Одеський Національний Політехнічний Університет
Інститут комп'ютерних систем

Кваліфікаційна наукова
праця на правах рукопису

Тимченко Борис Ігорович

УДК 004.932.72'1

Дисертація

Нейромережеві методи аналізу планарних зображень в системах автоматизованого скринінгу

122 — Комп'ютерні науки
Інформаційні технології

Подається на здобуття наукового ступеня
доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

_____ Б. І. Тимченко

Науковий керівник: **Антощук Світлана Григорівна**,
доктор технічних наук, професор

Одеса — 2021

АНОТАЦІЯ

Тимченко Б. І. Нейромережеві методи аналізу планарних зображень в системах автоматизованого скринінгу. — Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 — Комп'ютерні науки. — ІКС ОНПУ, Одеса, 2021.

У вступі розглядаються комп'ютерні системи автоматизованого скринінгу, в яких виникає задача аналізу зображень на основі штучних нейронних мереж. Виявлено, що в таких системах основними методами аналізу планарних зображень є класифікація та семантична сегментація. Семантичною сегментацією називається задача присвоєння кожному пікселю зображення окремого заздалегідь визначеного класу. Наприклад, задача автоматизованого скринінгу є актуальною в медицині для виявлення та локалізації осередків раку, або захворювань сітківки ока. Також, такі системи використовуються на виробництві для виявлення дефектів продукції, і в кліматології для прогнозування змін клімату, чи локалізації стихійних лих.

В *першому розділі* дисертаційної роботи проведено аналіз методів розпізнавання, класифікації та семантичної сегментації планарних зображень, використання для цього методів глибокого навчання, та методів багатозадачного навчання. Проаналізовано метрики, що використовуються для оцінки достовірності в задачах класифікації та сегментації зображень. Виявлено, що достовірність прогнозів визначається як міра Дайса-Соренсена, або F1-міра. Наведено підходи використання багатозадачного навчання для класифікації та сегментації зображень в задачах автоматизованого скри-

нінгу.

Виявлено, що складність розробки систем автоматизованого скринінгу пов'язана з необхідністю вирішення наступних завдань:

1. Неузгодженість та незбалансованість навчальних наборів даних для задач автоматизованого скринінгу виникає через відносно малу кількість аномальних прикладів в популяціях та значний рівень похибки розмітників. Це створює проблему при навчанні глибоких нейронних мереж стандартними методами. Тому при побудові моделей класифікації та сегментації потрібно вирішувати задачу *навчання з урахуванням частково-помилкової розмітки.*
2. В задачах автоматизованого скринінгу важливо зменшення кількості хибно-позитивних результатів діагностики, через відносно малу розповсюдженість аномальних прикладів в популяціях та пов'язані з хибно-позитивними результатами подальші дії. Тому, при побудові систем класифікації та сегментації потрібно зосередитися на *зменшенні хибно-позитивних результатів.*
3. Попередні дослідження показали, що системи автоматизованого скринінгу, засновані на методах багатозадачного глибинного навчання, дозволяють підвищити достовірність класифікації та сегментації за рахунок спільного навчання на цих задачах. Однак, попередні дослідження *не враховують* можливої *комбінації результатів різних задач* для підвищення достовірності під час прогнозування.

В розглянутих системах автоматизованого скринінгу наявні нейромережеві підходи мають обмеження по достовірності прогнозів через нестачу точно розмічених тренувальних даних, оскільки отримання якісної розмітки від професіоналів є дуже дорогим, та іноді не є можливим в принципі.

Тому має місце *протиріччя* між підвищенням вимог щодо достовірності прогнозів нейромережевих моделей без підвищення витрат часу з одного боку, та необхідністю зниження витрат на отримання розмітки навчальних

даних.

Таким чином, виконано *перше завдання* дисертаційної роботи: проведено аналіз існуючих методів, алгоритмів та засобів аналізу планарних зображень в системах автоматизованого скринінгу і обґрунтувано вибір напрямку досліджень щодо побудови методу аналізу планарних зображень на основі нейронних мереж.

На основі зазначеного вище, сформульовано мету дисертаційної роботи, визначено об'єкт та предмет дослідження.

Метою роботи є підвищення достовірності класифікації та семантичної сегментації планарних зображень в системах автоматизованого скринінгу через розробку, удосконалення та розвиток методів аналізу планарних зображень на основі штучних нейронних мереж.

Об'єкт дослідження – Процес аналізу планарних зображень.

Предмет дослідження – Методи семантичної сегментації та класифікації планарних зображень, структури та моделі згорткових нейронних мереж.

У *другому розділі* виконано аналіз зашумлення розмітки в наборах даних для задач автоматизованого скринінгу який показав, що основними проблемами з розміткою даних в задачах скринінгу є:

1. Маски сегментації, що захоплюють сусідні з об'єктом пікселі;
2. Маски сегментації, що покривають об'єкт не повністю;
3. Відсутні маски сегментації для деяких об'єктів;
4. Присутні зайві маски на місцях, де немає об'єктів.

Таким чином, виконано *друге завдання* дисертаційного дослідження: здійснено аналіз зашумлення розмітки в наборах даних для задач автоматизованого скринінгу і обґрунтувано вибір напрямку досліджень щодо побудови моделей зашумлення розмітки в наборах даних.

На основі проведеного аналізу розроблено метод контрольованої генерації як зображень, так і масок сегментації що модифікуються відповідно до

випадково обраних недоліків для моделювання зазначених дефектів розмітки.

Контрольована генерація зображень виконується в два етапи: генерація фону та розташування на ньому довільної кількості об'єктів. Для генерації фону можуть бути використані як звичайні натуральні зображення, так і синтетичні текстури, чи заливка константним кольором.

В якості об'єктів використовуються зображення з наборів даних MNIST та FashionMNIST, що зумовлено трьома факторами:

— Можливість простого відділення об'єкта від фону;
— Наявність схожих елементів в різних класів (наприклад, цифри 1 та 7, або класи T-Shirt та Dress);
— Висока точність роботи сучасних нейронних мереж на цих наборах даних, що дозволяє сконцентруватися на впливі шуму в розмітці.

Помилки в розмітку штучно вводяться за допомогою випадкового застосування морфологічних операцій ерозії та дилатації з ядром випадкової форми до масок сегментації окремих об'єктів перед додаванням їх до загальної маски.

Окрім наборів даних зображень фону $\mathcal{X}_b$ та об'єктів $\mathcal{X}_f$, модель має наступні параметри:

1. Розмір зображення S_i в пікселях
2. Середній розмір об'єкта S_o в пікселях
3. Максимальне відхилення розміру об'єкта σ_o в відсотках
4. Максимальна кількість об'єктів на зображенні N
5. Ймовірності ерозії та дилатації маски кожного з об'єктів P_e та P_d
6. Розміри ядер ерозії та дилатації масок всіх об'єктів S_e та S_d

На основі моделі генеруються навчальний та тестовий набори даних, всі зображення в наборах генеруються незалежно.

Таким чином виконано *третє завдання* дисертаційного дослідження: розроблено параметричну модель набору даних із зашумленою розміткою

для задач класифікації та семантичної сегментації.

Можна сформулювати перший пункт наукової новизни: *вперше* запропоновано параметричну модель набору даних із зашумленою розміткою що відповідає характеристикам зашумлення в реальних задачах автоматизованого скринінгу.

Відповідно до проведеного аналізу, в *третьому розділі* запропоновано метод підвищення достовірності результатів класифікації та семантичної сегментації за рахунок використання багатозадачного навчання нейронних мереж.

Запропонований метод складається з двох етапів: етапу навчання нейронних мереж та етапу прогнозування результатів.

На етапі навчання нейронної мережі в умовах частково-помилкової розмітки навчальних даних запропоновано використати метод багатозадачного навчання з використанням задач, похідних від оригінальної. Запропонований метод спирається на припущення, що існує семантично-близька більш загальна задача, для якої розмітка навчальних даних є більш влучною, ніж для вихідної задачі.

На відміну від попередніх методів, що спираються на вивчення більш детальних семантично-близьких задач, в розробленому методі використання більш точних даних, для загальніших задач дозволяє покращити роздільність внутрішніх представлень нейронної мережі, що, в свою чергу, покращує результати на вихідній задачі. Також, оскільки задачі є семантично близькими, не відбувається конфлікту градієнтів, що є типовим при навчанні семантично-різнорідних задач.

Так, більш загальною до задачі семантичної сегментації є задача класифікації. В такому контексті, задача класифікації зводиться до задачі навчання за набором зразків: замість розмітки кожного з об'єктів для всіх класів на зображенні, зображення являє собою мішок з одним, чи декількома об'єктами та відповідним маркуванням, чи є об'єкти заданих класів

на зображенні.

Для зменшення впливу помилкової частини розмітки, запропоновано зміну до функції втрат при навчанні нейронних мереж: функцію втрат для задачі з менш точною розміткою обмежено зверху, таким чином, при навчанні на декількох задачах, градієнти від неправильної розмітки не впливають на процес навчання:

$$\mathcal{L}] = min(L, \theta)$$

де θ - поріг обмеження функції втрат.

Для функції обмеження зверху градієнт визначено лише на проміжку $(-\inf, \theta]$, тому для проміжку $(\theta, \inf)$ градієнт встановлено рівним нулю:

$$\nabla min(L, \theta) = \begin{cases} 1 & L \in (-\inf, \theta] \\ 0 & L \in (\theta, \inf) \end{cases}$$

Так, виконано *четверте завдання* дисертаційного дослідження: розвинуто нейромережевий метод аналізу планарних зображень за рахунок навчання на декількох семантично-близьких задачах одночасно, щоб підвищити точність класифікації та семантичної сегментації планарних зображень

Можна сформулювати *другий* пункт наукової новизни: *вперше* розроблено метод багатозадачного навчання штучних нейронних мереж для роботи в умовах частково помилкової розмітки навчального набору даних в задачах автоматизованого скринінгу

З використанням запропонованого методу багатозадачного навчання, розроблено метод об'єднання семантично-близьких задач на етапі прогнозування, для того, щоб підвищити достовірність класифікації та сегментації планарних зображень без підвищення витрат часу.

За наявності маски сегментації M та результату класифікації C з на-

вченої багатозадачної нейронної мережі отримання покращених результатів обчислюється як сумісна ймовірність приналежності пікселя масці та наявності відповідного класу на зображенні незалежно:

$$M_{refined} = \sigma(M) \cdot \sigma(C)$$

$$C_{refined} = \sum \sigma(M) \cdot \sigma(C)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Таким чином, виконано *п'яте завдання* дисертаційного дослідження: розробити метод об'єднання семантично-близьких задач на етапі прогнозування, для того, щоб підвищити достовірність класифікації та сегментації планарних зображень без підвищення витрат часу.

Можна сформулювати *третій* пункт наукової новизни: *вперше* розроблено метод об'єднання семантично-близьких задач на етапі прогнозування, для того, щоб підвищити достовірність класифікації та сегментації планарних зображень без підвищення витрат часу.

Для покращення можливостей інтерпретації прогнозів моделей, було вдосконалено метод сегментації важливих для класифікації ознак зображення за рахунок використання методів багатозадачного навчання, що дозволило його використання в умовах відсутності розмітки семантичної сегментації в навчальному наборі даних.

