

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»

Приладобудівний

(повна назва інституту/факультету)

Комп'ютерно-інтегрованих оптичних та навігаційних систем

(повна назва кафедри)

УДК

«До захисту  
допущено»

Завідувач кафедри

\_\_\_\_\_ Н. І. Бурау

(підпис)

(ініціали, прізвище)

“ ” \_\_\_\_\_ 2022р.

## Магістерська дисертація

зі спеціальності (спеціалізації): 151 Автоматизація та комп'ютерно-інтегровані технології

(код і назва спеціальності)

на тему: Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках

Виконала: студентка 2 курсу, групи ПО-11 мп

(шифр групи)

\_\_\_\_\_ Снагощенко Діана Михайлівна

(прізвище, ім'я, по батькові)

(підпис)

Науковий керівник проф., д.т.н. Колобродов В. Г.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Консультант стартап-проект проф., д.е.н. Бояринова К.О.

(назва розділу)

(науковий ступінь, вчене звання, прізвище, ініціали)

(підпис)

Рецензент доцент, к.т.н. Терещенко М.Ф.

(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали)

(підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.

Студентка \_\_\_\_\_

(підпис)

**Національний технічний університет України**  
**«Київський політехнічний інститут**  
**імені Ігоря Сікорського»**

Рівень вищої освіти – другий (магістерський)

Спеціальність – 151 – Автоматизація та комп'ютерно-інтегровані технології

Освітньо-професійна програма – Комп'ютерно-інтегровані системи та технології в приладобудуванні

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ Надія БУРАУ

«\_\_» \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ**

**на магістерську дисертацію студентки**

Снагощенко Діани Михайлівни

1. Тема дисертації «Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках», науковий керівник дисертації професор, д.т.н. Колобродов В.Г. затверджені наказом по університету від «\_\_» \_\_\_\_\_ 20\_\_ р. № \_\_\_\_\_

2. Термін подання студентом дисертації «06» грудня 2022 р.

3. Об'єкт дослідження: технологія розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках.

4. Предмет дослідження: методи сегментації та класифікації об'єктів на супутникових та аерофотознімках.

5. Перелік завдань, які потрібно розробити

- проведено вичерпний порівняльний аналіз архітектур сегментаційних мереж та кодувальників;
- досліджено набори даних для сегментації будинків на аерофотознімках;
- розроблене відповідне програмне забезпечення;

- проведено експерименти та навчено низку нейронних мереж;
- досліджено вплив функцій втрат на результат роботи моделі;
- проведено кількісний та візуальний аналіз отриманих результатів на тестовій вибірці обраного набору даних;
- розроблено стартап-проект «Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках»

6. Орієнтовний перелік ілюстративного матеріалу: сегментація зображень, семантична сегментація, сегментація об'єктів, паноптична сегментація, метрики оцінки якості роботи алгоритму, архітектури згорткових нейронних мереж (Unet, Unet++, Deeplab), кодувальники (ResNet, EfficientNet) та проведені експерименти.

#### 7. Орієнтовний перелік публікацій

- Снагощенко Д.М., РОЗПІЗНАВАННЯ (СЕГМЕНТАЦІЯ) ОБ'ЄКТІВ НА СУПУТНИКОВИХ ТА АЕРОФОТОЗНІМКАХ. Вимірювальна та обчислювальна техніка в технологічних процесах. – Статтю прийнято редакцією до друку в журналі №4 за 2022р. головним редактором журналу Валерієм Мартинюком.

- Snahoshchenko D.M., Kolobrodov V.G., OBJECT RECOGNITION AND SEGMENTATION ON SATELLITE AND AERIAL IMAGING. Global Society in Formation of New Security System and World Order: Proceedings of the 1st International Scientific and Practical Internet Conference, December 8-9, 2022. FOP Marenichenko V.V., Dnipro, Ukraine, 8-11 p.

- Snagoshchenko D.M., Vasylykivska I.O., OPTOELECTRONIC PROCESSOR FOR THE SMART HOME. Discovering New Horizons in Science and Prospects for Implementation of Innovations: Proceedings of the 1st International Scientific and Practical Internet Conference, July 7-8, 2022. FOP Marenichenko V.V., Dnipro, Ukraine, 105-107 p.

#### 8. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
Стартап	проф., д.е.н., Бояринова К.О		

9. Дата видачі завдання 08 вересня 2022

## Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Грунтовне ознайомлення з предметною областю	08.09.2022	
2	Визначення структури магістерської дисертації; вивчення літератури, пошук додаткової літератури	09.09.2022	
3	Робота над першим розділом магістерської дисертації	20.09.2022	
4	Проведення наукового дослідження; робота над другим розділом магістерської дисертації	01.10.2022	
5	Проведення наукового дослідження; робота над статтею за результатами наукового дослідження	26.10.2022	
6	Робота над третім розділом магістерської дисертації; підготовка статті та тез за результатами наукового дослідження; проведення експериментів	08.11.2022	
7	Робота над четвертим розділом магістерської дисертації – стартап-проект	19.11.2022	
8	Оформлення магістерської дисертації	04.12.2022	

Студентка

\_\_\_\_\_  
(підпис)Д.М. Снагощенко  
(ініціали, прізвище)

Науковий керівник дисертації

\_\_\_\_\_  
(підпис)В.Г. Колобродов  
(ініціали, прізвище)

## РЕФЕРАТ

### **Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках**

Обсяг роботи – 129

Кількість ілюстрацій – 68

Кількість таблиць – 29

Кількість додатків – 0

Кількість джерел за переліком посилань – 65

#### **Актуальність теми.**

Можливість виявляти нові будівлі безпосередньо з супутника зображення особливо корисні в регіонах, де кількість населення змінюється дуже швидко а також у віддалених і масштабних районах, де перепис цих нових будівель часто виконується вручну і швидко стає неактуальним. Обробка аерофотознімків також може знайти важливе застосування в оцінці будівельних пошкоджень при стихійних лихах, що дозволяє сформулювати адекватну відповідь у цільових областях. Нарешті, це може бути дуже корисним для виробників сонячних панелей, які хочуть оцінити корисну поверхню даху на певній ділянці.

Величезні об'єми зображень щодня знімаються бортовими або космічними платформами, і цей обсяг все ще зростає. Така кількість даних робить ручне опрацювання зображень дуже ресурсозатратним, отже зростає необхідність застосування нейронних мереж. Основною задачею автоматичного розпізнавання зображень є призначення семантичного класу або мітки кожному пікселю, тобто перетворення вхідних даних до семантично значущої растрової карти (яка в подальшому може підлягати додатковій обробці за допомогою, наприклад, векторизації або полігонізації). У більшості випадків анотовані дані для навчання класифікатора генеруються вручну окремо для кожного нового проекту, що є трудомістким і дорогим процесом. Мануальну анотацію потрібно повторювати кожного разу, коли змінюється завдання, географічне розташування, характеристики камери (сенсора) чи навіть погодні умови на зображенні, тому процес погано масштабується.

**Мета і задачі дослідження.** Метою даної роботи є проведення порівняльного аналізу архітектур нейронних мереж, кодувальників та метрик та наборів даних для задачі сегментації, дослідження та модифікація функції втрат глибокої згорткової нейронної мережі покращення тренування моделі а також створення, тренування та оцінка якості роботи глибокої згорткової нейронної мережі для сегментації об'єктів на супутникових або аерофотознімках.

**Для досягнення вказаної мети було розв'язано такі задачі:**

- проведено вичерпний порівняльний аналіз архітектур сегментаційних мереж та кодувальників;
- досліджено набори даних для сегментації будинків на аерофотознімках;
- розроблене відповідне програмне забезпечення;
- проведено експерименти та навчено низку нейронних мереж;
- досліджено вплив функцій втрат на результат роботи моделі;
- проведено кількісний та візуальний аналіз отриманих результатів на тестовій вибірці обраного набору даних;
- розроблено стартап-проект «Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках»

*Об'єктом дослідження є технологія розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках.*

*Предметом дослідження є методи сегментації та класифікації об'єктів на супутникових та аерофотознімках.*

**Методи дослідження.**

Для вирішення поставленої задачі було використано наступні методи:

- методи агрегації, збору та анотації зображень;
- методи чисельної оптимізації (для оптимізації параметрів глибокої нейронної мережі);
- методи розробки програмного забезпечення (для програмної реалізації продукту);
- методи комп'ютерної графіки (для роботи з зображеннями та генерації аугментацій);

- методи глибокого навчання (для побудови та навчання глибокої згорткової нейронної мережі).

**Наукова новизна одержаних результатів** складається з таких положень:

- проведено вичерпний порівняльний аналіз архітектур нейронних мереж, кодувальників, функцій втрат та метрик та наборів даних для задачі сегментації;
- модифіковано функції втрат глибокої згорткової нейронної мережі для покращення сегментації об'єктів на супутникових та/або аерофотознімках.

**Практичне значення одержаних результатів.**

- для розпізнавання будинків в регіонах, де кількість населення змінюється дуже швидко а також у віддалених і масштабних районах, де перепис цих нових будівель часто виконується вручну і швидко стає неактуальним;
- для виробників сонячних панелей для оцінки корисної поверхні даху та кількості сонячного світла на певній ділянці.;
- для оцінки пошкоджень при стихійних лихах та військових конфліктах;
- наразі існує потреба оцінки пошкоджень в деокупованих регіонах (де є можливість зробити знімки з дрона) та тимчасово окупованих регіонах (де наявні тільки супутникові знімки).

**Публікації.**

- Снагощенко Д.М., РОЗПІЗНАВАННЯ (СЕГМЕНТАЦІЯ) ОБ'ЄКТІВ НА СУПУТНИКОВИХ ТА АЕРОФОТОЗНІМКАХ. Вимірювальна та обчислювальна техніка в технологічних процесах. – Статтю прийнято редакцією до друку в №4 за 2022р. головним редактором журналу Валерієм Мартинюком.

- Snahoshchenko D.M., Kolobrodov V.G., OBJECT RECOGNITION AND SEGMENTATION ON SATELLITE AND AERIAL IMAGING. Global Society in Formation of New Security System and World Order: Proceedings of the 1st International Scientific and Practical Internet Conference, December 8-9, 2022. FOP Marenichenko V.V., Dnipro, Ukraine, 8-11 p.

- Snagoshchenko D.M., Vasytkovska I.O., OPTOELECTRONIC PROCESSOR FOR THE SMART HOME. Discovering New Horizons in Science and Prospects for

Implementation of Innovations: Proceedings of the 1st International Scientific and Practical Internet Conference, July 7-8, 2022. FOP Marenichenko V.V., Dnipro, Ukraine, 105-107 p.

**Ключові слова:** супутникові знімки, аерофотознімки, глибока згортова нейронна мережа, глибоке навчання, методи комп'ютерного зору, функції втрат, сегментація об'єктів.



## ABSTRACT

### **Object recognition and segmentation on satellite and aerial imaging**

Number of pages – 129

Number of figures – 68

Number of tables – 29

Number of applications – 0

Number of references – 65

#### **Topic relevance.**

The ability to detect new buildings directly from satellite imagery is particularly useful in regions where the population is changing very quickly, as well as in remote and large-scale areas where the census of these new buildings is often done manually and quickly becomes obsolete. Aerial image processing can also find important applications in the assessment of building damage during natural disasters, allowing for the formulation of an adequate response in target areas. Finally, it can be very useful for solar panel manufacturers who want to estimate the usable roof surface area of a particular site.

Huge volumes of images are taken every day by airborne or space-based platforms, and this volume is still growing. This amount of data makes manual processing of images very resource-intensive, so the need for the use of neural networks is increasing. The main task of automatic image recognition is to assign a semantic class or label to each pixel, i.e. to transform the input data into a semantically meaningful bitmap (which can be further processed using, for example, vectorization or polygonization). In most cases, the annotated data for classifier training is manually generated separately for each new project, which is a time-consuming and expensive process. Manual annotation needs to be repeated every time the task, geographic location, camera (sensor) characteristics, or even weather conditions on the image change, so the process does not scale well.

#### **Research goal and objectives.**

The purpose of this work is to conduct a comparative analysis of neural network architectures, encoders and metrics and data sets for the task of segmentation, research and modification of the loss function of a deep convolutional neural network, improvement of

model training, as well as creation, training and evaluation of the quality of a deep convolutional neural network for segmentation of objects on satellite or aerial photographs.

**To accomplish this goal, the following objectives were reached:**

- a comprehensive comparative analysis of architectures of segmentation networks and encoders was carried out;
- datasets for segmentation of buildings in aerial photographs were investigated;
- developed appropriate software;
- conducted experiments and trained a number of neural networks;
- the impact of loss functions on the model performance was investigated;
- a quantitative and visual analysis of the obtained results was carried out on a test sample of the selected data set;
- a startup project was developed «Object recognition and segmentation on satellite and aerial imaging»

*Object of research* there is a technology for object recognition (segmentation) on satellite and aerial photographs.

*Subject of research* there are methods of segmentation and classification of objects on satellite and aerial photographs.

**Methods of research.**

The following methods were used to solve the problem:

- methods of aggregation, collection and annotation of images;
- numerical optimization methods (to optimize the parameters of a deep neural network);
- software development methods (for software implementation of the product);
- computer graphics methods (for working with images and generating augmentations)
- deep learning methods (for building and training a deep convolutional neural network).

**Scientific contribution** consists of the following:

- a comprehensive comparative analysis of neural network architectures,

encoders, loss functions and metrics and data sets for the segmentation task was carried out;

- modified the loss functions of a deep convolutional neural network to improve the segmentation of objects on satellite and/or aerial images.

#### **Practical value of obtained results.**

- to recognize buildings in regions where the population is changing very quickly, as well as in remote and large-scale areas, where the census of these new buildings is often done manually and quickly becomes irrelevant;

- for solar panel manufacturers to estimate the useful surface of the roof and the amount of sunlight in a certain area;

- to assess damages in natural disasters and military conflicts;

- currently there is a need for damage assessment in de-occupied regions (where it is possible to take pictures from a drone) and temporarily occupied regions (where only satellite images are available).

#### **Publications.**

- Снагощенко Д.М., РОЗПІЗНАВАННЯ (СЕГМЕНТАЦІЯ) ОБ'ЄКТІВ НА СУПУТНИКОВИХ ТА АЕРОФОТОЗНІМКАХ. Вимірювальна та обчислювальна техніка в технологічних процесах. – Статтю прийнято редакцією до друку в №4 за 2022р. головним редактором журналу Валерієм Мартинюком.

- Snahoshchenko D.M., Kolobrodov V.G., OBJECT RECOGNITION AND SEGMENTATION ON SATELLITE AND AERIAL IMAGING. Global Society in Formation of New Security System and World Order: Proceedings of the 1st International Scientific and Practical Internet Conference, December 8-9, 2022. FOP Marenichenko V.V., Dnipro, Ukraine, 8-11 p.

- Snagoshchenko D.M., Vasytkovska I.O., OPTOELECTRONIC PROCESSOR FOR THE SMART HOME. Discovering New Horizons in Science and Prospects for Implementation of Innovations: Proceedings of the 1st International Scientific and Practical Internet Conference, July 7-8, 2022. FOP Marenichenko V.V., Dnipro, Ukraine, 105-107 p.

**Keywords:** satellite images, aerial images, deep convolutional neural network, deep learning, computer vision methods, loss functions, object segmentation.

## ЗМІСТ

ВСТУП.....	14
РОЗДІЛ 1. АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ СЕГМЕНТАЦІЇ ЗОБРАЖЕНЬ.....	15
1.1. ЗАДАЧІ СЕГМЕНТАЦІЇ .....	15
1.1.1. Сегментація зображень .....	15
1.1.2. Семантична сегментація .....	15
1.1.3. Сегментація об'єктів .....	16
1.1.4. Паноптична сегментація .....	16
1.1.5. Прикладне застосування.....	17
1.2. МЕТРИКИ ОЦІНКИ ЯКОСТІ РОБОТИ АЛГОРИТМУ .....	18
1.2.1. Піксельна точність .....	18
1.2.2. Індекс Жаккара .....	19
1.2.3. Коефіцієнт подібності Дайса.....	21
1.3. АРХІТЕКТУРИ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ.....	22
1.3.1. Unet .....	22
1.3.2. Deeplab.....	24
1.3.3. Unet++ .....	26
1.4. ІСНУЮЧІ ПРОГРАМНІ РІШЕННЯ.....	30
1.4.1. MMSegmenation .....	30
1.4.2. Detectron2 .....	32
1.4.3. Segmentation models pytorch .....	33
ВИСНОВКИ ДО РОЗДІЛУ 1.....	35
РОЗДІЛ 2. АНАЛІЗ АРХІТЕКТУРИ НЕЙРОННОЇ МЕРЕЖІ.....	36
2.1. МОДИФІКАЦІЇ АРХІТЕКТУРИ НЕЙРОННОЇ МЕРЕЖІ U-NET .....	36
2.1.1. Деконволюція.....	36
2.1.2. Симетричність кодувальника і декодувальника .....	38
2.1.3. Глибоке контрольоване навчання.....	40
2.2. КОДУВАЛЬНИКИ.....	44
2.2.1. ResNet .....	44
2.2.2. EfficientNet .....	51

2.3. Функції втрат.....	56
2.3.1. Ентропія.....	56
2.3.2. Перехресна ентропія .....	59
2.3.3. Бінарна перехресна ентропія.....	60
2.3.4. Бінарна перехресна ентропія для задачі семантичної сегментації.....	61
2.3.5. Фокальна функція втрат .....	61
ВИСНОВКИ ДО РОЗДІЛУ 2.....	66
РОЗДІЛ 3 ВИПРОБУВАННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ.....	67
3.1. НАБОРИ ДАНИХ ДЛЯ НАВЧАННЯ ТА ВАЛІДАЦІЇ МОДЕЛІ .....	67
3.1.1. Набір даних Inria.....	67
3.2. АУГМЕНТАЦІЇ ДАНИХ.....	70
3.3. ЕКСПЕРИМЕНТИ .....	78
3.3.1. Кодувальники.....	78
3.3.2. Архітектури.....	87
3.3.3. Функції втрат .....	92
3.4. Обчислювальні потужності .....	97
ВИСНОВКИ ДО РОЗДІЛУ 3.....	98
РОЗДІЛ 4. РОЗРОБКА СТАРТАП ПРОЄКТУ «РОЗПІЗНАВАННЯ (СЕГМЕНТАЦІЯ) ОБ'ЄКТІВ НА СУПУТНИКОВИХ ТА АЕРОФОТОЗНІМКАХ» .....	102
4.1. ОПИС ІДЕЇ ПРОЄКТУ ТА ТЕХНОЛОГІЧНИЙ АУДИТ ІДЕЇ ПРОЄКТУ.....	102
4.2. АНАЛІЗ РИНКОВИХ МОЖЛИВОСТЕЙ ЗАПУСКУ СТАРТАП ПРОЄКТУ.....	105
4.3. РОЗРОБЛЕННЯ РИНКОВОЇ СТРАТЕГІЇ ПРОЄКТУ.....	113
4.4. РОЗРОБЛЕННЯ МАРКЕТИНГОВОЇ ПРОГРАМИ СТАРТАП-ПРОЄКТУ.....	116
ВИСНОВКИ ДО РОЗДІЛУ 4.....	120
ВИСНОВКИ.....	122
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ .....	124

## ВСТУП

Можливість виявляти нові будівлі безпосередньо з супутника зображення особливо корисні в регіонах, де кількість населення змінюється дуже швидко а також у віддалених і масштабних районах, де перепис цих нових будівель часто виконується вручну і швидко стає неактуальним. Обробка аерофотознімків також може знайти важливе застосування в оцінці будівельних пошкоджень при стихійних лихах, що дозволяє сформулювати адекватну відповідь у цільових областях. Нарешті, це може бути дуже корисним для виробників сонячних панелей, які хочуть оцінити корисну поверхню даху на певній ділянці.

Величезні об'єми зображень щодня знімаються бортовими або космічними платформами, і цей обсяг все ще зростає. Така кількість даних робить ручне опрацювання зображень дуже ресурсозатратним, отже зростає необхідність застосування нейронних мереж. Основною задачею автоматичного розпізнавання зображень є призначення семантичного класу або мітки кожному пікселю, тобто перетворення вхідних даних до семантично значущої растрової карти (яка в подальшому може підлягати додатковій обробці за допомогою, наприклад, векторизації або полігонізації). У більшості випадків анотовані дані для навчання класифікатора генеруються вручну окремо для кожного нового проекту, що є трудомістким і дорогим процесом. Мануальну анотацію потрібно повторювати кожного разу, коли змінюється завдання, географічне розташування, характеристики камери (сенсора) чи навіть погодні умови на зображенні, тому процес погано масштабується.

## РОЗДІЛ 1. АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ СЕГМЕНТАЦІЇ ЗОБРАЖЕНЬ

### 1.1. Задачі сегментації

#### 1.1.1. Сегментація зображень

Сегментація зображення [2] — це техніка комп'ютерного зору та обробки зображень, яка передбачає групування або позначення подібних областей або сегментів у зображенні на піксельному рівні. Мітка класу або маска представляє кожен сегмент пікселів.

У сегментації зображень зображення складається з двох основних підзадач: об'єкти і сутності. Об'єкти відповідають злічуваним предметам на зображенні (наприклад люди, квіти, птахи, тварини тощо). Для порівняння, сутності представляють собою аморфні області (або повторювані патерни) подібного матеріалу, який незліченний (наприклад, дорога, небо та трава).

#### 1.1.2. Семантична сегментація

Семантична сегментація [2] вивчає незліченні сутності на зображенні. Вона аналізує кожен піксель зображення та призначає унікальну мітку класу на основі текстури, яку він представляє.

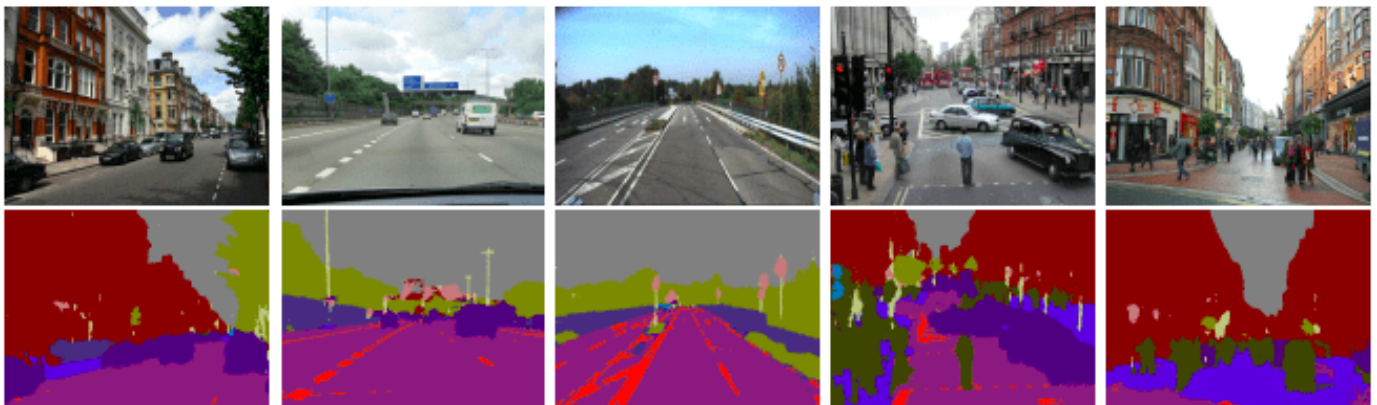


Рисунок 1.1 – Приклади семантичної сегментації

Семантична сегментація призначає унікальні мітки класу для кожної з текстур або категорій. Однак результат семантичної сегментації не може відрізнити або порахувати, наприклад, два автомобілі чи трьох пішоходів окремо. Зазвичай використовувані методи семантичної сегментації включають SegNet [3], U-Net [4], DeconvNet [5] і FCN [6].

### 1.1.3. Сегментація об'єктів

Сегментація об'єктів зазвичай стосується задач комп'ютерного зору, пов'язаних саме зі злічувальними об'єктами. Він може виявити кожен об'єкт або екземпляр класу, присутній на зображенні, і призначити йому власну маску або обмежуючий прямокутник з унікальним ідентифікатором.

Зазвичай використовувані методи сегментації екземплярів: Mask R-CNN [7], Faster R-CNN [8], PANet [9] і YOLACT [10]. Рисунок 2 демонструє приклади роботи нейронної мережі для сегментації об'єктів.

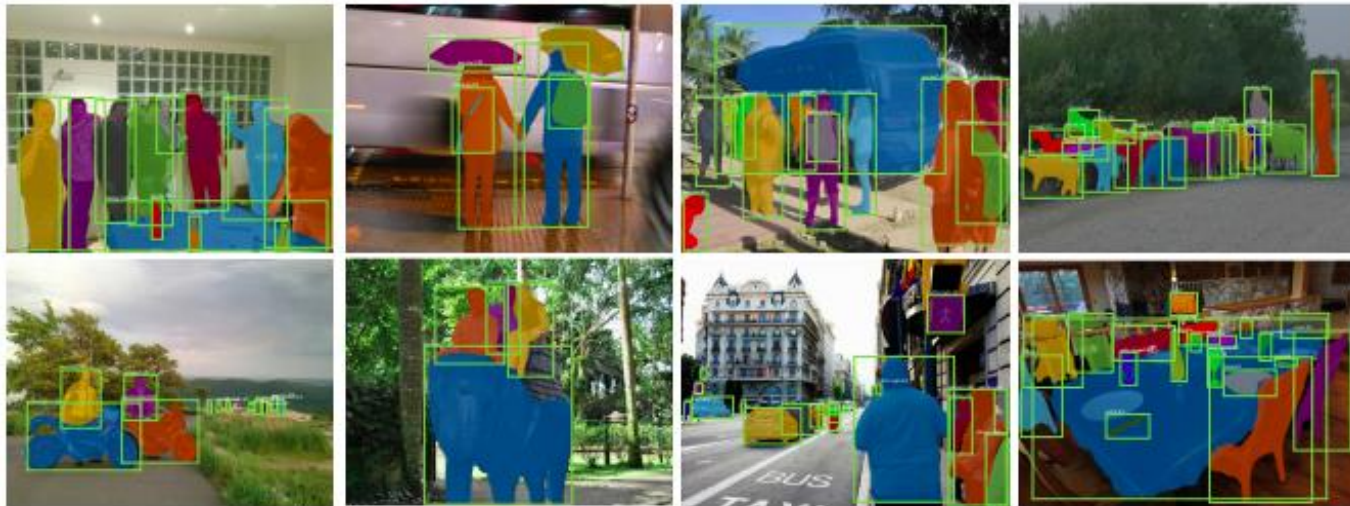


Рисунок 1.2 – Приклади сегментації екземплярів

### 1.1.4. Паноптична сегментація

Паноптична сегментація [2] – поєднання підходів семантичної сегментації та сегментації об'єктів. Він представляє уніфікований підхід до сегментації зображення,



коли кожному пікселю сцени присвоюється семантична мітка (через семантичну сегментацію) та унікальний ідентифікатор екземпляра (через сегментацію об'єктів).

Паноптична сегментація призначає кожному пікселю лише одну пару семантичної мітки та ідентифікатора екземпляра. Однак об'єкти можуть мати пікселі, що перекриваються. У цьому випадку паноптична сегментація усуває невідповідність, віддаючи перевагу екземпляру об'єкта, оскільки пріоритетом є ідентифікація саме об'єктів, а не сутностей. Рисунок 3 демонструє різні виявлення паноптичної сегментації.

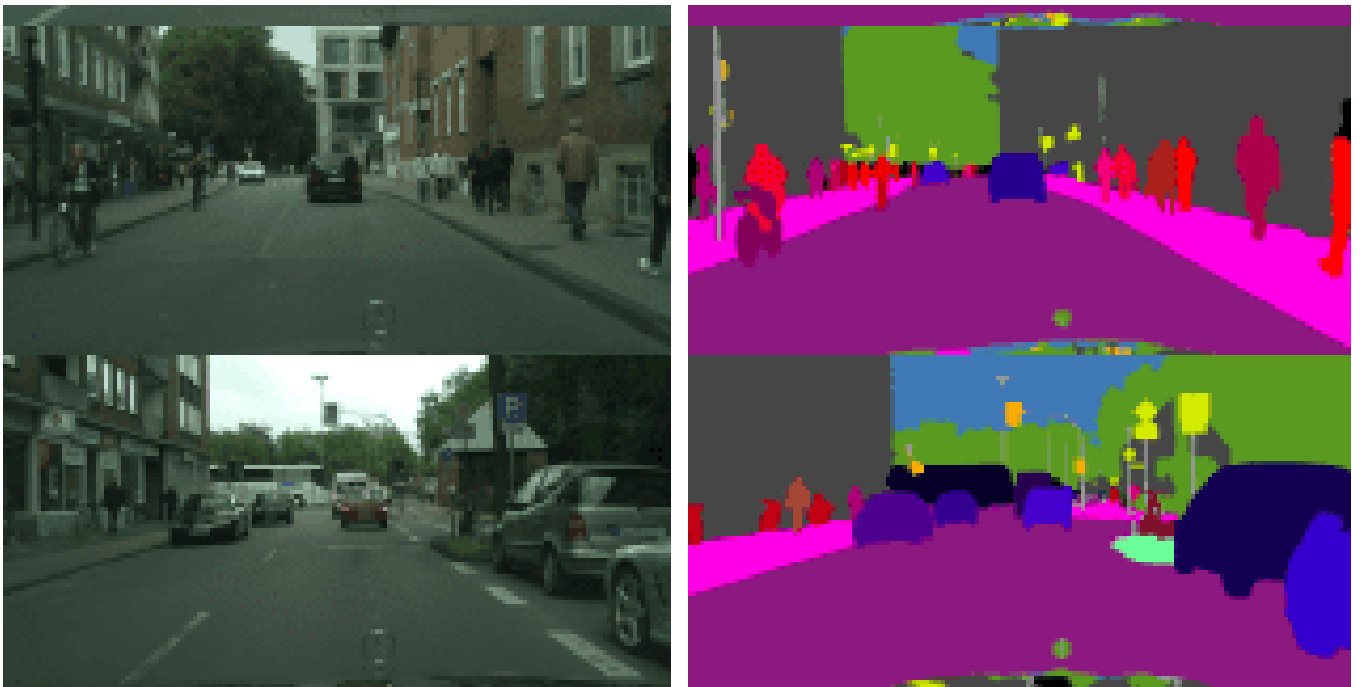


Рисунок 1.3 – Приклади паноптичної сегментації.

Більшість моделей паноптичної сегментації базуються на методі Mask R-CNN [7]. Їїго магістральні архітектури включають UPSNet [11], FPSNet [12], EPSNet [13] і VPSNet [14].

### 1.1.5. Прикладне застосування

Усі три методи сегментації зображень мають застосування в комп'ютерному зорі та обробці зображень. Разом вони дають змогу створювати безліч реальних додатків, допомагаючи людству збільшити його когнітивну пропускну здатність [2].

Деякі реальні програми для семантичної сегментації та сегментації об'єктів включають:

- Автономні транспортні засоби або самокеровані автомобілі: тривимірна семантична сегментація дозволяє транспортним засобам краще розуміти навколишнє середовище, ідентифікуючи різні об'єкти на вулиці. У той же час сегментація об'єктів ідентифікує кожен екземпляр об'єкта окрема, щоб забезпечити більшу глибину для розрахунку швидкості та відстані.
- Аналіз медичних зображень: обидва методи мають змогу ідентифікувати пухлини та інші аномалії під час МРТ, КТ та рентгенівського сканування.
- Супутникові або аерофотознімки: обидва методи сегментації допомагають з картографування світу з космосу або висоти. Вони можуть окреслювати об'єкти світу, наприклад, річки, океани, дороги, сільськогосподарські поля, будівлі тощо.

Паноптична сегментація виводить візуальне сприйняття автономних транспортних засобів на новий рівень. Вона продукує маски з точністю на рівні пікселів, що дозволяє безпілотним автомобілям приймати більш точні рішення. Крім того, паноптична сегментація набуває все більшого застосування у аналізі медичних зображень, анотації даних, доповненні даних, дистанційному зондуванні БПЛА (безпілотних літальних апаратів), відеоспостереженні та підрахунку людей у натовпі. У всіх областях паноптична сегментація забезпечує більшу глибину та точність під час прогнозування масок і обмежуючих прямокутників.

## **1.2. Метрики оцінки якості роботи алгоритму**

### **1.2.1. Піксельна точність**

Піксельна точність [15] (Pixel Accuracy – PA) — це метрика семантичної сегментації, яка позначає відсоток пікселів, які точно класифікуються на зображенні. Ця метрика обчислює відношення кількості правильно класифікованих пікселів до загальної кількості пікселів у зображенні:

$$PA = \frac{\sum_{j=1}^k n_{jj}}{\sum_{j=1}^k t_j} \quad (1.1)$$

Де  $n_{jj}$  – загальна кількість пікселів, класифікованих і позначених як клас  $j$ . Інакше кажучи,  $n_{jj}$  відповідає загальній кількості істинно позитивних (True Positive) результатів для класу  $j$ .  $t_j$  – це загальна кількість пікселів, позначених як клас  $j$ .

Оскільки в семантичній сегментації можуть бути присутніми кілька класів, усереднена піксельна точність пікселів (mean pixel accuracy – mPA) обчислює середню точність класу як:

$$mPA = \frac{1}{k} \sum_{j=1}^k \frac{n_{jj}}{t_j} \quad (1.2)$$

Піксельна точність (PA) та усереднена піксельна точність (mPA) є інтуїтивно зрозумілими та інтерпретованими метриками оцінки якості. Однак високий показник піксельної точності напряду не означає найкраще продуктивність сегментації, особливо у випадку незбалансованих наборів даних. У цьому випадку, коли клас домінує на зображенні, а деякі інші класи складають лише невелику частину зображення, лише правильна класифікація домінуючого класу (зазвичай фону) дасть високий показник піксельної точності.

### 1.2.2. Індекс Жаккара

Індекс Жаккара, більш відомий як Intersection-Over-Union (IoU), є однією з найбільш поширених метрик якості для оцінки продуктивності моделей семантичної сегментації.

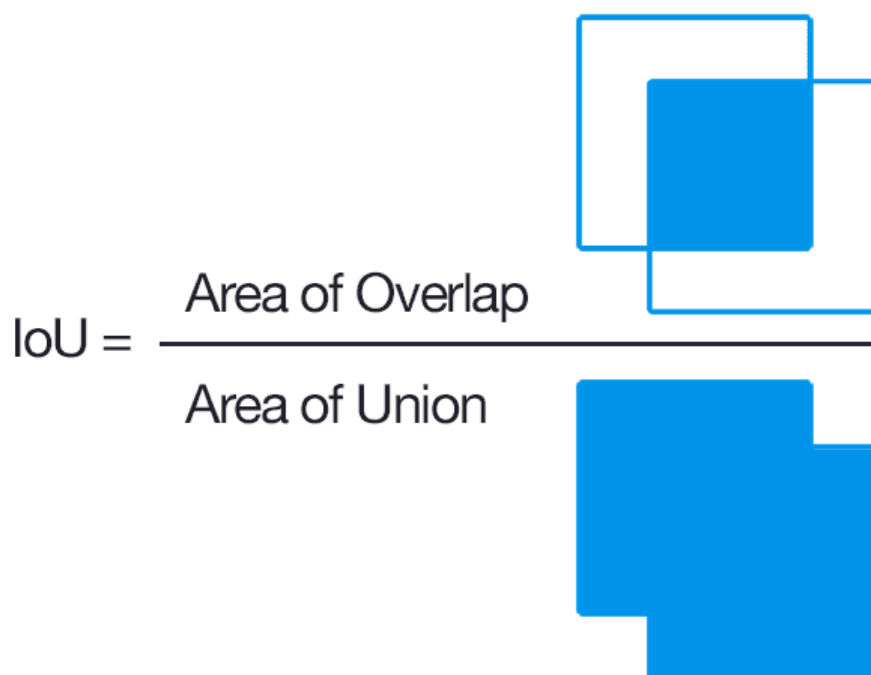


Рисунок 1.4 – IoU

IoU — це площа перекриття (перетину) між прогнозованою сегментацією та істинною сегментацією, поділена на площу об'єднання між прогнозованою сегментацією та істинною сегментацією, як показано на попередньому рисунку. Цей показник коливається від 0 до 1 (0–100%), де 0 означає відсутність перекриття, а 1 означає ідеальне перекриття сегментації – прогнозована сегментаційна маска попіксельно збігається з істинною.

В математичній нотації це матиме наступний вигляд:

$$\text{IoU}(A, B) = \frac{A \cap B}{A \cup B} \quad (1.3)$$

, де  $A$  – прогнозована моделлю сегментаційна маска, а  $B$  – істинна сегментаційна маска.

Для бінарної (два класи) або багатокласової семантичної сегментації зображення середнє значення метрики IoU обчислюється шляхом усереднення IoU кожного класу таким самим чином як для усередненої попіксельної точності.

### 1.2.3. Коефіцієнт подібності Дайса

Для задачі бінарної сегментації, Коефіцієнт подібності Дайса або F1-міра [16] між двома сегментаційними масками – це подвоєна площа перекриття, поділена на загальну кількість пікселів обох масок.

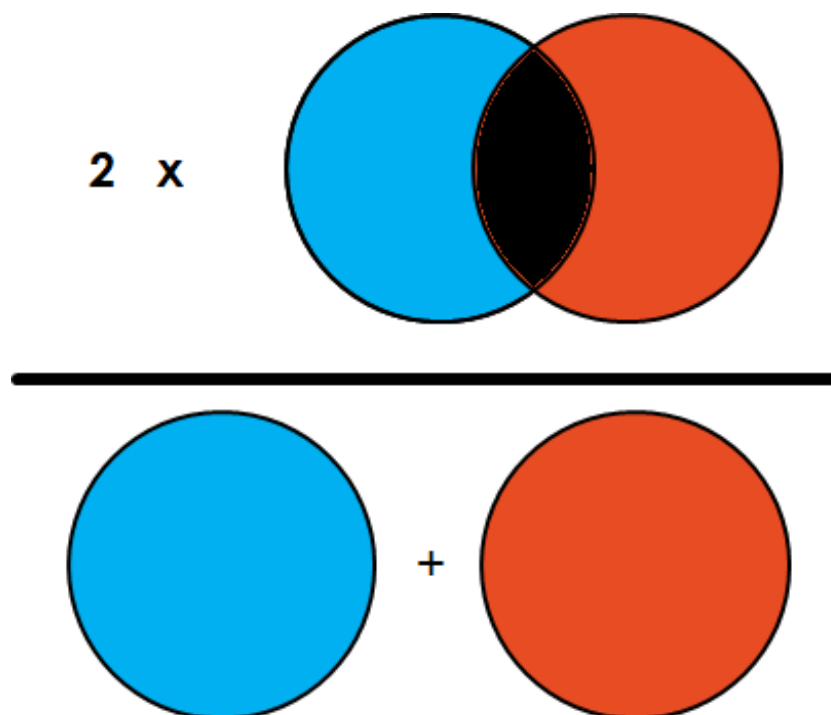


Рисунок 1.5 – Коефіцієнт подібності Дайса

В математичній нотації це матиме наступний вигляд:

$$\frac{2 \cdot |X \cap Y|}{|X| + |Y|} \quad (1.4)$$

, де  $X$  – прогнозована моделлю сегментаційна маска, а  $Y$  – істинна сегментаційна маска.

