним сподівання і стандартним відхиленням рівними 0.5.

### 3.2 Результати сегментації без використання GAN

Для розв'язання задачі семантичної сегментації даної вибірки супутникових знімків було використано архітектуру UNet. Для того, щоб ми мали змогу перевірити якість моделі використовуючи метрики, і при цьому отримали достовірні значення, тобто уникнули явища перенавчання, вибірку було поділено навпіл на навчальну і тестову.

У якості енкодера у використаній варіації UNet було обрано ResNet-34, що є компромісом між обчислювальною потужністю, необхідною для навчання подібних моделей та складністю мережі.

Дана вибірка є незбалансованою, що можна побачити на статистиках по сумарній кількості пікселів для кожного класів на рисунку 1.3 (мітки класів відповідають назвам, які можна знайти у таблиці 3.2). Тож для подолання проблем, які пов'язані з цим, ми спробували застосувати різні функції помилок, а саме зважені та не зважені Cross-Entropy та Focal Loss. У якості вагових коефіцієнтів для кожного класу були обрані нормовані зворотні відношення сумарної кількості пікселів класу до кількості усіх пікселів у тренувальній вибірці.

У якості оптимізатора було обрано добре відомий, один з найбільш застосовуваних та ефективних, алгоритм Adam [10] зі швидкістю навчання  $2 \cdot 10^{-4}$ . Розмір батчу дорівнював 64.

**Таблиця 3.1** – Глобальні метрики точності сегментації для реальної вибірки

Метрика	CE Loss		Focal Loss	
	w/o W	W	w/o W	W
Accuracy	0.76	0.738	<b>0.762</b>	0.755
$\kappa$	0.72	0.697	<b>0.722</b>	0.716
IoU	0.356	0.352	0.363	<b>0.368</b>

У результаті 500-та епох навчання було отримані значення метрик,

таких як точність (Assurasy), міра Жаккара (IoU) та каппа Коена ( $\kappa$ ), для різних варіацій функції помилки. Вони наведені у таблиці 3.1 (w/o W означає не зважену помилку, W – зважену).

Що стосується якості сегментації за кожним з класів, то слід звернутися до таких метрик як Producer Assurasy PA та User Assurasy UA, наведених у таблиці 3.2.

**Таблиця 3.2** – Метрики точності сегментації за класами для реальної вибірки

Назва класу	PA				UA			
	CE Loss		Focal Loss		CE Loss		Focal Loss	
	w/o W	W	w/o W	W	w/o W	W	w/o W	W
Штучні об'єкти	0.64	<b>0.66</b>	0.649	0.652	0.664	0.59	<b>0.68</b>	0.668
Зернові культури	0.822	0.772	<b>0.825</b>	0.791	0.718	0.712	0.721	0.734
Ріпак	0.285	0.285	0.272	<b>0.297</b>	<b>0.533</b>	0.404	0.533	0.458
Гречка	0.012	0.038	0.032	<b>0.047</b>	0.151	0.059	<b>0.165</b>	0.1
Кукурудза	0.882	0.825	<b>0.887</b>	0.861	0.837	<b>0.853</b>	0.834	0.85
Буряк	0.27	0.356	0.324	<b>0.367</b>	0.472	0.417	<b>0.537</b>	0.435
Соняшник	<b>0.887</b>	0.848	0.881	0.866	0.83	0.83	0.839	<b>0.844</b>
Соя	0.581	<b>0.64</b>	0.595	0.608	0.715	0.584	<b>0.716</b>	0.662
Інші культури	0.09	<b>0.207</b>	0.09	0.201	0.222	0.162	<b>0.233</b>	0.192
Ліс	0.915	0.827	<b>0.918</b>	0.897	0.894	<b>0.931</b>	0.897	0.913
Необроблювані землі	0.726	0.592	<b>0.741</b>	0.686	0.683	<b>0.723</b>	0.692	0.716
Відкритий ґрунт	0.464	0.694	0.494	<b>0.696</b>	<b>0.573</b>	0.392	0.57	0.438
Вода	<b>0.967</b>	0.964	0.962	0.959	0.939	0.927	<b>0.94</b>	<b>0.94</b>
Болото	0.315	<b>0.5</b>	0.308	0.434	0.461	0.287	<b>0.476</b>	0.369
Ячмінь	0.205	<b>0.292</b>	0.206	0.28	0.309	0.265	<b>0.317</b>	0.3
Горох	0.01	<b>0.018</b>	0.014	0.016	0.101	0.03	<b>0.135</b>	0.075
Трави	0.011	<b>0.051</b>	0.016	0.038	0.083	0.05	0.1	<b>0.111</b>
Сади, парки, лісополоси	0.51	0.496	<b>0.522</b>	0.495	0.491	0.433	<b>0.498</b>	0.486
Виноградники	0.039	0.344	0.01	<b>0.345</b>	0.169	0.09	<b>0.197</b>	0.144