В основі запропонованого методу лежить ітеративне уточнення карти ознак сегментації за допомогою направлення градієнтів від задачі класифікації:

Для карти ознак сегментації M та результату класифікації C обчислюються уточнені карти ознак сегментації та класифікації $M_{refined}$ та $M_{refined}$ відповідно:

$$M_{refined} = \sigma(M) \circ C$$

$$C_{refined} = \frac{\sum_h^H \sum_w^W M_{refined(hw)}}{\sum_h^H \sum_w^W \sigma(M_{hw})}$$

де $\sigma(x)$ - логістична функція.

Таке ітеративне уточнення спонукає нейронну мережу до призначення високих значень ($\sigma(M) \to 1$) для важливих ознак, що є спільними на зображеннях з навчального набору даних.

На етапі прогнозування використовується як задача класифікації, так і задача сегментації. Тільки у тому випадку, коли вихід класифікації перевищує заданий поріг T_c, виконується процедура декодування сегментації.

Оскільки в даному випадку маска сегментації M є неперервною, а її масштаб визначається процесом навчання, поріг бінаризації сегментації T_s може бути різним для різних зображень. Для вибору оптимального порогу T_s на кожному з вхідних зображень, в процесі пост-обробки запропоновано використати адаптивну бінарізацію за методом Оцу, щоб уникнути необхідності калібрації прогнозів нейронної мережі.

Після виконання бінаризації, над отриманою маскою $M_t = M > T_s$ виконується морфологічна операція ерозії з квадратним ядром розміру 1% від розміру зображення для того, щоб позбутися малих можливо хибнопозитивних регіонів маски:

$$M_e = M_t \ominus k$$

де k - ядро ерозії.

Таким чином, виконано *шосте* завдання дисертаційного дослідження: вдосконалено метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки семантичної сегментації в навчальному наборі даних.

Це відповідає *четвертому* пункту наукової новизни: *удосконалено* метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки для сегментації в навчальному наборі даних

У *четвертому розділі* описано розроблені інструментальні засоби, що реалізують запропоновані методи.

Інструментальні засоби розроблено мовою програмування *Python* з використанням фреймворку автоматичного диференціювання *PyTorch*.

Таким чином, виконано *сьоме* завдання дисертаційного дослідження: спроектовано та розроблено інструментальні засоби, що виконують розроблені моделі зашумлених даних, та методи навчання нейронних мереж та прогнозування результатів.

У *п'ятому розділі* описано випробування розробленого методу та інструментальних засобів в рамках експериментів:

Експерименти проводилися з використанням платформи для змагань з наук про дані Kaggle, кожний експеримент є окремим змаганням:

— Сегментація дефектів сталевих листів. Набір даних *Severstal: Steel Defect Detection*. Підвищення F1-міри на 4.2%.

— Сегментація формацій хмар. Набір даних *Understanding Clouds from Satellite Images*. Підвищення F1-міри на 3.9%

— Класифікація стадії осередків діабетичної ретинопатії. Набір даних *APTOS 2019 Blindness Detection*. Підвищення F1-міри 2.1%;

— Класифікація наявності раку шкіри та сегментація родимок. Набір даних *SIIM-ISIC Melanoma Classification*. Підвищення F1-міри 3.5%

Розроблені в роботі методи та інструментальні засоби отримали **впровадження** в програмний продукт SafetyRadar компанії *VITech Lab*.

Таким чином, мета дисертаційного дослідження досягнута і виконано *восьме* завдання, проведення експериментальних досліджень розроблених методів.

Ключові слова: аналіз зображень, автоматизований скринінг, глибокі нейронні мережі, багатозадачне машинне навчання, класифікація, семантична сегментація, функції втрат, неточна розмітка.

ЗМІСТ

Вступ **16**

Розділ 1. **Аналіз методів, алгоритмів і програмних засобів обробки планарних зображень** **21**

1.1. Аналіз задачі автоматизованого скринінгу 21

1.1.1. Задачі автоматизованого скринінгу в медицині 21

1.1.2. Задача автоматизованого скринінгу на виробництві . 26

1.1.3. Задачі автоматизованого скринінгу в метеорології . . 26

1.2. Аналіз методів та алгоритмів аналізу зображень 27

1.3. Методи глибинного машинного навчання 28

1.3.1. Згорткові нейронні мережі 28

1.3.2. Визначні сучасні архітектури ШНМ класифікації зображень 31

1.3.3. Визначні архітектури ШНМ для сегментації зображень 40

1.4. Методи багатозадачного глибинного машинного навчання . 47

1.4.1. Визначення багатозадачного машинного навчання . . 48

1.4.2. Трансферне навчання 49

1.4.3. Жорсткий розподіл параметрів 49

1.4.4. М'який розподіл параметрів 51

1.4.5. Дистиляція прогнозів моделі 53

1.4.6. Багатозадачне навчання в задачах класифікації та сегментації 53

1.5. Аналіз методів оцінки якості класифікації та сегментації зображень в задачі автоматизованого скринінгу 57

1.5.1. Помилки 1 і 2 роду та матриця невідповідностей . . . 57

1.5.2. Влучність, чутливість та специфічність 58
1.5.3. Міра F1 та індекс подібності Соренсена 59
1.5.4. Каппа Коена 59
1.6. Висновки та постановка задачі дослідження 60

Розділ 2. **Модель шумних даних в задачі семантичної сегментації** **63**
2.1. Аналіз зашумлення розмітки сегментації в задачі автоматизованого скринінгу 63
2.2. Моделювання зашумлення розмітки сегментації в задачі автоматизованого скринінгу 63
2.2.1. Алгоритм генерації наборів даних 63
2.2.2. Види згенерованих зображень 66
2.2.3. Висновки до другого розділу 72

Розділ 3. **Методи поєднання задач глибинного багатозадачного навчання** **73**
3.1. Метод багатозадачного навчання ШНМ в умовах частково помилкової розмітки семантичної сегментації 73
3.1.1. Генерація похідної задачі до задачі семантичної сегментації 73
3.1.2. Автоматична фільтрація помилкової розмітки 74
3.1.3. Архітектура багатозадачної нейронної мережі 77
3.1.4. Метод зниження хибно-позитивних результатів семантичної сегментації за рахунок використання задачі класифікації 81
3.2. Метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки для сегментації в навчальному наборі даних 84
3.2.1. Напівавтоматичне навчання ШНМ в задачі сегментації 85

3.2.2. Прогнозування результатів сегментації 86
3.2.3. Висновки до третього розділу 87

Розділ 4. **Інструментальні засоби** **88**
4.1. Узагальнена структура системи автоматизованого скринінгу 88
4.2. Узагальнена структура системи навчання глибоких нейронних мереж 88
4.3. Узагальнена структура прогнозування за допомогою глибоких нейронних мереж 88

Розділ 5. **Експериментальні дослідження в різних задачах скринінгу** **89**
5.1. Експерименти на синтетичних даних 89
5.1.1. Метод оцінки прогнозів нейронної мережі 89
5.1.2. Структура нейронних мереж 90
5.1.3. Параметри експериментів 90
5.2. Експерименти на реальних даних 95
5.2.1. Класифікація стадії осередків діабетичної ретинопатії 95
5.2.2. Детальний опис процедури навчання 96
5.3. Класифікація раку шкіри та сегментація родимок 98
5.3.1. Опис набору даних SIIM-ISIC Melanoma Classification 98
5.3.2. Детальний опис структури нейронної мережі 100
5.3.3. Детальний опис процедури навчання 100
5.3.4. Результати експерименту 100
5.4. Класифікація та сегментація формацій хмар 100
5.4.1. Опис набору даних Understanding Clouds from Satellite Images 100
5.4.2. Детальний опис структури нейронної мережі 101
5.4.3. Результати експерименту 102

Висновки **103**

Список використаних джерел **104**

ВСТУП

Обґрунтування вибору теми дослідження. В наш час активно розвиваються засоби превентивного управління в різних сферах людського життя. Сучасна медицина та виробництво активно розвиваються та стають все складнішими. Сучасна кліматологія здатна прогнозувати повені та урагани. Раннє діагностування захворювань, завчасне виявлення дефектів у виробах, та своєчасне прогнозування руху хмар дозволяє зменшувати витрати, зберігати природні ресурси та вчасно реагувати на зміни в довкіллі.

Задача автоматизованого скринінгу полягає в виявленні прихованих проблем на ранньому етапі без участі людини, поки ціна реагування на них є низькою. Серед небажаних ефектів скринінгу - можливість помилкової діагностики, створення неправдивого почуття впевненості у відсутності проблеми. З цих причин скринінгові дослідження повинні мати достатню чутливість і допустимий рівень специфічності.

Програмно-апаратні засоби відіграють одну з основних ролей в задачі автоматизованого скринінгу. Так, за допомогою фотокамер неперервно досліджуються сталеві листи для пошуку дефектів. Використання комп'ютеризованих офтальмоскопів дозволяє збільшити обсяги тестування захворювань ока в країнах, що розвиваються. Спеціальні супутники передають зображення земної поверхні в реальному часі, що допомагає вченим прогнозувати погоду.

Класифікація та семантична сегментація зображень є важливою складовою систем автоматизованого скринінгу і дозволяє виділяти об'єкти певного типу. Так, класифікація може показати на наявність проблеми, а семантична сегментація - на її локалізацію, що спрощує подальшу роботу людини.

Особливою популярністю в задачах обробки зображень користуються штучні нейронні мережі (ШНМ), які на вході отримують зображення, а на виході можуть давати як мітку класу, так і карту сегментації. Однією з основних проблем при роботі з ШНМ є необхідність великої кількості добре розмічених даних для навчання.

На жаль, в задачах скринінгу часто немає можливості отримати достатньої кількості розмічених зображень, які б мали всю необхідну розмітку для виконання задачі, однак існують набори даних з розміткою для схожих задач. Можливість використання таких наборів даних для виконання близьких задач дозволяє значно розширити області використання ШНМ.

Підвищення точності класифікації та сегментації планарних зображень в системах автоматизованого скринінгу є актуальною науковою задачею.

Зв'язок роботи з науковими програмами, планами, темами, грантами. Вказується, в рамках яких програм, тематичних планів, наукових тематик і грантів, зокрема галузевих, державних та/або міжнародних, виконувалося дисертаційне дослідження, із зазначенням номерів державної реєстрації науково-дослідних робіт і найменуванням організації, де виконувалася робота.

Мета і завдання дослідження. Метою дослідження є підвищення достовірності класифікації та семантичної сегментації планарних зображень через розробку, удосконалення та розвиток методів обробки планарних зображень на основі штучних нейронних мереж.

Для досягнення поставленої мети в роботі вирішено наступні задачі:

— Проведено аналіз існуючих методів, алгоритмів та засобів обробки планарних зображень в системах автоматизованого скринінгу

— Розроблено метод багатозадачного навчання штучних нейронних мереж для роботи в умовах частково помилкової розмітки навчального набору

— Розроблено метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки семантичної сегментації в навчальному наборі даних
— Удосконалено архітектуру штучної нейронної мережі для багатозадачного навчання
— Розвинуто нейромережевий метод обробки планарних зображень, який за рахунок навчання на декількох задачах одночасно підвищить точність класифікації та семантичної сегментації планарних зображень.
— Спроектовано та розроблено інструментальні засоби, що виконують розроблений метод
— Проведено випробування розробленого методу та інструментальних засобів в рамках експерименту

Об'єкт дослідження. Процес програмної обробки планарних зображень

Предмет дослідження. Методи семантичної сегментації та класифікації планарних зображень, структури та моделі згорткових нейронних мереж.