Коефіцієнт подібності Дайса, аналогічно до індексу Жаккара, коливається від 0 до 1 (0–100%), де 0 означає відсутність перекриття, а 1 означає ідеальне перекриття сегментації – прогнозована сегментаційна маска попіксельно збігається з істинною.

Коефіцієнт подібності Дайса дуже подібний до індексу Жаккара. Вони позитивно корелюють, тобто якщо один скаже, що модель А краща за модель В в сегментуванні зображення, то інший скаже те саме.

Загалом індекс Жаккара має тенденцію сильніше, ніж коефіцієнт подібності Дайса, штрафувати окремі випадки поганої класифікації більше, навіть якщо обидві метрики кажуть, що ці окремі випадки сегментовані погано. Таким чином, коефіцієнт подібності Дайса, як правило, вимірює щось ближче до усередненої оцінки якості роботи моделі, тоді як індекс Жаккара більше фокусується на саме на якості роботи моделі на найскладніших зображеннях (найгірших випадках сегментації).

### **1.3. Архітектури згорткових нейронних мереж**

#### **1.3.1. Unet**

Оригінальна архітектура нейронної мережі U-Net було розроблена для семантичної сегментації біомедичних зображень [17]. Навчена нейронна мережа дозволила вивести якість сегментації до по-піксельної точності.

Нейронна мережа U-Net побудована на основі повністю згорткової мережі (Fully Convolutional Network) і складається з двох основних частин: кодувальника і декодувальника.

Кодувальник (encoder) виділяє ознаки різних рівнів через послідовність двовимірних згортків, активаційних функцій (в даному випадку ReLU – rectified linear unit) та шарів максимізаційного агрегування (max pooling), що дозволяє охоплювати контекст кожного пікселя.

Наступним кроком декодувальник (decoder) підвищує дискретизацію результату, щоб збільшити роздільну здатність виділених кодувальником ознак зображення.

В архітектурі Unet вводиться поняття пропускних з'єднань (skip-connections), що додаються між блоками кодувальника і декодувальника на відповідних рівнях роздільної здатності, що дозволяє прогнозувати локалізацію так само добре, як і

контекст.

Таким чином декодувальник складається з послідовності зворотніх згорток і конкатенацій із відповідною картою ознак кодувальника, після чого йдуть звичайні згортки та активації ReLU. Кількість каналів ознак (feature channels) в кодувальнику подвоюється на кожному рівні зниження дискретизації (зменшення роздільної здатності).

На рис. 1.6 наведено схематичне зображення нейронної мережі U-Net з оригінальної статті [4]:

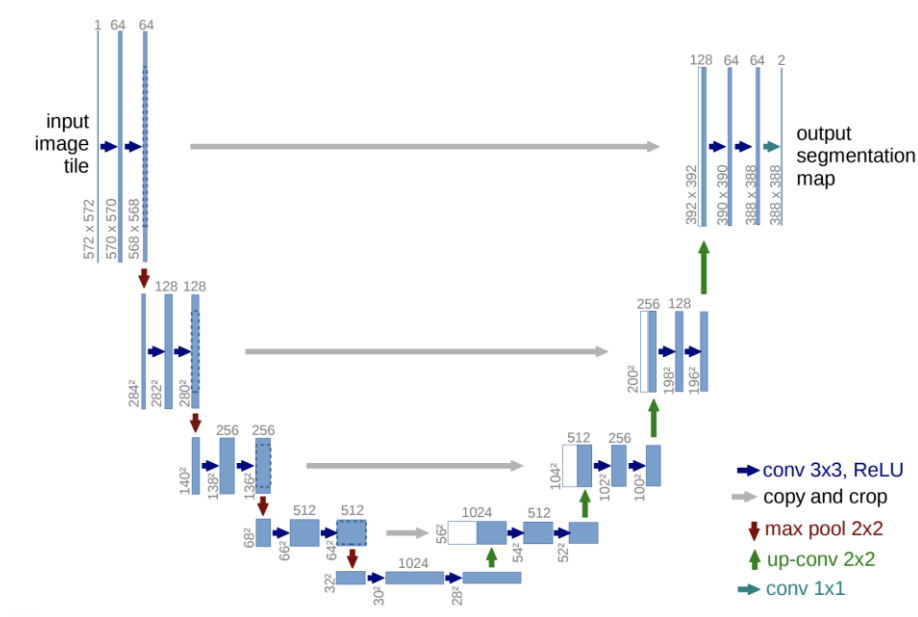


Рисунок 1.6 – Схематичне зображення мережі U-Net

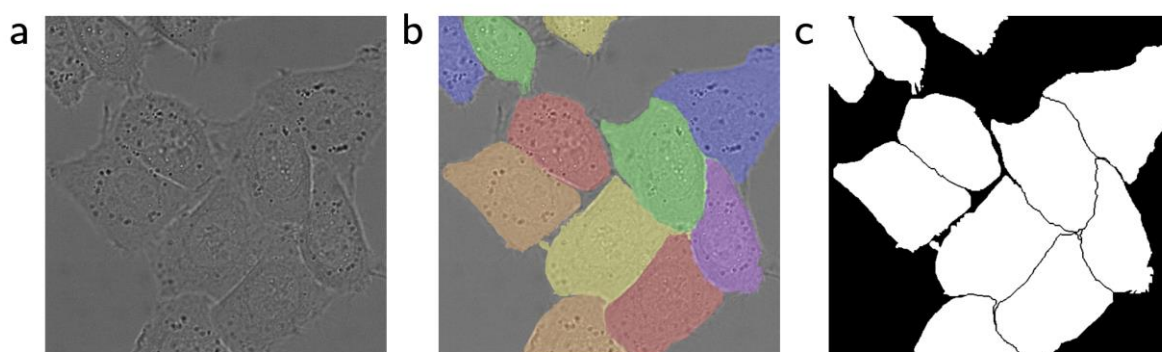


Рисунок 1.7 – Приклад результату роботи мережі U-Net для медичинських зображень. а – вхідне зображення, б – істинна анотація, с – прогнозована анотація.

### 1.3.2. Deeplab

DeepLab [18] – сімейство архітектур нейронних мереж [19-21] для семантичної сегментації зображень, що була розроблена дослідниками з компанії Google та являє собою комбінацію двох раніше відомих методів: глибоких згорткових нейронних мереж і повнозв’язаних умовних випадкових полів (Conditional Random Fields – CRF).

Схематичне зображення нейронної мережі DeepLab з оригінальної статті [19] зображено на рис. 1.8.

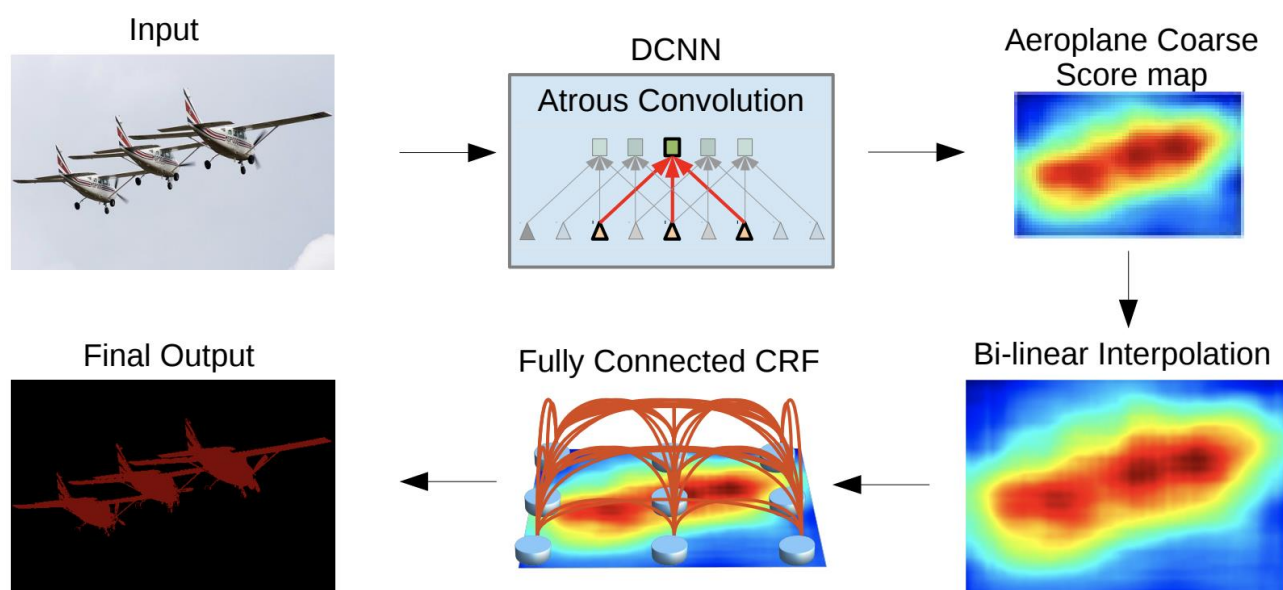


Рисунок 1.8 – Схематичне зображення мережі DeepLab

Перша частина мережі має класичну структуру згорткової нейронної мережі із послідовністю операцій згортки, активаційних функцій (ReLU) та шарів максимізаційного агрегування для отримання інформації про особливості на кожному рівні.

Додатково вводиться новий тип згортки: так звані атросні згортки (atrous convolution) або розширені згортки (dilated convolution). Ідея розширених згортки полягає в тому, щоб розширити рецептивне поле згорткових фільтрів за допомогою коефіцієнту розширення (звідси і назва). Коефіцієнт розширення збільшує рецептивне поле (розширює контекст), який бере до уваги згортковий фільтр і в той же час дозволяє зберегти ту саму кількість параметрів уникаючи збільшення



обчислювальних витрат, просто пропускаючи пікселі зображення, які стоять між цільовими пікселями.

Приклад розширених згорток з різними коефіцієнтами розширення зображено на рис. 1.9.

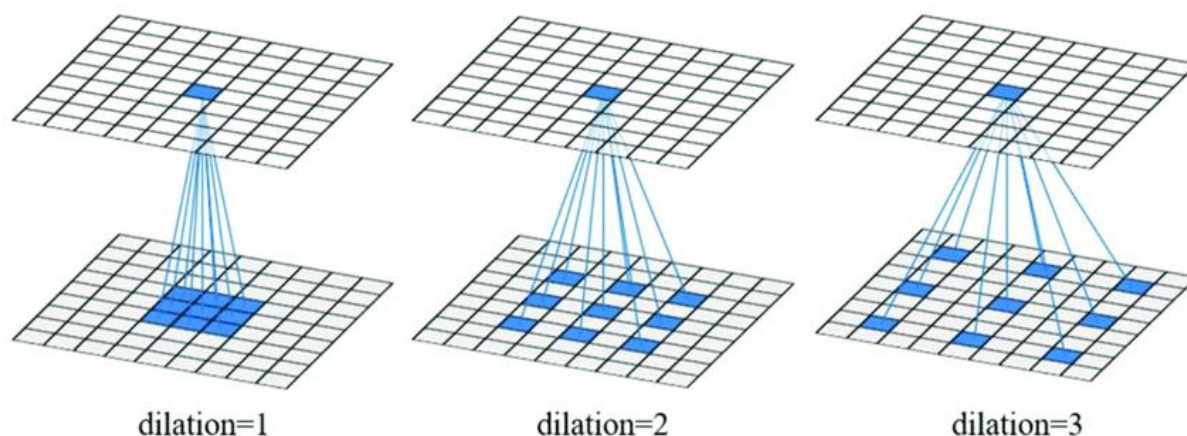


Рисунок 1.9 – Схематичне зображення розширених згорток для різних коефіцієнтів розширення (dilation)

Розширена згортка з коефіцієнтом розширення = 1 є фактично звичайною двовимірною згорткою. При однаковій кількості параметрів згортки ( $3 \times 3 = 9$ ) розмір рецептивного поля зростає з 9 (коефіцієнт розширення = 1) до 25 (коефіцієнт розширення = 2) і 49 (коефіцієнт розширення = 3)

Використання подібних глибоких згорткових нейронних мереж із розширеними згортками дозволяє вирішити проблему зменшення просторової роздільної здатності глибоких кодувальників ознак, дозволяючи в той же час отримати більшу вихідну матрицю та уникнути надмірного підвищення дискретизації.

Умовне випадкове поле (CRF) складається з набору з'єднаних вузлів (наприклад, у цьому випадку пікселів зображення), яким присвоєно мітку/значення на основі міток/значення їхніх прямих сусідів. Умовні випадкові поля часто використовуються для зменшення шуму та виділення меж об'єктів у зашумлених картах сегментації, і їх використання в DeepLab сприяє вирішенню цих самих проблем. Завдяки повторному застосуванню цього методу до вихідних даних

глибокої згорткової нейронної мережі дана карта сегментації поступово стає більш згладженою та досягає чіткіших меж між областями зображення. Необхідна обмежена кількість ітерацій, щоб отримати точну карту сегментації, у випадку DeepLab – це 10 ітерацій.

Так як Умовні Випадкові Поля в даному випадку розглядаються як метод постобробки, що ітеративно покращує вихідні дані згорткової нейронної мережі, загальну архітектуру не можна використовувати для наскрізного (повністю диференційованого) навчання та передбачення в реальному часі.

На рис. 1.10 наведено результати роботи нейронної мережі DeepLab на зображеннях з набору даних PASCAL-Context [22]. Зліва-направо: вхідне зображення, істинна анотація, результати роботи до блоку CRF, результати роботи після блоку CRF.

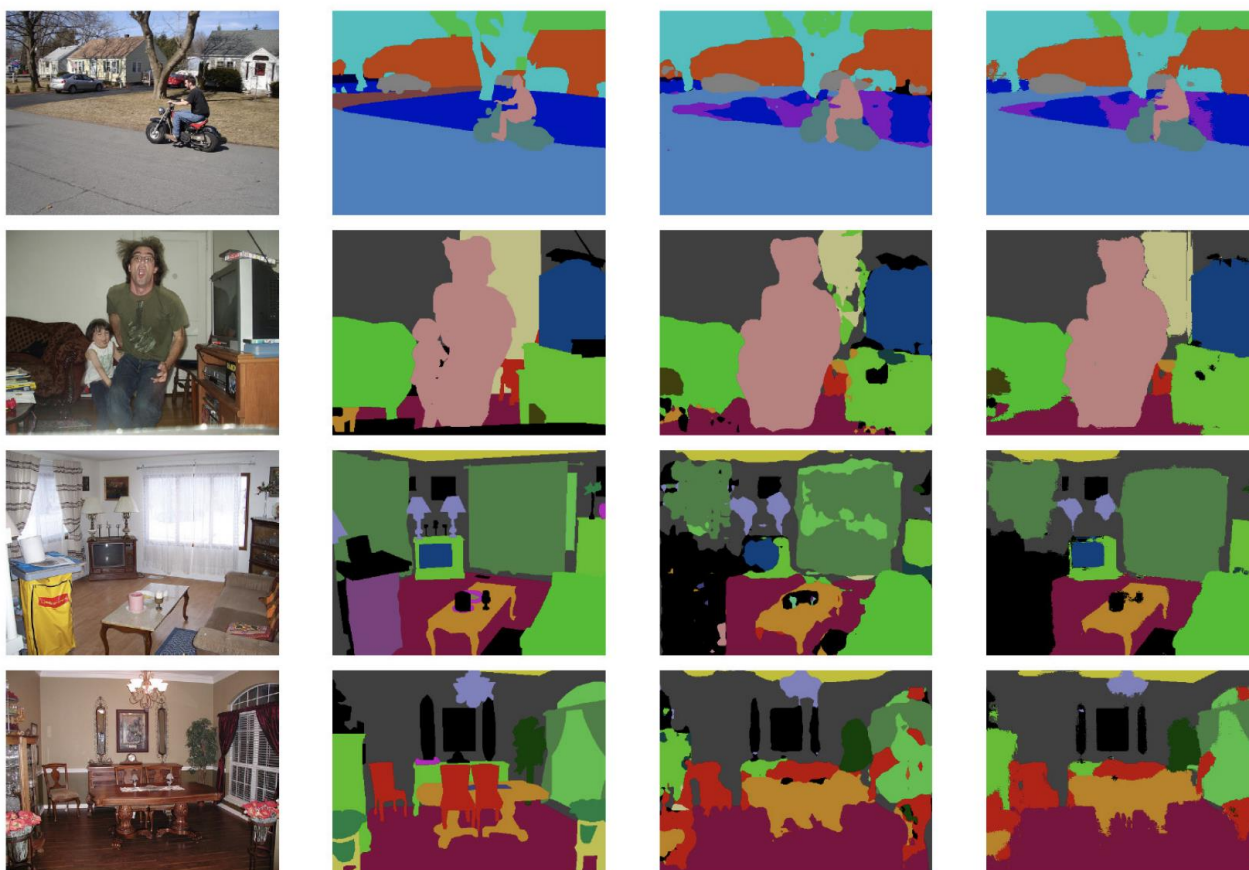


Рисунок 1.10 – Результати роботи мережі DeepLab

### 1.3.3. Unet++

Архітектура нейронної мережі U-Net++ є еволюційним продовженням оригінальної архітектури U-Net.

UNet++ [23] складається з кодувальника та декодувальника, що поєднані за допомогою серії вкладених щільних згорткових блоків (замість базових пропускних з'єднань, використаних в U-Net). Основна ідея UNet++ в тому, аби подолати семантичний розрив між картами ознак кодувальника і декодувальника до етапу злиття.

Наприклад, семантичний розрив між  $(X^{0,0}, X^{1,3})$  нівелюється за допомогою щільного блоку з трьома шарами згортки. На схематичному зображенні, чорний колір позначає оригінальну архітектуру U-Net, зелений і синій показують щільні згорткові блоки пропускних з'єднань, а червоний колір вказує на глибоке контрольоване навчання (Deep Supervision). Червоний, зелений і синій компоненти відрізняють UNet++ від U-Net.

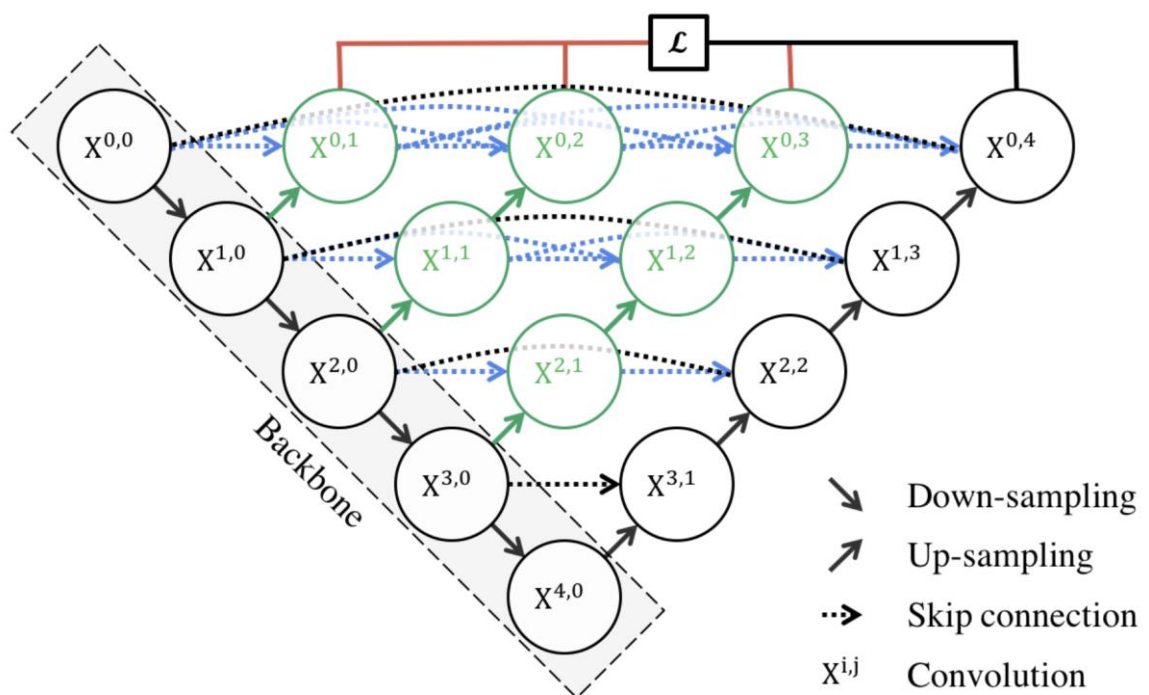


Рисунок 1.11 – Схематичне зображення мережі U-Net++

В оригінальній мережі U-Net карти ознак кодувальника безпосередньо передаються в декодувальник шляхом їх конкатенації з відповідними картами ознак,

однак у UNet++ вони проходять щільний блок згортки, кількість шарів згортки якого залежить номеру блоку кодувальника на номеру блоку декодувальника. По своїй суті, блок щільної згортки наближає семантичний рівень карт ознак кодувальника до семантичного рівня карт ознак, які очікуються у декодувальнику. Гіпотеза полягає в тому, що оптимізаційний процес матиме кращу збіжність, якщо отримані карти ознак кодувальника та відповідні карти ознак декодувальника будуть семантично подібні.

Також в статті запропоноване використання глибокого контрольованого навчання (Deep Supervision [24]), що дозволяє моделі працювати в двох режимах:

- точний режим, в якому виходи з усіх гілок сегментації усереднюються;
- швидкий режим, у якому остаточна карта сегментації вибирається лише з однієї гілки сегментації, вибір якої визначатиме ступінь спрощення моделі та збільшення швидкості (рис. 1.12).

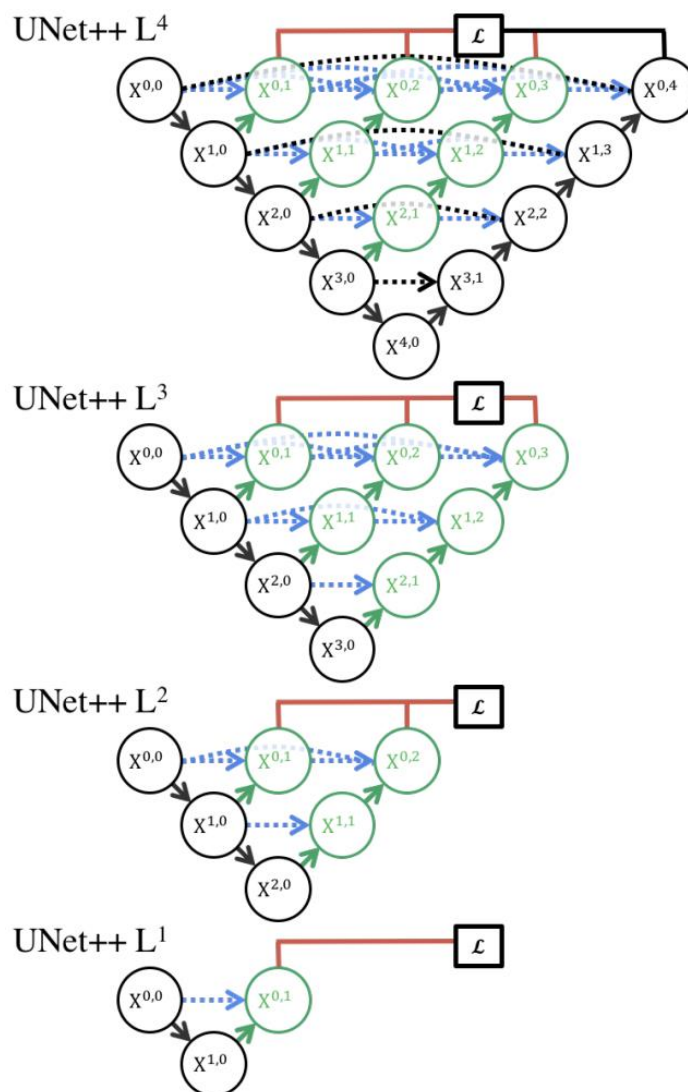


Рисунок 1.12 – Рівні спрощення нейронної мережі UNet++

На рис. 1.13 наведено результати роботи нейронної мережі UNet++ на медичинських зображеннях. Зліва-направо: вхідне зображення, істинна анотація, результати роботи до блоку мережі UNet, результати роботи мережі UNet++.

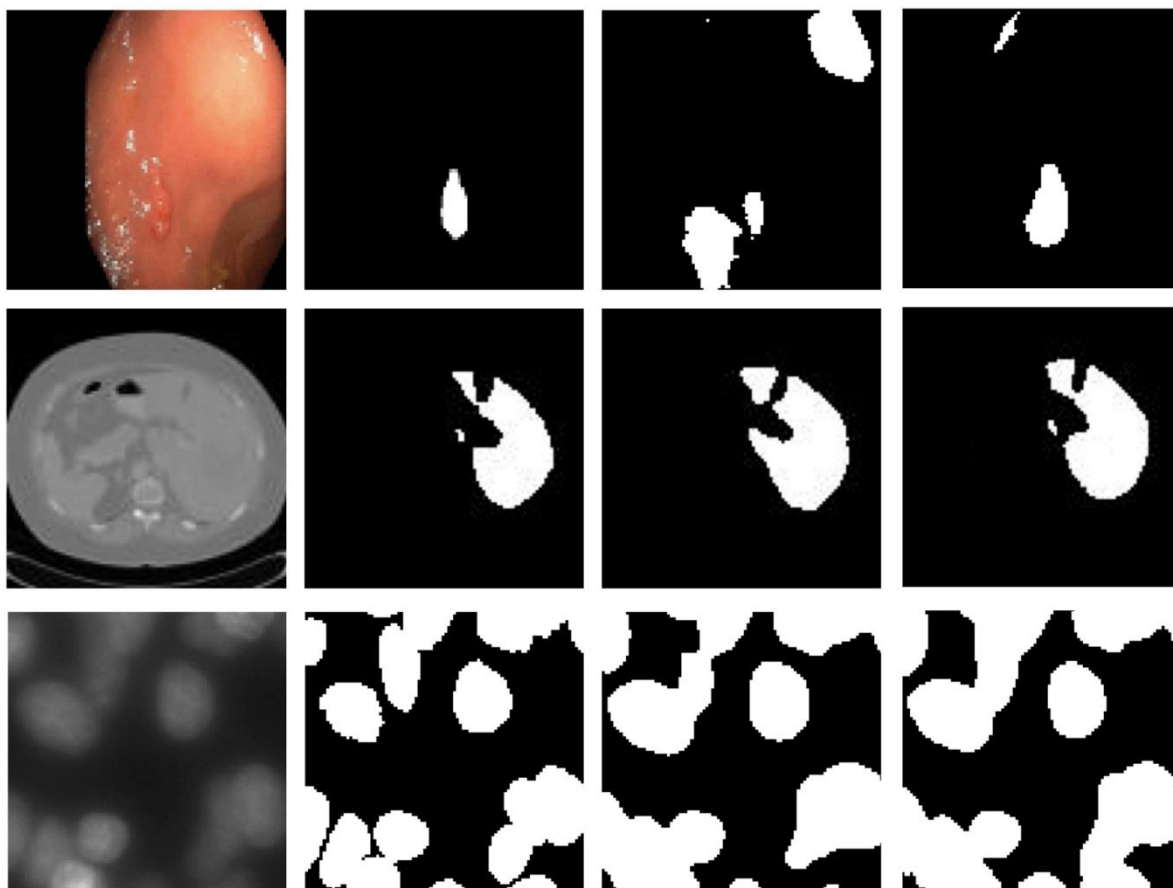


Рисунок 1.13 – Порівняння результатів роботи нейронних мереж UNet та UNet++

## 1.4. Існуючі програмні рішення

### 1.4.1. MMSegmentation

MMSegmentation [25] – одна з найпоширеніших систем алгоритмів комп'ютерного зору з відкритим кодом в епоху глибокого навчання розроблена дослідниками з OpenMMLab [26]. Вона спрямована на:

- надання високоякісних бібліотек, аби зменшити труднощі при повторній реалізації алгоритму;
- створення ефективних ланцюжків інструментів розгортання, націлених на різноманітні серверні модулі та кінцеві пристрої
- побудову міцної основи для досліджень і розвитку комп'ютерного зору

- подолання розриву між академічними дослідженнями та промисловими застосуваннями за допомогою повного пакету інструментів.



Рисунок 1.14 – Бібліотеки та модулі OpenMMLab

На основі PyTorch [27] OpenMMLab розробляє MMEngine [28] для забезпечення універсального механізму навчання та оцінювання якості роботи моделей глибокого навчання для комп'ютерного зору, а також MMCV [29] для забезпечення операторів нейронних мереж і перетворень даних, що служить основою всього проекту. З моменту випуску OpenMMLab випустив понад 30 бібліотек зору, реалізував понад 300 алгоритмів та містить понад 2000 попередньо навчених моделей.

MMSegmentation — це набір інструментів для семантичної сегментації цифрових зображень з відкритим вихідним кодом на основі PyTorch. Це частина проекту OpenMMLab.

Основні особливості:

- Уніфіковане тестування – надання уніфікованого інструментарію тестування для різних методів семантичної сегментації.
- Модульний дизайн – розкладання структури семантичної сегментації на різні компоненти, і надання можливості легко побудувати індивідуальну структуру семантичної сегментації, комбінуючи різні модулі.

- Підтримка багатьох методів і сучасних архітектур глибоких нейронних мереж – набір інструментів безпосередньо підтримує популярні та сучасні моделі семантичної сегментації, наприклад U-Net [4], PSPNet [30], DeepLabV3, PSANet, DeepLabV3+ [31] тощо.
- Висока ефективність – швидкість навчання є вищою або порівнянною з іншими подібними бібліотеками.

### 1.4.2. Detectron2

Detectron2 [32] — це бібліотека нового покоління від команди дослідників з Meta (Facebook) AI Research [33], яка надає найсучасніші алгоритми виявлення об'єктів та сегментації зображень. Ця бібліотека є наступником Detectron [34] і Maskrcnn-Benchmark [35]. Вона підтримує цілу низку дослідницьких проєктів комп'ютерного зору та виробничих програм у Facebook.

Основні особливості:

- PyTorch : оригінальний Detectron було реалізовано в Caffe2 [36]. PyTorch надає більш інтуїтивно зрозумілу модель імперативного програмування, яка дозволяє дослідникам і практикам швидше відтворювати дизайн моделі та проводити експерименти. Оскільки Detectron2 переписаний з нуля в PyTorch, користувачі тепер можуть скористатися підходом PyTorch до глибокого навчання, а також великою та активною спільнотою, яка постійно вдосконалює PyTorch.
- Модульна розширювана конструкція : у Detectron2 представлена модульна конструкція, яка дозволяє користувачам підключати користувацькі реалізації модулів майже до будь-якої частини системи виявлення об'єктів або сегментації. Це означає, що багато нових дослідницьких проєктів можна швидко реалізувати з чітким відокремленням між основною бібліотекою Detectron2 і новою реалізацією дослідження. Дослідники продовжують вдосконалювати модульну розширювану конструкцію, впроваджуючи нові моделі та відкриваючи нові шляхи, за допомогою яких можна зробити Detectron2 більш гнучким.



- Нові моделі та функції : Detectron2 включає всі моделі, які були доступні в оригінальному Detectron, наприклад Faster R-CNN, Mask R-CNN, RetinaNet [37] і DensePose [38]. Він також містить велику кількість нових моделей, зокрема Cascade R-CNN [39], Panoptic FPN [40] і TensorMask [41].
- Нові завдання комп'ютерного зору: Detectron2 підтримує низку завдань, пов'язаних із виявленням об'єктів. Як і оригінальний Detectron, він підтримує виявлення об'єктів за допомогою обмежуючих прямокутників і масок сегментації об'єктів, а також передбачення пози людини. Крім того, Detectron2 додає підтримку семантичної та паноптичної сегментації, завдання, яке поєднує як семантичну сегментацію, так і сегментацію об'єктів.
- Якість реалізації : переписування Detectron2 з нуля дозволило переглянути низькорівневі проектні рішення та вирішити кілька проблем реалізації, що були присутні в оригінальному Detectron.
- Швидкість і масштабованість : перемістивши весь навчальний процес на графічний процесор, вдалося зробити Detectron2 швидшим за оригінальний Detectron для переважної більшості стандартних моделей. Крім того, тепер легко розподілити навчання між кількома серверами з GPU, що значно спрощує масштабування навчання на дуже великих наборах даних.
- Detectron2go [42]: Дослідники комп'ютерного зору з Meta (Facebook) AI Research реалізували додатковий рівень програмного забезпечення Detectron2go, щоб полегшити розгортання нових передових моделей на виробництві. Ці функції включають стандартні навчальні робочі процеси з власними наборами даних, мережева квантизація та перетворення моделі в оптимізовані формати для хмарного та мобільного розгортання.

### 1.4.3. Segmentation models pytorch

Основні особливості цієї бібліотеки [43]:

- API високого рівня (лише два рядки коду для створення нейронної мережі)

- 9 архітектур моделей для бінарної та багатокласової сегментації (включаючи U-Net, U-Net++, MA-Net [44], та LinkNet [45])
- 119 доступних кодувальників (і підтримка понад 400 кодувальників від бібліотеки timm) серед яких ResNet [46], DenseNet [47], Inception [48], EfficientNet [49], MobileNet [50-52], VGG [53] , та інші
  - Усі кодувальники мають попередньо натреновані ваги для швидшої та кращої збіжності оптимізаційної моделі
  - Реалізовані найбільш поширені функції втрат та метрики оцінки якості роботи моделей

Не зважаючи на багато менші масштаби реалізації, порівняно з MMSegmentation та Detectron2, саме Segmentation Models Pytorch є найпоширенішою та найпростішою в застосуванні бібліотекою для семантичної сегментації зображень. Її широко застосовують як і для швидкого прототипування для конкурсів з машинного навчання (таких як Kaggle [54]), так і для великої продуктивних рішень.

## ВИСНОВКИ ДО РОЗДІЛУ 1

Найбільш часто заснованими метриками для оцінки якості роботи моделі семантичної сегментації є індекс Жаккара та Коефіцієнт подібності Дайса. Незважаючи на їх подібність та корельованість, їх зазвичай використовують разом для перевірки як усередненої оцінки якості роботи моделі, так і якості роботи моделі на найскладніших зображеннях.

Що стосується метрики піксельної точності, через відсутність стійкості і наданню переваги домінуючим класам при роботі з незбалансованими наборами даних, дана метрика не знайшла широкого застосування для задачі семантичної сегментації зображень.

Не зважаючи на багато менші масштаби реалізації, порівняно з MMSegmentation та Detectron2, саме Segmentation Models Pytorch є найпоширенішою та найпростішою в застосуванні бібліотекою для семантичної сегментації зображень. Її широко застосовують як і для швидкого прототипування для конкурсів з машинного навчання (таких як Kaggle [54]), так і для великої продуктивних рішень.

## РОЗДІЛ 2. АНАЛІЗ АРХІТЕКТУРИ НЕЙРОННОЇ МЕРЕЖІ

### 2.1. Модифікації архітектури нейронної мережі U-Net

#### 2.1.1. Деконволюція

Коли нейронні мережі передбачають семантичну маску зображення високої роздільної здатності, вони часто створюють їх з високорівневих семантичних ознак низької роздільної здатності [55]. Це дозволяє мережі відтворити грубу маску об'єкта, а вже згодом збільшити її роздільну здатність до розмірів вхідного зображення.

Для переходу від маски з низькою роздільною здатністю до більш високої, зазвичай використовується операція деконволюції (транспонованої згортки). Шари деконволюції дозволяють моделі використовувати кожен піксель на меншій масці для того, щоб відтворити квадрат у на більшій масці.

Проте деконволюція має нерівномірне перекриття, коли розмір ядра згортки (розмір вихідного вікна) не ділиться націло на крок рухомого вікна. Не зважаючи на те, що нейронна мережа могла б ретельніше вивчати ваги, щоб уникнути нерівномірного перекриття, на практиці нейронним мережам досить важко повністю уникнути цього ефекту.

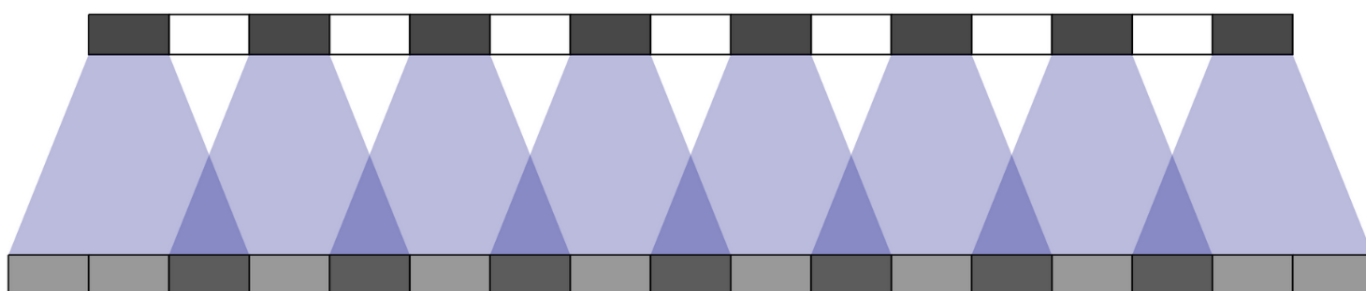


Рисунок 2.1 – Одновимірний випадок. Ядро згортки = 3, крок = 2

Подібний паттерн перекриття також формується у двовимірному випадку (власне, у випадку з зображенням). Нерівномірні перекриття на двох вісях додатково накладаються, створюючи характерний шаховий візерунок різного масштабу.