Як можна побачити, за усіма метриками, які є глобальними перевагу має нейромережа архітектури UNet, яка при навчанні використовувала не зважену функцію помилки FocalLoss, а саме вона переважає у Accurasy та капі Коена, у той же час відставання за метрикою IoU від лідера не значне.

Що стосується результатів метрик за кожним окремим класом, то тут бачимо той факт, що як і очікувалось ті класі, які є міноритарними, тобто кількість їх пікселів мала відносно інших класів, мають значно нижчі показники, аніж мажоритарні. Але і для цих метрик мережа, що навчалась з не зваженим Focal Loss видає переважно кращі результати, особливо, якщо звернути увагу на User Accurasy, де така варіацію UNet переважає для 11 класів, більшість з яких саме міноритарні. Значення метрики Producer Accurasy теж є достатньо високими: за 5 класами підхід з використання цієї функції помилки є найкращими (у той час як лідер переважає у 7 класах), а для інших теж показує гарні результати.

### **3.3 Результати сегментації з використанням аугментованого набору даних**

Як ми переконались, незбалансованість класів дуже істотно впливає на якість семантичної сегментації супутникових знімків. Тож спробуємо доповнити нашу вибірку, таким чином збалансувавши її. Для цього використаємо розроблений нами процес аугментації за допомогою генеративно-змагальних мереж.

#### **3.3.1 Генерація доповнених вибірок**

Перш за все були навчені GAN архітектури Pix2Pix та її варіацію (будемо позначати її Pix2Pix + HD), що використовує ті покращення, що описані у розділі 2, а саме: декілька дискримінаторів та функції помилки

узгодженості ознак (2.5) та узгодженості внутрішніх представлень навченої мережі класифікації (2.6).

Навчання відбувалося протягом 500 епох з використання оптимізатора Adam [10] зі швидкістю навчання  $2 \cdot 10^{-4}$ . У якості параметрів експоненційного середнього  $(\beta_1, \beta_2)$  використовувались значення 0.5, 0.999 відповідно. Розмір батчу був невеликий і дорівнював 8, бо великі значення можуть приводити до зниження якості згенерованих зображень.

Приклади справжніх зображень (лише червоний, зелений та синій канал) та тих, які генерують навченні генеративно-змагальні мережі наведені на рисунках 3.1 – 3.3.