Методи дослідження. Для проведення досліджень застосовуються методи комп'ютерного зору для попередньої обробки зображень, методи математичної оптимізації для навчання штучних мереж.

Змістовно відзначається, що саме досліджувалось кожним методом; обґрунтовується вибір методів, що забезпечують достовірність отриманих результатів та висновків.

Наукова новизна отриманих результатів.

вперше:

— Розроблено метод багатозадачного навчання штучних нейронних мереж для роботи в умовах частково помилкової розмітки навчаль-

ного набору даних в задачах автоматизованого скринінгу

— Розроблено метод зниження дисперсії прогнозів нейронної мережі за рахунок використання результатів семантично-близьких задач при багатозадачному навчанні

— Розроблено метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки для сегментації в навчальному наборі даних

удосконалено:

— Архітектуру штучної нейронної мережі для багатозадачного навчання

отримав подальший розвиток:

— Нейромережевий метод обробки планарних зображень, який за рахунок навчання на декількох задачах одночасно підвищує точність класифікації та семантичної сегментації планарних зображень.

Практичне значення отриманих результатів. Надаються відомості про використання результатів досліджень або рекомендації щодо їх практичного використання.

Особистий внесок здобувача. Якщо у дисертації використано ідеї або розробки, що належать співавторам, разом з якими здобувачем опубліковано наукові праці, обов'язково зазначається конкретний особистий внесок здобувача в такі праці або розробки; здобувач має також додати посилання на дисертації співавторів, у яких було використано результати спільних робіт.

Апробація матеріалів дисертації. Основні результати дослідження доповідалися на наукових конференціях різного рівня та наукових семінарах. Це такі конференції:

Структура та обсяг дисертації. Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків.

Робота містить X сторінок основного тексту, Y рисунків та Z таблиць. Загальний обсяг дисертації – N сторінок, список літературних джерел налічує K найменувань. Дисертація містить I додатків, розміщених на J сторінках.

РОЗДІЛ 1
АНАЛІЗ МЕТОДІВ, АЛГОРИТМІВ І ПРОГРАМНИХ ЗАСОБІВ ОБРОБКИ ПЛАНАРНИХ ЗОБРАЖЕНЬ

1.1. Аналіз задачі автоматизованого скринінгу

Задача автоматизованого скринінгу постає, коли виникає необхідність ідентифікації рідкісних випадків в великій популяції. Метою скринінгу є раннє виявлення аномальних екземплярів в популяції, що дозволяє забезпечити раннє реагування з розрахунку на завчасне полегшення негативних наслідків. Попри те, що скринінг сприяє ранньому реагуванню, існує можливість гіпердіагностики або помилкової діагностики, а також створення неправдивого почуття упевненості у відсутності проблеми. З цих причин скринінгові дослідження повинні мати достатню чутливість і допустимий рівень специфічності [**?**]. Також, в задачі автоматизованого скринінгу важливою є локалізація ознак можливих проблем, для подальшого оцінювання людьми. [**?**]

1.1.1. Задачі автоматизованого скринінгу в медицині. Серед багатьох застосувань скринінгу в медицині, особливо нагальною є потреба в методах скринінгу для агресивних прихованих хвороб в регіонах, де в більшості населення немає доступу до професійного лікаря. Так, агресивними хворобами, для яких утруднене діагностування на ранніх стадіях є діабетична ретинопатія та рак шкіри.

1.1.1.1. Скринінг раку шкіри. Рак шкіри є найбільш розповсюдженим видом злоякісних пухлин, і саме меланома є причиною більшості смертей від раку. Світова проблема захворюваності на меланому стрімко

зростала за останні 50 років і стала проблемою, з якою намагаються боротися багато вчених з різних країн.

Меланома є п'ятим за поширеністю раком серед чоловіків та шостим за поширеністю раком серед жінок [1]. Подібно до інших типів раку, ранні та легкі стадії візуально навряд чи можна розрізнити. В наш час, дерматологи оцінюють кожну родинку пацієнта, щоб виявити незвичні осередки або такі, що, швидше за все, є злоякісними. Якщо меланому помітити вчасно, її можна вилікувати незначними оперативними втручаннями.

Огляд літератури Недавні дослідження в галузі автоматичного виявлення злоякісних утворень пов'язані з найсучаснішими підходами до глибокого навчання в розпізнаванні зображень. Набагато менше робіт використовують класичне машинне навчання та сконструйовані дослідниками ознаки.

Тут наведено найвпливовіші роботи в цій галузі. Так, Mustafa et al. [7] створив підхід з підібраними вручну ознаками (GrabCut для сегментації раку) та методом опорних векторів для класифікації ракових уражень. Також, Nasiri et al. [8] згенерував похідні зображення за допомогою декількох алгоритмів і використав на них метод k-найближчих моделей сусідів для вирішення завдання.

Брінкер та ін. [9] проводив експерименти із попередньо навченими на наборі даних ImageNet згортковими нейронними мережами, такими як ResNet-50, для класифікації ранніх стадій меланоми. Автори використали 4204 перевірених біопсією зображення меланоми та звичайних родинок. Крім того, були інтегровані новітні на той момент методи глибокого навчання: різні темпи навчання для різних частин нейронної мережі, зменшення темпу навчання на основі функції косинуса, стохастичний градієнтний спуск з перезапуском для того щоб уникнути локальних мінімумів.

Коделла та ін. [10] запропонували систему сегментації та класифікації меланоми за дермоскопічними зображеннями шкіри. Для класифікації

хвороб вони застосували ансамбль останніх методів машинного навчання, включаючи глибокі залишкові мережі, згорткові нейронні мережі тощо. Вони довели, що ансамблі здатні давати кращі результати, ніж моделі окремо.

Насірі та ін. [4] запропонували класифікацію уражень шкіри за допомогою глибокого навчання для раннього виявлення меланоми в системі міркувань на основі прецедентів (англ. case-based reasoning). Цей підхід був використаний для отримання схожих вхідних зображень із бази даних прецедентів запропонованої системи DePicT Melanoma Deep-CLASS для підвищення точності рекомендацій щодо запитуваної проблеми (наприклад, зображення родинки). Їх метод, що заснований на глибоких згорткових нейронних мережах, генерує ознаки з зображень, щоб використовувати їх у процесі пошуку в базі даних. Інтеграція цього підходу до DePicT Melanoma CLASS значно покращила ефективність класифікації зображень та якість рекомендаційної частини системи.

Дослідження в галузі багатозадачного навчання також проводили Сонг та ін. [5]. Вони запропонували нейронну мережу, яка може одночасно виконувати завдання детекції, класифікації та сегментації уражень шкіри, не вимагаючи додаткових етапів попередньої обробки або подальшої обробки. Подібну роботу представили Чен та співавт. [6], вони використали багатозадачну мережу U-Net для задачі детекції та сегментації. Янг та ін. [11] запропонував більш складну багатозадачну модель, яка одночасно вирішує завдання сегментації уражень та дві незалежні задачі бінарної класифікації, використовуючи спільності та відмінності між завданнями.

1.1.1.2. Скринінг діабетичної ретинопатії. Діабетична ретинопатія є одним із найбільш загрозливих ускладнень діабету, при якому пошкодження сітківки викликає сліпоту. Вона пошкоджує кровоносні судини тканини сітківки, викликаючи витік рідини та погіршення зору. Поряд із захворюваннями, що призводять до сліпоти, такими як катаракта і глау-

кома, ретинопатія є одним із найпоширеніших захворювань, згідно зі статистикою США, Великобританії та Сінгапуру.

Лікарі встановили чотири стадії діабетичної ретинопатії:

— Легка непроліферативна ретинопатія, найраніша стадія, коли можуть виникати лише мікроаневризми;
— Помірна непроліферативна ретинопатія, стадію якої можна описати втратою здатності кровоносних судин до транспортування крові через набряк з прогресуванням захворювання;
— Важка непроліферативна ретинопатія призводить до обмеженого кровопостачання сітківки через підвищений набряк великої кількості кровоносних судин;
— Проліферативна діабетична ретинопатія - це запущена стадія, коли фактори росту, що виділяються сітківкою, активізують проліферацію нових кровоносних судин, зростаючи вздовж оболонки сітківки в склоподібному тілі, заповнюючи око.

Щонайменше 56% нових випадків можна уникнути за допомогою належного та своєчасного лікування та скринінгу очей. Однак, початкова стадія цього захворювання не має помітних для пацієнта ознак, і виявити його на ранній стадії є справжньою проблемою. Більш того, добре навчені діагности іноді не можуть вручну оцінити стадію за діагностичними зображеннями очного дна пацієнта.

Огляд літератури

Багато дослідницьких зусиль було присвячено проблемі раннього виявлення діабетичної ретинопатії. Перш за все, дослідники намагалися використовувати класичні методи комп'ютерного зору та машинного навчання, щоб забезпечити відповідне рішення цієї проблеми.

Наприклад, Priya et al. запропонували підхід на основі комп'ютерного зору для виявлення діабетичної ретинопатії за допомогою кольорових зображень очного дна. Автори створили набір ознак із вихідного зображе-

ння, використовуючи класичні методи обробки зображень, і використали метод опорних векторів для бінарної класифікації. Їх метод досяг чутливості 98%, специфічності 96% та точності 97% на тестовому наборі з 250 зображень.

Крім того, інші дослідники намагалися використати інші моделі для багатокласової класифікації, наприклад, застосовуючи аналіз головних компонент до зображень та використовуючи дерева рішень, Баєсові класифікатори або метод найближчих сусідів conde з найкращими результатами 73.4% точності та 68.4% для F-міри, використовуючи набір даних із 151 зображення з різною роздільною здатністю.

Зі зростанням популярності підходів, заснованих на глибокому навчанні, з'явилися методи, які застосовують глибокі згорткові нейронні мережі до цієї проблеми. Пратт та ін. pratt розробили архітектуру нейронної мережі та використали аугментацію даних, яка може ідентифікувати складні ознаки захворювання, пов'язані із задачею класифікації, такі як мікроаневризми, ексудат та крововиливи в сітківку ока, і, отже, автоматично діагностувати стадію захворювання. Цей метод досяг чутливості 95% і точності 75% на 5000 валідаційних зображень. Крім того, є й інші роботи про використання глибоких штучних нейронних мереж від інших дослідників lam, li.

Asiri та ін. провели аналіз значної кількості доступних методів та наборів даних, висвітливши їх плюси та мінуси asiri. Крім того, автори вказали на проблеми, які слід вирішити при розробці та вивченні ефективних та надійних алгоритмів глибокого навчання для різних проблем діагностики діабетичної ретинопатії, та звернули увагу на напрямки подальших досліджень.

Інші дослідники також намагалися здійснити трансферне навчання за допомогою згорткових нейронних мереж. Хагош та ін. hagos спробував навчити InceptionNet V3 для класифікації 5 класів з нейронною мережею,

натренованою на наборі даних ImageNet і досяг точності 90,9Сарки та ін. sarki провів дослідження з навчання різних архітектур, зокрема ResNet50, Xception, DenseNets та VGG за допомогою попереднього навчання на наборі даних ImageNet і досяг найкращої точності 81.3%. Обидві групи дослідників використовували набори даних, які надавали APTOS та Kaggle.

1.1.2. Задача автоматизованого скринінгу на виробництві.