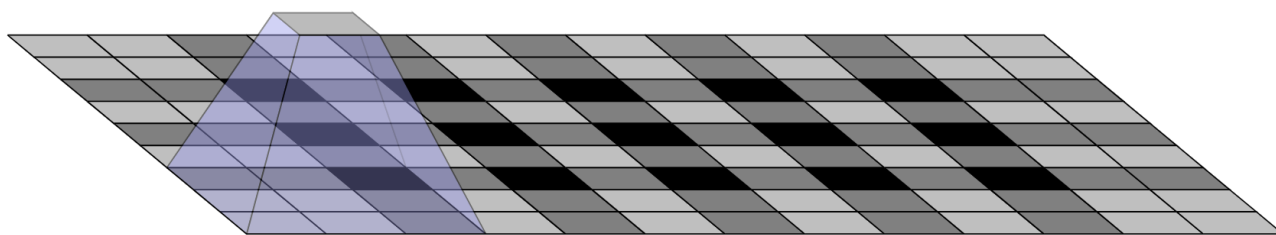


Рисунок 2.2 – Двовимірний випадок. Ядро згортки = 3, крок = 2

Фактично, нерівномірне перекриття має тенденцію бути більш екстремальним для двовимірного випадку. Оскільки два патерни накладаються (множаться), нерівність зводиться в квадрат. Наприклад, в одному вимірі деконволюція з розміром ядра 3 і кроком 2 має деякі виходи з подвійною кількістю вхідних даних, порівняно з іншими. В двовимірному ж випадку, деякі пікселі перекриваються вже до чотирьох разів.

Проте сучасні нейронні мережі для сегментації зображень майже ніколи не містять лише один блок збільшення розмірності. Зазвичай використовують кілька рівнів деконволюції під час створення сегментаційних масок вхідного розміру, ітеративно відтворюючи велику маску використовуючи маски з попередніх блоків декодувальника з нижчою роздільною здатністю. Хоча ці складені деконволюції могли б нівелювати артефакти, на практиці вони ще більше перекриваються, створюючи артефакти на різних масштабах.

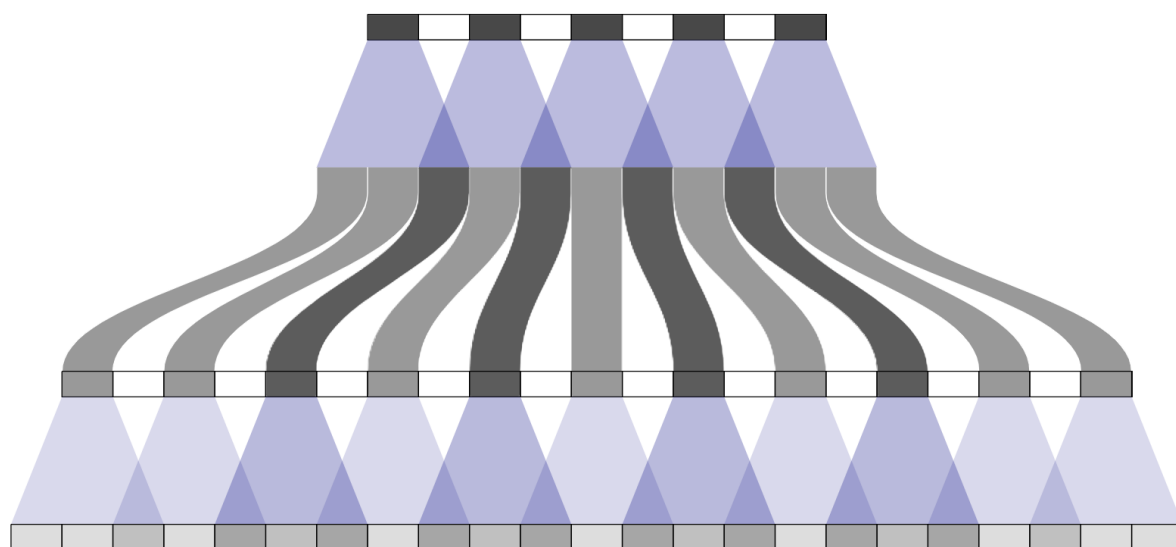


Рисунок 2.3 – Одновимірний випадок. Дві послідовні операції деконволюції. Ядро згортки = 3, крок = 2

Для того, щоб уникнути подібних артефактів, в сучасних архітектурах відокремлюють операцію збільшення розмірності від операції згортки. Операція збільшення розмірності зазвичай замінюється інтерполяцією (інтерполяцією найближчого сусіда або білінійною інтерполяцією), а потім додається додатковий згортковий шар. Це виглядає як природний підхід, що подібний до методів, що гарно себе проявили в задачах генерації зображень та збільшення розмірної здатності.

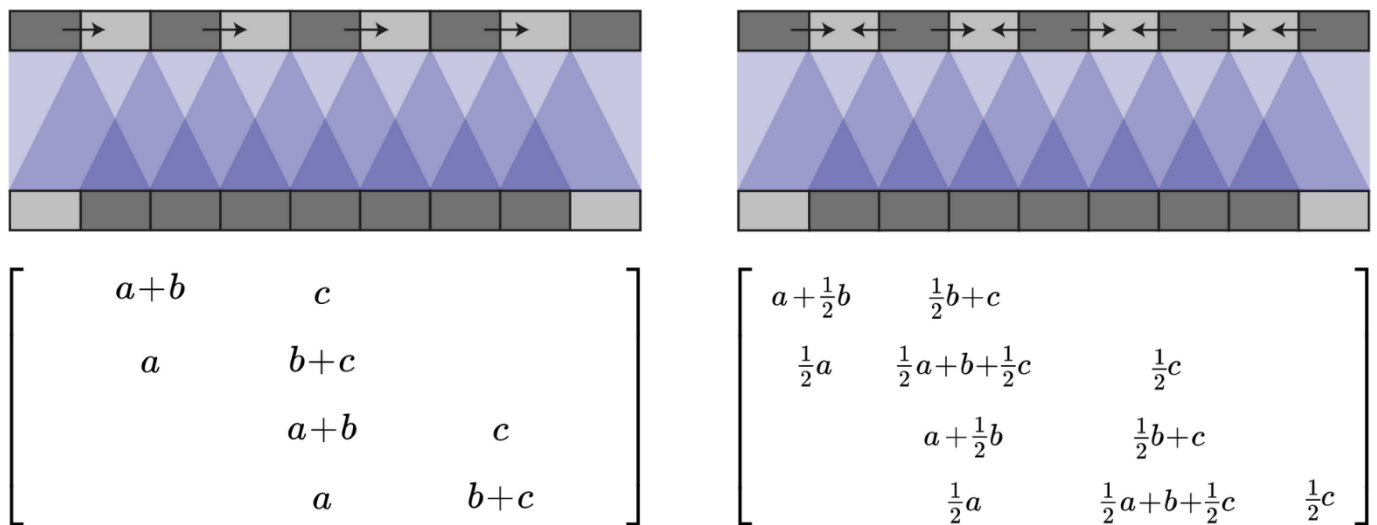


Рисунок 2.4 – Одновимірний випадок. Інтерполяція найближчого сусіда та білінійна інтерполяція

### 2.1.2. Симетричність кодувальника і декодувальника

На наступному схематичному зображенні архітектури U-Net з оригінальної статті можна побачити, що роздільна здатність відповідних блоків кодувальника і декодувальника не співпадають точно. Це відбувається через специфічні параметри ядра згортки та кроку для згорткових шарів цієї архітектури.

Через це, для реалізації пропускних з'єднань (сірі стрілки) і для співставлення розмірів карт ознак, з'являється необхідність вирізання центральної частини відповідного блока кодувальника. Наприклад, з карти ознак другого блоку кодувальника, роздільної здатності 280x280, вирізається центральна частина

розміром 200x200 для подальшої конкатенації з відповідною картою ознак передостаннього блоку декодувальника розміру 200x200.

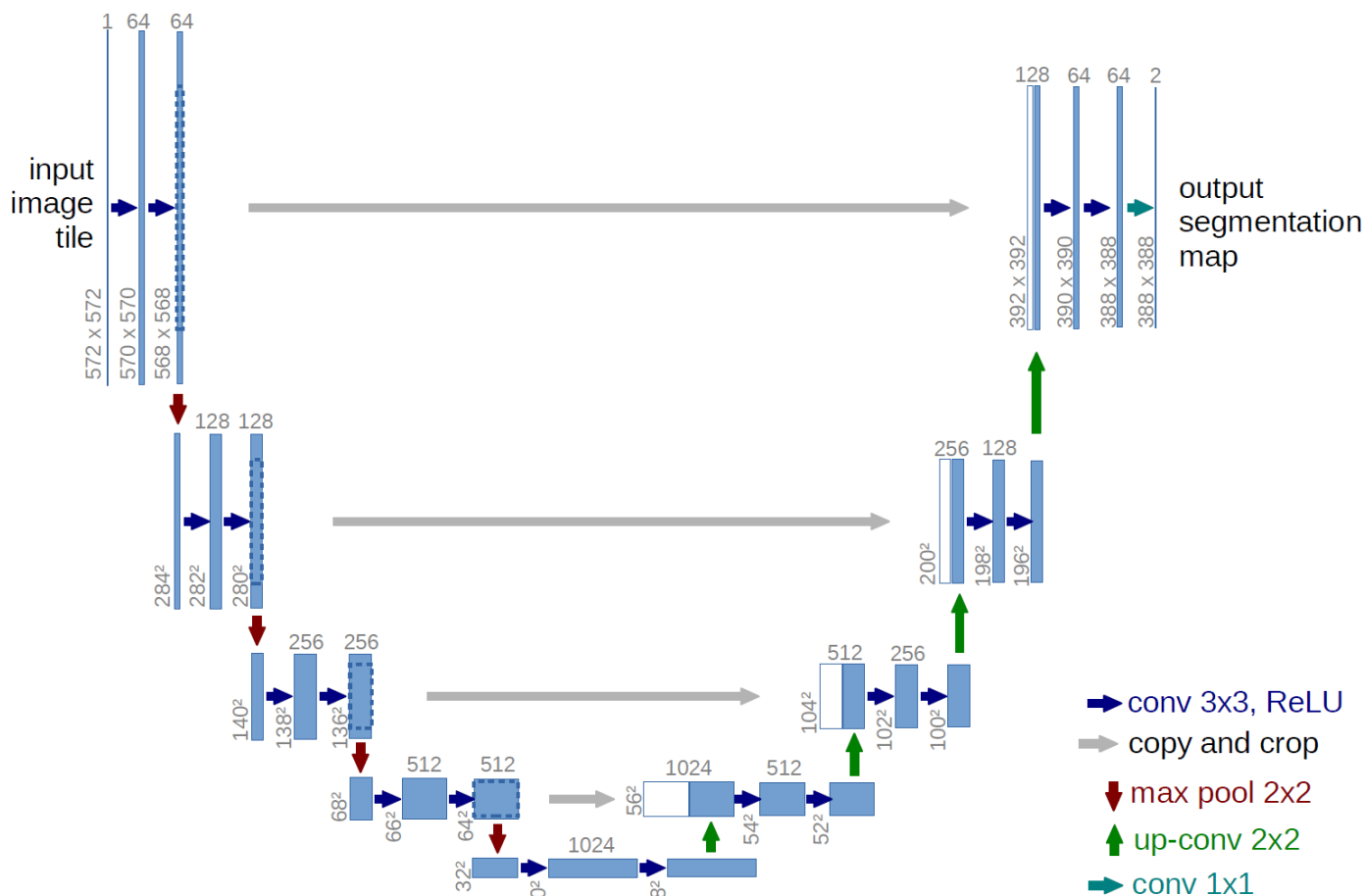


Рисунок 2.5 – Схематичне зображення архітектури U-Net з оригінальної статті

Подібна операція може призвести до погіршення роботи моделі на межах зображення через відкидання відповідних ознак на кожному з рівнів нейронної мережі.

В сучасних архітектурах, необхідність подібної операції (центрального вирізання) нівелюється підбором відповідних параметрів ядра згортки, кроку, а також паддингу (обрамлення вхідної карти ознак нульовими пікселями) для згорткових шарів. За рахунок цього операція зменшення розмірності відбувається виключно в шарах максимізаційного агрегування (max pooling), в свою чергу просторовий розмір карти ознак (до та після згорткового шару) залишається сталим.

Також на практиці для збільшення ефективності обчислень на графічному процесорі всі розміри (як просторовий розмір, так і кількість каналів карт ознак)

зображень, карт ознак, та результуючої маски є степенями двійки (16, 32, 64, 128, ...). Приклад такої мережі з симетричними розмірами відповідних блоків кодувальника і декодувальника зображено на рис. 2.6.

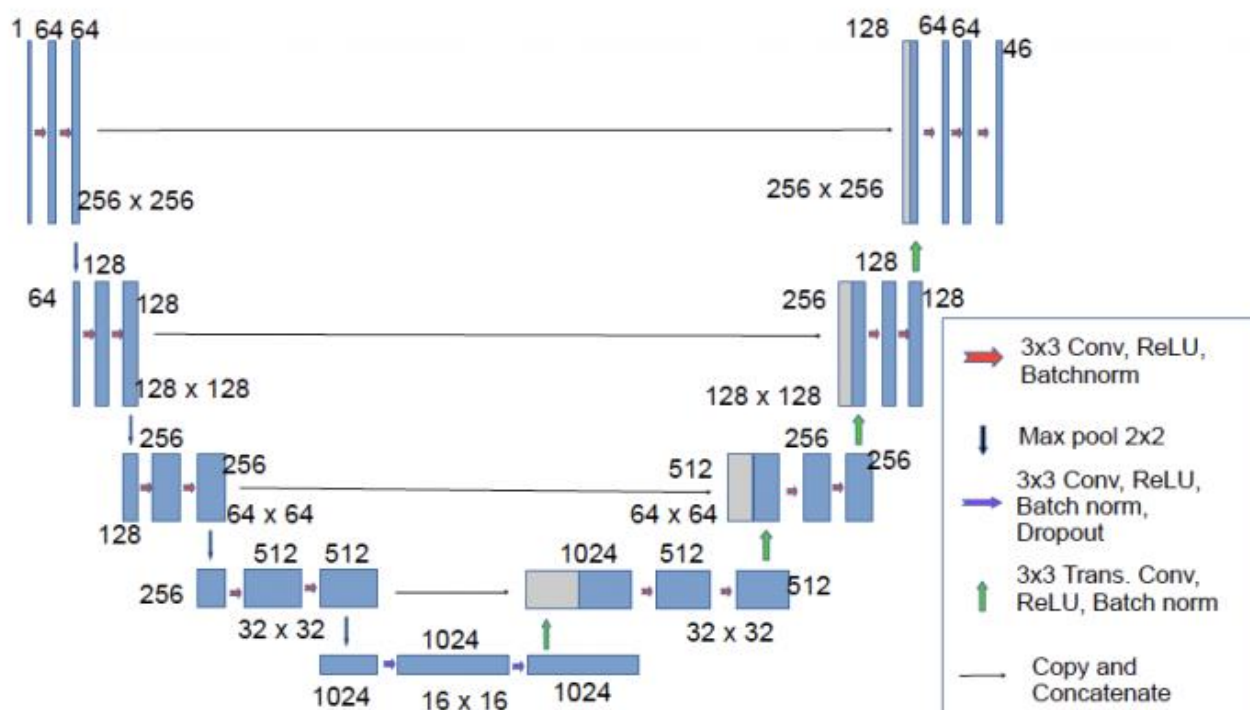


Рисунок 2.6 – Схематичне зображення архітектури U-Net з симетричними розмірами відповідних блоків кодувальника і декодувальника

### 2.1.3. Глибоке контрольоване навчання

В розділі про архітектуру нейронної мережі U-Net++ ми вже згадували про глибоке контрольоване навчання. Проте цей підхід почав застосовувати для тренування нейронних мереж набагато раніше.

Основною ідеєю глибокого контрольованого навчання є передбачення результуючих сегментаційних масок на різних рівнях декодувальника. Це автоматично означає також і підрахунок функцій втрат та додаткову оптимізацію цих самих блоків декодувальника [24].

Розглянемо попереднє схематичне зображення архітектури U-Net з симетричними розмірами відповідних блоків. Результуюча сегментаційна маска тут



лише одна - P1 розміром  $256 \times 256 \times 2$  (де  $256 \times 256$  - просторова розмірність маски, а 2 - кількість класів/каналів). Відповідна до P1 істинна сегментаційна маска позначена як T1, а відповідна функція втрат позначена як Loss1.

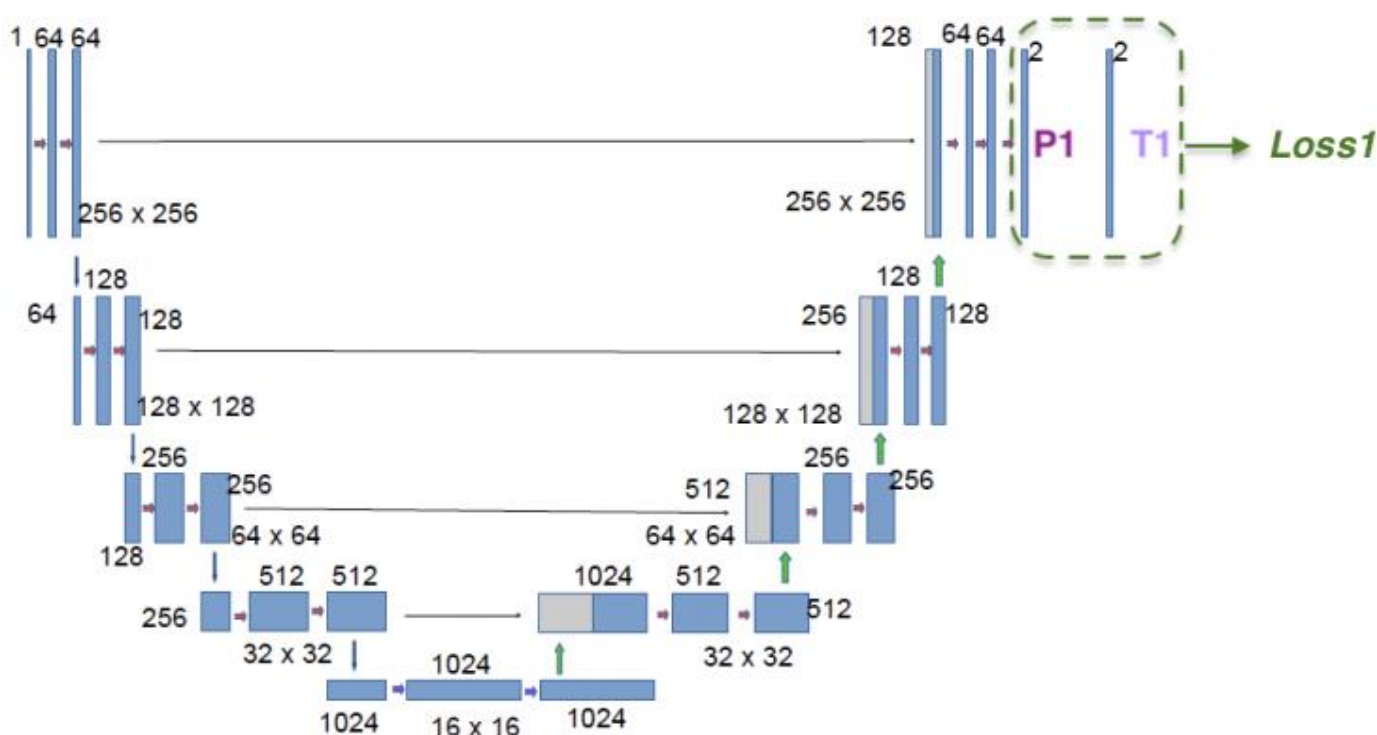


Рисунок 2.7 – Схематичне зображення архітектури U-Net без застосуванням глибокого контрольованого навчання

У випадку ж використання глибокого контрольованого навчання, щонайменше ще один блок декодувальника має сегментаційний вихід. Для простоти пояснення, в усіх прикладах буде використовуватись лише один додатковий блок декодувальника. В нашому випадку це буде P2 розміром  $128 \times 128 \times 2$  (де  $128 \times 128$  - просторова розмірність маски, а 2 - кількість класів/каналів). Застосувати глибоке контрольоване навчання в даному випадку можна щонайменше двома способами.

У першому випадку (наступний рисунок) ми можемо згенерувати істинну сегментаційну маску T2 (відповідного до P2 розміру) шляхом зниження просторової розмірності маски T1 вдвічі (з  $256 \times 256 \times 2$  до  $128 \times 128 \times 2$ ) використовуючи білінійну інтерполяцію або інтерполяцію найближчого сусіда.

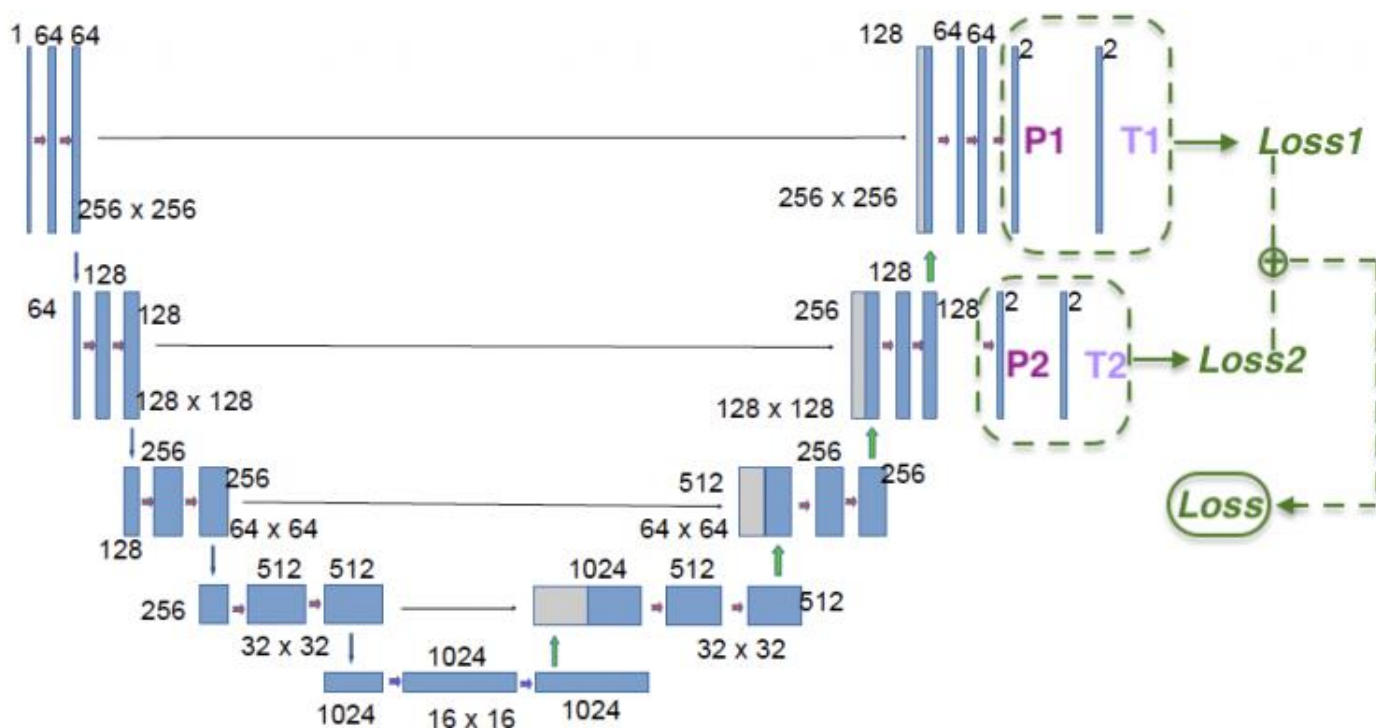


Рисунок 2.8 – Схематичне зображення архітектури U-Net з застосуванням глибокого контрольованого навчання. Випадок 1

Тепер використовуючи передбаченну моделлю маску P2 та істинну маску T2 ми можемо підрахувати відповідне значення функції втрат Loss2. Результуючу функцію втрат всієї нейронної мережі можна буде підрахувати як:

$$Loss = 0.5 * Loss1 + 0.5 * Loss2 \quad (2.1)$$

В загальному ж випадку функція втрат всієї нейронної мережі має наступний вигляд:

$$Loss = \alpha * Loss1 + \beta * Loss2 \quad (2.2)$$

, де alpha та beta є гіперпараметрами, що можуть бути підібрані під час тренування.

У другому ж випадку (наступний рисунок) замість того, щоб генерувати істинну сегментаційну маску меншої розмірності (T2), ми вдвічі збільшимо роздільну здатність передбаченої сегментаційної маски P2 до розміру маски P1 (з 128x128x2 до

256x256x2) використовуючи білінійну інтерполяцію або інтерполяцію найближчого сусіда.

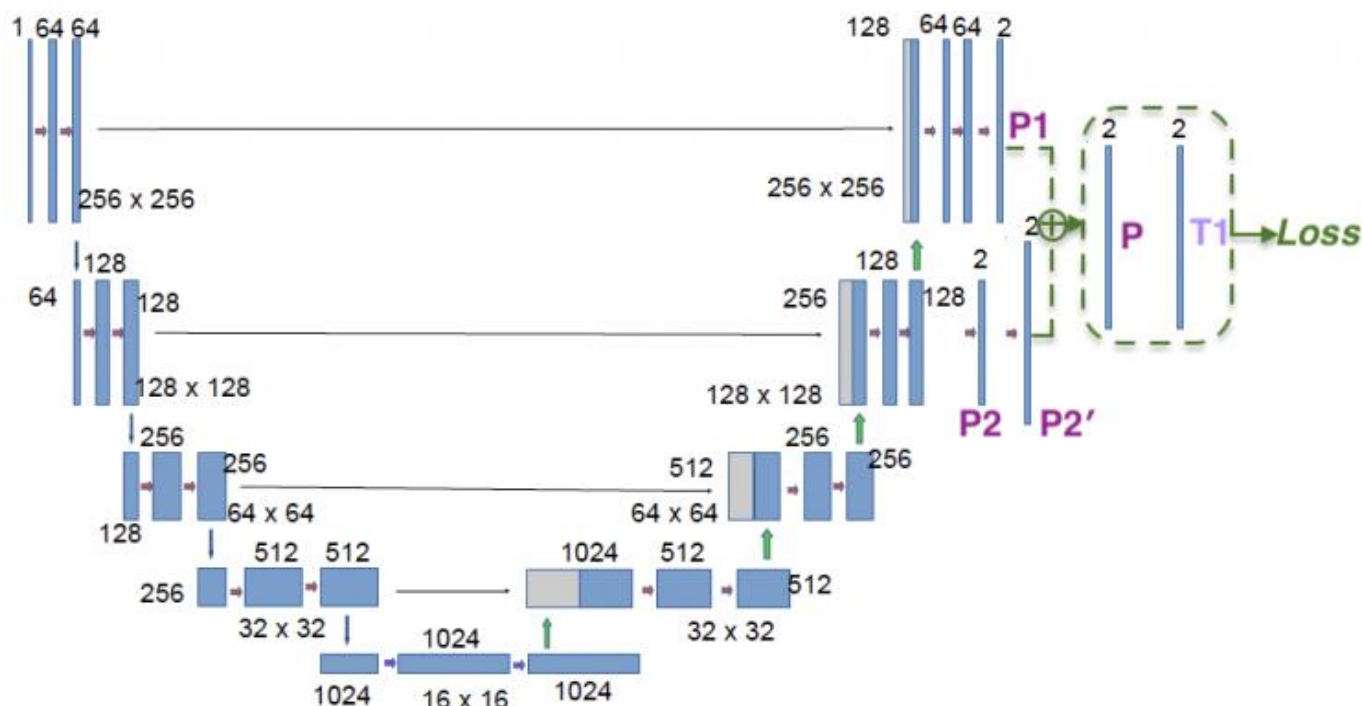


Рисунок 2.9 – Схематичне зображення архітектури U-Net з застосуванням глибокого контрольованого навчання. Випадок 2

Далі ми об'єднаємо отриману маску  $P2'$  з маскою  $P1$  застосовуючи операцію попіксельного додавання (поелементне додавання матриць). Результуюча маска  $P$  передбачена нейронною мережею матиме вигляд:

$$P = 0.5 * P1 + 0.5 * P' \quad (2.3)$$

Результуючу функцію втрат всієї нейронної мережі можна буде підрахувати використовуючи передбачену сегментаційну маску  $P$  та відповідну істинну сегментаційну маску  $T1$ .

Підхід глибокого контрольованого навчання допомагає кращій та швидшій збіжності оптимізаційної задачі, а також покращує точність семантичної сегментації. В той же час, глибоке контрольоване навчання сприяє покращенню результатів

сегментації (як візуальних, так і кількісних) шляхом поєднання сегментаційних карт різної глибини - з різних блоків декодувальника різної роздільної здатності.

## 2.2. Кодувальники

### 2.2.1. ResNet

Глибокі нейронні мережі мають змогу виявляти велику кількість патернів або ознак із навчальних даних та вивчити дуже складні, і в той же час значущі представлення [56].

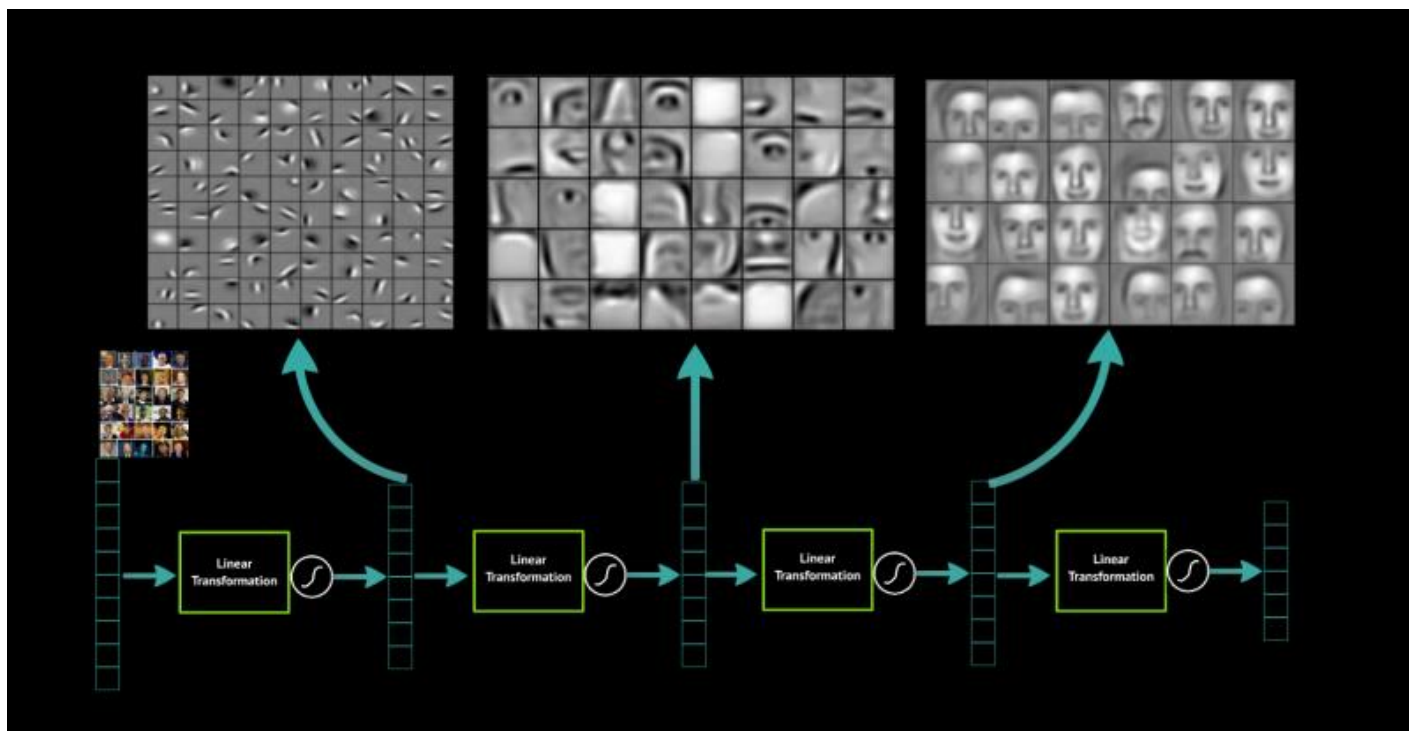


Рисунок 2.10 – Виявлення, агрегація на навчання ознак

Виявленням цих особливостей або патернів можна завдячувати глибині нейронної мережі, оскільки вони більш сприйнятливі до виявлення таких на пізніших (глибших) рівнях мережі. Оскільки складність задач, які на той час намагалися вирішувати нейронними мережами, ставала (і досі стає) все більшою, дослідники почали використовувати все глибші моделі для досягнення більшої точності

результатів. Було помічено, що глибокі нейронні мережі працюють краще, ніж їх не глибокі аналоги.

Без значної глибини нейронна мережа не матиме змоги у складний спосіб інтегрувати ознаки (патерни) різних рівнів, щоб навчатися на навчальних даних. Цей висновок призводить до того, що складні проблеми можуть бути вирішені лише за допомогою впровадження дійсно глибоких моделей навчання (нейронні мережі, що мають більше 50 прихованих шарів). Дослідження проводилися з різними моделями глибокого навчання глибиною до 100 шарів, щоб досягти більшої точності натренованої мережі.

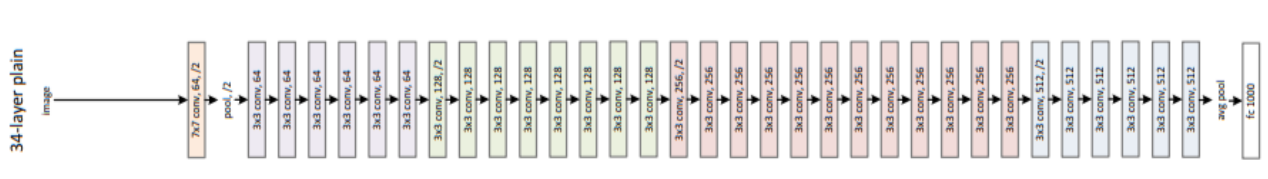


Рисунок 2.11 – Звичайна модель глибокого навчання з 34 прихованими шарами.

Зображення з оригінальної статті ResNet [56]

Теоретично, зі збільшенням кількості шарів у звичайній нейронній мережі вона має ставати все кращою у задачі розпізнаванні складних ознак і особливостей, що має в свою чергу приводити до кращої точності та навчання. Однак, всупереч поширеній думці, виявилось, що така глибока модель була неефективною для забезпечення очікуваного результату. Крім того, точність навчання починала падати після певного моменту навчання нейронної мережі. Перешкодою для відповіді на вищезгадане запитання та розуміння такою розбіжності між теоретичними засадами та практичними результатами була також проблема розмивання/вибуху градієнтів. Вони перешкоджали збіжності оптимізаційних від самого початку тренування нейронної мережі, що робило модель нестабільною в її здатності навчатися точно та ефективно. Однак цю проблему частково вирішують за допомогою нормалізованої ініціалізації ваг мережі, а також проміжних рівнів нормалізації. Це дозволило нейронним мережам з більшою кількістю прихованих шарів збігатися для стохастичного градієнтного спуску та зворотнього поширення градієнтів. Проте навіть після вирішення проблеми розмивання/вибуху градієнтів було виявлено, що точність навчання продовжувала

падати, при поступовому збільшенні кількості прихованих шарів глибокої нейронної мережі. Подібну поведінку можна прослідкувати на схематичному зображенні (рис. 2.12).

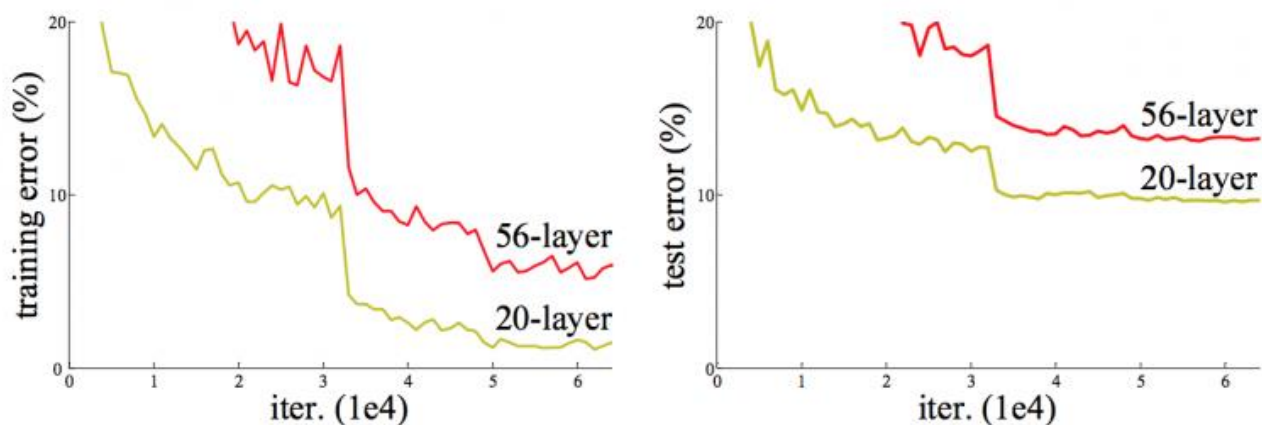


Рисунок 2.12 – Відсоткова помилка нейронної мережі на тренувальній та тестовій вибірках. Нейронні мережі з 20-ма та 56-ма прихованими шарами. Зображення з оригінальної статті ResNet [56]

Було помічено, що нейронна мережа з більшою кількістю шарів - 56 шарів призводить до більшої помилки навчання, на відміну від мережі, яка має набагато меншу кількість шарів - 20 шарів, що призводить до вищих помилок тестування. Можна припустити, що це може бути результатом надмірного навчання. Однак це не так, оскільки глибші мережі показують вищу помилку саме на навчальній, а не лише на тестовій підвбірці. Надмірне навчання, як правило, відбувається, коли помилки на навчальній підвбірці значно кращі (нижчі), ніж помилки на тестовій підвбірці. Це називається проблемою деградації. Зі збільшенням глибини мережі, її точність насичується (модель вивчає всі ознаки ще до досягнення останніх найглибших шарів нейронної мережі), а потім починає швидко погіршуватися при додаванні нових шарів. Щоб краще зрозуміти, чому результат виглядає несподіваним і не очевидним, розглянемо наступний приклад. Припустимо, у нас є нейронна мережа з «n» шарами, яка дає помилку навчання «x». Тепер розглянемо більш глибоку нейронну мережу з «m» ( $m > n$ ) шарів. Коли ми навчаємо цю мережу, ми очікуємо, що вона працюватиме принаймні так само добре, як і попередня модель (n-рівнів), оскільки перші «n» шарів із загальної кількості «m» шарів дадуть таку ж точність, і якщо модель вимагає більш

складне представлення, то решта «m-n» шарів вивчать його, а якщо навчання більше не потрібне, решта цих «m-n» шарів поводитимуться як тотожне відображення, тобто переводитиме значення кожного елементу попереднього шару нейронної мережі в таке ж саме значення наступного шару. Звідси можна зробити висновок, що нейронна мережа з «m» шарами дасть помилку навчання «y» ( $y \leq x$ ). Але на практиці цього не відбувається, і більш глибокі нейронні мережі не обов'язково дають менші помилки навчання «x».