**Рисунок 3.1** – Приклади справжніх супутникових знімків



**Рисунок 3.2** – Приклади згенерованих Pix2Pix зображень

Як ми бачимо, на згенерованих прикладах присутня деяка розмитість та інші незначні артефакти, проти вони виглядають наближеними до справжніх. Причому візуально знімки згенеровані



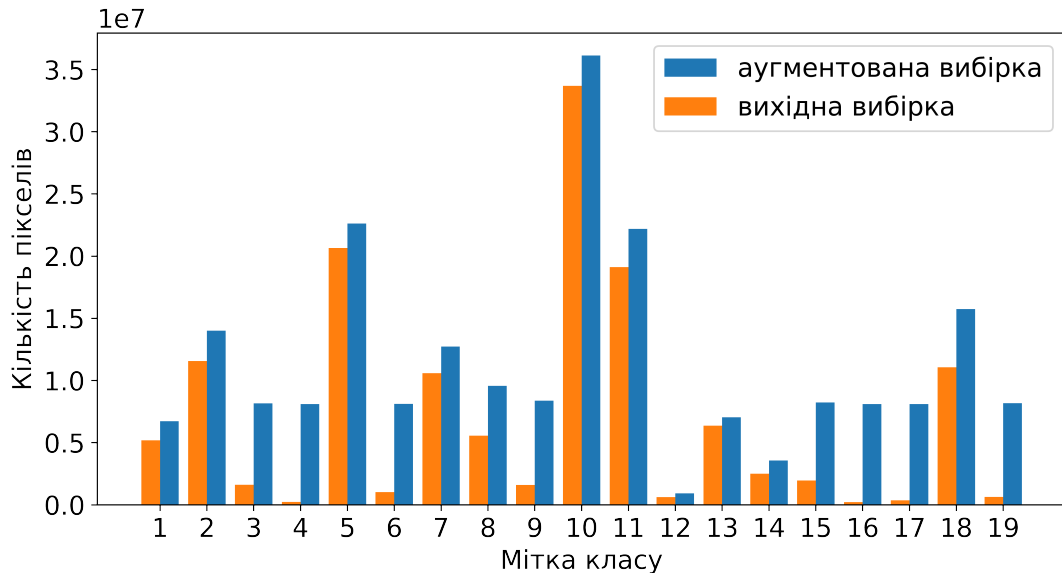
**Рисунок 3.3** – Приклади згенерованих Pix2Pix + HD зображень

модифікацією Pix2Pix виглядають краще. До того ж, як ми вже зазначали, кінцевою нашою метою є саме покращення якості сегментації, а не генерація ідеальних зображень.

Наступним етапом було застосування алгоритму 2.2 генерації штучних масок. Як можна побачити на рисунку 3.1 у досліджуваному наборі даних дуже сильний дисбаланс у кількості пікселів для кожного з класів, тому якщо ми будемо намагатися зробити повністю рівномірний розподіл, то нам потрібно буде згенерувати дуже багато штучних зображень, що може тільки зменшити метрики якості семантичної сегментації. Тож було обрано наступний підхід: за цільові значення кількості пікселів обрано 70% від кількості пікселів класу «Зернові культури». У якості параметру  $l$  алгоритму було обрано 2. При чому у прикладних застосуваннях нас цікавить підвищення якості класифікації не усіх класів, а лише окремих, як наприклад нас не цікавить розпізнавання штучних об'єктів, у тому числі будинків та автомобілів, якщо ми вирішуємо задачі, які пов'язані з сільськогосподарськими культурами. Тож ми використали властивість розробленого алгоритму, яка полягає у тому, що він дозволяє обирати класи, які будуть доповнюватись, та доповнили наступні класи: ріпак, гречка, буряк, соя, інші культури, виноградники, трави, горох, ячмінь.

Усього було згенеровано 1252 нових приклади кожної з навчених генеративно-змагальних мереж, і усі вони були додані до вихідної тренувальної вибірки, таким чином ми отримали аугментовану вибірку.

Статистика по кількості пікселів у доповненій вибірці наведена на рисунку 3.4. Як можна побачити, класи, які нас особливо цікавлять, стали набагато більш збалансовані.