1.1.3. Задачі автоматизованого скринінгу в метеорології. Однією з найцінніших особливостей визначення кліматичної моделі Землі є поведінка хмар. Однак дослідження їх поведінки є однією з найскладніших частин, оскільки вимагає досконалого розуміння всіх процесів в атмосфері. Класифікація різних типів організації хмар допомагає покращити розуміння цих хмар, що, в свою чергу, допоможе нам побудувати кращі кліматичні моделі.

Дослідники з Інституту метеорології імені Макса Планка зібрали найбільший набір даних, що складається з приблизно 10 000 фотографій хмар в видимому спектрі з супутників Terra і Aqua MODIS. Завдяки краудсорсинговій спільноті Zooniverse, вони створили анотований набір даних, в якому хмари позначені чотирма типами масок: цукор, квітка, гравій, риба (англ. Sugar, Flower, Gravel, Fish). Однак, через недосконалу процедуру маркування фотографій, та через те, що маркування було зроблене непрофесіоналами, розмітка є частково помилковою. Так, маски сегментації класів містять багато пікселів, що належать до фону, та деякі хмари не мають масок сегментації.

Огляд літератури Щоб описати широкий спектр дослідницьких робіт, розглянуто роботи, що використовують як супутникові зображення в видимому спектрі, так і в інших діапазонах: мультиспектральні та інфрачервоні.

Один із методів виявлення хмар був розроблений Чжу та співавт. [2], во-

ни запропонували метод під назвою Fmask (функція маски) для виявлення хмар та їхніх тіней на зображеннях із супутника Landsat 7. Fmask використовує підходи, засновані на правилах на основі фізичних властивостей хмар, щоб відокремити потенційні хмарні регіони від чистого неба. В якості вхідних даних метод використовує інформацію із семидіапазонних датчиків Enhanced Thematic Mapper (ETM) та Enhancet Themtic Mapper Plus (ETM+), якими обладнаний супутник Landsat. Альтернативний підхід був запропонований Harb et al. [3], автори проаналізували мультиспектральні дані середньої роздільної здатності з супутників програми CBERS. Даний алгоритм використовує набір математичних операцій над спектральними смугами, щоб покращити видимість хмар та їхніх тіней.

Ці методи дають точні результати, але в значній мірі залежать від моделей датчиків (оскільки вони є методами, заснованими на правилах), а пропоновані рішення не є масштабованими до інших типів датчиків.

Hu et al. [4] представили більш загальне рішення, використовуюч методи комп'ютерного зору для виявлення декількох низькорівневих ознак, таких як колір, особливості текстури тощо. Для оцінки піксельних масок автори використали класичні алгоритми машинного навчання. Озкан та ін. застосували глибокі нейронні мережі (такі як Feature Pyramid Network) для сегментації хмар із низькоорбітальних RGB-зображень супутників Gokturk-2 та RASAT.

1.2. Аналіз методів та алгоритмів аналізу зображень

Аналізом зображень називається отримання значущої інформації з цифрових зображень за допомогою методів обробки зображень.

Основними методами аналізу зображень є:

— Виділення ознак

— Класифікація зображень

— Сегментація зображень

— Детекція об'єктів

— Розпізнавання паттернів

Зазвичай, першим кроком аналізу зображень є виділення ознак.

Характерним для задачі скринінгу є клас планарних зображень. Особливістю планарних зображень є постійний масштаб об'єктів, та можливість знехтувати перспективними спотвореннями [**?**].

1.3. Методи глибинного машинного навчання

Глибинне машинне навчання - це частина більш широкого сімейства методів машинного навчання, заснованих на штучних нейронних мережах (ШНМ). Алгоритми глибинного машинного навчання моделюють високорівневі абстракції за допомогою глибинного графу з декількох шарів, що побудовані з лінійних, чи нелінійних перетворень.

Штучні нейронні мережі — це обчислювальні системи, натхнені біологічними нейронними мережами. ШНМ ґрунтується на сукупності з'єднаних вузлів, що називають штучними нейронами (аналогічно до біологічних нейронів у головному мозку тварин). Кожне з'єднання (аналогічне синапсу) між штучними нейронами може передавати сигнал від одного до іншого. Штучний нейрон, що отримує сигнал, може обробляти його, й потім сигналізувати іншим нейронам, приєднаним до нього [**?**].

1.3.1. Згорткові нейронні мережі. Революція в розпізнавання образів була зроблена за допомогою згорткових нейронних мереж (ЗНМ). Раніше, для задач розпізнавання використовувалися фільтри, які обиралися вручну, а після них використовувався простий класифікатор. Велика перевага нейронних мереж, полягає в тому, що потрібні лише тренувальні дані. На основі даних, фільтри і класифікатори навчаються автоматично. Це стало особливо потужним методом в завданнях розпізнавання зображень.

Дані з зображень захоплюється за допомогою операції згортки. Використовуючи згорткові ядра для сканування цілого зображення, потрібно вивчити порівняно небагато параметрів відносно повнозв'язних штучних нейронних мереж. Окрім безпосередньо операції згортки, сучасні ЗНМ використовують нелінійні функції активації, операції підвибірки та нормалізації.

Згорткові шари. Згорткові шари в глибоких нейронних мережах є основним блоком, що дозволяє формувати карти ознак.

Активації. Функції активації забезпечують нелінійність нейронної мережі, виконуючи нелінійне перетворення карти ознак, отриманої від згорткового, або повнозв'язного шару.

Найбільш популярними функціями активації в даний час є [?]:

— Сигмоподібна (Sigmoid) [?]
— Гіперболічний тангенс (TanH) [?]
— Лінійний ректифікатор (ReLU) [?]
— Нещільний лінійний ректифікатор (LeakyReLU) [?]
— SoftPlus [?]
— Доповнена тотожна функція (Bent identity) [?]
— Swish [?]
— Mish [?]

Графіки зазначених функцій зображено на рис. 1.1.

Підвибірки. Шари підвибірки використовуються для зменшення просторової роздільної здатності та агрегації просторових ознак. В сучасних нейронних мережах використовуються шари підвибірки з функцією максимуму та середнього арифметичного. Також існують шари підвибірки з іншими функціями, але вони не набули загального застосування [?, ?, ?].

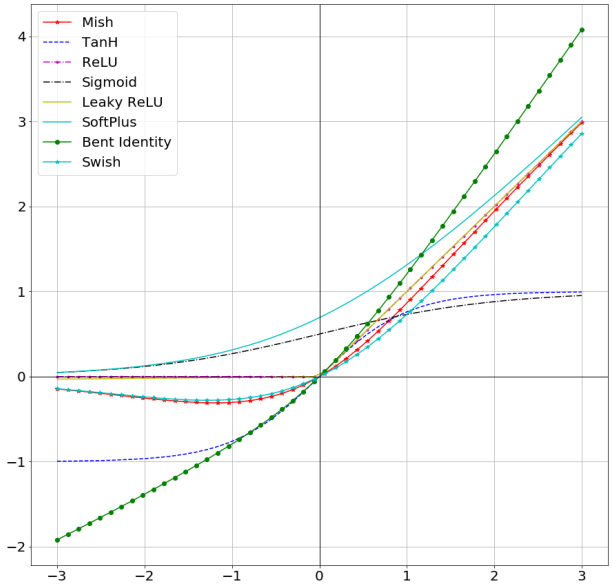

Рис. 1.1: Графіки популярних функцій активації [?]

Нормалізації. В сучасних нейронних мережах, шари нормалізації використовуються для стабілізації процесу навчання. Розподіл входів кожного шару змінюється під час навчання, оскільки змінюються параметри попередніх шарів. Це уповільнює навчання, вимагаючи нижчих темпів навчання та ретельної ініціалізації параметрів.

Нормалізація проводиться над картами ознак, або їх частинами, приводячи їх до розподілу з середнім значенням 0 та стандартним відхиленням 1. Різні типи нормалізацій які застосовуються в нейронних мережах включають:

— Пакетна нормалізація (batch normalization)[?]

— Пошарова нормалізація (layer normalization)[?]

— Поекземплярна нормалізація (instance normalization) [?]

— Групова нормалізація (group normalization) [?]

Візуалізація різних типів нормалізації зображена на рис. 1.2. Тут N - кількість карт ознак в пакеті, C - розмірність каналів карти ознак, H, W - просторова розмірність одного каналу. Синім виділено набори ознак, на яких обчислюється середнє значення та стандартне відхилення.

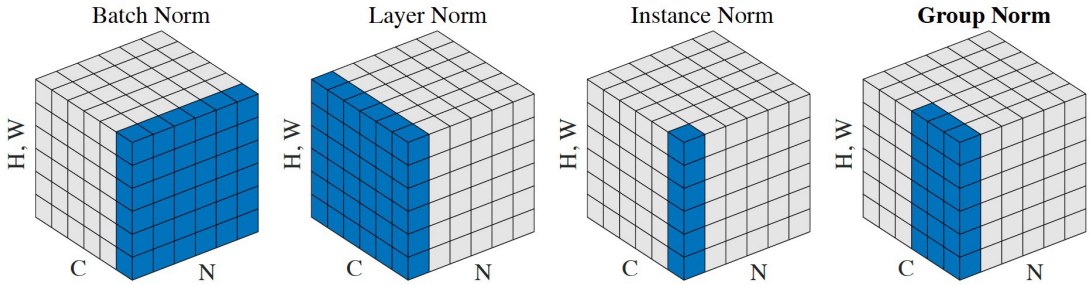

Рис. 1.2: Візуалізація різних типів нормалізації [?]

1.3.2. Визначні сучасні архітектури ШНМ класифікації зображень.

Нейронна мережа AlexNet. AlexNet є однією з найперших вдалих спроб застосування штучних згорткових нейронних мереж до зображень у великому масштабі. На відміну від більш ранніх робіт, автори досягли великого прориву на великому наборі даних ImageNet [?].

Архітектура нейронної мережі є логічним продовженням розробок Я. Лекуна [?] для розпізнавання рукописних літер за допомогою згорткових нейронних мереж. AlexNet складається з восьми шарів, перші п'ять з яких - згорткові, а останні три - повнозв'язними шарами. AlexNet використовує функцію активації ReLU, яка демонструвала покращені результати навчання в порівнянні сигмоподібною. Для збільшення поля зору, автори використали великий розмір ядра в перших згорткових шарах (11 та 5 пікселів відповідно)

Для того, щоб мати можливість навчання не великій кількості зображень порівняно високої роздільної здатності, автори розробили алгоритм паралельного навчання на декількох графічних процесорах (GPU).

Також, автори вперше показали можливість використання карт ознак як міри подібності для вхідних зображень та необхідність регуляризації для запобігання перенавчанню.

AlexNet вважається однією з найвпливовіших статей, опублікованих в

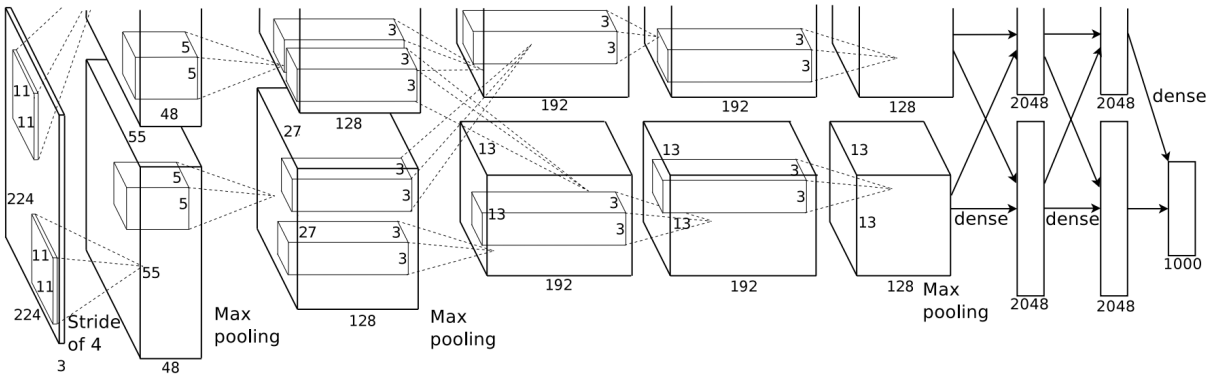

Рис. 1.3: Граф нейронної мережі AlexNet [**?**]

галузі комп'ютерного зору, завдяки чому було опубліковано багато інших публікацій із використанням CNN та GPU для прискорення глибокого навчання.