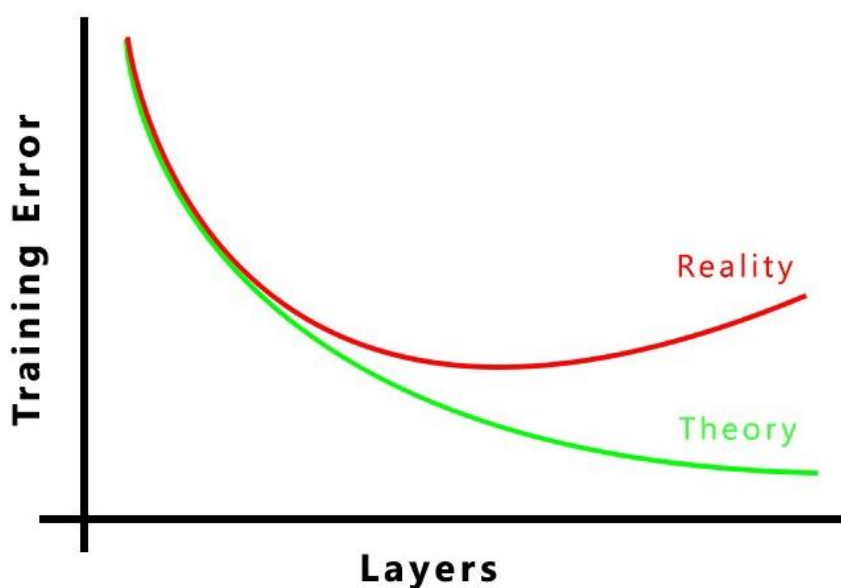


Рисунок 2.13 – Схематичне порівняння теоретичної помилки нейронної мережі та помилки, що зазвичай отримується на практиці на тренувальній вибірці при збільшенні кількості прихованих шарів

Дослідники компанії Microsoft представили структуру залишкового навчання (ResNet), щоб полегшити навчання мереж, які є значно глибшими, ніж раніше, нівелюючи проблему їх деградації. Вони довели на практиці, що нейронні мережі ResNet легше оптимізувати та вони можуть мати вищу точність на значних глибинах. Як показано раніше, останні шари в більш глибоких мережах не можуть вивчити тотожне перетворення, яке потрібна для передачі до наступного шару вхідних значень поточного шару. У залишкових мережах замість того, щоб сподіватися, що шари відповідають бажаному відображенню, ми даємо змогу цим шарам відповідати залишковому відображенню. Раніше для звичайних нейронних мереж бажаним

відображенням було  $H(x)$ . Проте зараз шарам нейронної мережі дозволили відповідати залишковому відображенню  $F(x) = H(x) - x$ , оскільки виявилось, що значно легше оптимізувати саме залишкове відображення, а не повноцінне.

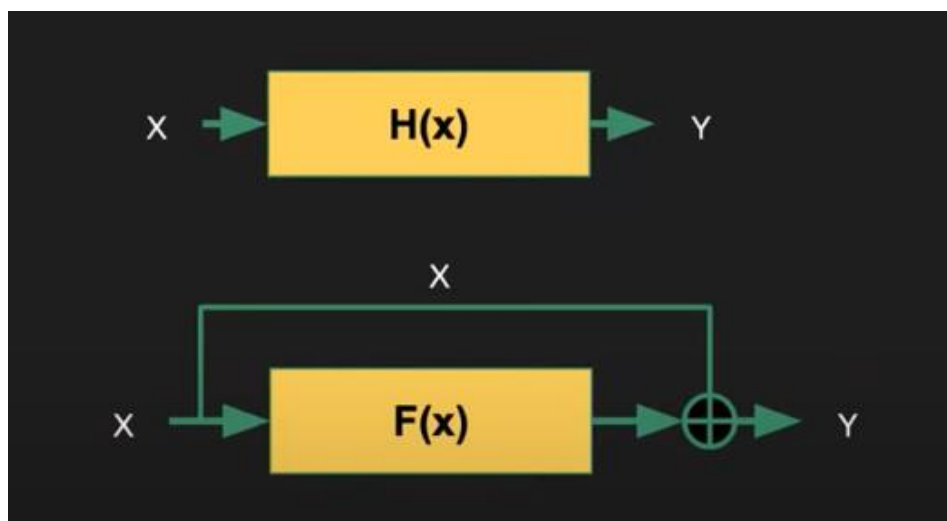


Рисунок 2.14 – Схематичне порівняння повноцінного та залишкового відображень

Цей метод прокидання даних з одного рівня на інший називається з'єднаннями швидкого доступу або пропускними з'єднаннями. Цей підхід дозволяє легко передавати дані між шарами, не перешкоджаючи здатності до навчання дуже глибоких нейронних мереж. Перевага застосування цього типу пропускних з'єднань полягає в тому, що якщо будь-який шар погіршує продуктивність моделі, його буде замінено тотожним перетворенням.

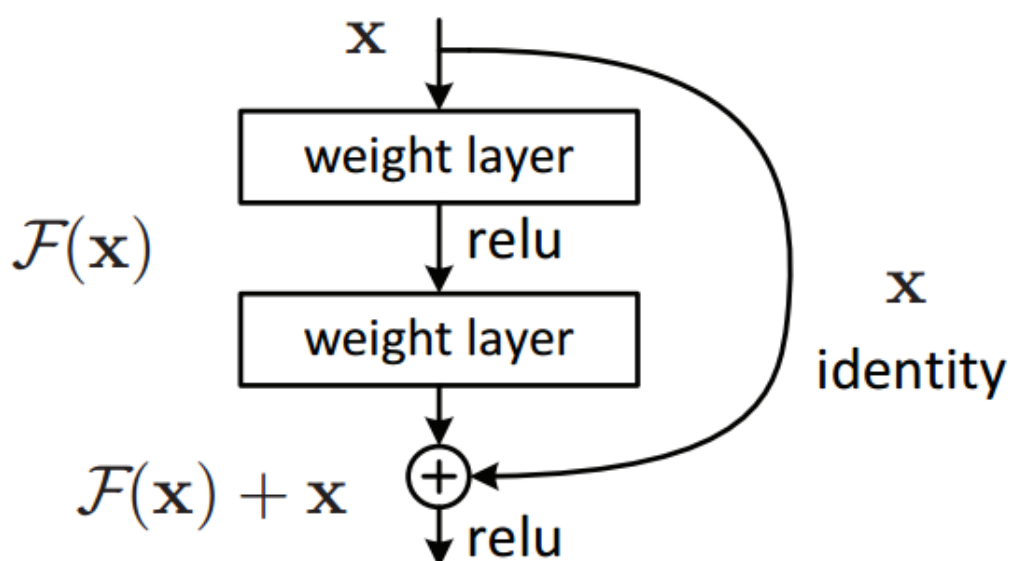


Рисунок 2.15 – З'єднаннями швидкого доступу або пропускними з'єднаннями.

Зображення з оригінальної статті ResNet



Інтуїція, що стоїть за пропусковими з'єднаннями, полягає в тому, що мережі легше навчитися перетворювати значення  $f(x)$  на нуль, щоб воно поведилося як тотожна функція, аніж навчитися вести себе як тотожна функція самостійно, намагаючись щоб знайти правильний набір ваг згорткових шарів, який призведе до потрібного результату.

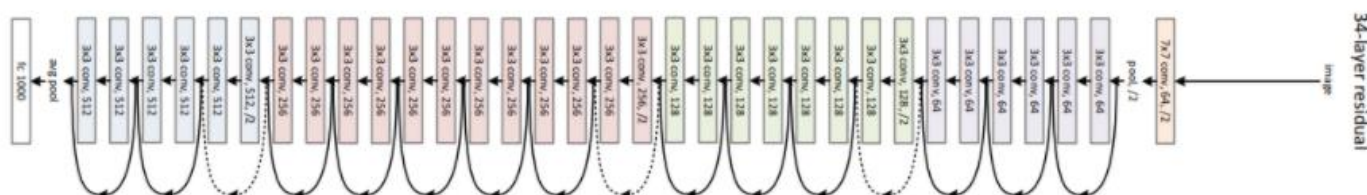


Рисунок 2.16 – Нейронна мережа ResNet з 34 прихованими шарами. Зображення з оригінальної статті ResNet

ResNet використовує два основні типи блоків для побудови всієї мережі:

1. Блок зі звичайним пропусковим з'єднанням (зазначений раніше)

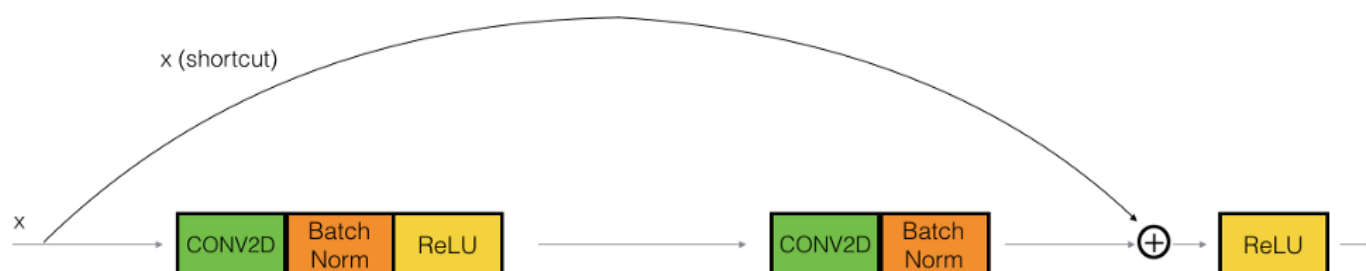


Рисунок 2.17 – Схематичне зображення блоку зі звичайним пропусковим з'єднанням

2. Блок зі згортковим пропусковим з'єднанням

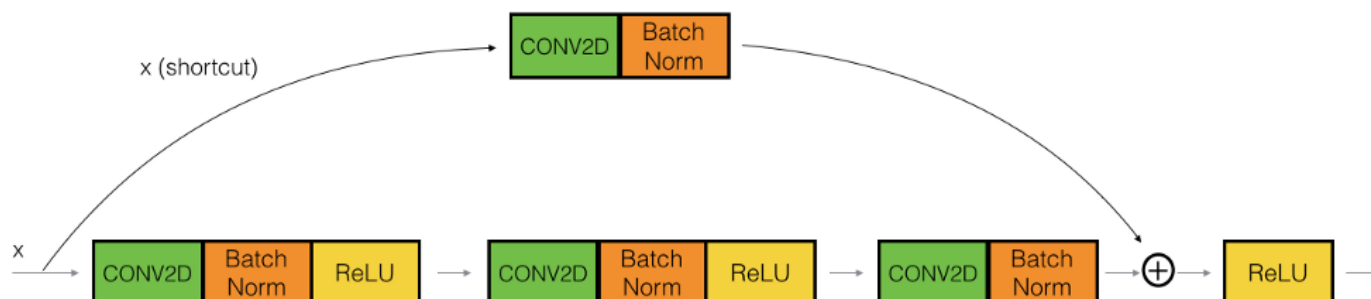


Рисунок 2.18 – Схематичне зображення блоку зі згортковим пропусковим з'єднанням

Блоки зі згортковими пропусковими з'єднаннями допомагають модифікувати та реструктурувати вхідні дані так, щоб вихід попереднього шару відповідав вихідним розмірам поточному шару, щоб в подальшому їх можна було поелементно додати. Ці компоненти допомагають досягти кращої та швидшої оптимізації та вищої точності моделей глибокого навчання. Практичні результати показують ефект використання пропускових з'єднань ResNet поверх звичайних шарів на графіку (рис. 2.19).

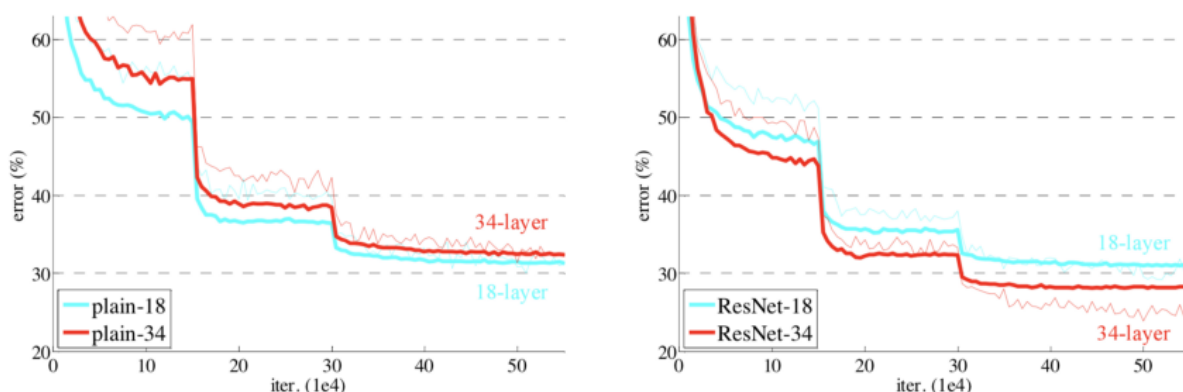


Рисунок 2.19 – Порівняння відсоткової помилки для звичайних нейронних мереж (ліворуч) та нейронних мереж ResNet (праворуч) різної глибини. Зображення з оригінальної статті ResNet [56]

Як бачимо, при збільшенні кількості шарів в нейронній мережі ResNet, точність зростає (відповідно похибка зменшується). Натомість в звичайних нейронних мережах, при збільшенні кількості прихованих шарів, точність починає зменшуватись, що демонструється в табл. 2.1.

Таблиця 2.1

Порівняння відсоткової помилки для звичайних нейронних мереж (ліворуч) та нейронних мереж ResNet (праворуч) різної глибини. Менше значення відповідає кращим результатам

К-сть шарів	Звичайна нейронна мережа	ReNet
18	27,95	27,88
34	38,54	<b>25,03</b>

Отже, ми можемо зробити висновок, що ResNet, безсумнівно, є значним покращенням в глибокому навчанні. Завдяки пропусковим з'єднанням він відкрив доступ іншим дослідникам та розробникам до значно глибших моделей нейронних мереж, що, у свою чергу, призвело до значно кращих результатів роботи цих моделей.

### 2.2.2. EfficientNet

Зазвичай глибокі згорткові нейронні мережі розробляються з фіксованою наперед визначеними вимогами/обмеженнями щодо обчислювальних ресурсів. Пізніше за наявності ресурсів для досягнення кращої точності роботи збільшують розмір нейронної мережі. До прикладу, модель ResNet-18, яку ми розглядали в попередньому розділі, можна збільшити до моделі ResNet-200, додавши ще більше прихованих шарів до вихідної моделі. У більшості випадків ця техніка масштабування моделі допомагає забезпечити кращу точність роботи нейронної мережі для більшості наборів даних для порівняльного аналізу. Але звичайні методи масштабування моделі дуже різноманітні. Деякі з них масштабуються по ширині, а деякі – по глибині. Деякі моделі просто отримують на вхід зображення з більшою роздільною здатністю, сподіваючись отримати кращі результати. Ця техніка випадкового масштабування моделей вимагає ручного налаштування та багато затрат робочого часу, що дуже часто призводить лише до незначного покращення якості роботи нейронної мережі, або ж і до повної відсутності покращень.

Архітектура нейронної мережі EfficientNet [57] використовує техніку під назвою складений коефіцієнт для масштабування моделей простим, але ефективним способом. Замість випадкового збільшення ширини, глибини чи роздільної здатності складене масштабування рівномірно масштабує кожен вимір, використовуючи певний фіксований набір коефіцієнтів масштабування. Використовуючи метод масштабування та AutoML [58], автори статті EfficientNet розробили сім моделей різної розмірності, які перевершують точність більшості сучасних згорткових нейронних мереж і мають набагато кращу ефективність з точки зору ресурсів.

Для розробки методу складеного масштабування автори статті EfficientNet систематично вивчали вплив кожного методу масштабування на продуктивність і ефективність моделі. Вони прийшли до висновку, що хоча масштабування окремих вимірів допомагає підвищити продуктивність моделі, саме збалансоване масштабування в усіх трьох вимірах — ширині, глибині та роздільній здатності вхідного зображення — враховуючи змінні доступні ресурси, найефективніше покращує загальну продуктивність моделі. Схематичну візуалізацію методу складеного масштабування показано на рис. 2.20.

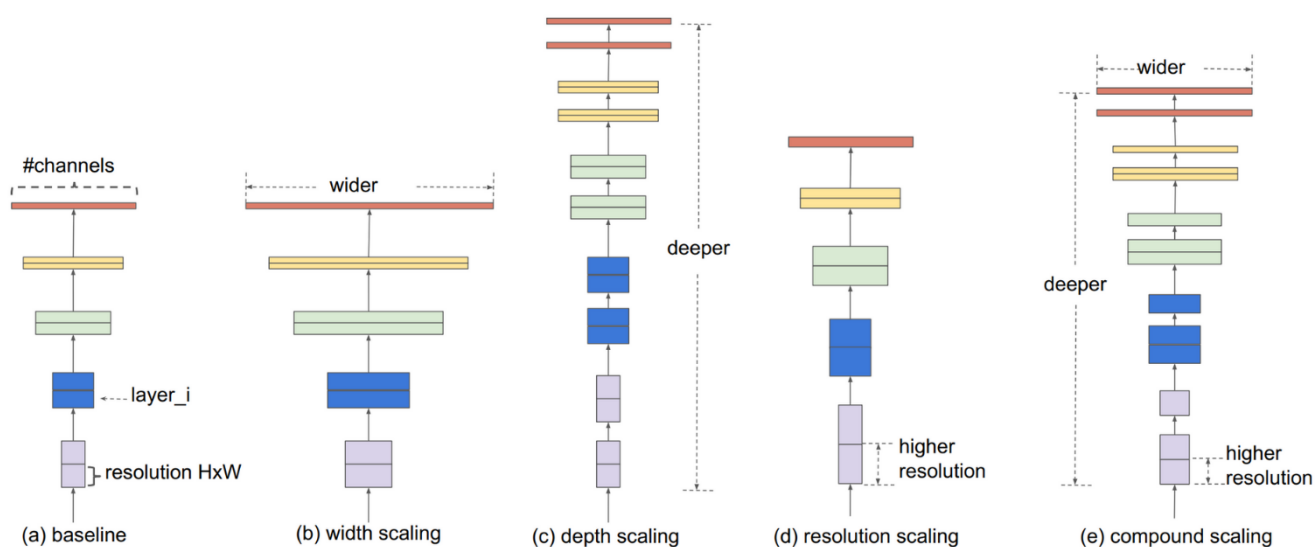


Рисунок 2.20 – Рисунок 1. Різні методи масштабування та складене масштабування.

Зображення з оригінальної статті EfficientNet

Метод складеного масштабування базується на ідеї збалансування розмірів ширини, глибини та роздільної здатності вхідного зображення шляхом масштабування з постійним співвідношенням.

Інтуїція для глибоких згорткових нейронних мереж наступна: якщо вхідне зображення більше, то мережі потрібно більше шарів для збільшення рецептивного поля та більше каналів для захоплення більш дрібних паттернів на більшому зображенні. Техніка складеного масштабування також допомогла підвищити ефективність і точність моделі попередніх архітектур нейронних мереж, таких як

MobileNet і ResNet, приблизно на 1,4% і 0,7% точності на наборі даних ImageNet, відповідно, порівняно з іншими методами випадкового масштабування.

Архітектура нейронної мережі EfficientNet базується на базовій мережі, розробленій шляхом пошуку нейронної архітектури з використанням структури AutoML MNAS. Модель налаштована для отримання максимальної точності, але також штрафується, якщо мережа дуже важка з точки зору обчислень. Він також штрафується за повільний час висновку (inference time), коли мережі потрібно багато часу, щоб зробити прогнози. Архітектура використовує мобільну інвертовану згортку, подібну до MobileNet V2, але більшу за розмірами. Цю базову архітектуру було збільшено методом складеного масштабування для отримання сімейства моделей EfficientNet (EfficientNet-B0, EfficientNet-B1, ..., EfficientNet-B7).

На рисунку 2 показано точність роботи EfficientNet у порівнянні з іншими архітектурами глибоких згорткових нейронних мереж. Найбільша модель EfficientNet - EfficientNet-B7 отримала найкращий результат на наборах даних ImageNet і CIFAR-100. Він отримав приблизно 84,4% top-1 і 97,3% top-5 точності на ImageNet. Крім того, розмір моделі був у 8,4 разів меншим і в 6,1 разів швидшим, ніж попередня найкраща модель. Він отримав точність 91,7% на наборі даних CIFAR-100 і точність на рівні 98,8% на наборі даних Flowers.

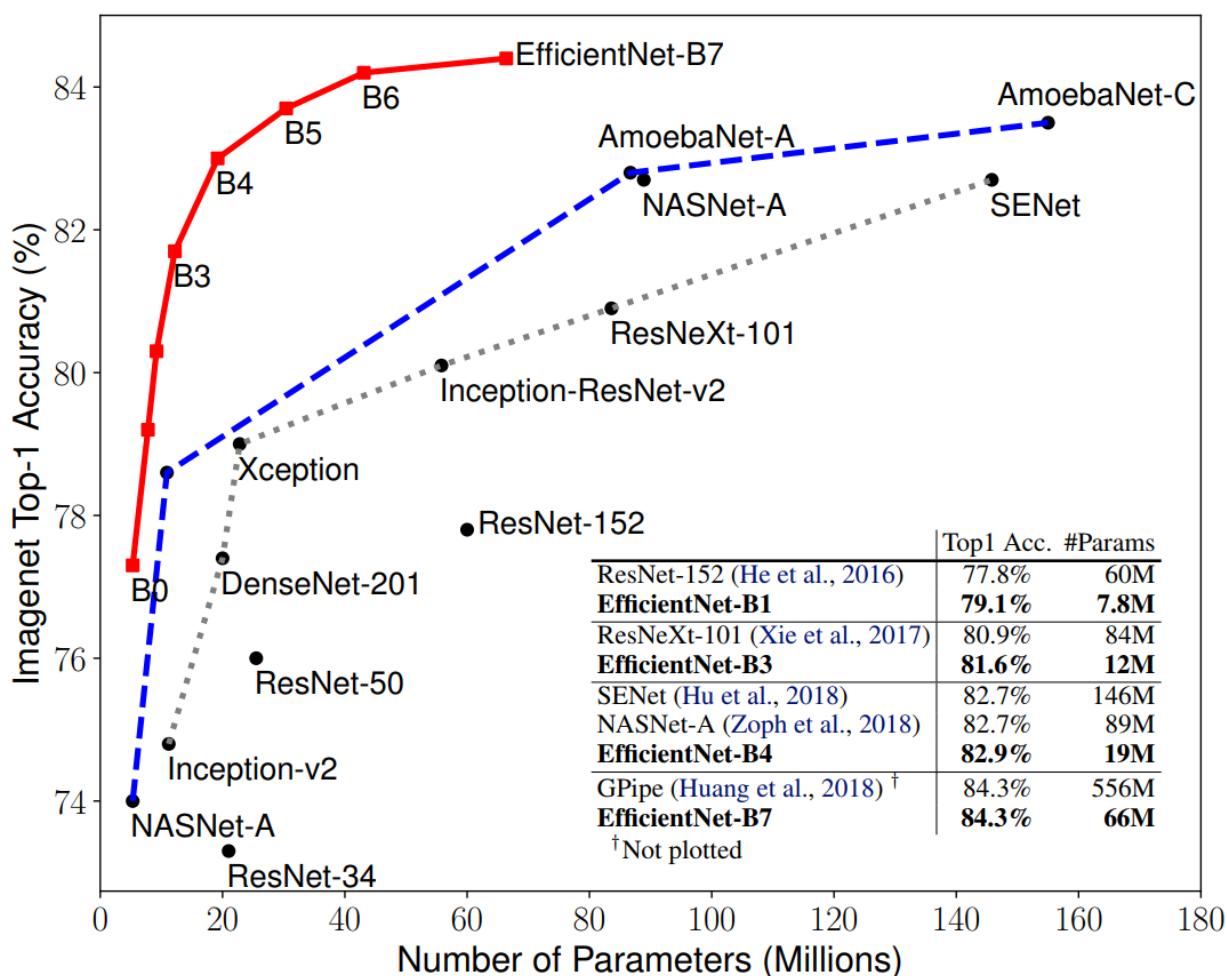


Рисунок 2.21 – Розмір і продуктивність EfficientNet у порівнянні з іншими архітектурами глибоких нейронних мереж на наборі даних ImageNet. Зображення з оригінальної статті EfficientNet

Як показано на рисунку 3, модель також забезпечує кращі карти активації класів, що більш сфокусовані на відповідних регіонах з більшою кількістю деталей об'єктів, що допомагає покращенню інтерпретованості роботи згортової нейронної мережі.

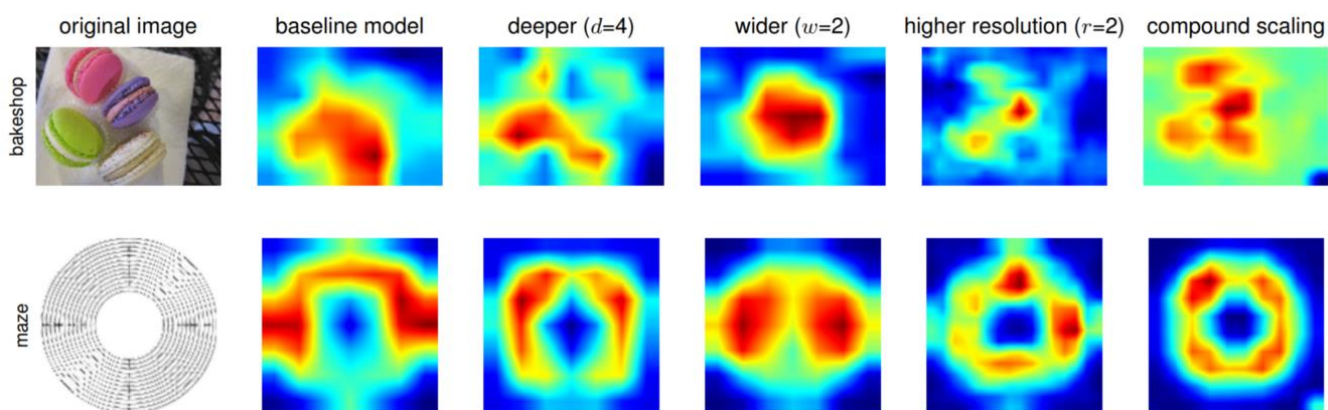


Рисунок 2.22 – Метод складеного масштабування (останній стовпчик) дозволяє масштабованій моделі краще зосереджуватися на більш релевантних регіонах з більшою кількістю деталей цільового об'єкта. Зображення з оригінальної статті EfficientNet

Як показано в таблиці 1, результати роботи глибокої згорткової нейронної мережі EfficientNet перевершують усі попередні архітектури у більшості наборів даних порівняльного аналізу. Метод складеного масштабування також можна використовувати для ефективного масштабування інших архітектур нейронних мереж. Це дозволяє масштабувати моделі EfficientNet таким чином, щоб досягти найкращої точності з на порядок меншою кількістю параметрів і швидкодії FLOPS (операцій з рухомою комою на секунду) на ImageNet та інших широко використовуваних наборах даних для порівняння якості роботи класифікаційних нейронних мереж (табл. 2.2).

Таблиця 2.2

Результати роботи нейронної мережі EfficientNet на валідаційних наборах даних. Моделі EfficientNet досягають найкращої точності для 5 із 8 наборів даних, водночас в середньому містять в 9,6 разів менше параметрів. Зображення з оригінальної статті EfficientNet

	Model	Comparison to best public-available results					Comparison to best reported results					
		Acc.	#Param	Our Model	Acc.	#Param(ratio)	Model	Acc.	#Param	Our Model	Acc.	#Param(ratio)
CIFAR-10	NASNet-A	98.0%	85M	EfficientNet-B0	98.1%	4M (21x)	†Gpipe	<b>99.0%</b>	556M	EfficientNet-B7	98.9%	64M (8.7x)
CIFAR-100	NASNet-A	87.5%	85M	EfficientNet-B0	88.1%	4M (21x)	Gpipe	91.3%	556M	EfficientNet-B7	<b>91.7%</b>	64M (8.7x)
Birdsnap	Inception-v4	81.8%	41M	EfficientNet-B5	82.0%	28M (1.5x)	GPipe	83.6%	556M	EfficientNet-B7	<b>84.3%</b>	64M (8.7x)
Stanford Cars	Inception-v4	93.4%	41M	EfficientNet-B3	93.6%	10M (4.1x)	‡DAT	<b>94.8%</b>	-	EfficientNet-B7	94.7%	-
Flowers	Inception-v4	98.5%	41M	EfficientNet-B5	98.5%	28M (1.5x)	DAT	97.7%	-	EfficientNet-B7	<b>98.8%</b>	-
FGVC Aircraft	Inception-v4	90.9%	41M	EfficientNet-B3	90.7%	10M (4.1x)	DAT	92.9%	-	EfficientNet-B7	<b>92.9%</b>	-
Oxford-IIIT Pets	ResNet-152	94.5%	58M	EfficientNet-B4	94.8%	17M (5.6x)	GPipe	<b>95.9%</b>	556M	EfficientNet-B6	95.4%	41M (14x)
Food-101	Inception-v4	90.8%	41M	EfficientNet-B4	91.5%	17M (2.4x)	GPipe	93.0%	556M	EfficientNet-B7	<b>93.0%</b>	64M (8.7x)
Geo-Mean						<b>(4.7x)</b>						<b>(9.6x)</b>

## 2.3. Функції втрат

### 2.3.1. Ентропія

Розглянемо задачу класифікації, де зображення класифікується як собака, кіт, кінь або гепард зображену на рис. 2.23.

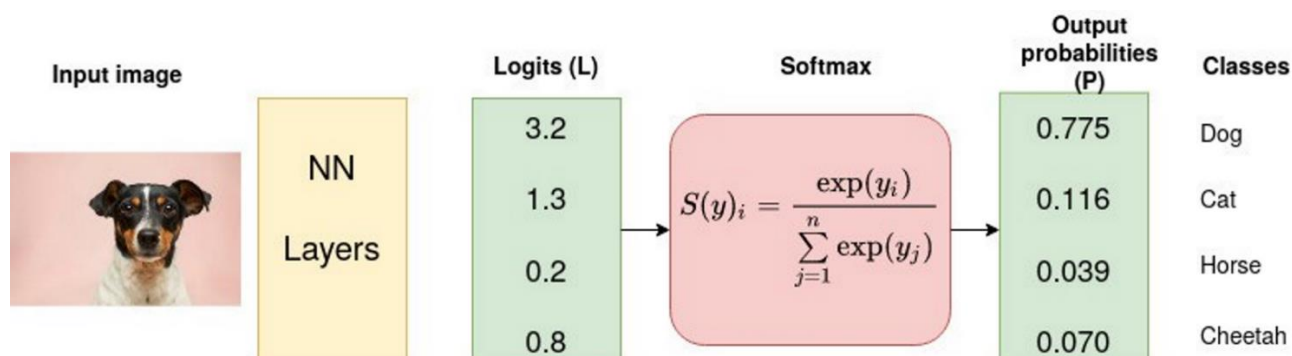


Рисунок 2.23 – Схематичне зображення класифікаційної нейронної мережі

На наведеному вище малюнку функція Softmax перетворює логіти в ймовірності. Метою перехресної ентропії [59] є визначення вихідних ймовірностей (P) і вимірювання відстані від істинних значень (як показано на рис. 2.24).



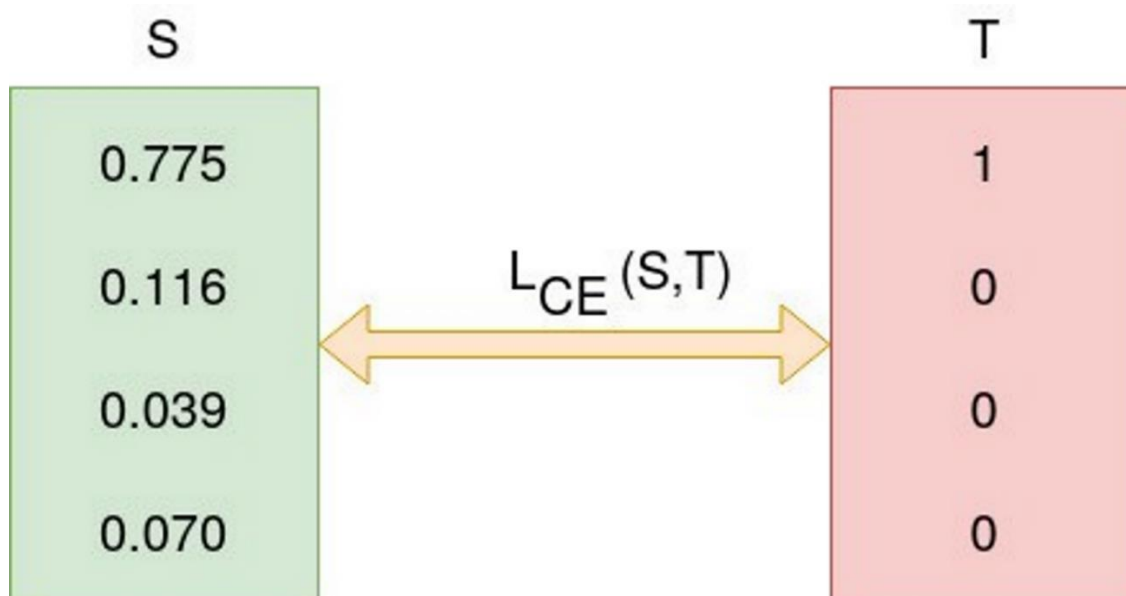


Рисунок 2.24 – Передбачені та істинні значення ймовірностей

У наведеному вище прикладі істинне значення вектору  $T$   $[1, 0, 0, 0]$  для класу «собака», але вивід моделі наступний  $[0.775, 0.116, 0.039, 0.070]$ .

Мета полягає в тому, щоб вихід моделі був якомога ближчим до істинних значень (розмітки). Під час навчання моделі вагові коефіцієнти моделі ітеративно коригуються з метою мінімізації функції втрат – крос-ентропії. Процес коригування ваг – це і є процес навчання моделі. Оскільки модель тренується ітеративно і втрати мінімізуються поступово, ми вважаємо, що модель навчається.

Концепція перехресної ентропії, що бере свій початок у галузі теорії інформації, була представлена Клодом Шенноном в 1948 році.

Ентропія випадкової величини  $X$  — це рівень невизначеності, або його усередненого інформаційного вмісту.

Для  $p(x)$  — розподілу ймовірностей випадкової величини  $X$  ентропія визначається наступним чином:

$$H(X) = \begin{cases} - \int_x p(x) \log p(x), & \text{якщо } X \text{ — неперервна величина} \\ - \sum_x p(x) \log p(x), & \text{якщо } X \text{ — дискретна величина} \end{cases} \quad (2.4)$$

Чим більше значення ентропії  $H(x)$ , тим більша невизначеність розподілу ймовірностей, а чим менше значення, тим менша невизначеність.

Розглянемо наступний приклад з трьома «контейнерами» з фігурами (трикутники та кола):

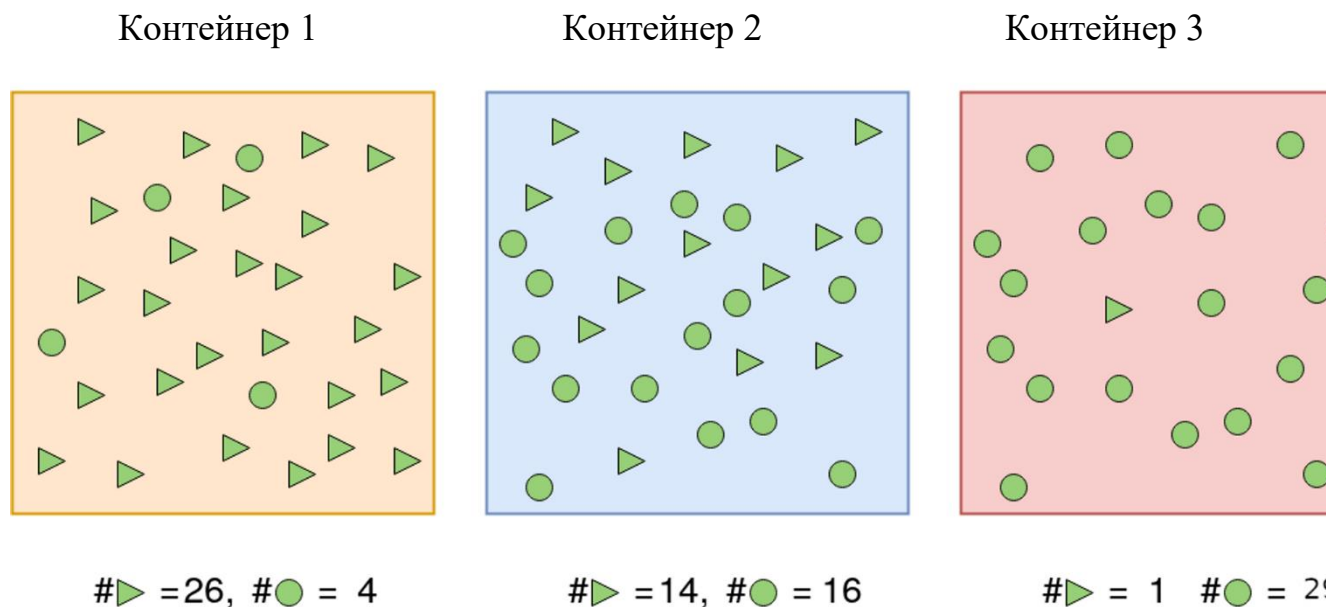


Рисунок 2.25 – Схематичний приклад з трьома контейнерами

Контейнер 1: ймовірність вибрати трикутник становить  $26/30$ , а ймовірність вибрати коло –  $4/30$ . З цієї причини вірогідність вибору однієї фігури та/або невибору іншої більш певна.