**Рисунок 3.4** – Порівняння кількості пікселів для різних класів у досліджуваному та аугментованому наборах даних

### 3.3.2 Аналіз якості семантичної сегментації при застосуванні доповнених вибірок

Маючи аугментовану вибірку ми маємо змогу перейти до перевірки нашої гіпотези про те, що генеративно-змагальні мережі здатні допомогти підвищити якість семантичної сегментації. Для цього, щоб порівнювати з результатами на вихідній навчальній вибірці, ми теж навчаємо мережу архітектури UNet з енкодером ResNet-34. Усі параметри навчання були обрані такі ми же, окрім кількості епох, яка була зменшена до 300 через те, що ми працюємо з більшою вибіркою.

Результати у вигляді метрики точності по класам наведені у таблиці 3.3.



**Таблиця 3.3** – Порівняння метрик точності сегментації за кожним класом для доповненої та вихідної вибірок

Назва класу	РА			UA		
	Вихідна вибірка	Аугментована вибірка		Вихідна вибірка	Аугментована вибірка	
		Pix2Pix	Pix2Pix + HD		Pix2Pix	Pix2Pix + HD
Штучні об'єкти	0.649	0.652	<b>0.6795</b>	<b>0.68</b>	0.656	0.587
Зернові культури	0.825	0.827	<b>0.829</b>	<b>0.721</b>	0.719	0.71
Ріпак	0.272	0.276	<b>0.292</b>	0.533	<b>0.539</b>	0.537
Гречка	0.032	<b>0.036</b>	0.03	0.165	0.227	<b>0.232</b>
Кукурудза	<b>0.887</b>	0.875	0.863	0.834	<b>0.846</b>	0.798
Буряк	0.324	0.333	<b>0.34</b>	0.537	0.563	<b>0.57</b>
Соняшник	0.881	0.88	<b>0.886</b>	0.839	<b>0.85</b>	0.841
Соя	0.595	<b>0.627</b>	0.596	0.716	0.711	<b>0.718</b>
Інші культури	0.09	0.092	<b>0.101</b>	<b>0.233</b>	0.221	0.23
Ліс	<b>0.918</b>	0.91	0.91	0.897	<b>0.9</b>	0.898
Необроблювані землі	<b>0.741</b>	0.736	0.728	<b>0.692</b>	0.67	<b>0.692</b>
Відкритий ґрунт	<b>0.494</b>	0.373	0.382	0.57	<b>0.571</b>	0.562
Вода	0.962	<b>0.965</b>	0.953	<b>0.94</b>	0.937	0.929
Болото	<b>0.308</b>	0.295	0.287	0.476	0.446	<b>0.477</b>
Ячмінь	0.206	<b>0.207</b>	0.201	0.317	0.315	<b>0.334</b>
Горох	<b>0.014</b>	0.009	0.01	<b>0.135</b>	0.133	0.132
Трави	0.016	<b>0.021</b>	0.019	0.1	0.114	<b>0.117</b>
Сади, парки, лісополоси	<b>0.522</b>	0.506	0.511	<b>0.498</b>	0.487	0.484
Виноградники	0.01	0.026	<b>0.027</b>	0.197	0.1443	<b>0.21</b>

Найкращі показники має Pix2Pix + HD, особливо, що стосується тих класів, якість розпізнавання яких ми прагнули підвищити. Особливою уваги заслуговує той факт, що подекуди показники вдалось збільшити більш ніж у 1.5 рази, як наприклад для гречки та виноградників.

Якщо ж казати про глобальні метрик , то вони наведені у таблиці

3.4. За цією категорією метрик навчання на аугментованій вибірці теж показало результати, які краще за навчання тільки на реальних даних.

**Таблиця 3.4** – Порівняння глобальних метрик точності сегментації для доповненої та вихідної вибірок

Метрика	Вихідна вибірка	Аугментована вибірка	
		Pix2Pix	Pix2Pix + HD
Accuracy	<b>0.762</b>	<b>0.762</b>	0.755
$\kappa$	0.722	0.722	<b>0.723</b>
IoU	0.363	<b>0.367</b>	<b>0.367</b>

Отримані результати свідчать про те, що якщо і відбувається незначне погіршення у точності класифікації мажоритарних класів, то воно повністю нівелюється значними покращеннями для міноритарних, а саме тих, яких ми і прагнули покращити.