Нейронна мережа VGGNet. VGGNet є інкрементальним покращенням архітектури AlexNet, в якому адресовано проблему продуктивності та систематизовано залежність точності від глибини нейронної мережі. Покращення продуктивності в порівнянні з AlexNet досягається заміною великих розмірів ядра згортки (11 і 5 у першому та другому згортковому шарі, відповідно) кількома згортками з розміром ядра 3x3 одна за одною. Також, за рахунок використання більшої кількості нелінійних активацій, підвищується дикримінаційна здатність нейронної мережі. Так, три згортки з ядром 3x3 одна за одною та кроком 1 мають той самий розмір поля зору що й одна згортка з розміром ядра 7x7, але кількість задіяних параметрів становить $3 \cdot (3^2C^2)$ порівняно з 7^2C^2 параметрів для ядер розміром 7x7. Тут C - це кількість вхідних та вихідних каналів згорткового шару.

Автори порівнюють декілька нейронних мереж, які відрізняються лише кількістю шарів. Вихідна таблиця порівняння архітектур зазначена на рис. 1.4.

В даній роботі було продемонстровано, що глибина нейронної мережі вигідна для точності класифікації, а також того, що можна досягти високої

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Рис. 1.4: Порівняння архітектур VGGNet різної глибини [?]

точності за допомогою звичайної архітектури згорткової нейронної мережі.

Сімейство архітектур нейронних мереж Inception. Мережа Inception [?]стала важливою віхою в розвитку згорткових нейронних мереж. До цієї роботи, згорткові нейронні мережі являли собою послідовні операції згортки, активації та підвибірки. До сімейства Inception входять нейронні мережі, що були спеціально розроблені для підвищення точності та продуктивності.

Автори розглядають проблему варіативності розміру об'єктів на зображеннях: вибір правильного розміру ядра для операції згортки стає важким. Більше ядро краще підходить для виділення інформації, яка розповсюджена більш глобально, а менше ядро - для інформації, яка розповсюджується локально.

Для вирішення цієї проблеми, автори використовують декілька операцій згортки з різними розмірами ядра в одному шарі нейронної мережі. Оскільки виконання згорток з великою кількістю каналів є довгою операцією, автори зменшують кількість каналів за допомогою згорток з розміром ядра 1x1. Один такий шар називається *блоком Inception*, архітектура якого зображена на рис. 1.5

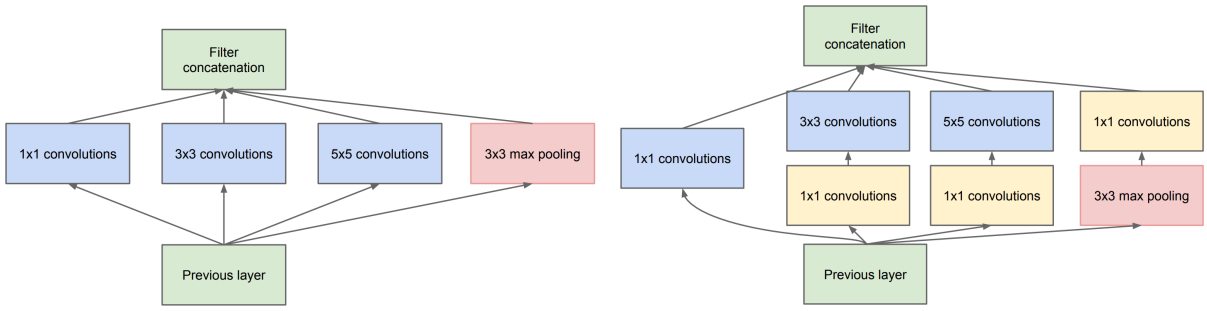

Рис. 1.5: Блок-схема Inception v1 до та після розкладу[?]

Нейронна мережа GoogLeNet [?] складалася з 22 шарів в глибину та була найглибшою на той час. Для того, щоб запобігти затуханню градієнтів, використовувалися додаткові класифікатори на менш глибоких рівнях,

результати яких було додано до функції втрат.

В наступних роботах [?], автори продовжують ідею зменшення кількості операцій, для чого вводять розкладання згорток з великим розміром ядра на декілька менших згорток. Так, автори пропонують розкладання згортки з ядром розміру 5x5 на дві згортки з розміром ядра 3x3, а згортки з розміром ядра 3x3 на дві згортки розміру 1x3 та 3x1 відповідно. Це дозволяє зменшити кількість параметрів в нейронній мережі при збереженні глибини, зменшуючи необхідні ресурси для обчислення.

Схема блоків Inception v2 та v3 зображено на рис. 1.6

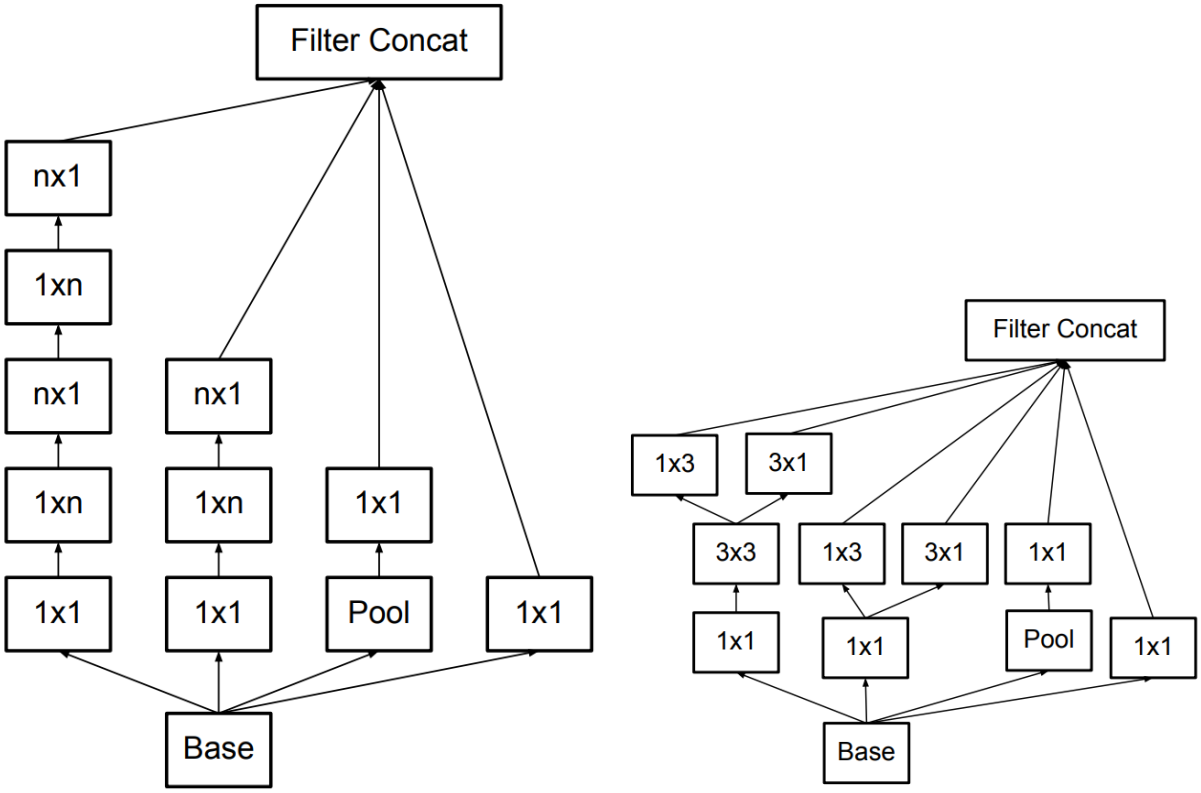

Рис. 1.6: Блок-схема Inception v2 до та після розкладу[?]

В нейронній мережі Inception v2 вперше було застосовано метод пакетної нормалізації [?], що дозволило значно пришвидшити навчання. Також, було показано, що додаткові класифікатори не мають впливу на затухання градієнтів, але мають властивості регуляризації, особливо на останніх стадіях навчання.

Сімейство архітектур нейронних мереж ResNet. Сімейство нейронних мереж ResNet було засновано на новому методі залишкового навчання, що був представлений в тій самій роботі, що й архітектура нейронної мережі [**?**]. Автори розв'язують проблему того, що збільшення глибини нейронної мережі зменшує точність як на навчальному, так і на тестувальному наборі даних.

В роботі показано, що метод стохастичного градієнтного спуску не придатний до оптимізації глибоких нейронних мереж. Так, наведено експеримент, в якому між шарами неглибокої нейронної мережі використано шари тотожного відображення або додаткові згорткові шари з функцією активації. Показано, що для більш глибокої нейронної мережі має існувати рішення з точністю не меншою, ніж для неглибокої нейронної мережі, але, на практиці, точність більш глибокої нейронної мережі залишається нижчою. Для збільшення глибини нейронних мереж без втрати точності, в роботі запропоновано метод залишкового навчання.

Нехай $\mathcal{H}_l(x_{l-1})$ - це відображення, що задається декількома шарами нейронної мережі, де x_{l-1} - це вхідний тензор до першого з цих шарів. Якщо припустити, що декілька нелінійних шарів можуть апроксимувати будь-яку нелінійну функцію, то логічно припустити, що вони можуть апроксимувати і залишкову функцію: $\mathcal{F}_l(x_{l-1}) = \mathcal{H}_l(x_{l-1}) - x_{l-1}$. Тоді, оригінальне відображення стає $\mathcal{H}_l(x_{l-1}) = \mathcal{F}_l(x_{l-1}) + x_{l-1}$. Хоча ці відображення є еквівалентними, алгоритму оптимізації «простіше» таким чином вивчити тотожну функцію, оскільки для цього достатньо наблизити ваги шарів нейронної мережі до нуля.

В роботі автори пропонують блоки нейронних мереж, що складаються з декількох згорткових шарів з функцією активації ReLU та об'єднання з вхідним тензором першого шару. Схема таких блоків зображена на рис. 1.7

Завдяки використанню таких блоків, стала можливою побудова надзвичайно глибоких нейронних мереж. В експериментах, автори запропонували

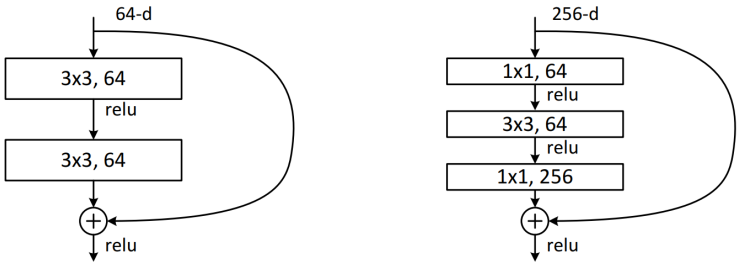

Рис. 1.7: Схема залишкових блоків нейронних мереж ResNet18 та ResNet50 [?]