Контейнер 2: ймовірність вибрати трикутник становить  $14/30$  і  $16/30$  в іншому випадку. Тобто ймовірність вибрати будь-яку конкретну фігуру майже 50-на-50. Впевненість у виборі заданої фігури менша, ніж у випадку з першим контейнером.

Контейнер 3: фігура, вибрана з контейнера 3, швидше за все, буде колом. Ймовірність вибрати коло становить  $29/30$ , а ймовірність вибрати трикутник —  $1/30$ . Цілком ймовірно, що обрана фігура буде колом.

Обчислимо ентропію, щоб підтвердити наші твердження щодо надійності вибору даної форми:

Ентропія для першого контейнера:

$$\begin{aligned}
H(X) &= - \sum_x p(x) \log p(x) \\
&= - [p(x_1) \log_2(p(x_1)) + p(x_2) \log_2(p(x_2))] \\
&= - \left[ \frac{26}{30} \log_2 \left( \frac{26}{30} \right) + \frac{4}{30} \log_2 \left( \frac{4}{30} \right) \right] = 0.5665
\end{aligned} \tag{2.5}$$

Ентропія для другого контейнера:

$$\begin{aligned}
H(X) &= - \sum_x p(x) \log p(x) \\
&= - [p(x_1) \log_2(p(x_1)) + p(x_2) \log_2(p(x_2))] \\
&= - \left[ \frac{14}{30} \log_2 \left( \frac{14}{30} \right) + \frac{16}{30} \log_2 \left( \frac{16}{30} \right) \right] = 0.9968
\end{aligned} \tag{2.6}$$

Ентропія для третього контейнера:

$$\begin{aligned}
H(X) &= - \sum_x p(x) \log p(x) \\
&= - [p(x_1) \log_2(p(x_1)) + p(x_2) \log_2(p(x_2))] \\
&= - \left[ \frac{1}{30} \log_2 \left( \frac{1}{30} \right) + \frac{29}{30} \log_2 \left( \frac{29}{30} \right) \right] = 0.2108
\end{aligned} \tag{2.7}$$

Як і очікувалося, ентропія для першого і третього контейнерів менша, ніж для другого. Це пояснюється тим, що ймовірність вибору заданої фігури більш певна в контейнерах 1 і 3, ніж у 2.

### 2.3.2. Перехресна ентропія

Також зустрічаються інші назви, такі як логарифмічна або логістична функція втрат. Для підрахунку перехресної ентропії [59] кожна прогнозована ймовірність класу порівнюється з фактичним істинним значенням класу 0 або 1, і розраховується оцінка / втрата, яка штрафує прогнозовану ймовірність на основі того, наскільки вона

далеко від фактичного істинного значення. Дана функція втрат є логарифмічною за своєю природою, що дає великий штраф для великих відмінностей, близьких до 1, і малий штраф для малих відмінностей, близьких до 0.

Перехресна ентропія використовується для коригування ваг моделі під час навчання. Мета полягає в тому, щоб оптимізувати функцію втрат (мінімізувати втрати), так як менші втрати ведуть до кращої моделі. Ідеальна модель має перехресну ентропію рівну 0. На практиці ж, на жодному тренувальному наборі даних для реальної задачі неможливо досягти такого значення, тому більш точне формулювання – ідеальна модель має перехресну ентропію близьку до 0.

Перехресна ентропія визначається як:

$$L_{CE} = - \sum_{i=1}^n t_i \log(p_i) \quad (2.8)$$

для  $n$  класів, де  $t_i$  - істинне значення, а  $p_i$  - softmax ймовірності для  $i$ -го класу.

### 2.3.3. Бінарна перехресна ентропія

Для бінарної класифікації [59] (завдання класифікації з двома класами — 0 і 1) ми маємо бінарну перехресну ентропію, визначену як:

$$\begin{aligned} L &= - \sum_{i=1}^2 t_i \log(p_i) = -[t_1 \log(p_1) + t_2 \log(p_2)] \\ &= -[t \log(p) + (1 - t) \log(1 - p)] \end{aligned} \quad (2.9)$$

де  $t_i$  – істинне значення, а  $p_i$  – softmax ймовірності для  $i$ -го класу

Бінарна перехресна ентропія – зазвичай обчислюється як середня перехресна ентропія для всіх прикладів даних, тобто:

$$L = -\frac{1}{N} \left[ \sum_{j=1}^N [t_j \log(p_j) + (1 - t_j) \log(1 - p_j)] \right] \quad (2.10)$$

для  $N$  зображень тренувальної вибірки, де  $t_j$ - істинне значення, а  $p_j$ - softmax ймовірності для  $j$ -го зображення.

### 2.3.4. Бінарна перехресна ентропія для задачі семантичної сегментації

Всі попередні викладки, щодо перехресної ентропії та бінарної перехресної ентропії були охоплювали задачу класифікації зображень. Проте нам вже відомо, що задача семантичної сегментації зображень являє собою задачу попиксельної класифікації.

Тобто замість того, щоб передбачити для зображення одну мітку класу, ми передбачаємо відповідну мітку класу для кожного пікселя даного зображення. Наприклад, для зображення розміром  $512 \times 512$  ми маємо передбачити  $512 \times 512 = 262144$  міток класів.

Тому для задачі семантичної сегментації зображень функція втрат бінарної перехресної ентропії [59] матиме наступний вигляд:

$$L = -\frac{1}{N} \left[ \sum_{j=1}^N \frac{1}{M} \sum_{i=1}^M [t_{ji} \log(p_{ji}) + (1 - t_{ji}) \log(1 - p_{ji})] \right] \quad (2.11)$$

де  $N$  - кількість зображень тренувальної вибірки,  $M$  - кількість пікселів зображення,  $t_{ji}$  – істинне значення, а  $p_{ji}$  – softmax ймовірності для  $i$ -го пікселя  $j$ -го зображення.

### 2.3.5. Фокальна функція втрат

Штрафування моделі однаково для всіх класів є недоліком, особливо коли існує нерівність або дисбаланс у появі класів на зображенні [60]. Типове зображення для

семантичної сегментації (або ж сегментації об'єктів) складається з багатьох пікселів фону (клас = 0) і невеликої кількості пікселів переднього плану (клас = 1), тобто пікселів, що містять об'єкт інтересу. Це призводить до проблеми класового дисбалансу.

Проблема дисбалансу класів впливає на модель семантичної сегментації (або сегментації об'єктів наступним чином):

- Відсутність навчання через легкі негативні зразки: Якщо побудувати нейронну мережу та почати її тренувати, вона швидко навчиться класифікувати негативні (фонові) пікселі на базовому рівні. З цього моменту збільшення кількості навчальних прикладів не покращать ефективність роботи моделі, тому що модель вже і так добре виконує свою роботу. Це робить навчання неефективним, оскільки більшість пікселів на зображенні є легкими негативними (фоновими) зразками, тобто їх легко класифікувати моделлю як фон, отже вони не сприяють покращенню навчання моделі.

- Коли невеликі втрати від таких легких негативних зразків підсумовуються для багатьох зображень, вклад від них домінує в загальній функції втрат, породжуючи вироджені моделі. Легко класифіковані пікселі фону складають більшу частину функцій втрат і домінують у градієнті. Це зміщує фокус моделі на правильну класифікацію лише фонових та легких пікселів. Іншими словами, вагові коефіцієнти нейронної мережі оновлюються таким чином, що продуктивність лише фонових та легких пікселів стає все кращою.

Фокальна функція втрат одночасно вирішує обидві ці проблеми:

1. проблему дисбалансу між легкими та складними зразками;
2. проблему дисбалансу між позитивними пікселями (об'єктами) та негативними пікселями (фоном).

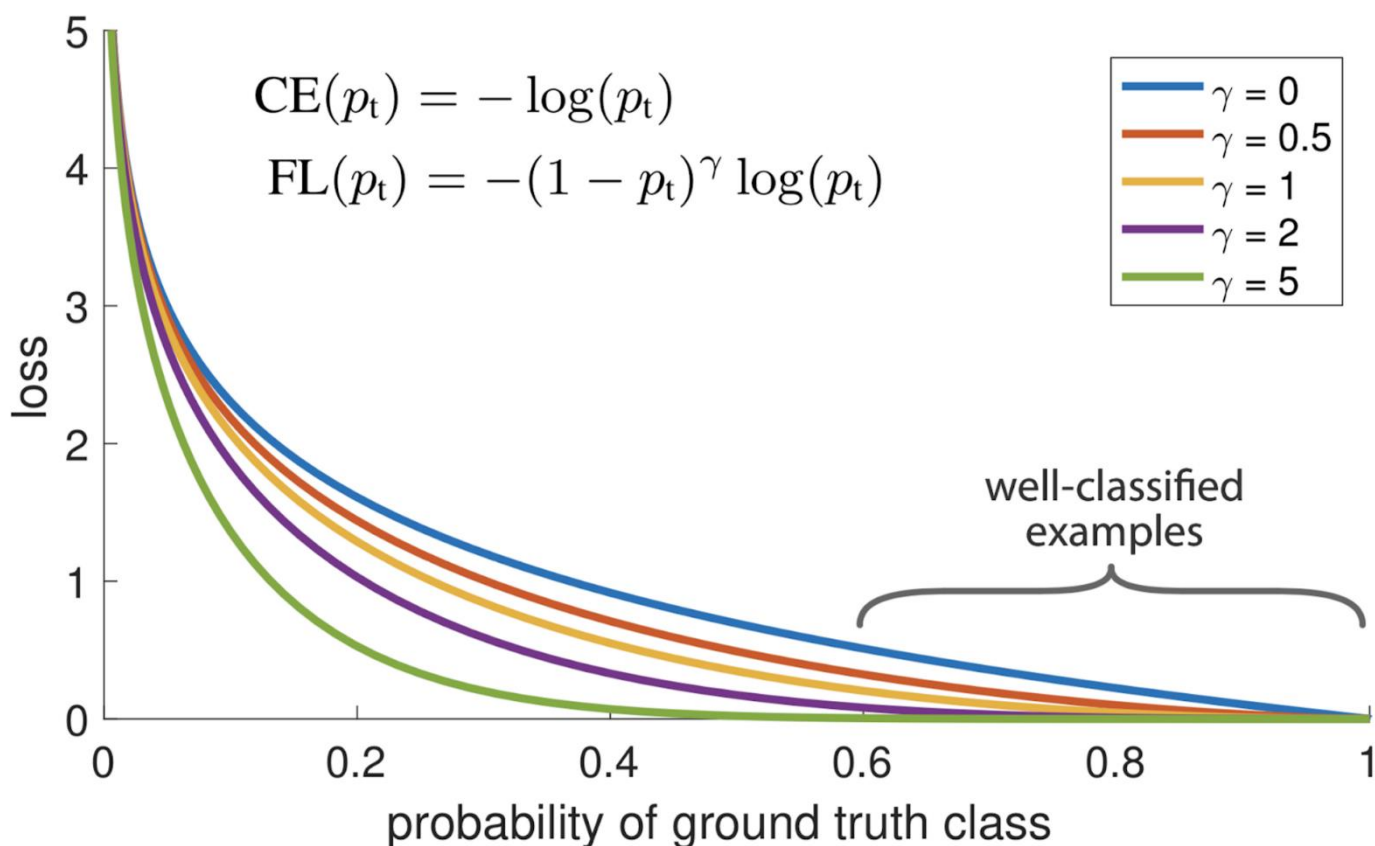


Рисунок 2.26 – Фокальна функція втрат

Було введено новий параметр, коефіцієнт модуляції ( $\gamma$ ), щоб створити покращену функцію втрат. Це можна інтуїтивно зрозуміти із зображення вище. Коли  $\gamma=0$ , функція втрат є ідентичною стандартній перехресній ентропії, а діапазон прогнозованих ймовірностей  $p$ , для яких вклад у функцію втрат низький, приблизно обмежена відрізком  $[0,6, 1]$  (синя крива). Це область простих прикладів, класифікованих з високою ймовірністю. Іншими словами, навіть приклади, які легко класифікувати, несуть втрати з нетривіальною величиною. Саме для таких прикладі ми шукаємо спосіб зменшити їх внесок у загальну функцію втрат.

Тепер, коли ми збільшуємо значення  $\gamma$  (до 5), ми повільно розширюємо відрізок прогнозованих ймовірностей, де значення функції втрат є низьким до з  $[0,6, 1]$  до  $[0,3, 1]$  (зелена крива). Це означає, що ми «розширюємо» або «послаблюємо» наші критерії до добре класифікованих прикладів шляхом збільшення значень параметра  $\gamma$ .

Коефіцієнт модуляції зменшує внесок в загальну функцію втрат від простих прикладів і розширює діапазон передбачених ймовірностей, які сприяють низькому значенню втрат.

Балансування складних і легких випадків використовуючи коефіцієнт модуляції  $\gamma$ :

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (2.12)$$

Додавши параметр  $\gamma$ , ми змогли розрізнити легкі та складні приклади. Проте до досі не є вирішеною проблема класового дисбалансу позитивних і негативних прикладів.

Для цього було введено ваговий параметр ( $\alpha$ ).  $\alpha_t$  — зважений термін, значення якого дорівнює  $\alpha$  для позитивного класу (пікселів об'єкту) та  $1-\alpha$  для негативного класу (пікселів фону). Параметр  $\alpha$  зазвичай дорівнює зворотній частоті цільового класу, тобто якщо на усіх зображеннях конкретного набору даних пікселі, що відповідають будинкам, становлять 10% усіх пікселів, то значення відповідного параметру дорівнює 0.9 для позитивного класу (будинки) та, відповідно, 0.1 для негативного класу (фон).

Усунення дисбалансу між цільовим класом і класом фону використовуючи ваговий параметр  $\alpha$ :

$$FL(p_t) = -\alpha_t \log(p_t) \quad (2.13)$$

Об'єднавши дві версії фокальних втрат, зазначені вище, ми отримаємо:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (2.14)$$

Таким чином, фокальна функція втрат призначена для усунення дисбалансу легких та складних прикладів шляхом зменшення ваги простих прикладів таким чином, щоб їхній внесок у загальну функцію втрат був невеликим, навіть якщо їх



кількість велика. Фокальна функція втрат зосереджує навчання моделі на вивчені підмножини складних прикладів. Водночас фокальна функція втрат усуває проблему дизбалансу між цільовим та фоновим класом.

## ВИСНОВКИ ДО РОЗДІЛУ 2

Основною ідеєю глибокого контрольованого навчання є передбачення результуючих сегментаційних масок на різних рівнях декодувальника. Це автоматично означає також і підрахунок функцій втрат та додаткову оптимізацію цих самих блоків декодувальника.

ResNet, безсумнівно, є значним покращенням в глибокому навчанні. Завдяки пропусковим з'єднанням він відкрив доступ іншим дослідникам та розробникам до значно глибших моделей нейронних мереж, що, у свою чергу, призвело до значно кращих результатів роботи цих моделей.

Моделі EfficientNet досягають найкращої точності для 5 із 8 наборів даних, водночас в середньому містять в 9,6 разів менше параметрів.

Ентропія випадкової величини  $X$  — це рівень невизначеності, або його усередненого інформаційного вмісту. Чим більше значення ентропії  $H(x)$ , тим більша невизначеність розподілу ймовірностей, а чим менше значення, тим менша невизначеність.

Перехресна ентропія використовується для коригування ваг моделі під час навчання. Мета полягає в тому, щоб оптимізувати функцію втрат (мінімізувати втрати), так як менші втрати ведуть до кращої моделі. Ідеальна модель має перехресну ентропію рівну 0.

Бінарна перехресна ентропія зазвичай обчислюється як середня крос-ентропія для всіх прикладів даних.

Для задачі семантичної сегментації зображень функція втрат бінарної перехресної ентропії замість того, щоб передбачити для зображення одну мітку класу, ми передбачаємо відповідну мітку класу для кожного пікселя даного зображення.

Фокальна функція втрат одночасно вирішує обидві ці проблеми:

3. проблему дисбалансу між легкими та складними зразками;
4. проблему дисбалансу між позитивними пікселями (об'єктами) та негативними пікселями (фоном).

## РОЗДІЛ 3 ВИПРОБУВАННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

### 3.1. Набори даних для навчання та валідації моделі

#### 3.1.1. Набір даних Inria

Набір даних Inria Aerial Image Labeling [61] створений для основної задачі дистанційного зондування, а саме автоматичного попиксельного маркування аерофотознімків.

Особливості набору даних:

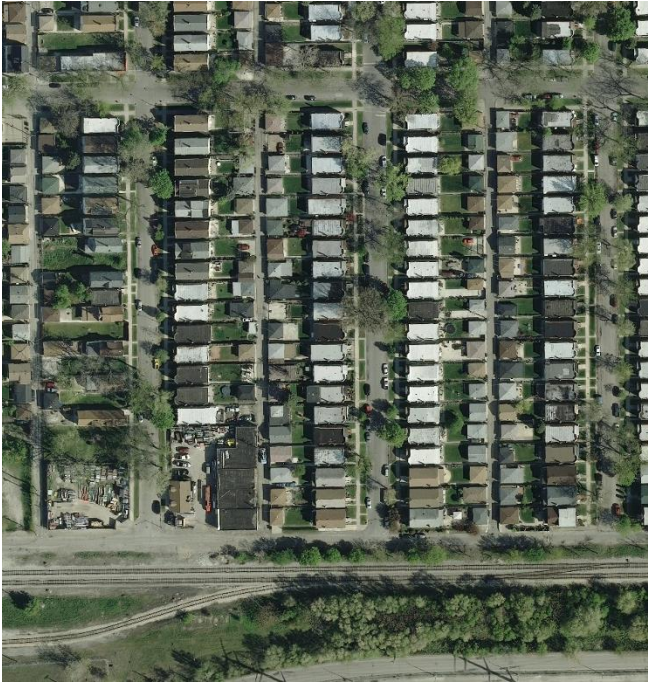
- Покриття 810 км<sup>2</sup> (405 км<sup>2</sup> для навчання та 405 км<sup>2</sup> для тестування);
- Кольоровий аерофотознімок з просторовою роздільною здатністю 0,3 метри на піксель;
- Розмічені дані для двох семантичних класів: будівля і фон (публічно доступні лише для навчальної підмножини).

Зображення охоплюють різні міські поселення, починаючи від густонаселених районів, наприклад, фінансовий район Сан-Франциско, і закінчуючи альпійськими містами, наприклад, Лієнц в австрійському Тіролі.

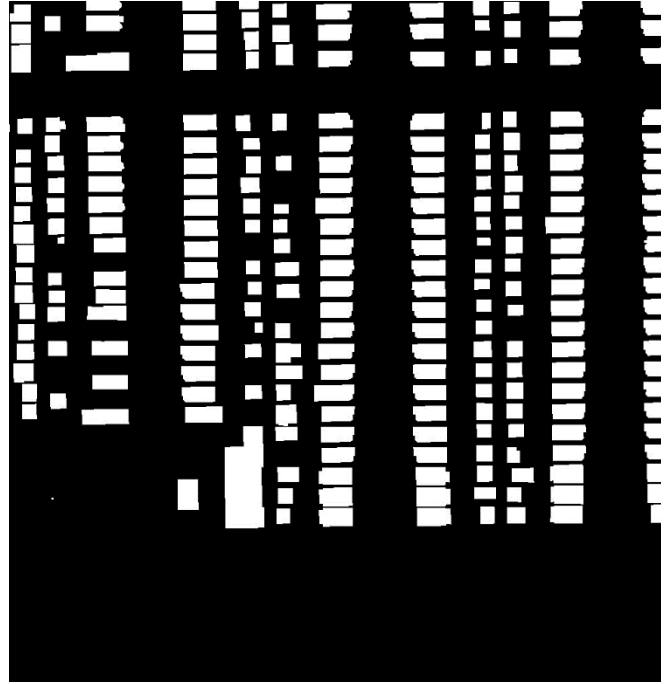
В оригінальному наборі даних замість поділу суміжних частин однакових зображень на навчальні та тестові підмножини, різні міста включені в кожен підмножин. Наприклад, зображення з Чикаго входять до навчального набору (а не до тестового), а зображення з Сан-Франциско включені до тестового набору (а не до навчального).

Кінцевою метою цього набору даних є оцінка узагальнюючої потужності методів: хоча зображення Чикаго можна використовувати для навчання, система повинна навчитися самостійно розмічати аерофотознімки з іншими регіонів, з різними умовами освітлення, з різним ландшафтом та в різні пори року.

Набір даних було створено шляхом об'єднання зображень, що є загальнодоступними, та розміткою будівель, що також знаходяться у відкритому доступі.

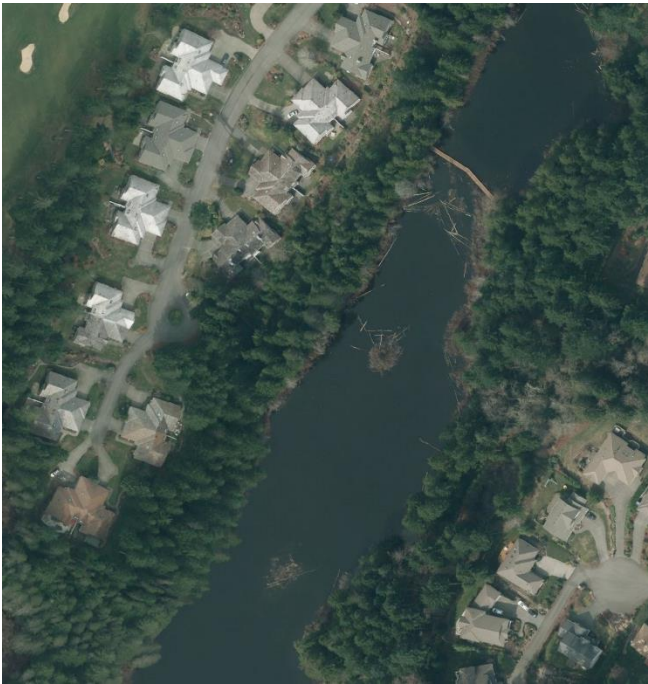


а

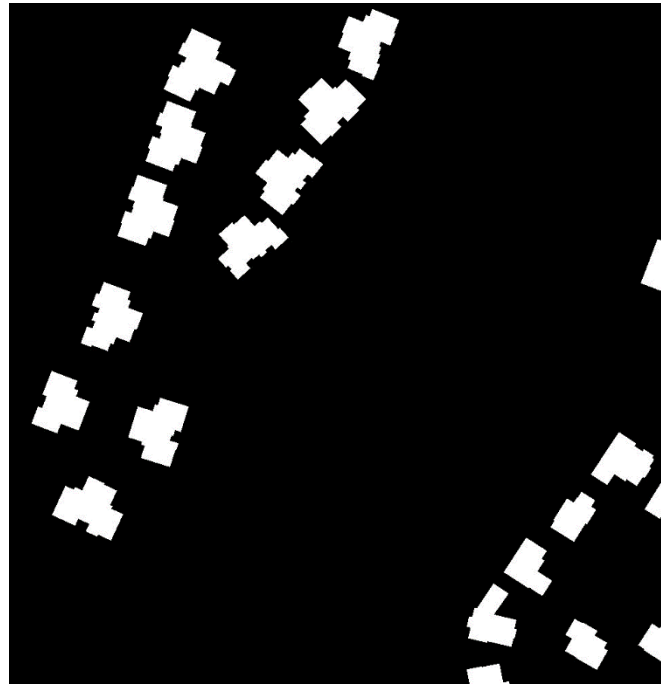


б

Рисунок 3.1 – Чикаго (а – аерофотознімок, б – розмітка)



а



б

Рисунок 3.2 – Кітсеп (а – аерофотознімок, б – розмітка)

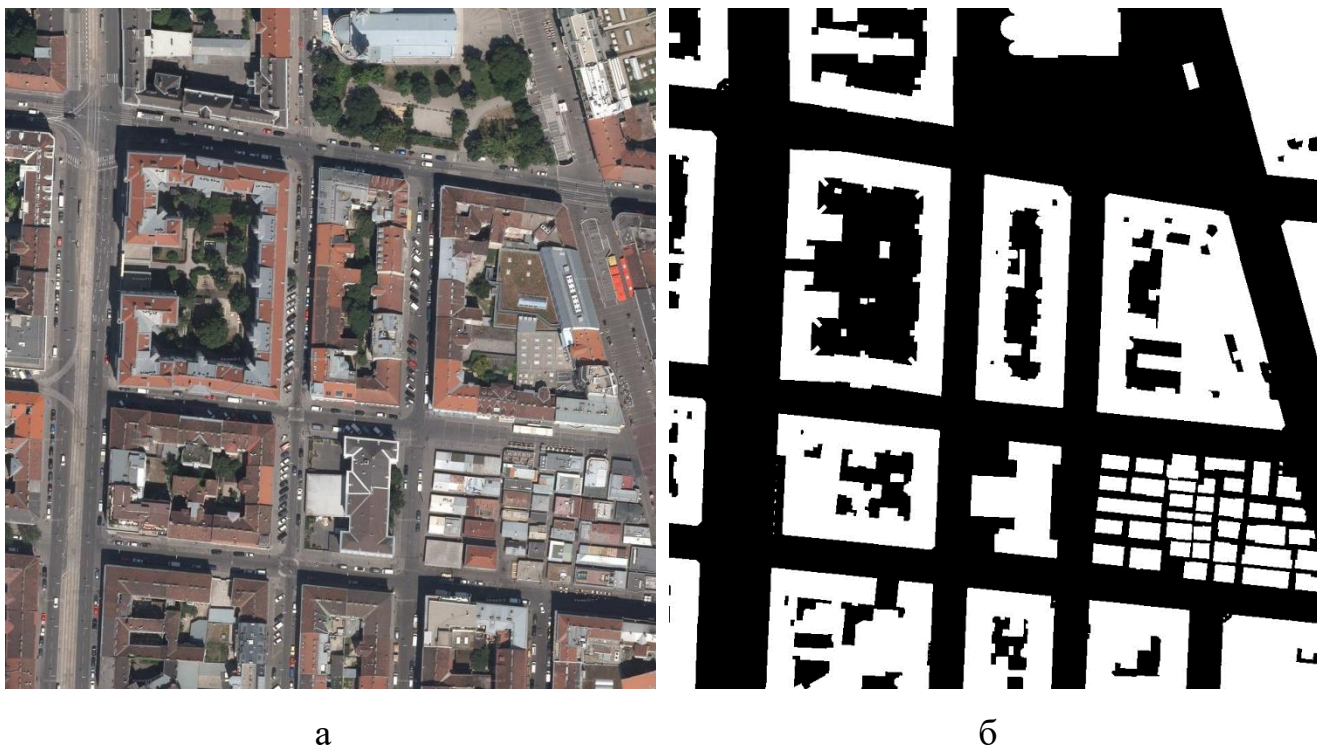


Рисунок 3.3 – Відень (а – аерофотознімок, б – розмітка)

Навчальний набір [62] містить 180 кольорових зображень розміром  $5000 \times 5000$ , які покривають поверхню  $1500 \text{ м} \times 1500 \text{ м}$  кожен (з роздільною здатністю 30 см на піксель). Є 36 плиток для кожного з наступних регіонів:

- Остін
- Чикаго
- Кітсеп
- Західний Тіроль
- Відень

Формат даних — GeoTIFF (TIFF із прив'язкою до координат, але зображення можна використовувати як будь-який інший TIFF). Файли називаються префіксом, пов'язаним із регіоном (наприклад, austin- або vienna-), після якого йде відповідний номер зображення (1-36). Довідкові дані знаходяться в іншій папці, а імена файлів точно відповідають назвам кольорових зображень. Для зображень з тренувального набору даних надано розмітку, що є одноканальними зображеннями зі значеннями 255 для класу будівлі та 0 для класу фону.

В якості валідаційної вибірки пропонують використовувати перші п'ять зображень з кожного регіону (наприклад, `austin{1-5}.tif`, `chicago{1-5}.tif`). Водночас, їх необхідно видалити навчальної вибірки, так як тренувальна та валідаційна вибірка не можуть перетинатися.

Тестова вибірка містить стільки ж зображень, скільки і навчальна (але не міститься відповідної розмітки). Тестова вибірка містить по 36 зображень для кожного з наступних регіонів:

- Беллінгем
- Блумінгтон
- Інсбрук
- Сан Франциско
- Східний Тіроль

В оригінальному конкурсі на основі даного набору даних, якість роботи натренованої нейронної мережі оцінюється за допомогою двох метрик - `Intersection-over-Union` (IoU) та попиксельна точність (`pixel accuracy`) для позитивного класу (будівля). Вони обчислюються як і окремо для кожного регіону (наприклад, Беллінгем, Сан-Франциско), так і для всієї тестової вибірки загалом.

### 3.2. Аугментації даних

Аугментації даних — це набір методів, що покращують розмір і якість навчальних наборів даних для машинного навчання, щоб на них можна було навчати нейронні моделі кращої якості.

Останні досягнення в технології глибокого навчання були зумовлені розвитком архітектур глибоких нейронних мереж, покращенням відеокарт та доступом до великого обсягу даних [63]. Однією з найскладніших проблем є узагальнююча можливість моделей глибокого навчання, яка описує різницю в якості моделі при оцінці на навчальних даних, які ця модель бачила раніше (під час тренування), порівняно з даними, які ця модель ніколи раніше не бачила (тестова вибірка). Моделі з поганою узагальненістю перенавчалися на тренувальній вибірці - через велику

кількість параметрів моделі (до кількох мільйонів), такі нейронні мережі можуть просто “запам’ятати” тренувальні дані замість того, аби вивчати певні закономірності і паттерни.

Щоб побудувати дійсно якісні моделі глибокого навчання, аугментації даних є дуже потужним методом запобігання перенавчанню, генеруючи більш повний і різноманітний набір можливих зображень, щоб мінімізувати відстань між вибірками для навчання та тестування.

Ранні експерименти, що демонструють ефективність від аугментації даних, походять від простих трансформацій зображень, наприклад, горизонтального перевертання, збільшення колірному простору та випадкового масштабування.

Для сучасних нейронних мереж для класифікації / сегментації / детекції зображень використовують різноманітні типи аугментацій:

- Геометричні перетворення: збільшення набору даних за допомогою розвороту по горизонталі або вертикалі, довільного кадрування та/або масштабування, повороту на випадковий кут, переносу для зміщення зображень ліворуч/праворуч/вгору/вниз.



а



б



в



г

Рисунок 3.4 – Геометричні аугментації. Розворот, транспонування (а – оригінальне зображення, б – горизонтальний розворот, в – вертикальний розворот, г – транспонування)

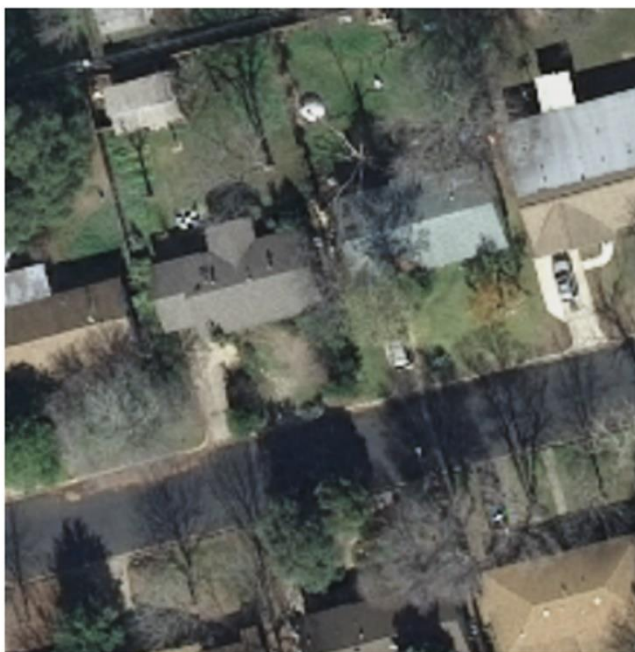


а



б





в



г

Рисунок 3.5 – Геометричні аугментації. Поворот, масштабування, перенесення (а – поворот на 10 градусів, б – поворот на 45 градусів, в – масштабування, г – горизонтальне та вертикальне перенесення)

- Введення шуму: випадковий шум, Гаусівський шум. Для аугментації даних застосовують навіть більш прикладні перетворення, такі як додавання дощу, туману, снігу, сонячних засвітів, тощо.



а



б



в



г



Рисунок 3.6 – Аугментації шуму. Гаусів шум, сонячні засвіти, дощ (а – оригінальне зображення, б – гаусів шум, в – сонячні засвіти, г – дощ)

- Спотворення кольору включає зміну яскравості, відтінку або насиченості зображень. Зміна розподілу кольорів або маніпулювання гистограмою колірних каналів RGB (червоний, зелений, синій) використовується для підвищення стійкості моделі до зміни освітлення.



а



б



в



г

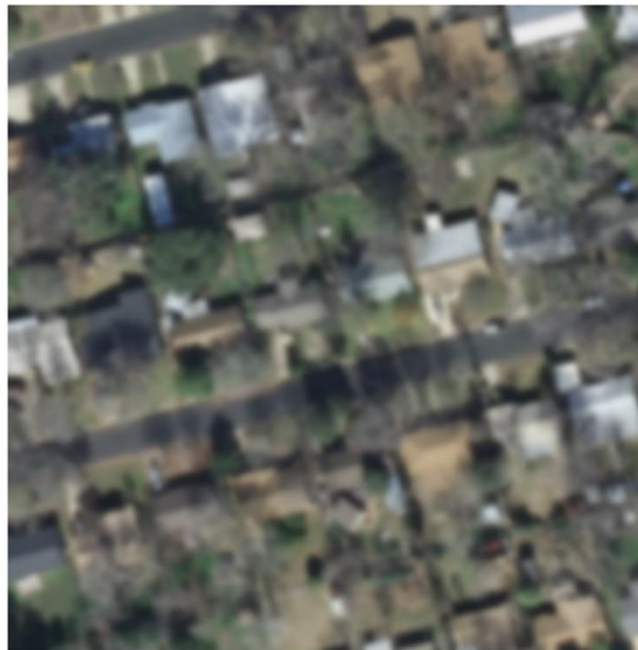
Рисунок 3.7 – Аугментації спотворення кольору (а – яскравість, б – контраст, в – RGB канали, г – насиченість)

- Згорткові фільтри використовують методи обробки зображень для підвищення різкості та розмивання зображень. Ці методи спрямовані на підвищення стійкості моделі до більш деталізованих зображень (щоб не перетреноувати модель

для нових камер) або підвищення стійкості до розмиття руху (для кращої роботи з відео).



а



б



в



г

Рисунок 3.8 – Застосування згорткових фільтрів. Розмиття, різкість (а – оригінальне зображення, б – гаусове розмиття, в – розмиття руху, г – різкість)

- Змішування зображень застосовує методи змішування різних зображень

шляхом усереднення їхніх піксельних значень для кожного каналу RGB або за допомогою випадкового кадрування та виправлення зображень. Не зважаючи на те, що цей метод суперечить інтуїції, він показав ефективність у підвищенні якості роботи моделей.

- Видалення (замальовування) певної частини інформації використовується для підвищення стійкості до оклюзії об'єктів (перекривання частини об'єкта іншим об'єктом, в нашому випадку перекривання будинків деревами, стовпами, лініями електропередач) при розпізнаванні та сегментації зображень, що призводить до помітного підвищення якості роботи моделі для прикладних задач.



а



б



в

Рисунок 3.9 – Видалення частини зображення (а – оригінальне зображення, б – видалення областей зображення (розмір=15 пікселів, кількість=5), в – видалення областей зображення (розмір=50 пікселів, кількість=3))

Аугментації даних для нейронних мереж комп'ютерного зору почали застосовувати ще з найперших експериментів та наукових статей. В статті 1998-го року LeNet-5 [64] відомого дослідника Яна ЛеКуна вже використовувались такі базові аугментації, як горизонтальна і вертикальна трансляція, масштабування (по одній або двом вісям), та поворот.

Такі відомі статті, як ImageNet та U-Net теж застосовували аугментації до навчальних наборів даних. Більш того, аугментації даних були однією з основних причин високої продуктивності роботи мережі U-Net. Автори цієї статті змогли натренувати мережу використовуючи всього 30 медичинських зображень розміром 512x512 використовуючи велику кількість різноманітних аугментацій.

### 3.3. Експерименти

#### 3.3.1. Кодувальники

Ми почнемо наші експерименти з пошуку найкращого кодувальника для даної задачі та заданого набору даних.

Для цього нам необхідно зафіксувати інші важливі параметри моделі:

- Архітектура: U-Net [4]
- Функція втрат: фокальна функція втрат
- Роздільна здатність вхідного зображення: 256x256
- Метрика оцінки якості: IoU
- Оптимізатор: Адам
- Аугментації: всі, перераховані в підрозділі 3.2

Ми будемо проводити експерименти з різними архітектурами та функціями втрат в наступних підрозділах

В цьому розділі ми проведимо порівняння між наступними кодувальниками:

- ResNet [46]
- MobileNet [51]
- EfficientNet [49]

Найперший експеримент було проведено з мережею з сімейства моделей ResNet [46], а саме ResNet-18 так як ця мережа і досі є гарною відправною точкою для більшості експериментів комп'ютерного зору.

В подальшому ми порівнюватимемо наступні показники якості роботи моделі:

- значення функції втрат на тренувальній вибірці
- значення функції втрат на валідаційній вибірці
- значення метрики на валідаційній метриці
- візуальні результати

Для найпершого порівняння ми також натренували нейронну мережу з кодувальником з сімейства моделей MobileNet [50-52], а саме MobileNet-V2 [51], що є порівняно малою архітектурою (2 мільйони параметрів проти 11 мільйони параметрів в мережі ResNet-18) розробленої для підвищення швидкості глибоких нейронних мереж на мобільних пристроях.

Слід зазначити, що для графіків функції втрат менші значення відповідають кращим результатам, так як ми намагаємось оптимізувати (мінімізувати) цю функцію. Для графіків метрики - навпаки: більші значення відповідають кращим результатам (чим більший відсоток перекриття між істинною розміткою та прогнозованими результатами - тим краще працює модель).

Значення метрики на тренувальній вибірці зазвичай не досліджуються взагалі, так як нас цікавлять результати роботи нейронної мережі саме на валідаційній вибірці (на тій підмножині даних які модель “не бачила” під час навчання). Проте значення функції втрат на тренувальній вибірці все ж таки досліджується. Це робиться насамперед для того, щоб з'ясувати, чи відбувається перенавчання (over-fitting) моделі на тренувальних даних (коли мережа ”запам’ятовує” тренувальні зображення замість того, щоб узагальнювати ознаки та патерни).