### Висновки до розділу 3

У даному розділі було проведено велику кількість експериментів з застосуванням супутникових даних Sentinel-2 для Київської області, взятих за період з 1 липня по 1 серпня 2021 року. Було детально описано методи отримання та попередньої обробки даних, які після цього були поділені на тренувальну та тестову вибірки, які використовувалися у подальшому.

Було проаналізовано вплив різних функцій помилки, а саме крос-ентропії та Focal Loss з використанням та без використання вагів. Було обраховані як глобальні метрики якості семантичної сегментації, так і для кожного з класів. У результаті було визначено, що найкращі показники має не зважений Focal Loss.

Генеративно-змагальні мережі архітектури Pix2Pix та її варіацію, що використовує дискримінатор та функції помилки із архітектури

Pix2PixHD були навчені на даному наборі даних. Якість згенерованих зображень модифікації виявилась більшою за стандартний Pix2Pix.

Врешті решт було застосовано, описаний у попередніх розділах, процес доповнення навчальної вибірки для тих класів, які мали особливе значення. Результати якості по класам дозволять стверджувати, що запропонований підхід дозволив значно підвищити якість міноритарних класів. При цьому деякі метрики для глобальних класів незначно погіршились. Якщо ж порівнювати різні GAN, то тут як і у випадку з якістю згенерованих зображень перевагу має модифікація Pix2Pix + HD. Глобальні метрики теж свідчать про те, що загальна якість семантичної сегментації при застосуванні аугментованих вибірок не погіршилась.

## ВИСНОВКИ

У результаті проведеної роботи була досліджена проблема семантичної сегментації супутникових знімків, яка має важливу роль у прикладних застосуваннях. Проте були виявлені та описані проблеми, які виникають під час розв'язку цієї задачі сучасними нейромережевими методами, а саме: складність створення великих вибірок та сильна незбалансованість класів. Причому, на відміну від інших сфер, отримати навчальні приклади, які могли б змінити розподіл класів у багатьох випадках для супутникових даних не є можливим.

Для подолання цих проблем було розглянуто можливість аугментації навчальної вибірки штучно згенерованими навчальними прикладами. Для цього була дослідженні моделі, які дозволяють вирішити цю задачу, а саме генеративно-змагальні мережі. При цьому, для вирішення проблеми незбалансованості класів стандартні GAN не кращий вибір, бо ми не можемо контролювати розподіл класів на згенерованих зображеннях. Тому були опрацьовані моделі image-to-image translation, зокрема Pix2Pix, які дозволяють генерувати супутникові знімки на основі масок. У свою чергу модифікація вже існуючих масок, тобто зміна одних класів на інші дозволяє скоригувати незбалансованість класів.

Таким чином було розроблено процес аугментації наборів даних супутникових знімків, ефективність якого була перевірена експериментальним чином (табл. 3.3, 3.4). Для генерації були використані різні модифікації архітектури Pix2Pix, у тому числі та, яка використовує дискримінатор та функції помилки з архітектури Pix2PixHD, яка показала кращі результати, як з точки зору метрик семантичної сегментації так і за якістю згенерованих зображень.

Також було проведено дослідження різних функції помилки, які застосовні при навчанні модифікації архітектури UNet, яка безпосередньо

і відповідає за семантичну сегментацію. Серед усіх досліджуваних підходів найкращим виявився той, що використовував не зважену функцію похибки Focal Loss. Його ж ми і застосовували при навчанні моделей вже на доповнених вибірках.

У результаті ми отримали підтвердження того, що запропонований підхід дозволяє значно підвищити якість семантичної сегментації міноритарних класів, при цьому якість класифікації мажоритарних класів, хоч у деяких випадках і стала нижчою, проте незначно.