архітектури, що складаються з 18, 34, 50, 101, 152 та 1202 згорткових шарів, та можуть бути оптимізовані стандартними методами (наприклад, методом стохастичного градієнтного спуску).

Сімейство архітектур нейронних мереж DenseNet. Архітектура нейронних мереж сімейства DenseNet [?] є логічним продовженням ідей, запропонованих в архітектурі ResNet. В даній архітектурі також використовується метод залишкового навчання, але, на відміну від архітектури ResNet, послідовні шари поєднані кожний з кожним.

Для того, щоб підвищити об'єм інформації, яким обмінюються шари нейронної мережі, автори пропонують ввести пряме з'єднання від кожного шару до всіх наступних:

$$x_l = \mathcal{H}([x_0, x_1, x_3 ... x_{l-1}]) \tag{1.1}$$

де $[x_0, x_1, x_3 ... x_{l-1}]$ - конкатенація всіх карт активації з шарів від 0 до $l-1$.

Така зміна архітектури дозволила останнім шарам перевикористовувати ознаки з попередніх, що, в свою чергу, дозволило зменишити кількість параметрів в нейронній мережі та підвищити точність.

Граф нейронної мережі DenseNet зображено на рис. 1.8.

На практиці, хоча архітектура дозволяє досягти більшої точності з використанням меншої кількості параметрів, на час виходу роботи, більшість реалізацій, заснованих на популярних фреймворках машинного навчання

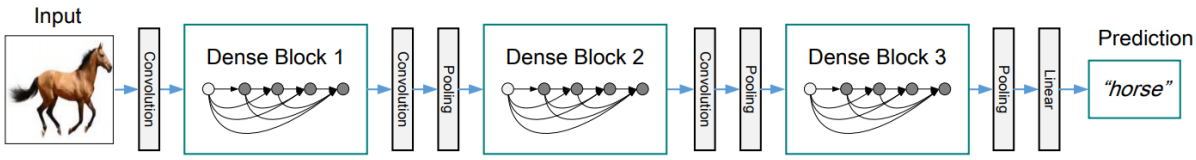

Рис. 1.8: Архітектура нейронної мережі DenseNet [?]

вимагали більшої кількості оперативної пам'яті через необхідність копіювання карт активацій для кожного з шарів. Таким чином, наївна реалізація архітектури DenseNet потребувала $\mathcal{O}(\frac{L(L-1)}{2})$ оперативної пам'яті. З розвитком фреймворків машинного навчання, була реалізована можливість перевикористання тензорів без копіювання, що дозволило знизити потребу оперативної пам'яті до $\mathcal{O}(L)$ [?].

Сімейство архітектур нейронних мереж ResNeXt. Архітектури сімейства ResNeXt [?] є логічним продовженням методів, застосованих в архітектурах Inception та ResNet. Основною ідеєю цієї архітектури є використання набору з декількох простих однакових трансформацій в просторі низької розмірності в кожному з шарів нейронної мережі.

Так, автори пропонують розділити згортковий шар на декілька паралельних трансформацій, результати з яких агрегуються:

$$\mathcal{H}(x_{l-1}) = \sum_{i=1}^{C} \mathcal{T}_i(x_{l-1}) \tag{1.2}$$

де $\mathcal{T}_i$ - це будь-яка нелінійна функція, що проектує вхід x_{l-1} до простору низької розмірності, трансформує його, та проектує результат назад в простір високої розмірності. Тут C - потужність множини трансформацій.

Графічну репрезентацію блоку ResNeXt зображено на рис. 1.9

В роботі автори проводять експерименти відносно важливості параметру C відносно інших, та показують, що збільшення C підвищує точність більше, ніж підвищення кількості каналів згорток. Також, в роботі показано, що такий метод здатний створювати кращі репрезентаціїї вхідних

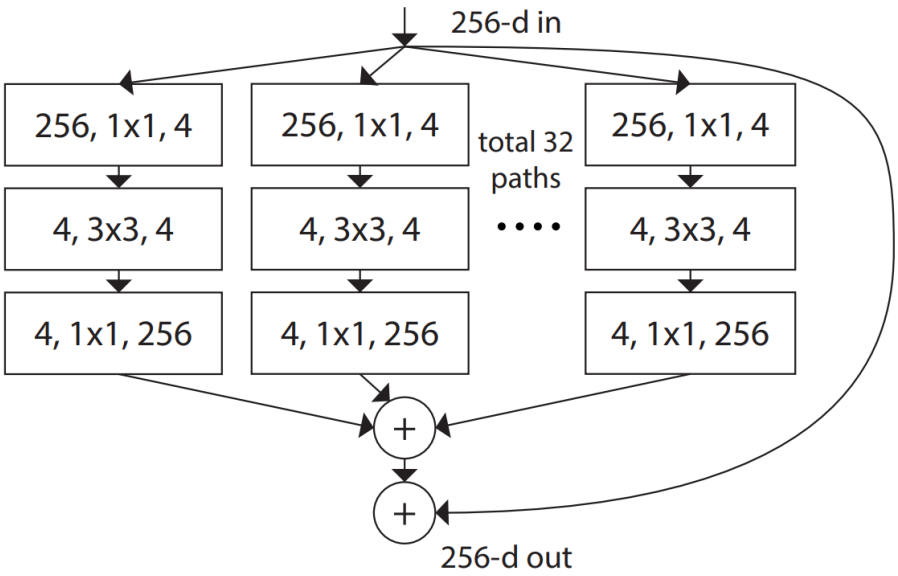

Рис. 1.9: Графічна репрезентація блоку ResNeXt [**?**]

даних.

Сімейство нейронних мереж EfficientNet. Сімейство нейронних мереж EfficientNet також є логічним продовженням архітектури ResNet. В даній роботі вперше розглядається узагальнений метод масштабування згорткових нейронних мереж для збільшення точності під заданий обчислювальний бюджет.

Автори пропонують метод, за допомогою якого можна масштабувати одночасно вхідну роздільну здатність зображення, глибину та ширину нейронної мережі. Нведено спостереження, що ці величини не є незалежними.

Інтуїтивно зрозуміло, що для зображень із більш високою роздільною здатністю необхідно збільшити глибину мережі, щоб більші збільшити поле зору та захопити більше пікселів у більшому зображенні. Відповідно, також необхідно збільшувати ширину мережі, коли роздільна здатність збільшується, для того, щоб мати можливість розпізнати більшу кількість ознак на зображеннях із високою роздільною здатністю.

Для цього, структура нейронної мережі формулюється як задача максимізації точності в залежності від ширини, глибини та роздільної здатності:

$$max\ Accuracy(\mathcal{N}(d, w, r)) : \tag{1.3}$$

$$\mathcal{N}(d, w, r) = \bigodot_{i=1..s} \hat{\mathcal{F}}_i^{d \cdot \hat{L}_i}(X_{(r \cdot \hat{H}_i, r \cdot \hat{W}_i, w \cdot \hat{C}_i)}) \tag{1.4}$$

де w, d, r - параметри масштабування ширини, глибини та роздільної здатності відповідно,

та $\hat{\mathcal{F}}_i$ - функція, що задається архітектурою шару,

$\hat{L}_i$, $\hat{H}_i, \hat{W}_i$ - розміри карти активації, а $\hat{C}_i$ - кількість каналів.

Також, пропонується визначення залежності між d, w, r:

$$d = \alpha^{\phi} \tag{1.5}$$

$$w = \beta^{\phi} \tag{1.6}$$

$$r = \gamma^{\phi} \tag{1.7}$$

так, що виконується наступне:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \tag{1.8}$$

$$\alpha \geqslant 1, \beta \geqslant 1, \gamma \geqslant 1 \tag{1.9}$$

де ϕ - множник, що задається в залежності від наявного обчислювального бюджету.

Завдяки цьому методу масштабування, автори представляють 8 варіантів нейронної мережі, що відрізняється необхідною точністю та продуктивністю: від найменшої до найбільшої.

1.3.3. Визначні архітектури ШНМ для сегментації зображень.

Повністю згорткова нейронна мережа. (*англ. Fully-convolutional network, FCN*) [**?**] - це перша нейронна мережа, що використовувала лише

шари згортки, активації та підвибірки, не використовуючи повнозв'язних шарів в задачі семантичної сегментації.

Основною перевагою цього підходу стала можливість навчання та прогнозування на зображеннях різного розміру, без зміни параметрів нейронної мережі.

В даній роботі вперше було застосовано індуктивний переніс знань з задачі класифікації на задачу сегментації. Для цього перші шари нейронної мережі було ініціалізовано за допомогою параметрів, отриманих після тренування на наборі даних ImageNet [**?**].

Для отримання масок сегментації, повнозв'язні шари було сконвертовано в згорткові, а також було додано інформацію з перших шарів для збереження просторових ознак. Граф нейронної мережі FCN зображено на рис. 1.10.

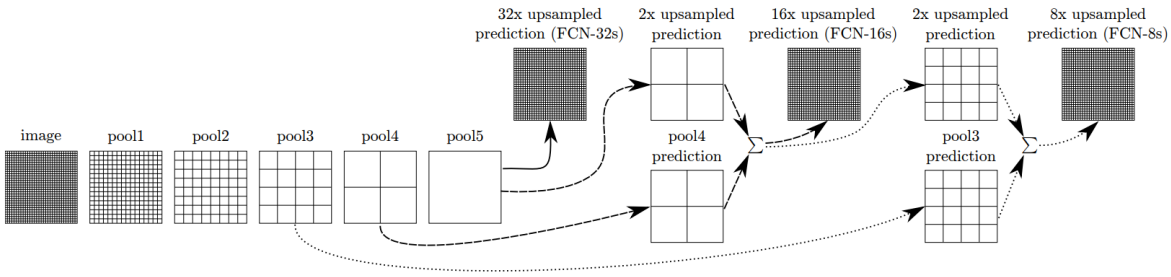

Рис. 1.10: Граф нейронної мережі FCN [**?**]

Тут, *pool[1-5]* - карти активації після відповідних операцій підвибірки, а *upsampled* - карти активації після білінійної інтерполяції.

Нейронна мережа SegNet. SegNet (*сегментаційна мережа*) [**?**]- перша повністю згорткова нейронна мережа, що використовувала архітектуру типу "енкодер-деокдер"для вирішення задачі семантичної сегментації. В даній архітектурі, енкодер стискає семантичне представлення до більш компактного простору. Роль декодера полягає у відображенні з компактного простору з низькою роздільною здатністю у карту актавації повного розміру для класифікації кожного пікселя.

Архітектура мережі енкодера топологічно ідентична 13 згортковим шарам мережі VGG16 [?].

Новина SegNet полягає в тому, як декодер підвищує розширення карт активаціїї низького розширення. Так, декодер використовує індекси операцій підвибірки, обчислені на відповідному шарі енкодера, для нелінійного підвищення розширення. Карти активації підвищенного розлирення виявляються розрідженими, для того, щоб зробити їх щільними, виконується декілька операцій згортки. Структура SegNet зображена на рисунку ??.