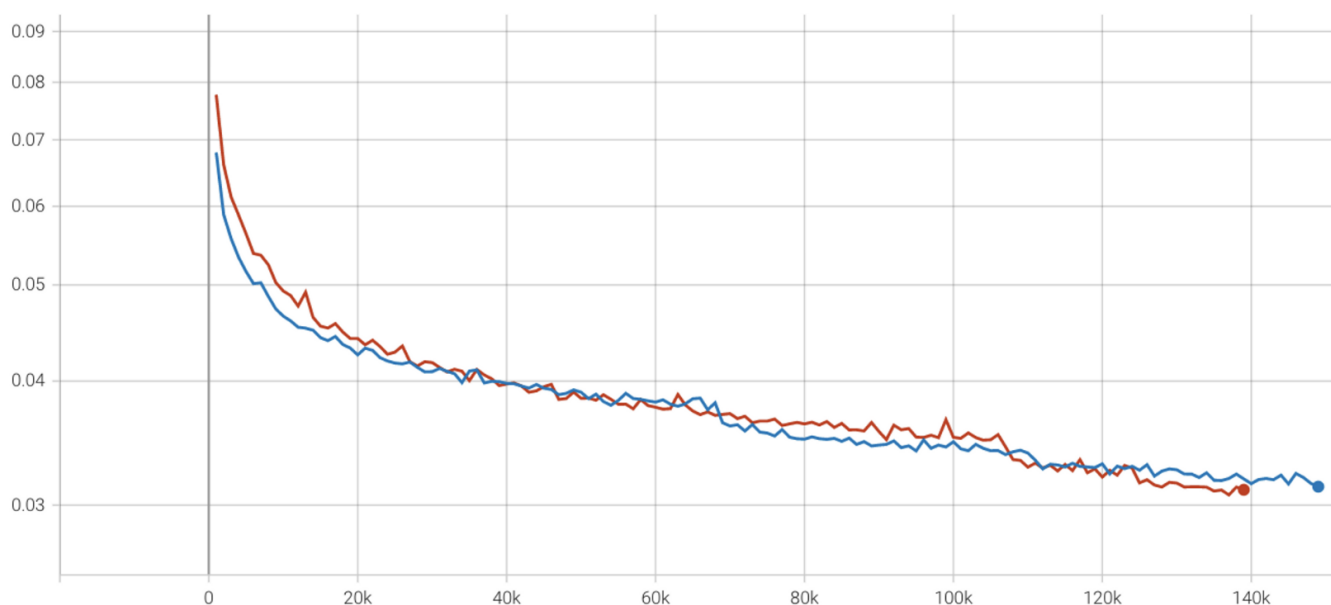


Рисунок 3.10 – Залежність функції втрат на тренувальній вибірці від номеру ітерації.  
ResNet-18 (синій) проти MobileNet-V2 (червоний)



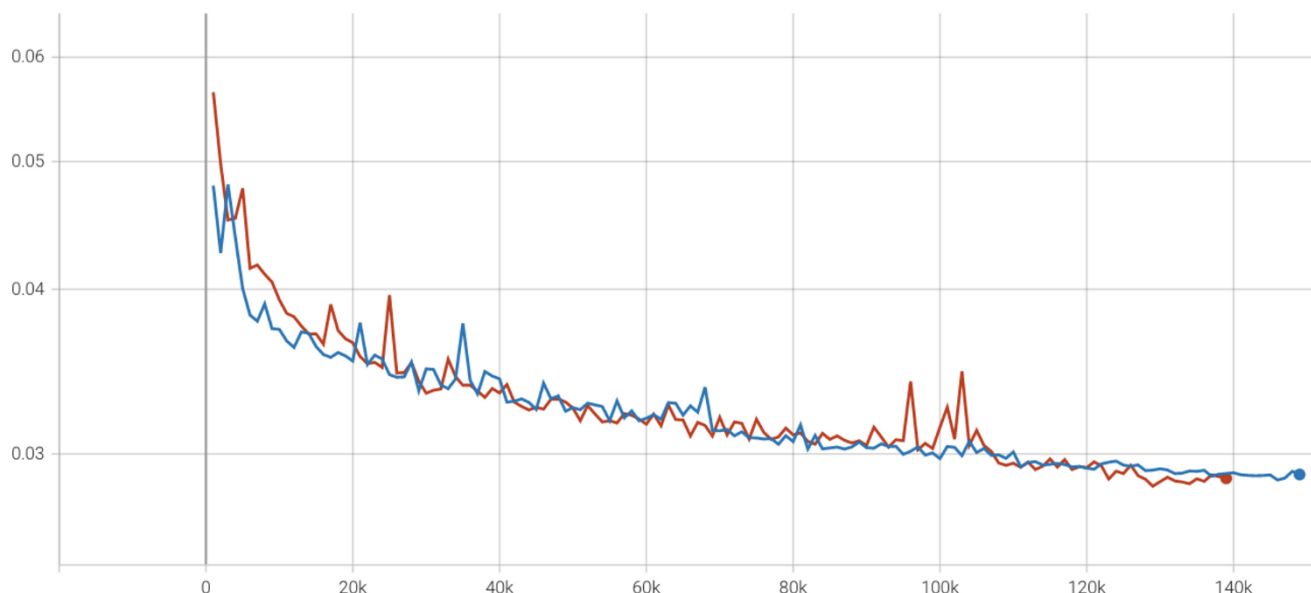


Рисунок 3.11 – Залежність функції втрат на валідаційній вибірці від номеру ітерації.  
ResNet-18 (синій) проти MobileNet-V2 (червоний)

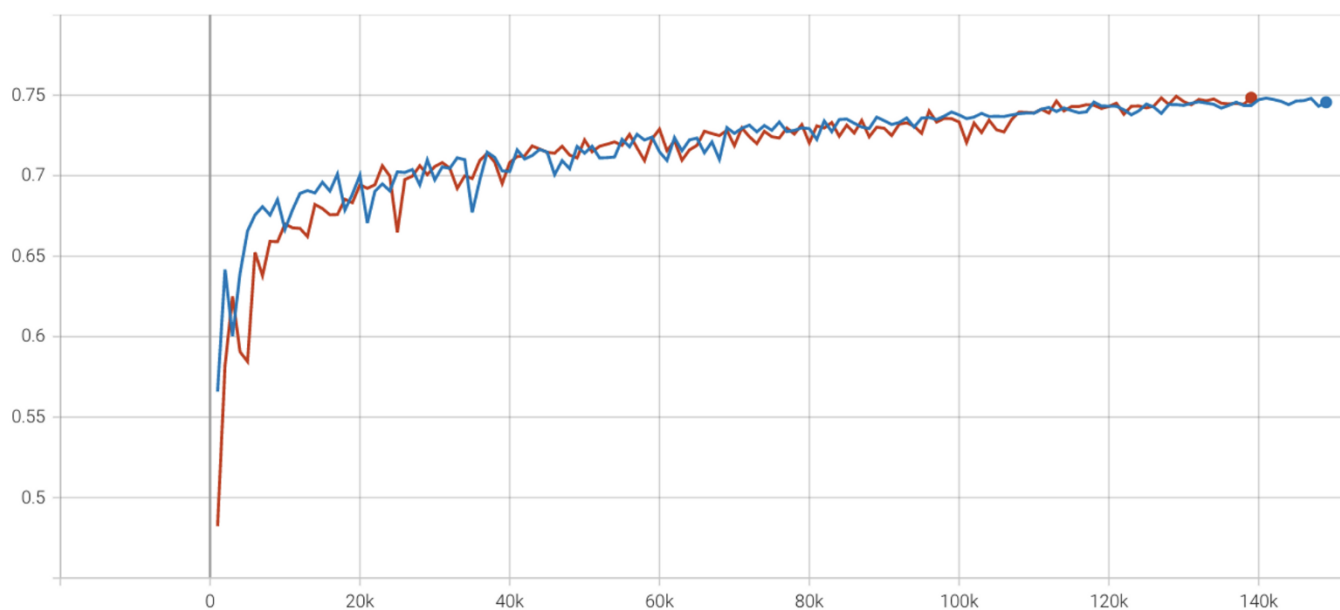


Рисунок 3.12 – Залежність метрики на валідаційній вибірці від номеру ітерації.  
ResNet-18 (синій) проти MobileNet-V2 (червоний)

Не зважаючи на великий розрив в кількості параметрів (різниця в 5.5 разів), згідно з усіма графіками, обидві моделі працюють майже ідентично. Значення метрики якості для кодувальників ResNet-18 та MobileNet-V2 на валідаційній вибірці **74.8%** та **74.9%** відповідно. То ж в подальших експериментах немає сенсу використовувати кодувальник ResNet-18, що містить набагато більшу кількість

параметрів, а отже і процес тренування, і процес виводу з використанням даної мережі займає більшу кількість часу.

В той же час, для кодувальника ResNet-18, в точці локального мінімуму значення функції втрат на валідаційній вибірці менше ( а отже краще) ніж значення функції втрат на тренувальній вибірці **0.02835** проти **0.03072**. Це може свідчи про недо-навчання (under-fitting) моделі, коли в усій нейронній мережі, або в кодувальнику недостатньо ємності для того щоб узагальнити увесь тренувальний набір даних.

Тому наступним кроком є використання більшої (з точки зору кількості параметрів) та кращої (новішої та потужнішої) мережі, а саме EfficientNet [49]. Ми розглянули в підрозділі 2.2.2 EfficientNet, що даний підхід використовує метод збалансованого масштабування в трьох вимірах — ширині мережі, глибині мережі та роздільній здатності вхідного зображення. Так як розмір вхідного зображення для всіх наших експериментів фіксований, ми можемо скористатися наступною таблицею відповідностей між роздільними здатностями вхідного зображення та розміром (версією) мережі EfficientNet (табл. 3.1).

Таблиця 3.1

Відповідність між роздільними здатностями вхідного зображення та розміром (версією) мережі EfficientNet:

Розмір мережі EfficientNet	Роздільна здатність вхідного зображення
EfficientNetB0	224
EfficientNetB1	240
EfficientNetB2	260
EfficientNetB3	300
EfficientNetB4	380
EfficientNetB5	456
EfficientNetB6	528

EfficientNetB7

600

В нашому випадку (роздільна здатність вхідного зображення = 256x256) найближче табличне значення - 260, отже оптимальним буде використовувати модель EfficientNetB2

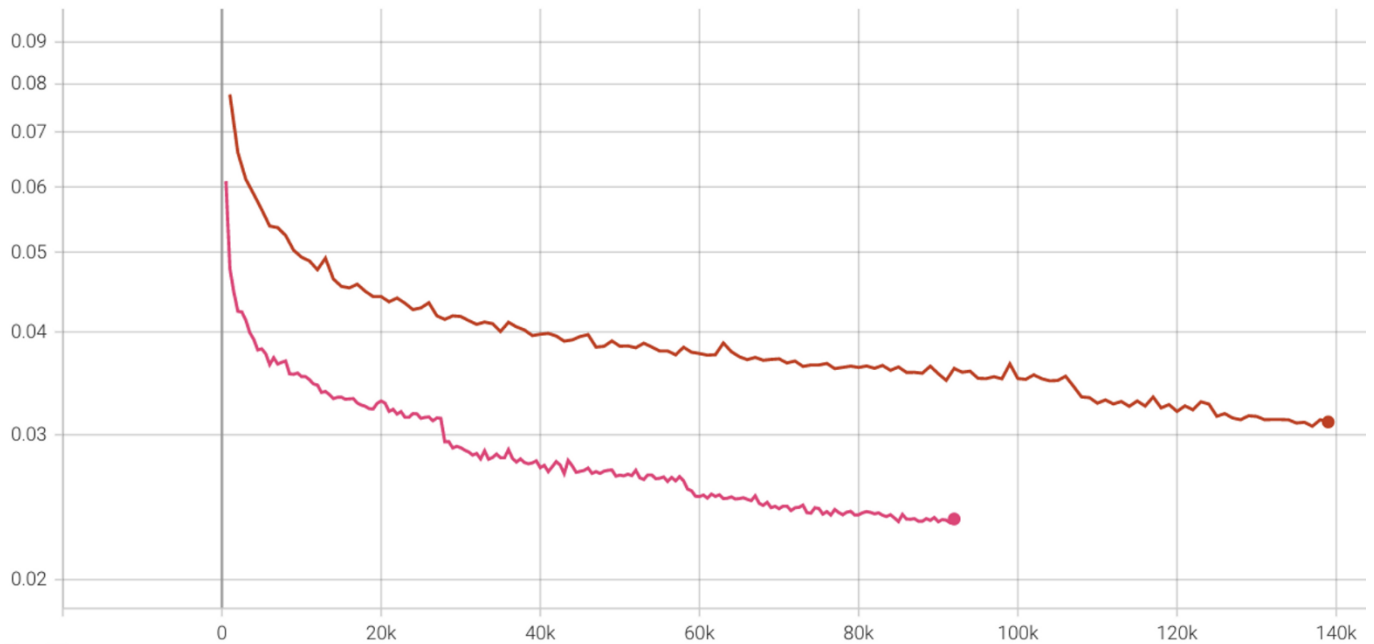


Рисунок 3.13 – Залежність функції втрат на тренувальній вибірці від номеру ітерації. EfficientNetB2 (рожевий) проти MobileNet-V2 (червоний)

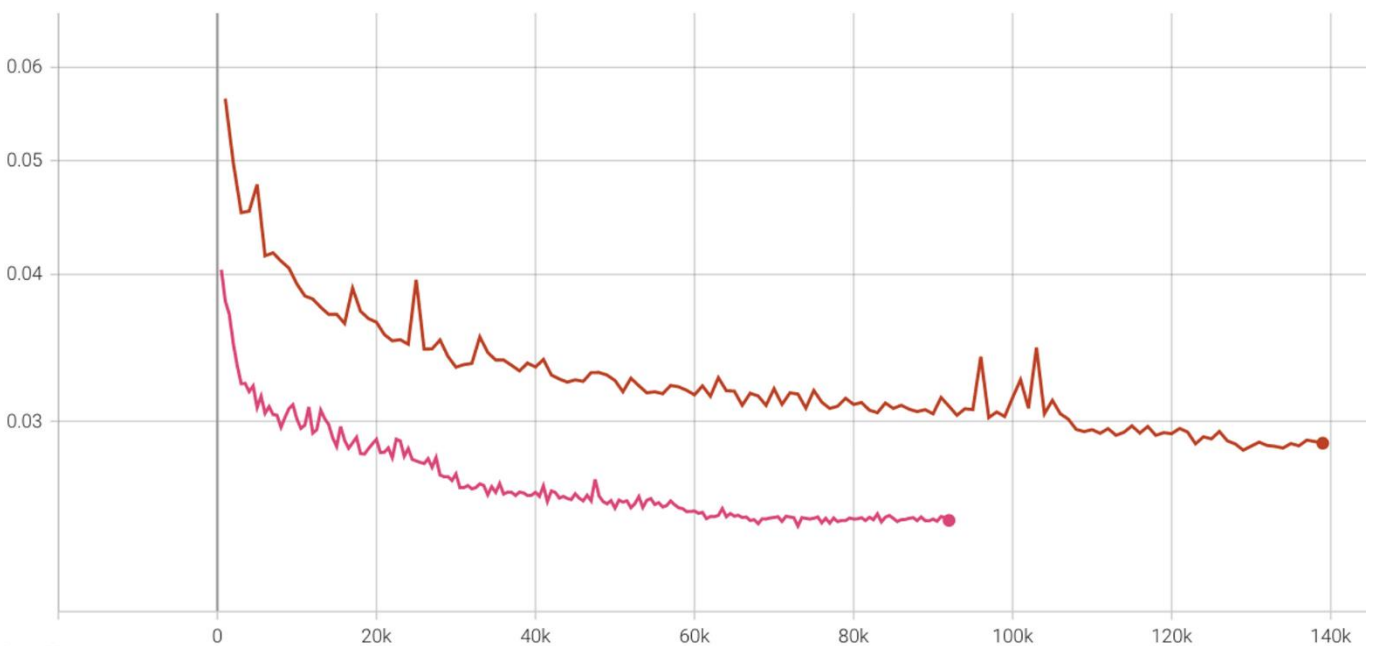


Рисунок 3.14 – Залежність функції втрат на валідаційній вибірці від номеру ітерації.

## EfficientNetB2 (рожевий) проти MobileNet-V2 (червоний)

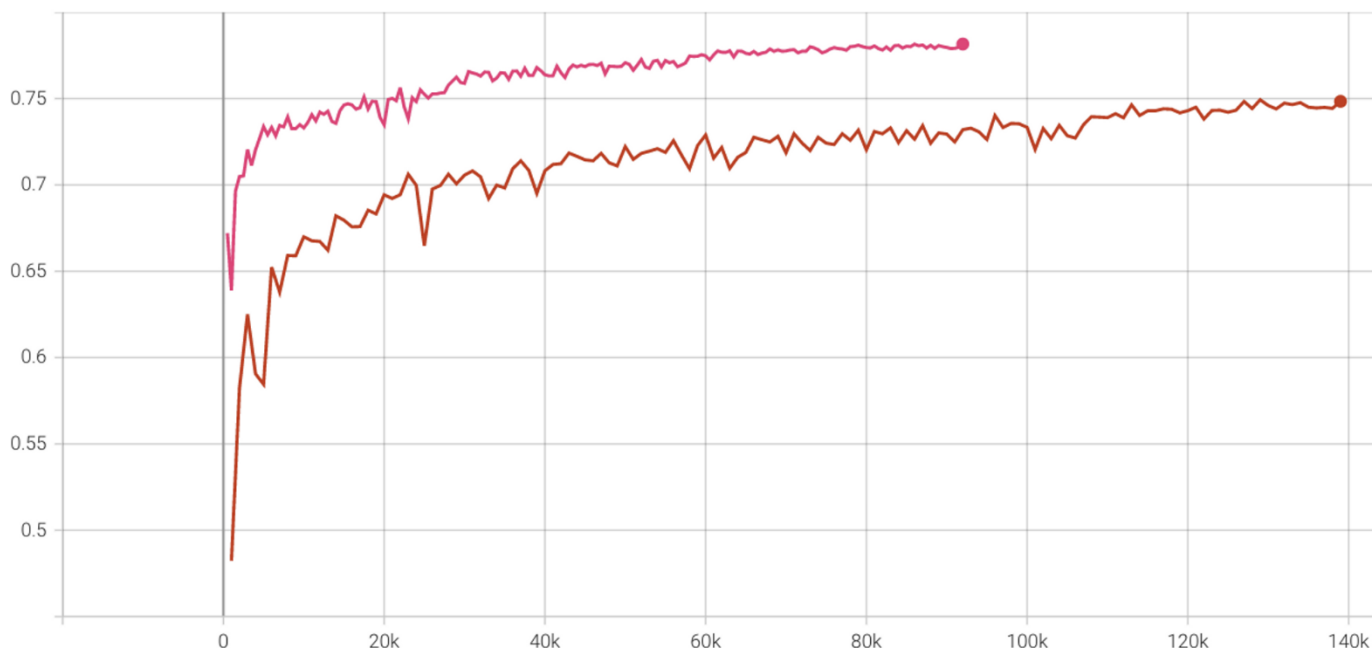


Рисунок 3.15 – Залежність метрики на валідаційній вибірці від номеру ітерації.

## EfficientNetB2 (рожевий) проти MobileNet-V2 (червоний)

Як видно з графіків вище архітектура нейронної мережі U-Net з кодувальником EfficientNetB2 значно переважає мережу U-Net з кодувальником MobileNet-V2 за усіма показниками, що розглядаються. За основною метрикою оцінки якості (IoU) кодувальник EfficientNetB2 кращий за MobileNet-V2 на **3.3%** (**78.2%** проти **74.9%**). Отже для всіх наступних експериментів ми використовуватимемо саме кодувальник EfficientNetB2.

Повний порівняльний аналіз кодувальників наведено в табл. 3.2.

Таблиця 3.2

Порівняльний аналіз кодувальників ResNet-18, MobileNet-V2, та EfficientNetB2

Кодувальник	кількість параметрів (мільйони)	функції втрат ↓ (валідаційна вибірка)	метрика ↑ (валідаційна вибірка)
ResNet-18	11	0.029	74.8%
MobileNet-V2	2	0.028	74.9%
EfficientNetB2	7	<b>0.025</b>	<b>78.2%</b>

Візуальний порівняльний аналіз кодувальників ResNet-18, MobileNet-V2, та EfficientNetB2 зображено на рис. 3.16.

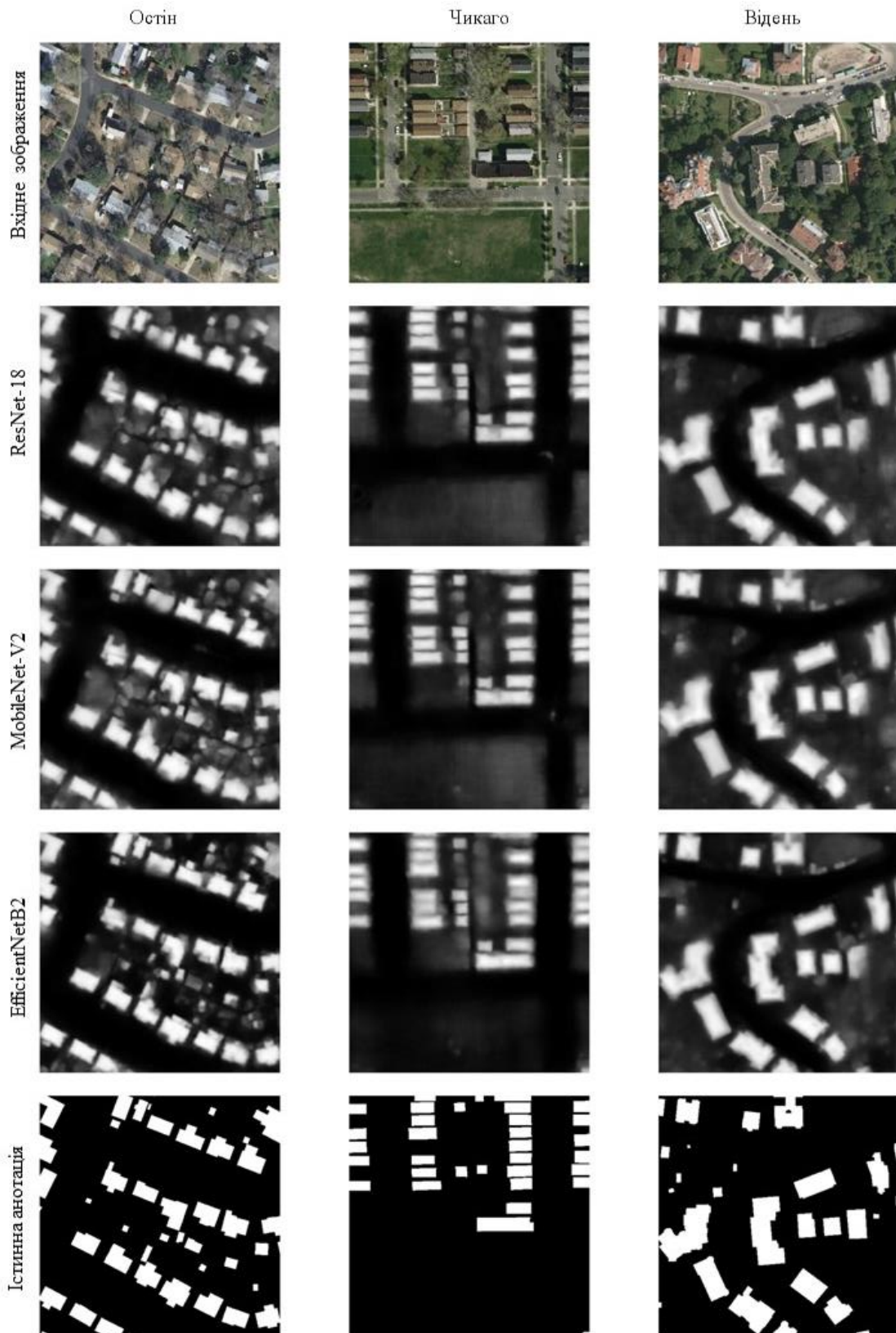
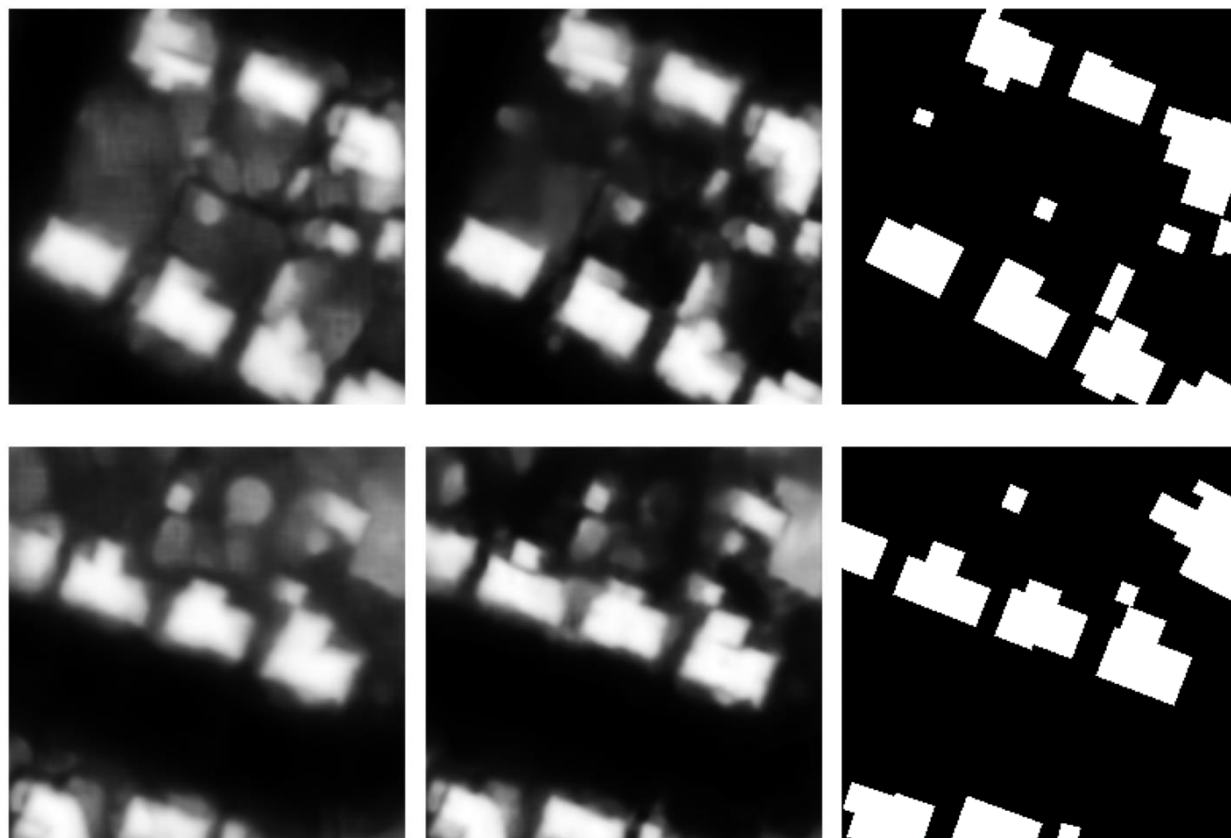


Рисунок 3.16 – Візуальний порівняльний аналіз кодувальників ResNet-18, MobileNet-V2, та EfficientNetB2

На рис. 3.17 представлено детальніші результати роботи моделей MobileNet-V2

та EfficientNetB2 для першого зображення (м. Остін).



MobileNet-V2

EfficientNetB2

істинна анотація

Рисунок 3.17 – Детальніші результати роботи моделей MobileNet-V2 та EfficientNetB2 для зображення з Остіну

Як можна побачити з результатів, кодувальник EfficientNetB2 не тільки досягає кращих кількісних результатів (оцінюючи метрики і значення функції втрат) а і кращих візуальних результатів. Результати, передбачені мережею U-Net з кодувальником EfficientNetB2, містять менше шумів та більш впевнені маски об'єктів.

### 3.3.2. Архітектури

Продовжимо наші експерименти з пошуку найкращої архітектури глибокої нейронної мережі для даної задачі та заданого набору даних.

Для цього нам необхідно зафіксувати інші важливі параметри моделі:

- Кодувальник: EfficientNetB2 [49] (обраний в попередньому підрозділі)

- Функція втрат: фокальна функція втрат
- Роздільна здатність вхідного зображення: 256x256
- Метрика оцінки якості: IoU
- Оптимізатор: Адам
- Аугментації: всі, перераховані в підрозділі 3.2

В цьому розділі ми проведимо порівняння між наступними кодувальниками:

- U-Net [4]
- U-Net++ [23]

Як і в попередньому підрозділі, ми порівнюватимемо наступні показники якості роботи моделі:

- значення функції втрат на тренувальній вибірці
- значення функції втрат на валідаційній вибірці
- значення метрики на валідаційній метриці
- візульні результати

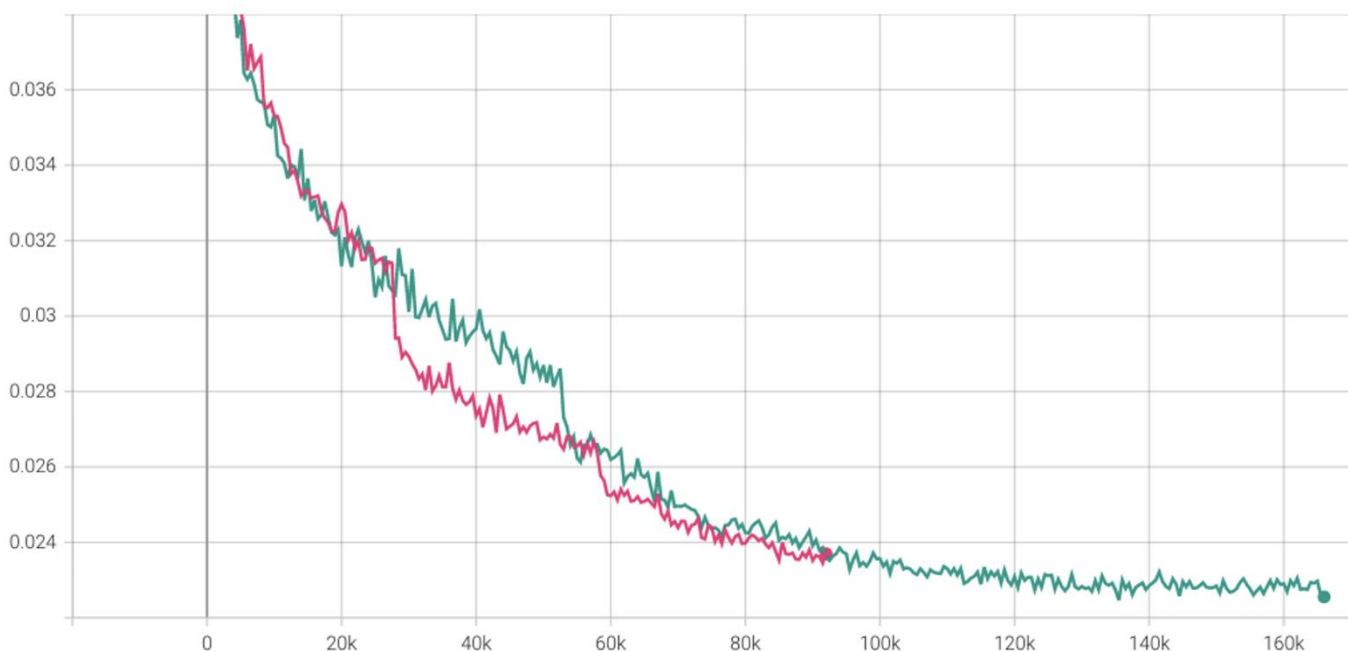


Рисунок 3.18 – Залежність функції втрат на тренувальній вибірці від номеру ітерації.

U-Net++ EfficientNetB2 (зелений) проти U-Net EfficientNetB2 (рожевий)



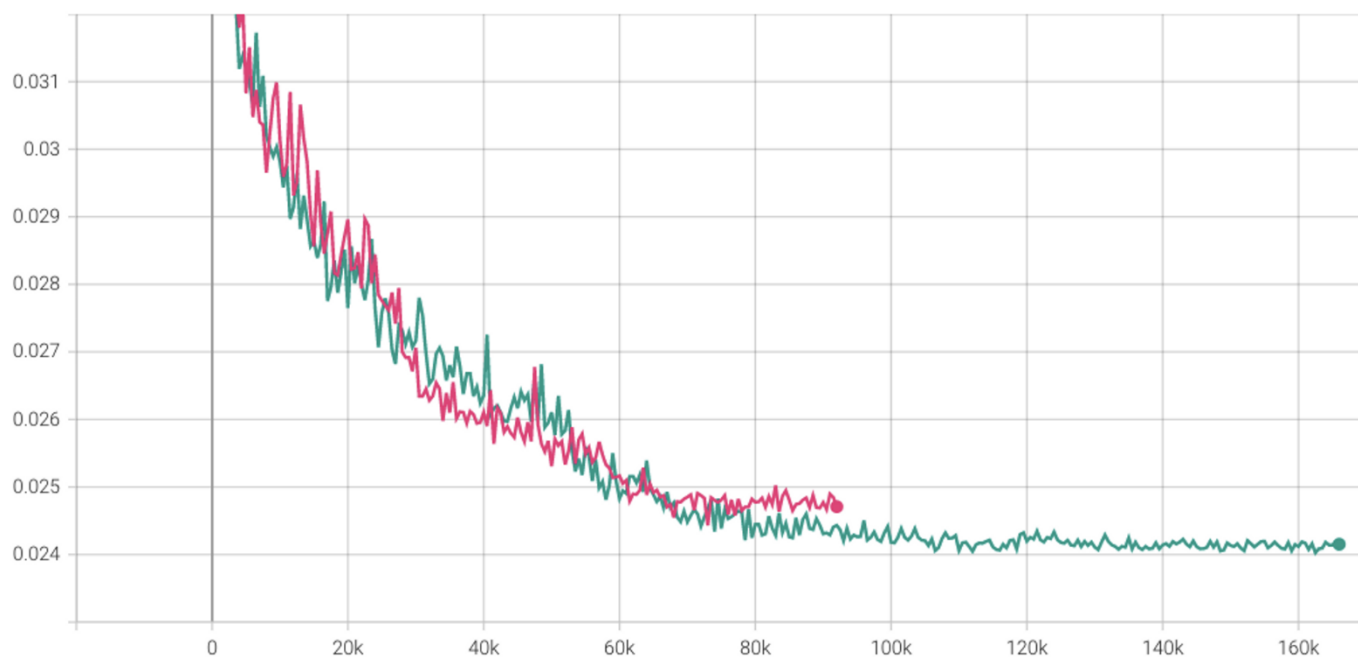


Рисунок 3.19 – Залежність функції втрат на валідаційній вибірці від номеру ітерації. U-Net++ EfficientNetB2 (зелений) проти U-Net EfficientNetB2 (рожевий)

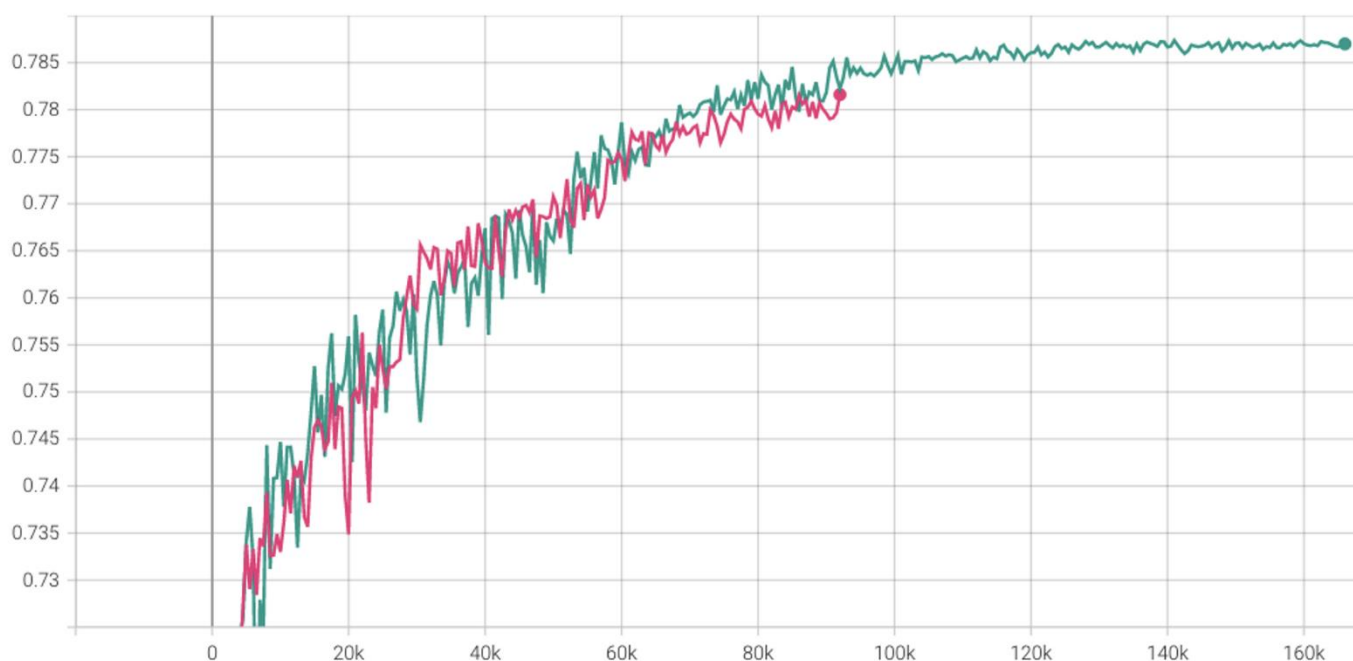


Рисунок 3.20 – Залежність метрики на валідаційній вибірці від номеру ітерації. U-Net++ EfficientNetB2 (зелений) проти U-Net EfficientNetB2 (рожевий)

Як видно з графіків вище архітектура нейронної мережі U-Net++ [23] з кодувальником EfficientNetB2 [49] переважає мережу U-Net [4] з тим самим

кодувальником за усіма показниками, що розглядаються. За основною метрикою оцінки якості (IoU) мережа U-Net++ краща за U-Net на **0.5%** (**78.7%** проти **78.2%**).

Окрім дого, зміна базової архітектури нейронної мережі дозволила тренувати модель більш ніж *вдвічі* довше (**15:53** проти **6:47** годин) довше завдяки більшій кількості параметрів (**10.4M** проти **10M**). Отже для всіх наступних експериментів ми використовуватимемо саме архітектуру U-Net++.