Очевидні і подальші напрямки досліджень у даній сфері. По-перше, це дослідження методів генерації, які б дозволили б врахувати той факт, що супутникові знімки однієї території у різні пори року значно відрізняються. По-друге, створення більш досконалих методів генерації масок. І врешті решт, покращення існуючих генеративно-змагальних мереж для генерації ще більш реалістичних штучних супутникових знімків.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Deep learning classification of land cover and crop types using remote sensing data / Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, Andrii Shelestov // IEEE Geoscience and Remote Sensing Letters. — 2017. — Vol. 14, no. 5. — Pp. 778–782.
2. Shumilo, Leonid. U-Net Model for Logging Detection Based on the Sentinel-1 and Sentinel-2 Data / Leonid Shumilo, Nataliia Kussul, Mykola Lavreniuk // 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. — 2021. — Pp. 4680–4683.
3. Jadon, Shruti. A survey of loss functions for semantic segmentation / Shruti Jadon // 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). — 2020. — Pp. 1–7.
4. Grandini, Margherita. Metrics for multi-class classification: an overview / Margherita Grandini, Enrico Bagli, Giorgio Visani // arXiv. — 2020. — (Препринт / arXiv:2008.05756).
5. Gradient-based learning applied to document recognition / Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner // Proceedings of the IEEE. — 1998. — Vol. 86, no. 11. — Pp. 2278–2324.
6. Trung, Nghia Duong. Automated Pneumonia Detection in X-Ray Images via Depthwise Separable Convolution Based Learning / Nghia Duong Trung, Tuyen Tran Ngoc, Hiep Xuan Huynh // Proc. FAIR-Fundamental And Applied IT Research. — 2019.
7. Ronneberger, Olaf. U-net: Convolutional networks for biomedical image segmentation / Olaf Ronneberger, Philipp Fischer, Thomas Brox // International Conference on Medical image computing and computer-assisted intervention / Springer. — 2015. — Pp. 234–241.

8. Goodfellow, Ian. Nips 2016 tutorial: Generative adversarial networks / Ian Goodfellow // arXiv. — 2016. — (Препринт / arXiv:1701.00160).

9. Generative adversarial nets / Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza et al. // Advances in neural information processing systems. — 2014. — Vol. 27.

10. Kingma, Diederik P. Adam: A method for stochastic optimization / Diederik P Kingma, Jimmy Ba // arXiv. — 2014. — (Препринт / arXiv:1412.6980).

11. Arjovsky, Martin. Towards principled methods for training generative adversarial networks / Martin Arjovsky, Léon Bottou // arXiv. — 2017. — (Препринт / arXiv:1701.04862).

12. Spectral normalization for generative adversarial networks / Takeru Miyato, Toshiki Kataoka, Masanori Koyama, Yuichi Yoshida // arXiv preprint arXiv:1802.05957. — 2018.

13. Stabilizing training of generative adversarial networks through regularization / Kevin Roth, Aurelien Lucchi, Sebastian Nowozin, Thomas Hofmann // Advances in neural information processing systems. — 2017. — Vol. 30.

14. Arjovsky, Martin. Wasserstein generative adversarial networks / Martin Arjovsky, Soumith Chintala, Léon Bottou // International conference on machine learning / PMLR. — 2017. — Pp. 214–223.

15. Image-to-image translation with conditional adversarial networks / Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A Efros // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2017. — Pp. 1125–1134.

16. High-resolution image synthesis and semantic manipulation with conditional gans / Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu et al. //

Proceedings of the IEEE conference on computer vision and pattern recognition.  
— 2018. — Pp. 8798–8807.

17. Sentinel-2: ESA's optical high-resolution mission for GMES operational services / Matthias Drusch, Umberto Del Bello, Sébastien Carlier et al. // Remote sensing of Environment. — 2012. — Vol. 120. — Pp. 25–36.