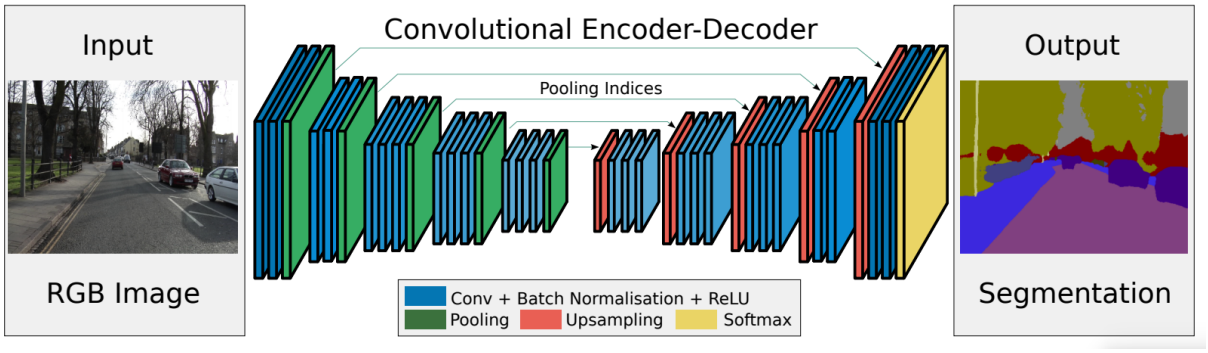

Рис. 1.11: Граф нейронної мережі SegNet [?]

Нейронна мережа PSPNet. Pyramid Scene Parsing Network (PSPNet) - це подальший розвиток ідеї FCN, в якому виправлено декілька значних проблем архітектури FCN. Серед основних проблем FCN автори відзначили:

— Відсутність можливості обробки інформації про контекст рівня зображення

— Різний розмір та вигляд різних об'єктів одного класу

Для того, щоб вирішити ці проблеми, автори запропонували додати до архітектури FCN модуль пірамідальної підвибірки (*англ. Pyramid Pooling Module*). Цей модуль складається з чотирьох операцій підвибірки з різними масштабом, що дозволяє захопити контекст зображення на різних рівнях.

Структура PSPNet та модуля підвибірки зображена на рисунку ??.

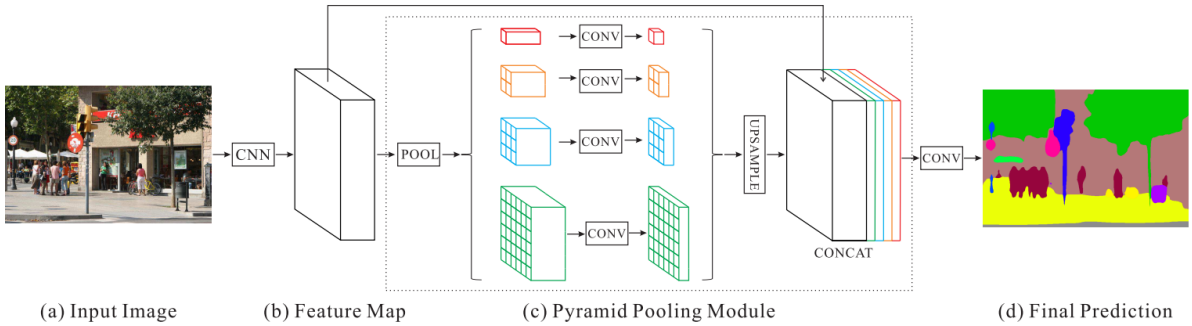

Рис. 1.12: Граф нейронної мережі PSPNet [**?**]

UNet. U-Net (*U-подібна мережа*) [**?**] - повністю згорткова нейронна мережа, що складається з симетричних енкодера та декодера, в яких, на відміну від SegNet замість індексів підвибірки, до декодера передаються карти активації енкодера напряму, через операцію конкатенації. Окрім спрощення кодування, така організація мережі дозволила спростити навчання, оскільки передча інформації стала повністю диференційованою.

Також, для забезпечення можливості навчання на малій кількості даних, використовується велика кількість різноманіних аугментацій вхідних зображень, щоб синтетично збільшити розмір набору даних.

Для більш чіткого розподілення близько розташованих об'єктів, автори запропонували спеціальну зважену функцію втрат, що забезпечувала більш чітке формування границь об'єктів.

Хоча, оригінально ця архітектура була запропонована для семантичної сегментації планарних мікроскопічних знімків, подальші дослідження показали її застосовність в широкому спектрі задач семантичної сегментації.

Структура нейронної мережі U-Net зображена на рисунку 1.13.

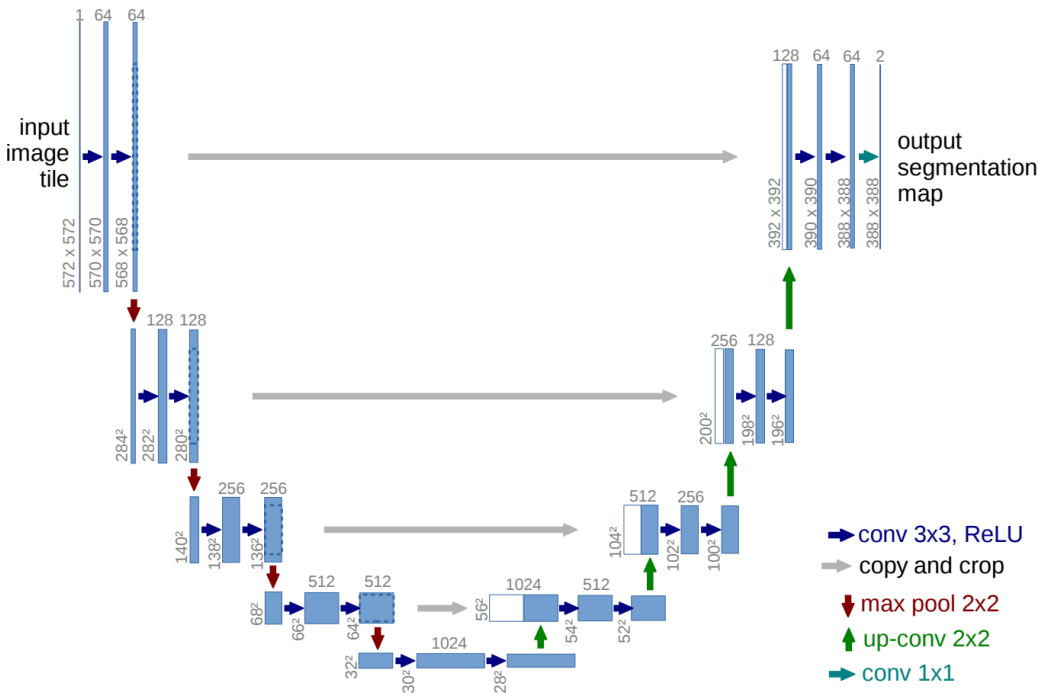

Рис. 1.13: Граф нейронної мережі U-Net [?]

LinkNet. Архітектура нейронної мережі LinkNet є подальшим розвитком архітектури U-Net, і також складається з енкодера та декодера, що поєднані через проміжні карти активації.

На відміну від U-Net, LinkNet використовує ResNet в якості енкодера, а також для поєднання карт активації енкодера та декодера використовує операцію суми, замість конкатенації. Це дозволило значно зменшити кількість параметрів, підвищити швидкість тренування при збереженні точності сегментації.

Також, автори LinkNet вперше використали індуктивний переніс з ImageNet виключно до енкодера, залишивши декодер випадково ініціалізованим на момент початку навчання.

Структура нейронної мережі LinkNet зображена на рисунку 1.14.

В подальших роботах, автори зазвичай називають свої архітектури U-Net-подібними, хоча, насправді, вони є ближчими до LinkNet. [?, ?, ?]

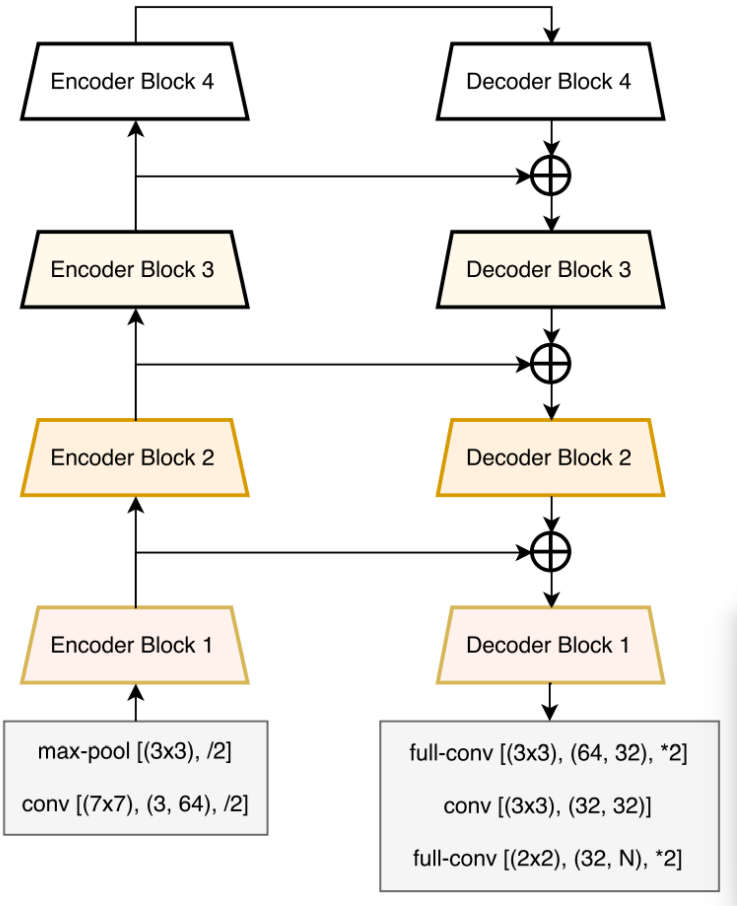

Рис. 1.14: Граф нейронної мережі LinkNet [?]

Feature Pyramid Network. Вперше розроблена для задач детекції об'єктів, Feature Pyramid Network (FPN) [?], була успішно застосована для задач семантичної сегментації [?].

Структура FPN складається з двох шляхів: на одному з них зменшується просторова та розмірність збільшується глибина карт ознак (шлях знизу вверх), на іншому - навпаки (шлях зверху-вниз). Шлях знизу-вверх представляє собою звичайну згорткову нейронну мережу (зазвичай, з сімейства ResNet). Шлях зверху-вниз представляє собою послідовність згорток та операцій підвищення розширення.

Для отримання карти активації, з якої буде отримано результат семантичної сегментації, всі рівні декодера оброблюються двома операціями згортки, після чого приводяться до однакового просторового розширення. З цієї карти ознак отримується фінальна карта сегментаціїї за допомогою операції згортки.

Структура нейронної мережі FPN для семантичної сегментації зображена на рисунку 1.15.