Повний порівняльний аналіз архітектур нейронних мереж наведено в табл. 3.3.

Таблиця 3.3

Повний порівняльний аналіз архітектур нейронних мереж

Архітект ура	Загальна кількість параметрів (мільйони)	Кодувальник	кількість параметрів (мільйони)	Триваліст ь тренуван ня моделі	функції втрат ↓ (валідаційн а вибірка)	метрика ↑ (валідаційн а вибірка)
U-Net	10	EfficientNet B2	7	6:47	0.025	78.2%
U-Net++	10,4	EfficientNet B2	7	15:53	0.024	78.7%

Візуальний порівняльний аналіз архітектур нейронних мереж U-Net та U-Net++ зображено на рис. 3.21.

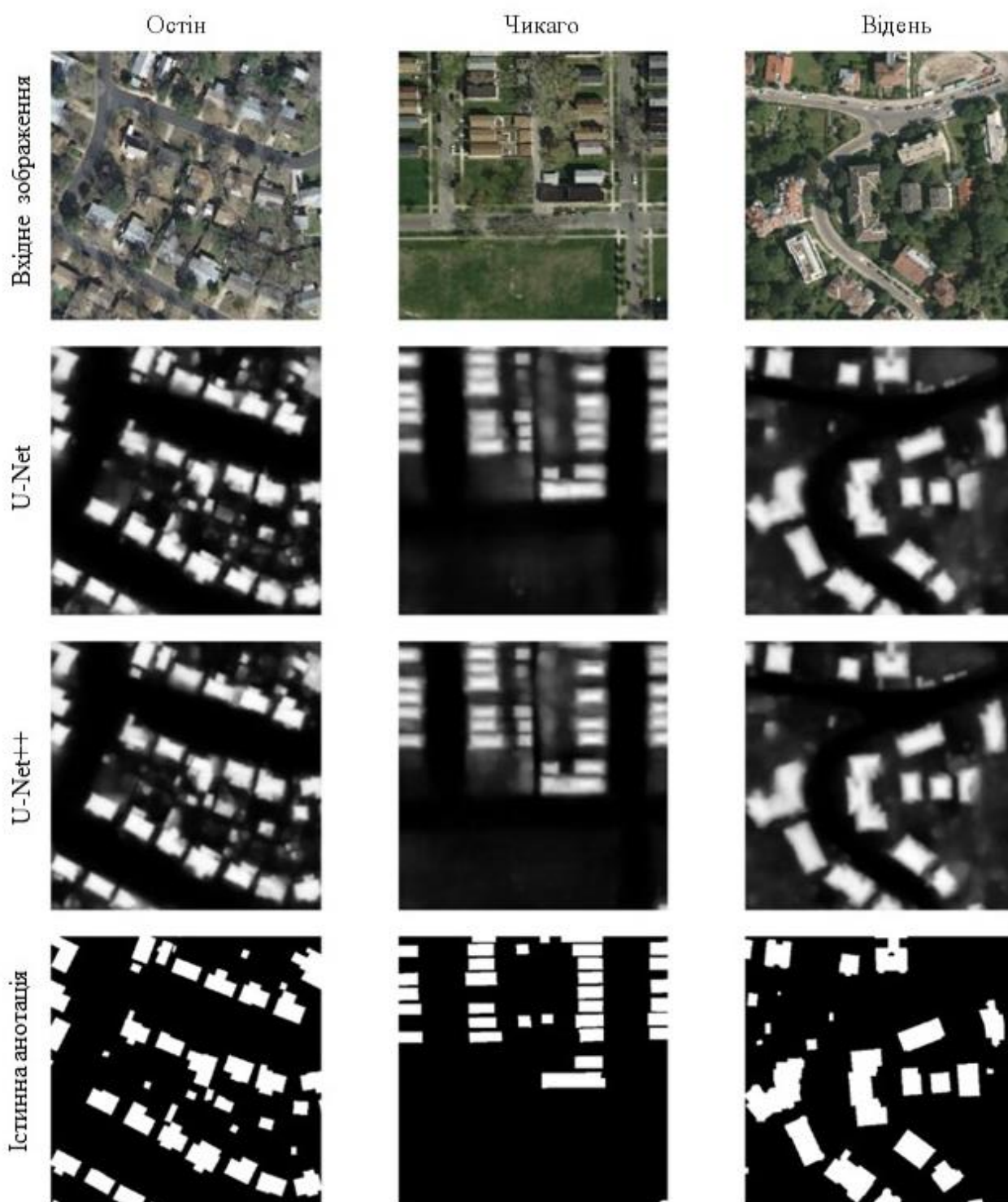


Рисунок 3.21 – Візуальний порівняльний аналіз архітектур нейронних мереж U-Net та U-Net++

На рис. 3.22 представлено детальніші результати роботи архітектур U-Net та U-Net++ для другого зображення (м. Чикаго).

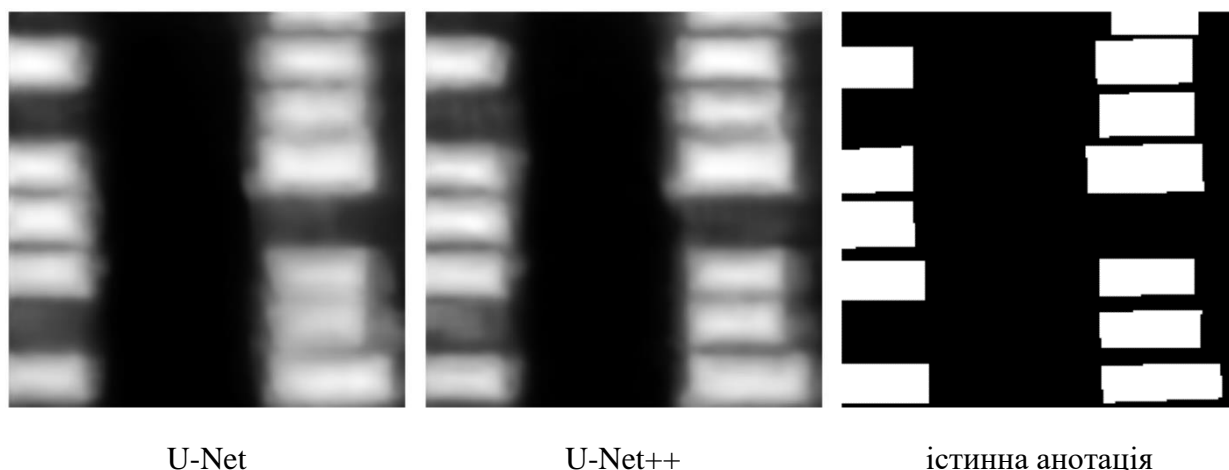


Рисунок 3.22 – Детальніші результати роботи архітектур U-Net та U-Net++ для зображення з Чикаго

Як можна побачити з результатів, архітектура нейронної мережі U-Net++ не тільки досягає кращих кількісних результатів (оцінюючи метрики і значення функції втрат) а і кращих візуальних результатів у порівнянні з архітектурою. Результати, передбачені мережею U-Net з кодувальником EfficientNetB2, містять чіткіші межі масковок об'єктів. Об'єкти, що знаходяться ближче один до одного розділені краще.

### 3.3.3. Функції втрат

Останньою серією експериментів буде підбір функції втрат для задачі сегментації об'єктів на супутникових знімках. В деяких задачах комп'ютерного зору підбір правильної або оптимальної функції втрат є визначним фактором для побудови якісної моделі глибокого навчання.

Для цього нам необхідно зафіксувати інші важливі параметри моделі:

- Архітектура: U-Net++ [23] (обрана в попередньому підрозділі)
- Кодувальник: EfficientNetB2 [49] (обраний в попередніх підрозділах)
- Роздільна здатність вхідного зображення: 256x256
- Метрика оцінки якості: IoU
- Оптимізатор: Адам
- Аугментації: всі, перераховані в підрозділі 3.2

В цьому розділі ми проведимо порівняння між наступними кодувальниками:

- Фокальна функція втрат [37] (Focal Loss)
- Бінарна перехресна ентропія [59] + функція втрат Дайса [16] (BCE+Dice Loss)

В цьому розділі ми порівнюватимемо виключно значення метрики на валідаційній метриці. Ми не можемо порівнювати значення функції втрат, так як функції втрат для цього експерименту різні, і до того ж вони мають різні області значень, які зображені на рис. 3.23 та на рис. 3.24.

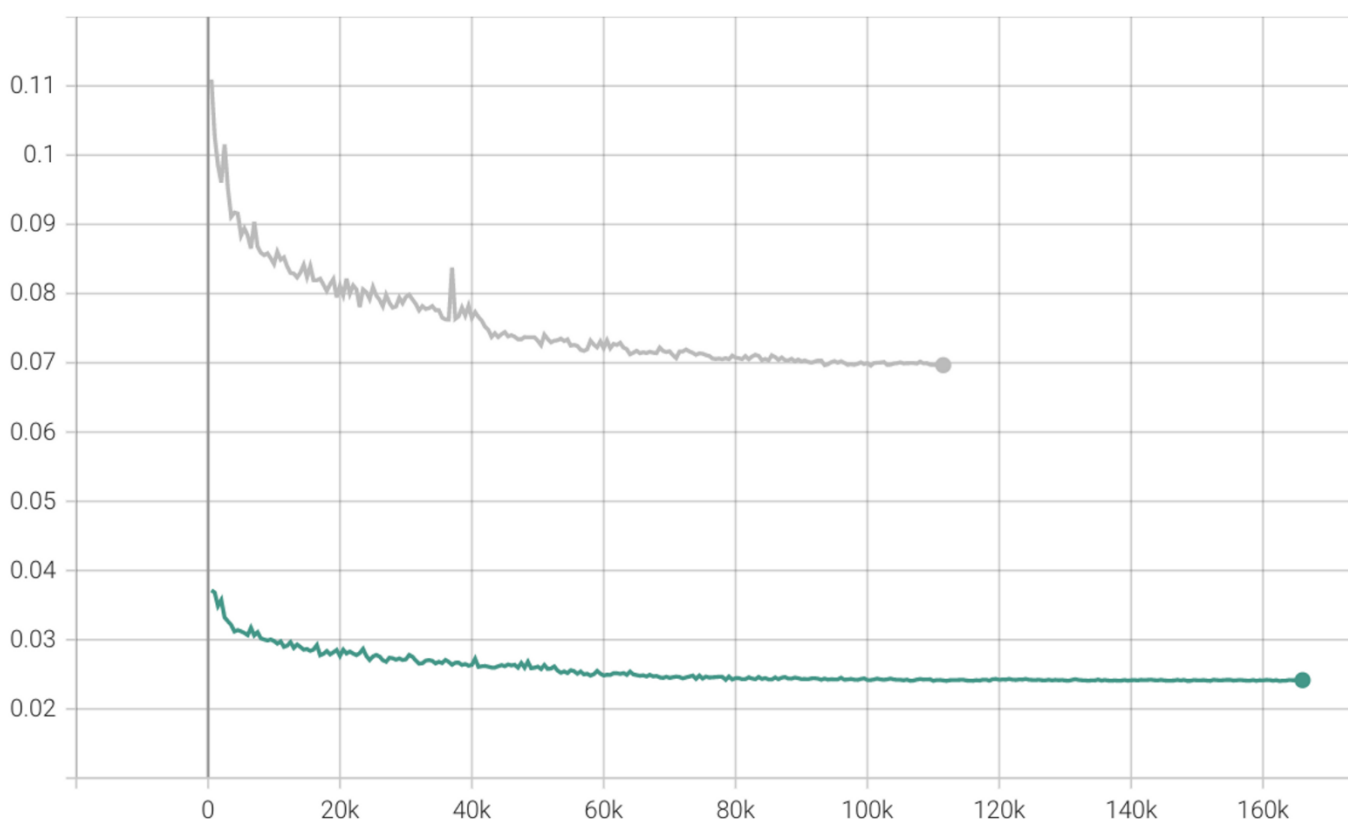


Рисунок 3.23 – Залежність функції втрат на валідаційній вибірці від номеру ітерації.

Фокальна функція втрат (зелений) проти Бінарна перехресна ентропія + функція втрат Дайса (сірий). Графік не є порівнянням, а ілюструє відмінність областей значень двох різних функцій втрат

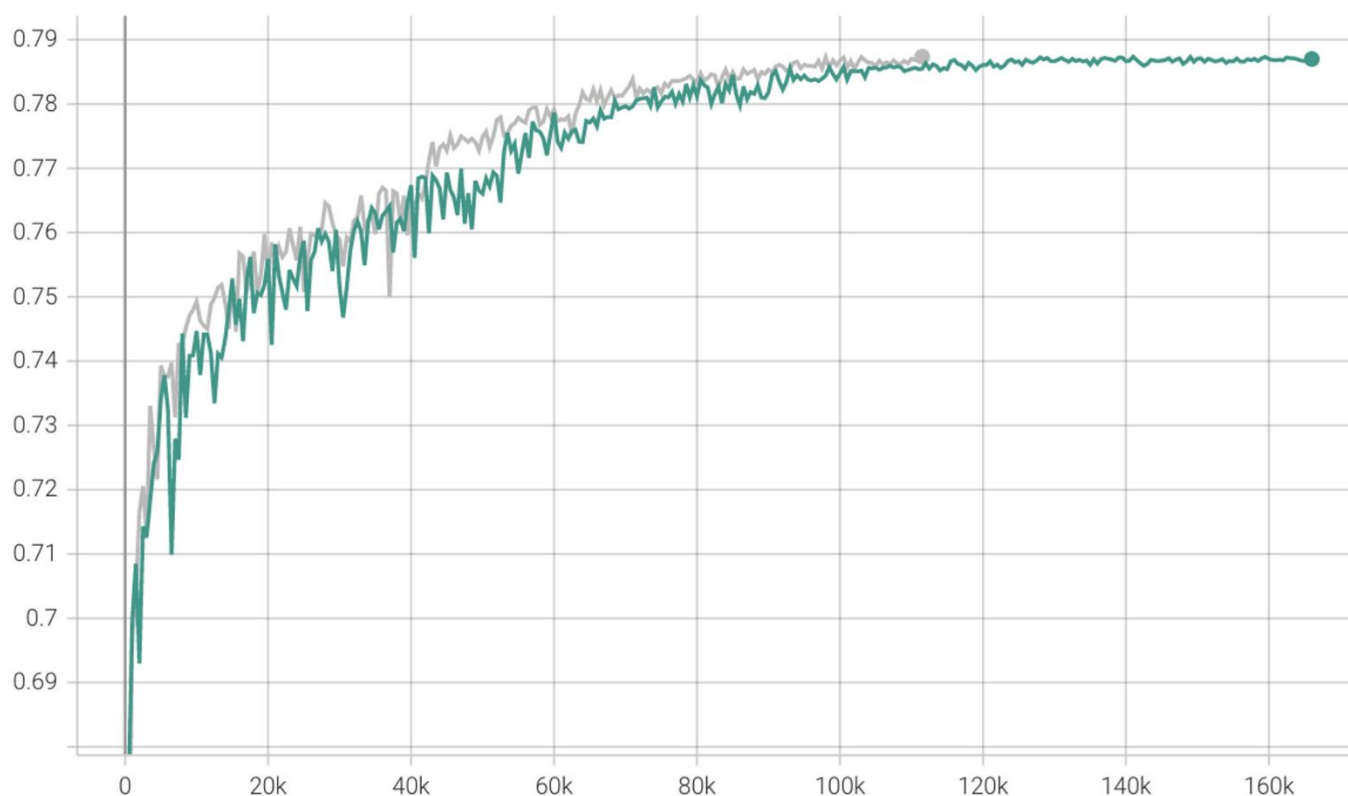


Рисунок 3.24 – Залежність метрики на валідаційній вибірці від номеру ітерації. Фокальна функція втрат (зелений) проти Бінарна перехресна ентропія + функція втрат Дайса (сірий)

Як видно з графіку порівняння значень метрик, зазначені функції втрат працюють майже ідентично. В максимальній точці, значення метрики дорівнює **78.7%** для обох функцій втрат.

Проте при аналогічних значеннях метрик, сирі (без ніякої пост-обробки) візуальні результати згенеровані мережею U-Net++ з використанням функції втрат BCE+Dice набагато кращі за результати роботи аналогічної мережі U-Net++, натренованої з використанням фокальної функції втрат, що зображено на рис. 3.25.

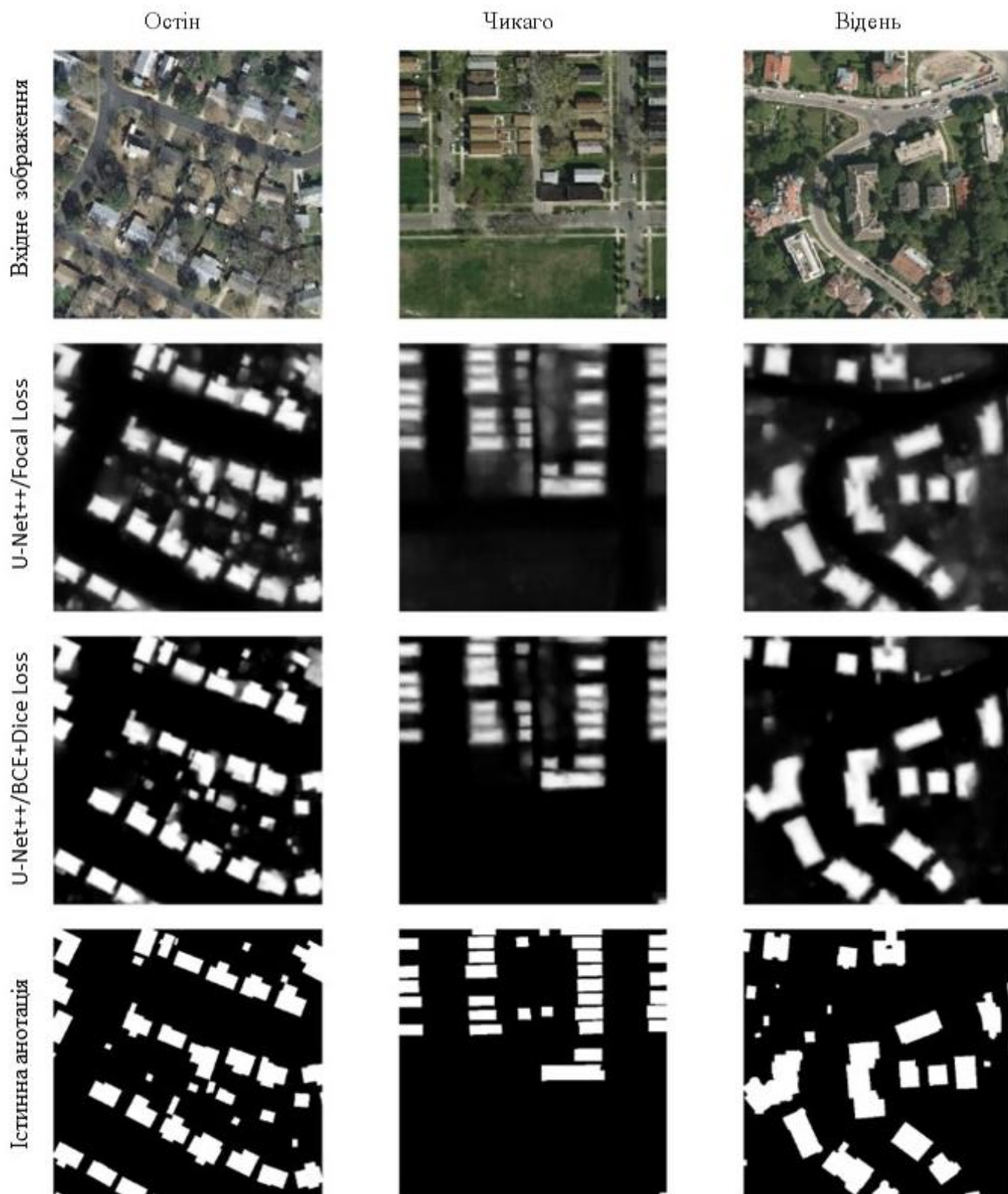
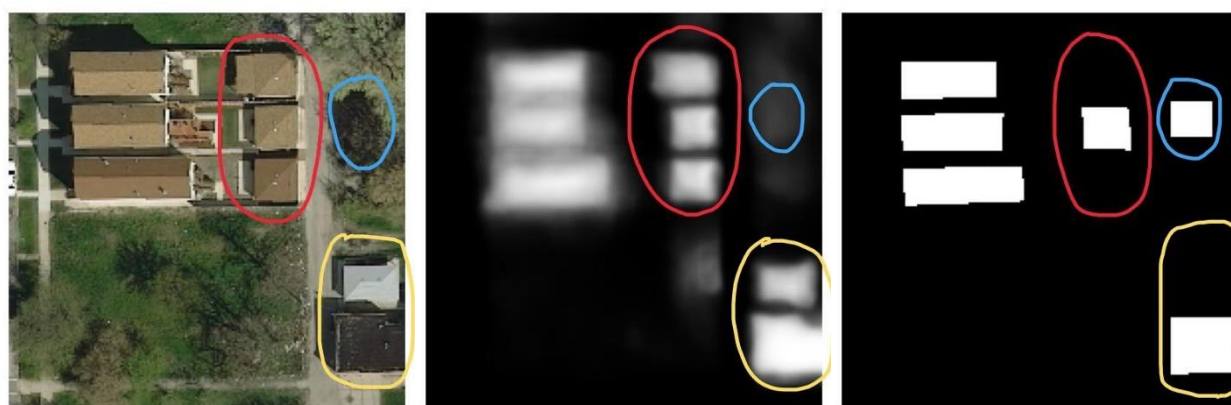


Рисунок 3.25 – Візуальний порівняльний аналіз архітектур нейронних мереж U-Net++ натренованих з використанням фокальної функції втрат (Focal Loss) та бінарної перехресної ентропії з функції втрат Дайса (BCE+Dice Loss)

Друга модель продукує ще менше шумів в регіонах невизначеності - шуми майже відсутні. Також дана модель генерує ще більш впевненні маски об'єктів та їх чіткіші межі.

Перевагу фокальної функції втрат над бінарною перехресною ентропією можна пояснити недостатньою якістю істинної розмітки. Фокальна функція втрат [37] сильніше штрафує ті ділянки зображення де модель помиляється сильніше. Проте якщо наша розмітка неправильна для цієї ділянки - ми помилково штрафуємо модель за правильну роботу. Це можна прослідкувати на зображенні з Чикаго:



вхідне зображення

результати моделі

“істинна” анотація

Рисунок 3.26 – Детальніші результати роботи архітектур U-Net++ з функцією втрат BCE+Dice Loss для зображення з Чикаго

В областях, обведених червоним та жовтим кольором, “істинна” анотація помилково не містить розмітки існуючих будинків, в той же час в області, обведеної синім, “істинна” анотація помилково містить розмітку будинку, якого насправді не існує.

Щодо використання функція втрат Дайса – дана функція втрат спрямована на генерування більш впевнених масок та чіткіших меж об'єктів.

Моделі, натреновані з функцією втрат Дайса, більше наголошують на великих областях переднього плану (власне об'єктах інтересу) і, таким чином, продукують відносно однорідні та більш впевнені ймовірності (яскравіший білий колір на наших візуалізаціях) для цих регіонів.



Дане зображення взяте з тестової вибірки, тобто значення метрики, що наразі дорівнює **78.7%**, насправді є значно вищим, і вже на даному етапі ми можемо стверджувати, що наша модель працює стабільніше за надану розмітку, що є дуже гарним знаком.

### **3.4. Обчислювальні потужності**

Обробка та підготовка навчального набору даних, а також розробка нейронної мережі, її тренування і валідація відбувалися на сервері з наступними характеристиками:

- Відеокарта: Nvidia GTX 1080 8Gb
- Процесор: AMD Ryzen Threadripper 2950X
- SSD: Samsung 970 EVO Plus 1TB
- Оперативна пам'ять: 2 x Kingston HyperX Fury 16Gb DDR4 3200
- Материнська плата: MSI X399 SLI PLUS

Додаткові обрахунки для найбільших моделей проводилися на віртуальній машині на основі відеокарти NVIDIA A100 Tensor Core GPU.

## ВИСНОВКИ ДО РОЗДІЛУ 3

В даному розділі було проведено експерименти для вибору оптимальних характеристик глибокої згорткової нейронної мережі для сегментації будинків для навчального набору даних Inria.

Експерименти проводилися для наступних складових нейронних мереж:

- Базова архітектура нейронної мережі:
  - U-Net [4]
  - U-Net++ [23]
- Модель кодувальника
  - ResNet-18 [46]
  - MobileNetv2 [51]
  - EfficientNetB2 [49]
- Функції втрат:
  - Фокальна функція втрат [37] (Focal Loss)
  - Бінарна перехресна ентропія [59] + функція втрат Дайса [16] (BCE+Dice Loss)

Інші параметри нейронної мережі, що були зафіксовані для усіх проведених експериментів:

- Роздільна здатність вхідного зображення: 256x256
- Метрика оцінки якості: IoU
- Оптимізатор: Адам
- Аугментації: всі, перераховані в підрозділі 3.2

В табл. 3.4 наведено загальні порівняння для всіх проведених експериментів.

Таблиця 3.4

Загальні порівняння для всіх проведених експериментів

Архітект ура	Загальна кількість параметрів (мільйони)	Кодувальник	кількість параметрів (мільйони)	функції втрат	метрика ↑ (валідаційн а вибірка)	Візуальна оцінка ↑ (0-10)

U-Net	14	ResNet-18	11	Focal Loss	74.8%	4
U-Net	5	MobileNet-V2	2	Focal Loss	74.9%	4
U-Net	10	EfficientNet B2	7	Focal Loss	78.2%	6
U-Net++	10,4	EfficientNet B2	7	Focal Loss	<b>78.7%</b>	7
U-Net++	10,4	EfficientNet B2	7	BCE+Dice Loss	<b>78.7%</b>	<b>9</b>

Візуальне порівняння першого базового експерименту (U-Net/ResNet-18/Focal Loss) та найкращого експерименту (U-Net++/EfficientNetB2/BCE+Dice Loss) зображено на рис. 3.27

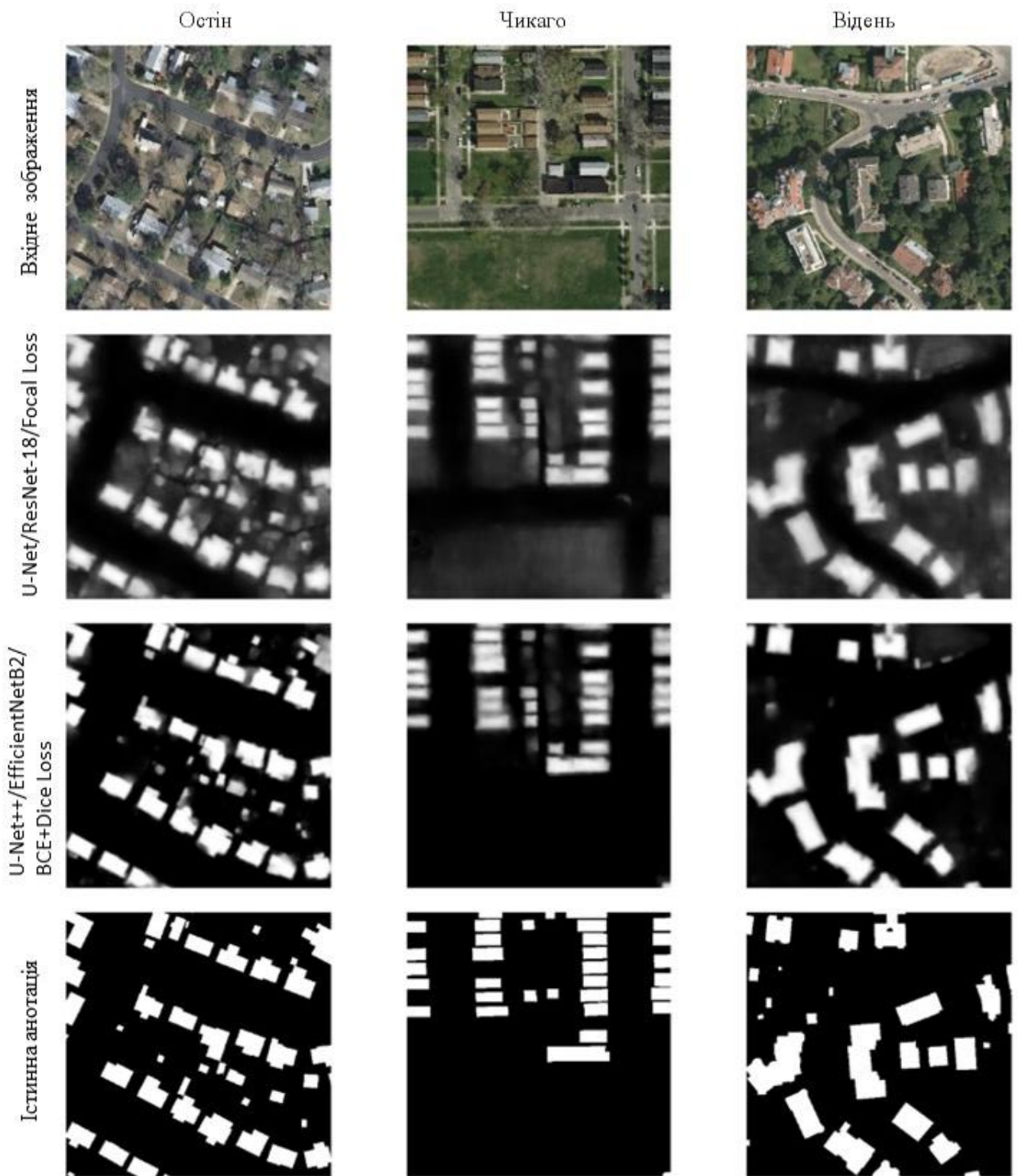
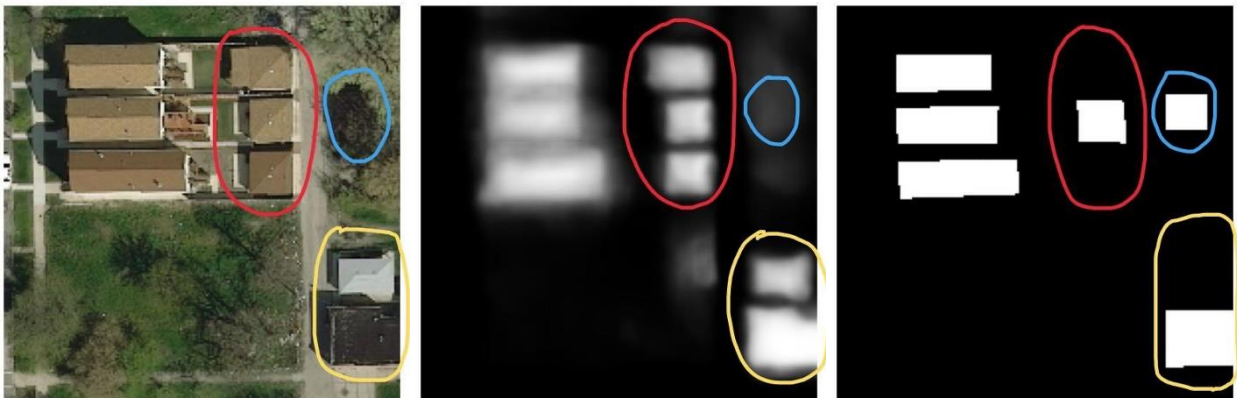


Рисунок 3.27 – Візуальний порівняльний аналіз експерименту U-Net/ResNet-18/Focal Loss та експерименту U-Net++/EfficientNetB2/VCE+Dice Loss

Також було показано, що значення метрики на валідаційній вибірці для найкращої моделі, що наразі дорівнює **78.7%**, насправді є значно вищим, і вже на даному етапі ми можемо стверджувати, що наша модель працює стабільніше за

надану розмітку, що є дуже гарним знаком.

Приклад переваги результатів роботи нейронної мережі над “істинними” анотаціями для зображення з Чикаго зображено на рис. 3.28.



вхідне зображення

результати моделі

“істинна” анотація

Рисунок 3.28 – Приклад переваги результатів роботи нейронної мережі над “істинними” анотаціями для зображення з Чикаго

## РОЗДІЛ 4. РОЗРОБКА СТАРТАП ПРОЄКТУ «РОЗПІЗНАВАННЯ (СЕГМЕНТАЦІЯ) ОБ'ЄКТІВ НА СУПУТНИКОВИХ ТА АЕРОФОТОЗНІМКАХ»

### 4.1. Опис ідеї проєкту та технологічний аудит ідеї проєкту

Розглянувши в попередніх розділах переваги використання нейронних мереж для розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках. В цьому розділі буде проведено аналіз стартап проєкту «Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках».

Ідея проєкту полягає в порівняльному аналізі нейронних мереж для розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках, модифікації архітектури і компонентів нейронних мереж для найоптимальнішого вирішення поставленої задачі, що уточнено наведено в табл. 4.1.

У табл. 4.1 зображено зміст ідеї та можливі базові потенційні ринки, в межах яких потрібно шукати групи потенційних клієнтів [65].

Таблиця 4.1.

Опис ідеї стартап проєкту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Знаходження найкращого методу для розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках	оцінка пошкоджень в деокупованих регіонах (де є можливість зробити знімки з дрона) та тимчасово окупованих регіонах (де наявні тільки супутникові знімки)	нижча ціна розробки, а отже нижча плата за користування
	оцінка пошкоджень при стихійних лихах та військових конфліктах	легкість масштабування
	сфера вироблення сонячних панелей для оцінки корисної поверхні даху та кількості сонячного світла на певній ділянці	можливість вибору одного з двох режимів роботи: <ul style="list-style-type: none"> <li>· швидкий з допустимою точністю;</li> <li>· повільний з найкращою точністю</li> </ul>
	розпізнавання будинків в регіонах, де кількість населення змінюється дуже швидко а також у віддалених і масштабних	генералізація: модель може працювати навіть для регіонів, що не використовувались для навчання моделі

	районах, де перепис цих нових будівель часто виконується вручну і швидко стає неактуальним	адаптивність: маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо...
--	--	---

Отже, пропонується нова модель для розпізнавання (сегментація) об'єктів, яку можливо використовувати для решти задач за наявності відповідних наборів даних це може бути абсолютно все що завгодно, що видно на супутникових чи аерофотознімках з дронів: будинки, річки, океани, дороги, сільсько-господарські поля, стадіони, заводи, літаки, кораблі, автомобілі, резервуар, бейсбольне поле, тенісний корт, баскетбольний майданчик, стадіон, гавань, міст, вертоліт, футбольне поле, басейн, контейнерний кран, аеропорт і вертолітний майданчик, навіть людей (тільки на аерофотознімках з дронів, якщо дозволяє роздільна здатність)

Далі проводимо аналіз потенційних техніко-економічних переваг ідеї порівняно із пропозиціями конкурентів:

- визначаємо перелік техніко-економічних властивостей та характеристик ідеї;
- визначаємо попереднє коло конкурентів (проектів-конкурентів) або товарів-замінників чи товарів-аналогів, що вже існують на ринку, та проводимо збір інформації щодо значень техніко-економічних показників для ідеї власного проекту та проектів-конкурентів відповідно до визначеного вище переліку;
- проводимо порівняльний аналіз показників: для власної ідеї визначено показники, що мають а) гірші значення (W, слабкі); б) аналогічні (N, нейтральні) значення; в) кращі значення (S, сильні) (табл. 4.2).

Таблиця 4.2.

Визначення сильних, слабких та нейтральних характеристик ідеї проєкту

№ п/п	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проєкт	Конкурент 1 <b>namR</b>	Конкурент 2 <b>RMCTeam</b>	Конкурент 3 <b>TRIUMPHroofing</b>			
1.	Вартість розробки продукту	50 000\$	200 000\$	175 000\$	167 000\$	-	+	-
2.	Вартість використання продукту	15\$ персональні, 50\$ корпоративні	40\$ персональні, 100\$ корпоративні	75\$	80\$	-	-	+
3.	Використання нейронних мереж	+	+	-	-	-	-	+
4.	Можливість масштабування	+	+	-	-	-	-	+
5.	Можливість адаптації до нового регіону	+	+	-	-	-	-	+
6.	Можливість узагальнення моделі на інші класи об'єктів	+	-	-	-	-	-	+
7.	Інвестиції	-	+	+	+	+	-	-
7.	Торгова марка	-	+	+	+	+	-	-

Сильною стороною даного проєкту є його універсальність, а саме використання нейронних мереж, можливість масштабування, можливість адаптації до нового регіону, можливість узагальнення моделі на інші класи об'єктів. Так-як стартап в розробці, інвестиції та торгова марка відсутні, що поки є слабкою стороною.

Проаналізувавши всі сильні, слабкі та нейтральні сторони можна зробити висновок, що даний стартап проєкт є конкурентоспроможним.



В межах даного підрозділу проводимо аудит технології за допомогою якої можна реалізувати ідею створення проєкту.

Визначення технологічної здійсненності ідеї проєкту передбачає аналіз складових які вказані в таблиці 4.3.

Таблиця 4.3.

Технологічна здійсненність ідеї проєкту

№ п/п	Ідея проєкту	Технології її реалізації	Наявність технологій	Доступність технологій
1.	Найоптимальніше вирішення для розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках	Нейронні мережі	+, з подальшим удосконаленням	Доступна
		Хмарні обчислення	+	Платний сервіс
		CUDA (програмно-апаратна архітектура паралельних обчислень, яка дозволяє істотно збільшити обчислювальну продуктивність завдяки використанню графічних процесорів фірми Nvidia)	+	Доступна
Обрана технологія реалізації ідеї проєкту: глибока нейронна мережа				

Проаналізувавши таблицю можна зробити висновок що наш проєкт «Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках» і спробувавши всі технології на практиці, можна зробити висновок, що саме глибока нейронна мережа є необхідною для реалізації стартапу, також доступна для використання, на відміну від технології хмарного обчислення.

## 4.2. Аналіз ринкових можливостей запуску стартап проєкту

Визначимо ринкові можливості, які можна використати під час ринкового впровадження проєкту, та ринкові загрози, які можуть перешкодити його реалізації.

Це дозволяє спланувати напрями розвитку проєкту із урахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проєктів конкурентів.

Спочатку проведемо аналіз попиту: наявність попиту, обсяг, динаміка розвитку

ринку (таблиця 4.4).

Таблиця 4.1.

Попередня характеристика потенційного ринку стартап-проєкту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	10
2	Загальний обсяг продаж, грн/ум.од	100 тис. ум. од.
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Пошук інвестицій
5	Специфічні вимоги до стандартизації та сертифікації	Відсутні
6	Середня норма рентабельності в галузі (або по ринку), %	25%

Середня норма рентабельності в галузі (або по ринку) порівнюється із банківським відсотком на вкладення менша. За умови, що останній є вищим, можливо, має сенс вкласти кошти в інший проєкт. За результатами аналізу таблиці робиться висновок щодо того, що ринок є привабливим для входження за попереднім оцінюванням.

Надалі визначаємо потенційні групи клієнтів, їх характеристики, та формуємо орієнтовний перелік вимог до товару для кожної групи (табл. 4.5).

Таблиця 4.2.

## Характеристика потенційних клієнтів стартап-проєкту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Необхідність застосування нейронних мереж для швидкої машинної обробки величезних об'ємів зображень	Військовий ринок, ринок сонячних панелей, це може бути будь-який ринок за наявності відповідних наборів даних (аерофотознімки з дронів, якщо дозволяє роздільна здатність)	Різний набір даних на яких спеціалізуються потенційні цільові групи клієнтів	Низька плата за користування, можливість вибору одного з двох режимів роботи (швидкий з допустимою точністю, повільний з найкращою точністю), маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) натренувати для виявлення інших класів об'єктів

За результатами аналізу характеристик потенційних клієнтів стартап-проєкту видно, що цільовою аудиторією може бути будь-який ринок за наявності відповідних наборів даних (аерофотознімки з дронів, якщо дозволяє роздільна здатність). Також продукт має забезпечувати низьку плату за користування, можливість вибору одного з двох режимів роботи (швидкий з допустимою точністю, повільний з найкращою точністю), маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) натренувати для виявлення інших класів об'єктів.

При застосуванні даної технології існують певні загрози. (таблиця 4.6). Такими загрозами можуть бути фактори з різних економічних сторін.

Таблиця 4.3.

## Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1.	Заборона збору/використання даних над приватними територіями	Деякі власники приватних територій (будинків, держ. установ)	Домовитись на пряму з власниками приватних територій
2.	Обмежена доступність супутникових знімків в певних регіонах	Провайдери, ліцензія, права на використання	Юридичні домовленості, підписання строгих угод про нерозголошення та нерозповсюдження даних
3.	Неактуальність даних	Остання версія фото застаріла	Постійне оновлення та слідкування за отриманими даними
4.	Якість зображень	роздільна здатність (метрів/піксель), комбінація поганих погодних умов, недостатнього освітлення, зашумленості отриманого зображення	Апаратне покращення за можливості Використання методів супер-резолюції (підвищення роздільної здатності та якості зображень використовуючи генеративні нейронні мережі)
5.	Конкурентний ринок досліджень нейронних мереж	Будь-яка дослідницька команда/центр, що працює над покращенням та застосуванням нейронної мережі, може зайти в сферу	Постійно розвивати компанію, набирати дослідницьку команду та використовувати останні дослідження та розробки в сфері глибоких нейронних мереж

За результатами аналізу факторів загроз можна виокремити наступні для даного проекту: заборона збору/використання даних над приватними територіями, обмежена доступність супутникових знімків в певних регіонах, неактуальність даних, погана якість зображень та конкурентний ринок досліджень нейронних мереж.

Але поряд із колом загроз існують і певні можливості (таблиця 4.7).

Таблиця 4.4.

## Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1.	Ринок в Україні	Створення ринку в Україні із-за відсутності аналогів на ньому	Вільний розвиток та простий вихід на даний ринок

2.	Інвестиції	Знаходження інвесторів	Участь у презентаціях, для залучення інвесторів
3.	Економічний	Збільшення ринку	Розширення ринку за рахунок іноземних замовників
4.	Попит	При появі нового ринку, з'являться потенційні користувачі	Пошук нових користувачів
5.	Міжнародний	Наявність ринку в світі	Вихід на світовий ринок

За результатами аналізу факторів можливостей загроз можна виокремити наступні для даного прєкту: Створення ринку в Україні із-за відсутності аналогів на ньому, Знаходження інвесторів, Розширення ринку за рахунок іноземних замовників, Пошук нових користувачів та Вихід на світовий ринок.

Наступним кроком необхідно провести ступеневий аналіз конкуренції на ринку, що наведено в табл. 4.8. Даний аналіз розкриває вид ринкової конкуренції та допомагає у виборі стратегії поведження на ринку.

Таблиця 4.5.

#### Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції: Україна – монополія, світ – монополістична	В Україні не існує аналогів, а у світі їх 10	Можливість виходу як на Український, так і на світовий ринок
2. За рівнем конкурентної боротьби – світовий	Існуючі конкуренти з інших країн	Збут на міжнародному ринку
3. За галузевою ознакою – міжгалузева	Маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо...	Застосування продукту в різних галузях
4. Конкуренція за видами товарів – товарно-родова	Конкуренція на рівні технології задоволення потреб.	Реалізація для задоволення різних потребу
5. За характером конкурентних переваг – нецінова	Забезпечується поліпшенням якісних характеристик продукту (якості, надійності, довговічності) або поліпшенням умов його реалізації	Головна конкурентна перевага – унікальність позиціонування.

6. За інтенсивністю – марочна	Існуючі конкуренти мають власну марку – namR, RMСteam, TRIUMPHroofing	Є стимул створити власну марку, або працювати під відомою маркою
-------------------------------	---	--

За результатами ступеневого аналізу конкуренції на ринку можна зробити висновок, що можливим кроком для входу на ринок є створення власної марки, або робота під відомою маркою. Також ринок демонструє, як мінімум, 10 конкурентів, але є один, який є головним в цій сфері – namR . Він має власний бренд, впізнаваність та немалу ціну за представлення своїх послуг, на відміну від нашого проекту.

Після аналізу конкуренції проведемо більш детальний аналіз умов конкуренції в галузі (табл. 4.9).

Таблиця 4.6.

Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	namR, RMСteam, TRIUMPHroofing	Компанії, які працюють в одній галузі, але є потенціал розширення на інші галузі	Впізнаваність, висока ціна на товари	Потреба в оцінці пошкоджень дахів	-
Висновки:	Прямі конкуренти працюють у вузькій спеціальності (дахи), і даний проект може реалізовуватись в різних спеціальностях. Також на українському ринку конкуренти відсутні, тому в Україні відсутня конкурентна боротьба, а в світі вона мінімальна.	Поки бар'єру входу на ринок нема, тому що відсутні конкуренти на вітчизняному ринку.	Для реалізації цього проекту не потрібні постачальники.	Потреба в обробці великої кількості зображень	Обмеження для входу на ринок та роботи в ньому через товари-замінники відсутні, так як такі товари відсутні на ринку.

За результатами аналізу конкуренції в галузі за М. Портером можна виокремити наступних конкурентів в світі: *namR*, *RMSteam*, *TRIUMPHroofing*. Всі вони працюють у вузькій спеціальності (дахи), а на вітчизняному ринку конкурентів поки немає.

Після всіх аналізів визначається та обґрунтовується перелік факторів конкурентоспроможності (табл. 4.10).

Таблиця 4.7.

## Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проєктів значущим)
1	Генералізація	модель може працювати навіть для регіонів, що не використовувались для навчання моделі
2	Адаптивність	маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо...
3	Вартість продукту	Ціна нижче ніж у конкурентів
4	Надійність	При необхідності дотренувати модельку для нових регіонів, або змінити сервер на хмарне середовище/інший сервер
5.	Глибокі нейронні мережі	Швидка обробка великої кількості даних різної якості

За результатами аналізу факторів конкурентоспроможності можна виокремити наступні для даного прекуту: генералізація, адаптивність, вартість продукту, надійність, глибокі нейронні мережі. Ці фактори показують, що цей продукт буде достойним конкурентом на ринку.

Таблиця 4.8.

Порівняльний аналіз сильних та слабких сторін «*namR*»

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з <i>namR</i>						
			-3	-2	-1	0	+1	+2	+3
1	Наявність патентів	16	+						
2	Масштабування	15					+		
3	Досвід роботи	17		+					
4	Технічна підтримка	19				+			
5.	Ціна	14						+	

З таблиць 4.10 та 4.11 бачимо, що фактори конкурентоспроможності суттєві та мають великий позитивний внесок при реалізації проєкту. Основною перевагою та головним досягненням є нижча ціна розробки, а отже нижча плата за користування; можливість вибору одного з двох режимів роботи (швидкий з допустимою точністю, повільний з найкращою точністю); легкість масштабування; генералізація – модель може працювати навіть для регіонів, що не використовувались для навчання моделі; адаптивність – маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо.

Таблиця 4.9.

## SWOT- аналіз стартап-проєкту

<p>Сильні сторони:</p> <ol style="list-style-type: none"> <li>1. Нижча ціна розробки, а отже нижча плата за користування.</li> <li>2. Можливість вибору одного з двох режимів роботи (швидкий з допустимою точністю, повільний з найкращою точністю).</li> <li>3. Легкість масштабування.</li> <li>4. Генералізація – модель може працювати навіть для регіонів, що не використовувались для навчання моделі.</li> <li>5. Адаптивність – маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо.</li> </ol>	<p>Слабкі сторони:</p> <ol style="list-style-type: none"> <li>1. Відсутність інвесторів.</li> <li>2. Відсутність команд для просування проєкту.</li> <li>3. Відсутність торгової марки.</li> </ol>
<p>Можливості:</p> <ol style="list-style-type: none"> <li>1. Створення ринку в Україні із-за відсутності аналогів на ньому.</li> <li>2. Знаходження інвесторів за допомогою участі у презентаціях, для їх залучення.</li> <li>3. Розширення ринку за рахунок іноземних замовників;</li> <li>4. Збільшення попиту на продукт (При появі нового ринку, з'являться потенційні користувачі).</li> <li>5. Вихід на світовий ринок</li> </ol>	<p>Загрози:</p> <ol style="list-style-type: none"> <li>1. Заборона збору/використання даних над приватними територіями.</li> <li>2. Обмеження доступу до супутникових знімків в певних регіонах.</li> <li>3. Неактуальність даних (застаріла версія даних).</li> <li>4. Якість зображень (низька роздільна здатність (метрів/піксель), комбінація поганих погодних умов, недостатнього</li> </ol>



	освітлення, зашумленості отриманого зображення). 5. Конкурентний ринок досліджень нейронних мереж (Будь-яка дослідницька команда/центр, що працює над покращенням та застосуванням нейронної мережі, може зайти в сферу).
--	--

За результатами SWOT-аналізу можна зробити наступний висновок:

Можливості і загрози є вірогідними результатами впливу факторів, які могли б статися, але якими дійсно можна скористатися (можливості) або виправити (загрози).

Таблиця 4.10.

Альтернативи ринкового впровадження стартап-проєкту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Стратегія нейтралізації ринкових загроз сильними сторонами стартапу	82%	1 рік
2	Стратегія компенсації слабких сторін стартапу наявними ринковими можливостями	60%	1-2 роки
3	Стратегія виходу з ринку	25%	2-3 роки
4	Просування проєкту	90%	6 місяців
5	Вихід на світовий ринок	80%	1 рік

За результатами аналізу альтернативи ринкового впровадження стартап-проєкту можна зробити наступний висновок: з зазначених альтернатив логічніше обрати стратегію просування проєкту та вихід на світовий ринок, тому що час їх реалізації менше, ніж у всіх інших представлених альтернатив.

### 4.3. Розроблення ринкової стратегії проєкту

Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів.

Таблиця 4.11.

## Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Військова сфера	Готові	Великий	Низька	Просто
2	Сфера вироблення сонячних панелей	Готові	Великий	Середня	Просто
3	Дослідницькі центри (оцінка пошкоджень при стихійних лихах)	Готові	Великий	Висока	Непросто
4	Регіональні компанії для перепису населення та будинків	Відносно готові	Середній	Низька	Просто
Які цільові групи обрано: Під час аналізу потенційних груп споживачів було прийнято рішення що компанія буде працювати із всіма представленими цільовими групами.					

За результатами аналізу потенційних груп споживачів ми обрали такі цільові групи: Військова сфера, Сфера вироблення сонячних панелей, Дослідницькі центри (оцінка пошкоджень при стихійних лихах), Регіональні компанії для перепису населення та будинків. Також є можливість розширення цільових груп, так як проєкт, маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо.

Для роботи в обраному сегменті ринку необхідно сформувати базову стратегію розвитку.

Таблиця 4.12.

## Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проєкту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*

1	адаптивність	Маркетинг	Низька ціна, легкість масштабування, генералізація	Стратегія диференціації
---	--------------	-----------	--	-------------------------

За результатами аналізу базової стратегії розвитку ми обрали альтернативу розвитку проєкту – адаптивність (маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об’єктів: дороги, водойми, поля, автомобілі, тощо).

Наступним кроком є вибір стратегії конкурентної поведінки (табл. 4.16).

Таблиця 4.13.

Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проєкт «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
1	В Україні – так	Проєкт буде шукати нових споживачів	Ні, буде більший перелік можливостей, та краща ціна	Стратегія позиціонування.

За результатами аналізу базової стратегії конкурентної поведінки ми обрали стратегію позиціонування. Також проєкт буде шукати нових споживачів, тому що на ринку України таких компаній ще нема.

На основі вимог споживачів з обраного сегменту до постачальника і продукту, а також в залежності від стратегії розвитку та стратегії конкурентної поведінки розробляємо стратегію позиціонування яка визначається у формування ринкової позиції, за яким споживачі мають ідентифікувати проєкт.

Таблиця 4.14.

Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проєкту	Вибір асоціацій, які мають сформувати комплексну позицію власного проєкту (три ключових)
1	Швидка обробка великої	Стратегія диференціації	Легке масштабування. Генералізація. Адаптивність.	Якість. Ціна.

	кількості даних			
--	-----------------	--	--	--

Результатом даного підрозділу є система рішень щодо ринкової поведінки компанії, вона визначає в якому напрямі буде працювати компанія на ринку.

#### 4.4. Розроблення маркетингової програми стартап-проєкту

Під час розроблення маркетингової програми першим кроком є розробка маркетингової концепції товару, який отримає споживач. У таблиці 4.18 підсумовуємо результати аналізу конкурентоспроможності товару.

Таблиця 4.15.

Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Нижча ціна розробки	Нижча плата за користування	Оптимальна ціна для придбання продукту
2	Легкість масштабування	Збільшення територій обробки інформації	
3	Можливість вибору режиму	Два режимів роботи	1. швидкий з допустимою точністю; 2. повільний з найкращою точністю
4	Генералізація	Модель може працювати навіть для регіонів, що не використовувались для навчання моделі	Можливість використовувати модель на будь-яких територіях, що не використовувались для навчання моделі
5	Адаптивність	Маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо	Можливість натренувати модуль за допомогою будь-яких класів об'єктів

За результатами аналізу ключових переваг концепції потенційного товару

можна зробити наступний висновок: що наш проєкт задовольняє всі потреби клієнтів.

Таблиця 4.16.

Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Модель має розпізнавати (сегментувати) об'єкти на супутникових та аерофотознімках		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. нижча ціна розробки, а отже нижча плата за користування	М	Е
	2. можливість вибору одного з двох режимів роботи (швидкий з допустимою точністю, повільний з найкращою точністю)	Нм	Тл
	3. легкість масштабування	Нм	Тл
	4. генералізація: модель може працювати навіть для регіонів, що не використовувались для навчання моделі	Нм	Тл
III. Товар із підкріпленням	5. адаптивність: маючи відповідний набір тренувальних даних модель (без жодних архітектурних змін) може бути натренована для виявлення інших класів об'єктів: дороги, водойми, поля, автомобілі, тощо...	Нм	Тл
	Якість: стандарт		
	Пакування: модель, глибока нейронна мережа		
Марка: відсутня			
До продажу ознайомлення клієнта			
Після продажу підтримка клієнта протягом підписки			
За рахунок чого потенційний товар буде захищено від копіювання: Патент, комерційна таємниця			

За результатами аналізу трьох рівнів моделі товару можна зробити наступний висновок: задумом проєкту є модель, яка має розпізнавати (сегментувати) об'єкти на супутникових та аерофотознімках, а його характеристики – низька ціна, масштабування, генералізація та адаптивність.

Наступним кроком є визначення цінових меж, якими необхідно керуватися при встановленні ціни на потенційний товар, це передбачає аналіз цін товарів конкурентів, та доходів споживачів продукту (табл. 4.20).

Таблиця 4.17.

## Визначення меж встановлення ціни

№ п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	40 – 100\$	40 – 150\$	0,5 млн \$/рік	10 – 60\$

За результатами аналізу рівня цін на товари-замінники та на товари-аналоги також рівня доходів цільової групи споживачів можна зробити наступний висновок, що для нашого проєкту можна визначити таку верхню та нижню межі встановлення ціни на товар/послугу, як 10 – 60\$.

Таблиця 4.18.

## Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Підписка	Маркетингова складова Пошук клієнтів	Канал нульового рівня	Сайт компанії, тендери

За результатами аналізу формування системи збуту видно, що специфікою закупівельної поведінки цільових клієнтів буде підписка на продукт, а функції збуту, які має виконувати постачальник товару буде маркетинг та пошук клієнтів.

Таблиця 4.19.

## Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Потенційні клієнти отримують інформацію через рекламу, наукові конференції, поради знайомих.	Інтернет. Форуми, галузеві виставки. Соц. мережі.	Таргет Наукові статті Маркетинг	Розповісти про продукт, знаходження потенційних клієнтів	Акцент на масштабування, генералізація та адаптивність проєкта

За результатами аналізу Концепція маркетингових комунікацій бачимо, що

Потенційні клієнти можуть отримувати інформацію через рекламу, наукові конференції, поради знайомих, а компанія, в свою чергу, буде робити все можливе для популяризації свого продукту.

Таблиця 4.23

## Початкові вкладення у стартап-проект

<i>Види витрат</i>	<i>Вартість</i>
НДДКР	-
Захист прав на об'єкти інтелектуальної власності	10 000
Створення прототипу, дослідження	50 000
Оренда приміщення	-
Просування	20 000
Витрати на команду	150 000
Замовлення послуг, менторство	5 000
Закупівля обладнання	100 000
Створення сайту стартапу	15 000
Орієнтована собівартість першого продукту	100 000
<i>РАЗОМ</i>	450 000
<i>Витрати, що бере на себе стартапер</i>	300 000
<i>Необхідні інвестиції для запуску стартапу та виробництва першої партії</i>	150 000

## ВИСНОВКИ ДО РОЗДІЛУ 4

У цьому розділі було розроблено стартап проєкт, а саме:

1. Опис ідеї проєкту, який полягає в порівняльному аналізі нейронних мереж для розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках, модифікації архітектури і компонентів нейронних мереж для найоптимальнішого вирішення поставленої задачі. Визначення сильних, слабких та нейтральних характеристик ідеї проєкту в порівнянні з існуючими конкурентами (namR, RMCteam, TRIUMPHroofing).

2. Технологічний аудит ідеї проєкту: для проєкту «Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках» саме глибока нейронна мережа є необхідною для реалізації стартапу, також доступна для використання, на відміну від технології хмарного обчислення.

3. Аналіз ринкових можливостей запуску стартап-проєкту: проведено попередню характеристика потенційного ринку стартап-проєкту (Кількість головних гравців, загальний обсяг продаж, динаміка ринку, наявність обмежень для входу на нього); проведено характеристика потенційних клієнтів стартап-проєкту, де виявлено основні потреби та вимоги споживачів, цільову аудиторію (військовий ринок, ринок сонячних панелей, це може бути будь-який ринок за наявності відповідних наборів даних (аерофотознімки з дронів, якщо дозволяє роздільна здатність)); фактори загроз та можливостей, які наведені в SWOT- аналіз стартап-проєкту; ступеневий аналіз конкуренції на ринку; аналіз конкуренції в галузі за М. Портером; обґрунтування факторів конкурентоспроможності; Альтернативи ринкового впровадження стартап-проєкту.

4. Розроблення ринкової стратегії проєкту, а саме визначення стратегії ринку та опис цільових груп потенційних споживачів.

5. Розроблення маркетингової програми стартап-проєкту, що включає в себе план заходів із ознайомлення цільової аудиторії з ідеєю проєкту, просування його на ринок для підвищення зацікавленості та прихильності потенційних споживачів, а саме визначення цінових меж, якими необхідно керуватися при



встановленні ціни на потенційний товар (10 – 60\$); система збуту буде відбуватися за допомогою підписки на продукт.

За результатами всіх аналізів, можна зробити висновок, що стартап-проект «Розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках» на ринку буде достойним конкурентом та знайде своїх споживачів.

## ВИСНОВКИ

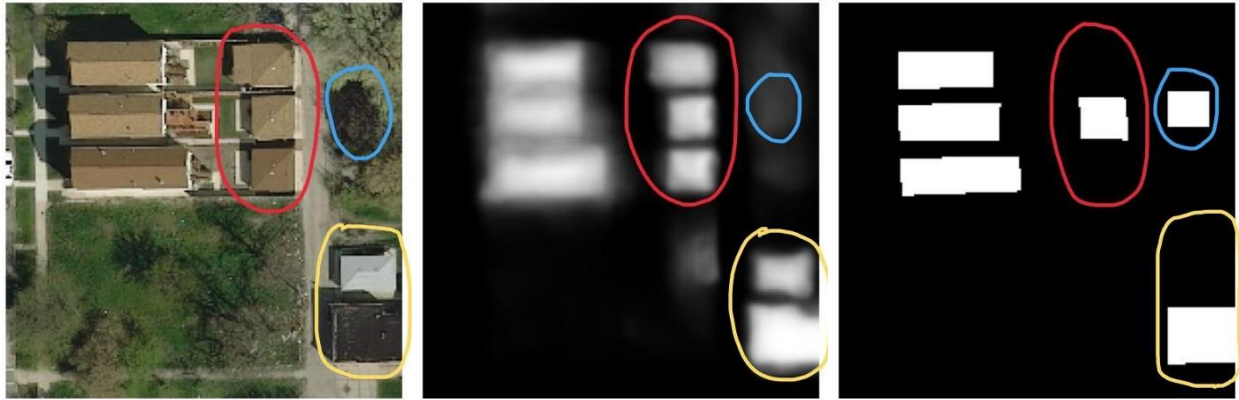
У даній магістерській дисертації було запропоновано розробку моделі для розпізнавання (сегментація) об'єктів на супутникових та аерофотознімках. Для її створення було:

- проведено вичерпний порівняльний аналіз архітектур сегментаційних мереж та кодувальників (Unet, Deeplab, Unet++, ResNet, EfficientNet), існуючі програмні рішення (MMSegmentation, Detectron2, Segmentation models pytorch).

Не зважаючи на багато менші масштаби реалізації, порівняно з MMSegmentation та Detectron2, саме Segmentation Models Pytorch є найпоширенішою та найпростішою в застосуванні бібліотекою для семантичної сегментації зображень. Її широко застосовують як і для швидкого прототипування для конкурсів з машинного навчання (таких як Kaggle [54]), так і для великої продуктивних рішень.

- досліджено набори даних для сегментації будинків на аерофотознімках (Набір даних Inria Aerial Image Labeling [61] створений для основної задачі дистанційного зондування, а саме автоматичного попиксельного маркування аерофотознімків);

- розроблене відповідне програмне забезпечення;
- проведено експерименти та навчено низку нейронних мереж;
- досліджено вплив функцій втрат на результат роботи моделі;
- проведено кількісний та візуальний аналіз отриманих результатів на тестовій вибірці обраного набору даних – в областях, обведених червоним та жовтим кольором, «істинна» анотація помилково не містить розмітки існуючих будинків, в той же час в області, обведеної синім, «істинна» анотація помилково містить розмітку будинку, якого насправді не існує.



вхідне зображення

результати моделі

“істинна” анотація

Щодо використання функція втрат Дайса – дана функція втрат спрямована на генерування більш впевнених масок та чіткіших меж об’єктів.

Моделі, натреновані з функцією втрат Дайса, більше наголошують на великих областях переднього плану (власне об’єктах інтересу) і, таким чином, продукують відносно однорідні та більш впевнені ймовірності (яскравіший білий колір на наших візуалізаціяї) для цих регіонів.

Дане зображення взяте з тестової вибірки, тобто значення метрики, що наразі дорівнює **78.7%**, насправді є значно вищим, і вже на даному етапі ми можемо стверджувати, що наша модель працює стабільніше за надану розмітку, що є дуже гарним знаком.

Також в магістерській дисертації було розроблено стартап-проект «Розпізнавання (сегментація) об’єктів на супутникових та аерофотознімках» за результатами всіх аналізів, можна зробити висновок, що стартап-проект на ринку буде достойним конкурентом та знайде своїх споживачів.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Бурау Н.І., Богдан Г.А., Півторак Д.О., Філіппова М.В. Магістерська дисертація. Організація виконання та захисту, вимоги до структури, змісту та оформлення: Навчальний посібник. Київ, 2022. 48 с
2. Semantic vs. Instance vs. Panoptic Segmentation [Електронний ресурс]. URL: <https://pyimagesearch.com/2022/06/29/semantic-vs-instance-vs-panoptic-segmentation/>
3. Badrinarayanan V., Kendall A., Cipolla R., Member S., IEEE, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv:1511.00561v3 [cs.CV] 10 Oct 2016. [Електронний ресурс]. URL: [1511.00561.pdf \(arxiv.org\)](https://arxiv.org/abs/1511.00561)
4. Ronneberger O., Fischer P., Brox T., U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1 [cs.CV] 18 May 2015. [Електронний ресурс]. URL: [U-Net: Convolutional Networks for Biomedical Image Segmentation \(arxiv.org\)](https://arxiv.org/abs/1505.04597)
5. Noh H., Hong S., Han B., Learning Deconvolution Network for Semantic Segmentation. arXiv:1505.04366v1 [cs.CV] 17 May 2015. [Електронний ресурс]. URL: [1505.04366.pdf \(arxiv.org\)](https://arxiv.org/abs/1505.04366)
6. Long J., Shelhamer E., Darrell T., Fully Convolutional Networks for Semantic Segmentation. arXiv:1411.4038v2 [cs.CV] 8 Mar 2015. [Електронний ресурс]. URL: [1411.4038.pdf \(arxiv.org\)](https://arxiv.org/abs/1411.4038)
7. He K., Gkioxari G., Dollár P., Girshick R., Mask R-CNN. arXiv:1703.06870v3 [cs.CV] 24 Jan 2018. [Електронний ресурс]. URL: [1703.06870.pdf \(arxiv.org\)](https://arxiv.org/abs/1703.06870)
8. Ren S., He K., Girshick R., Sun J., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497v3 [cs.CV] 6 Jan 2016. [Електронний ресурс]. URL: [1506.01497.pdf \(arxiv.org\)](https://arxiv.org/abs/1506.01497)
9. Liu S., Qi L., Qin H., Shi J., Jia J., Path Aggregation Network for Instance Segmentation. arXiv:1803.01534v4 [cs.CV] 18 Sep 2018. [Електронний ресурс]. URL:

[1803.01534v4.pdf \(arxiv.org\)](#)

10. Bolya D., Zhou C., Xiao F., Lee Y., YOLACT Real-time Instance Segmentation. arXiv:1904.02689v2 [cs.CV] 24 Oct 2019. [Электронный ресурс]. URL: [1904.02689.pdf \(arxiv.org\)](#)

11. Xiong Y., Liao R., Zhao H., Hu R., Bai M., Yumer E., Urtasun R., UPSNet: A Unified Panoptic Segmentation Network. arXiv:1901.03784v2 [cs.CV] 3 Apr 2019. [Электронный ресурс]. URL: [1901.03784.pdf \(arxiv.org\)](#)

12. Geus D., Meletis P., Dubbelman G., Fast Panoptic Segmentation Network. arXiv:1910.03892v1 [cs.CV] 9 Oct 2019. [Электронный ресурс]. URL: [1910.03892.pdf \(arxiv.org\)](#)

13. Chang C., Chang S., Hsiao P., Fu L., EPSNet: Efficient Panoptic Segmentation Network with Cross-layer Attention Fusion. arXiv:2003.10142v3 [cs.CV] 24 Dec 2020. [Электронный ресурс]. URL: [2003.10142.pdf \(arxiv.org\)](#)

14. Kim D., Woo S., Lee J.-Y., Kweon I., Video Panoptic Segmentation. arXiv:2006.11339v1 [cs.CV] 19 Jun 2020. [Электронный ресурс]. URL: [2006.11339.pdf \(arxiv.org\)](#)

15. Semantic scene segmentation for robotics [Электронный ресурс]. URL: <https://www.sciencedirect.com/topics/computer-science/pixel-accuracy>

16. Understanding DICE COEFFICIENT [Электронный ресурс]. URL: <https://www.kaggle.com/code/yerramvarun/understanding-dice-coefficient>

17. Chhor G., Bartolome Aramburu C., Bougdal-Lambert I., Satellite Image Segmentation for Building Detection using U-net. [Электронный ресурс]. URL: <http://cs229.stanford.edu/proj2017/final-reports/5243715.pdf>

18. The DeepLab Family [Электронный ресурс]. URL: <https://medium.com/@calleris.enrico/the-deeplab-family-70d8b98262b5>

19. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L., SEMANTIC IMAGE SEGMENTATION WITH DEEP CONVOLUTIONAL NETS AND FULLY CONNECTED CRFS. arXiv:1412.7062v4 [cs.CV] 7 Jun 2016. [Электронный ресурс]. URL: [1412.7062.pdf \(arxiv.org\)](#)

20. Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L., DeepLab:

Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. arXiv:1606.00915v2 [cs.CV] 12 May 2017. [Электронный ресурс]. URL: [1606.00915.pdf \(arxiv.org\)](https://arxiv.org/pdf/1606.00915)

21. Chen L.-C., Papandreou G., Schroff F., Adam H., Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv:1706.05587v3 [cs.CV] 5 Dec 2017. [Электронный ресурс]. URL: [1706.05587.pdf \(arxiv.org\)](https://arxiv.org/pdf/1706.05587)

22. PASCAL-Context Dataset [Электронный ресурс]. URL: <https://www.cs.stanford.edu/~roozbeh/pascal-context/>

23. Zhou Z., Siddiquee M.M.R., Tajbakhsh N., Liang J., UNet++: A Nested U-Net Architecture for Medical Image Segmentation. arXiv:1807.10165v1 [cs.CV] 18 Jul 2018. [Электронный ресурс]. URL: [1807.10165.pdf \(arxiv.org\)](https://arxiv.org/pdf/1807.10165)

24. Lee C.-Y., Xie S., Gallagher P., Zhang Z., Tu Z., Deeply-supervised nets. In Artificial Intelligence and Statistics, pages 562–570, 2015.

25. MMsegmentation [Электронный ресурс]. URL: <https://github.com/open-mmlab/msegmentation>

26. OpenMMLab [Электронный ресурс]. URL: <https://github.com/open-mmlab>

27. INSTALL PYTORCH [Электронный ресурс]. URL: <https://pytorch.org/>

28. MMEngine [Электронный ресурс]. URL: <https://github.com/open-mmlab/mengine>

29. OpenMMLab [Электронный ресурс]. URL: <https://github.com/open-mmlab/mmcv>

30. Zhao H., Shi J., Qi X., Wang X., Jia J., Pyramid Scene Parsing Network. arXiv:1612.01105v2 [cs.CV] 27 Apr 2017. [Электронный ресурс]. URL: [1612.01105.pdf \(arxiv.org\)](https://arxiv.org/pdf/1612.01105)

31. Chen L.-C., Zhu Y., Papandreou G., Schrof F., Adam H., Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. [Электронный ресурс]. URL: [1802.02611.pdf \(arxiv.org\)](https://arxiv.org/pdf/1802.02611)

32. Detectron2 [Электронный ресурс]. URL: <https://github.com/facebookresearch/detectron2>

33. Meta AI [Электронный ресурс]. URL: <https://ai.facebook.com/>

34. Detectron [Электронный ресурс]. URL: <https://github.com/facebookresearch/Detectron>
35. Faster R-CNN and Mask R-CNN in PyTorch 1.0 [Электронный ресурс]. URL: <https://github.com/facebookresearch/maskrcnn-benchmark>
36. Caffe2 [Электронный ресурс]. URL: <https://caffe2.ai/>
37. Lin T.-Y., Goyal P., Girshick R., He K., Dollar P., Focal Loss for Dense Object Detection. arXiv:1708.02002v2 [cs.CV] 7 Feb 2018. [Электронный ресурс]. URL: [1708.02002.pdf \(arxiv.org\)](https://arxiv.org/abs/1708.02002)
38. Guler R.A., Neverova N., Kokkinos I., DensePose: Dense Human Pose Estimation In The Wild. arXiv:1802.00434v1 [cs.CV] 1 Feb 2018. [Электронный ресурс]. URL: [1802.00434.pdf \(arxiv.org\)](https://arxiv.org/abs/1802.00434)
39. Cai Z., Vasconcelos N., Cascade R-CNN: Delving into High Quality Object Detection. arXiv:1712.00726v1 [cs.CV] 3 Dec 2017. [Электронный ресурс]. URL: [1712.00726.pdf \(arxiv.org\)](https://arxiv.org/abs/1712.00726)
40. Kirillov A., Girshick R., He K., Dollar P., Panoptic Feature Pyramid Network. arXiv:1901.02446v2 [cs.CV] 10 Apr 2019. [Электронный ресурс]. URL: [1901.02446.pdf \(arxiv.org\)](https://arxiv.org/abs/1901.02446)
41. Chen X., Girshick R., He K., Dollar P., TensorMask: A Foundation for Dense Object Segmentation. arXiv:1903.12174v2 [cs.CV] 27 Aug 2019. [Электронный ресурс]. URL: [1903.12174.pdf \(arxiv.org\)](https://arxiv.org/abs/1903.12174)
42. D2Go brings Detectron2 to mobile [Электронный ресурс]. URL: <https://ai.facebook.com/blog/d2go-brings-detectron2-to-mobile/>
43. Segmentation models [Электронный ресурс]. URL: [https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch)
44. Fan T., Wang G., Li Y., Wang H., MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. IEEE Access Digital Object Identifier 10.1109/ACCESS.2020.3025372. [Электронный ресурс]. URL: [IEEE Xplore Full-Text PDF:](https://ieeexplore.ieee.org/abstract/document/9123456)
45. Chaurasia A., Culurciello E., LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. arXiv:1707.03718v1 [cs.CV] 14 Jun 2017.

[Электронный ресурс]. URL: [1707.03718.pdf \(arxiv.org\)](https://arxiv.org/pdf/1707.03718)

46. He K., Zhang X., Ren S., Sun J., Deep Residual Learning for Image Recognition. arXiv:1512.03385v1 [cs.CV] 10 Dec 2015. [Электронный ресурс]. URL: [1512.03385.pdf \(arxiv.org\)](https://arxiv.org/pdf/1512.03385)

47. Huang G., Liu Z., Maaten L., Densely Connected Convolutional Networks. arXiv:1608.06993v5 [cs.CV] 28 Jan 2018. [Электронный ресурс]. URL: [1608.06993.pdf \(arxiv.org\)](https://arxiv.org/pdf/1608.06993)

48. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A., Going deeper with convolutions. arXiv:1409.4842v1 [cs.CV] 17 Sep 2014. [Электронный ресурс]. URL: [1409.4842.pdf \(arxiv.org\)](https://arxiv.org/pdf/1409.4842)

49. Tan M., Le Q.V., EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946v5 [cs.LG] 11 Sep 2020. [Электронный ресурс]. URL: [EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks \(arxiv.org\)](https://arxiv.org/pdf/1905.11946)

50. Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861v1 [cs.CV] 17 Apr 2017. [Электронный ресурс]. URL: [1704.04861.pdf \(arxiv.org\)](https://arxiv.org/pdf/1704.04861)

51. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C., MobileNetV2: Inverted Residuals and Linear Bottleneck. arXiv:1801.04381v4 [cs.CV] 21 Mar 2019. [Электронный ресурс]. URL: [1801.04381.pdf \(arxiv.org\)](https://arxiv.org/pdf/1801.04381)

52. Howard A., Sandler M., Chu G., Chen L.-C., Chen B., Tan M., Wang W., Zhu Y., Pang R., Vasudevan V., Le Q.V., Adam H., Searching for MobileNetV3. arXiv:1905.02244v5 [cs.CV] 20 Nov 2019. [Электронный ресурс]. URL: [1905.02244v5.pdf \(arxiv.org\)](https://arxiv.org/pdf/1905.02244v5)

53. Simonyan K., Zisserman A., Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556v6 [cs.CV] 10 Apr 2015. [Электронный ресурс]. URL: [1409.1556.pdf \(arxiv.org\)](https://arxiv.org/pdf/1409.1556)

54. Kaggle [Электронный ресурс]. URL: <https://www.kaggle.com/>

55. Deconvolution and Checkerboard Artifacts [Электронный ресурс]. URL: <https://distill.pub/2016/deconv-checkerboard/>



56. ResNet: A Simple Understanding of the Residual Networks [Електронний ресурс]. URL: <https://medium.com/swlh/resnet-a-simple-understanding-of-the-residual-networks-bfd8a1b4a447>
57. Understanding EfficientNet — The most powerful CNN architecture [Електронний ресурс]. URL: <https://medium.com/mllearning-ai/understanding-efficientnet-the-most-powerful-cnn-architecture-eaeb40386fad>
58. AutoML [Електронний ресурс]. URL: <https://cloud.google.com/automl>
59. Cross-Entropy Loss Function [Електронний ресурс]. URL: <https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e>
60. Focal Loss — What, Why, and How? [Електронний ресурс]. URL: <https://medium.com/swlh/focal-loss-what-why-and-how-df6735f26616>
61. Inria Aerial Image Labeling Dataset [Електронний ресурс]. URL: <https://project.inria.fr/aerialimagelabeling/>
62. Inria Aerial Image Labeling Dataset/Contest [Електронний ресурс]. URL: <https://project.inria.fr/aerialimagelabeling/contest/>
63. Image Data Augmentation for Computer Vision in 2022 (Guide) [Електронний ресурс]. URL: <https://viso.ai/computer-vision/image-data-augmentation-for-computer-vision/>
64. LeCun Y., Bottou L., Bengio Y., Haffner P., Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE ( Volume: 86, Issue: 11, November 1998). С. 2278 – 2324. [Електронний ресурс]. URL: [http://vision.stanford.edu/cs598\\_spring07/papers/Lecun98.pdf](http://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf)
65. Менеджмент стартап-проектів [Електронний ресурс] : навчальний наочний посібник для студентів спеціальностей 051 «Економіка», 073 «Менеджмент», 075 «Маркетинг» / О. А. Гавриш, К. О. Бояринова, М. О. Кравченко, К. О. Копішинська ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 37,15 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 435 с.