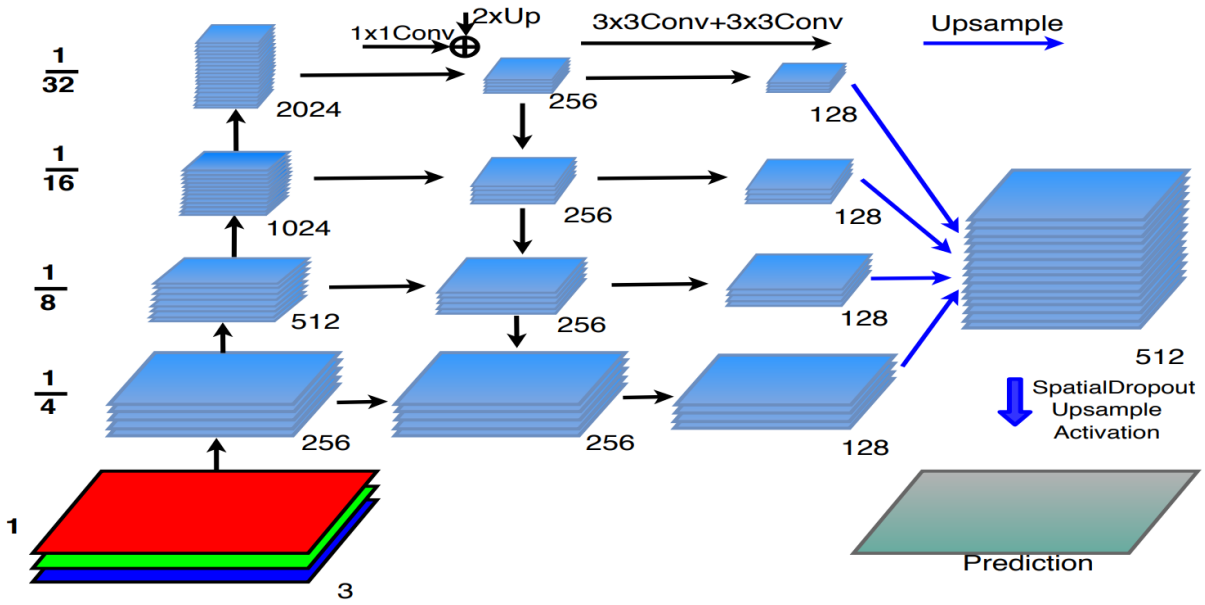

Рис. 1.15: Граф нейронної мережі FPN [?]

Сімейство нейронних мереж DeepLab. Нейронні мережі з сімейства DeepLab [?, ?] побудовані навколо одного алгоритму підвибірки з розширеної просторової піраміди (*англ. Atrous Spatial Pyramid Pooling (ASPP)* та архітектури енкодер-декодер. ASPP - це модуль для передискретизації карт активації в декількох масштабах до виконання операції згортки. Це дозволяє за одну операцію згортки виділити ознаки в різних масштабах, а також корисний контекст зображення.

Замість того, щоб насправді передискретизовувати карти ознак як в PSPNet, зміна масштабу виконується за допомогою фільтрів згортки різного розширення.

Структура блоку ASPP зображена на рисунку 1.16.

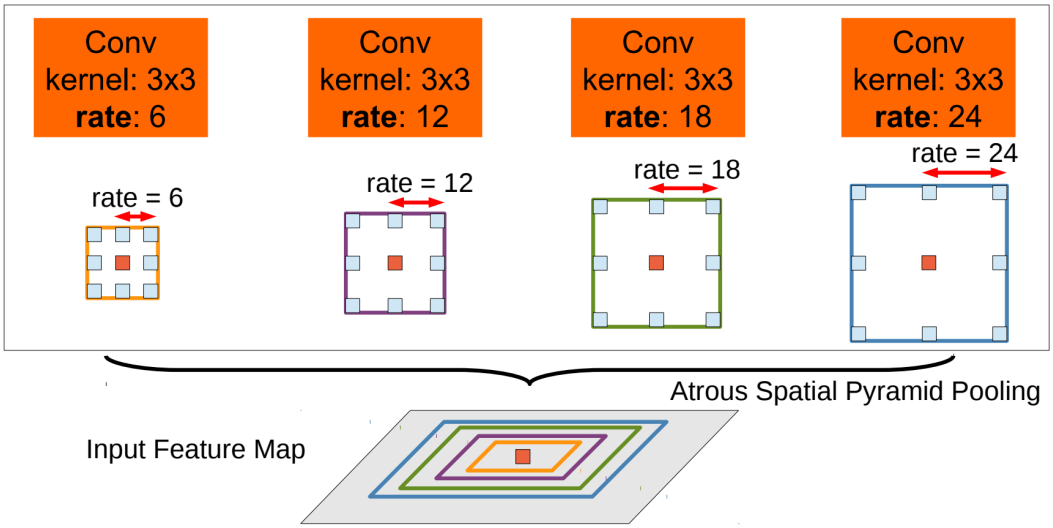

Рис. 1.16: Структура блоку ASPP в нейронних мережах DeepLab [?]

1.4. Методи багатозадачного глибинного машинного навчання

Багатозадачне машинне навчання - це один з підходів машинного навчання, в якому одночасно вирішуються кілька навчальних задач, використовуючи спільності та відмінності між ними. Кожна з таких задач може

бути загальною навчальною задачею, такою як навчання з учителем (наприклад, задача класифікації, або регресії), навчання без вчителя (кластеризація), напівавтоматичне навчання, завдання навчання з підкріпленням, або моделі на графах. Передбачається, що всі ці навчальні задачі або хоча б частина з них пов'язані одна з одною.

В цьому випадку виявлено [**?**], що спільне навчання на цих задачах може призвести до значного підвищення продуктивності в порівнянні з навчанням на кожній задачі окремо. Отже, багатозадачне навчання направлено на підвищення якості узагальнення кількох пов'язаних задач. Далі будуть розглядатися виключно задачі контрольованого навчання.

Подібно до людського навчання, корисно вивчати кілька навчальних задач разом, оскільки знання, що містяться в одній задачі, можуть бути використані іншими. Наприклад, людина, що вивчає математику та математичну статистику, може використати цей досвід для вивчення інших схожих галузей, наприклад, машинного навчання.

1.4.1. Визначення багатозадачного машинного навчання. Для n навчальних задач $T_1, T_2, \ldots T_n$, де всі задачі або їх підмножина пов'язані, багатозадачне навчання (БЗМН) має на меті допомогти покращити вивчення моделі для T_i, використовуючи знання, що містяться в усіх n або деяких з них задчах.

В літературі визначається декілька видів БЗМН:

Однорідне БЗМН: Для кожної з задач прогнозується лише один вихід. Наприклад, розпізнавання цифр MNIST зазвичай використовується для оцінки алгоритмів БЗМН. В цьому разі, воно розглядається як 10 завдань бінарної класифікації.

Неоднорідне БЗМН: Для кожної з задач прогнозується свій набір виходів. Наприклад, сучасні задачі детектування об'єктів потребують одночасного прогнозування класу об'єкта (задача класифікації), його положення

на зображенні та розмірів (задача регресії). Для задач контрольованого навчання, кожна задача T_i містить в собі набор даних D_i, що складається з m елементів...

Дослідники вважають, що коли різні задачі використовують однакові набори вхідних даних, БЗМН зводиться до задачі класифікації або регресії з кількома виходами. Однак, новіші дослідження визначають БЗМН на одному наборі вхідних даних, або незалежно від кількості наборів вхідних даних.

В методах БЗМН, заснованих на глибоких нейронних мережах, виділяють дві групи відповідно до способу розподілу параметрів між різними задачами.

1.4.2. Трансферне навчання. Трансферне навчання – це проблема машинного навчання, яка фокусується на збереженні знань, отриманих під час вирішення однієї задачі з подальшим застосуванням її результатів до іншої, але близької до неї. З практичної точки зору, повторне використання або передача інформації з раніше засвоєних завдань для вивчення нових може значно підвищити ефективність їх розв'язання [savchenko]. Деякі дослідники [] вважають багатозадачне машинне навчання підмножиною трансферного навчання.

На відміну від трансферного навчання, в багатозадачному навчанні немає різниці між задачами, та завдання стоїть підвищити продуктивність на всіх задачах одночасно.

1.4.3. Жорсткий розподіл параметрів. Зазвичай, архітектури згорткових нейронних мереж для багатозадачного навчання складаються зі спільного енкодера, який виділяє ознаки на вхідному зображенні (shared features), та окремих згорткових і/або повнозв'язних шарів для кожної з задач.

В найпершій роботі [?], що використовує цей принцип, було запропоновано архітектуру нейронної мережі Task-Constrained Deep Convolutional Network (TCDCN) для детекції ключових точок лиця використати додаткові задачі регресії пози голови та класифікації атрибутів. Архітектуру TCDCN зображено на рис. 1.17.

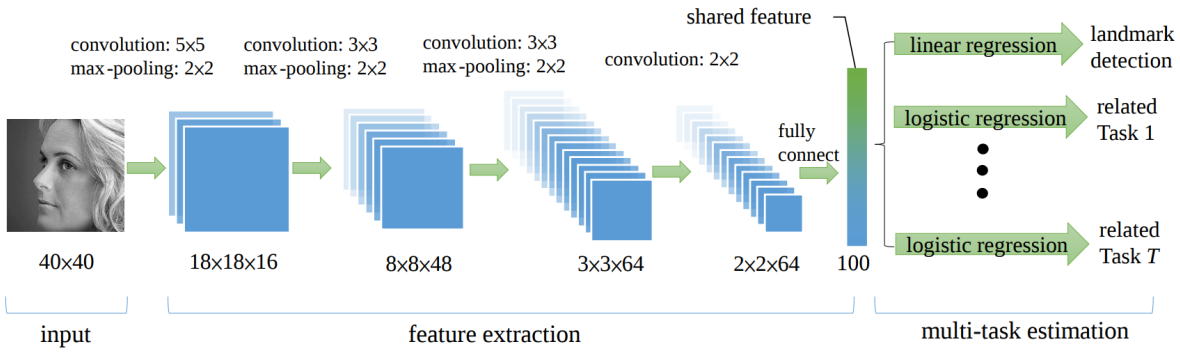

Рис. 1.17: Архітектура нейронної мережі TCDCN [?]

Модифікацією попереднього підходу є багатозадачні каскадні мережі ??, в яких результати одних задач використовуються як доповнення до вхідної інформації для наступних. Така організація шарів дозволяє підвищити точність на більш складних завданнях при правильному підборі черги задач за складністю. Загальну архітектуру зображено на рис. 1.18. Схожий підхід також використовується в деяких архітектурах нейронних мереж для розв'язання задачі детекції об'єктів [?, ?, ?]

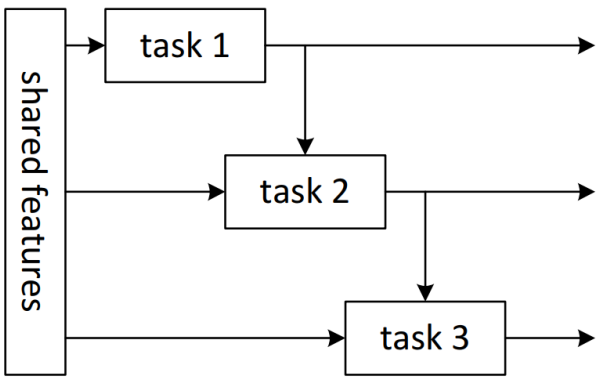

Рис. 1.18: Архітектура багатозадачної каскадної мережі [?]

Також, можлива інтеграція модулів уваги як до енкодера, так і до специфічних до задачі шарів. Таким чином, обчислення ознак відбувається не лише за рахунок параметрів енкодера, а і через специфічні до задачі модулі, що розташовані всередині мережі. Це дозволяє покращити репрезентації ознак для конкретних задач.

Можливою проблемою при багатозадачному навчанні на семантично-віддалених задачах є конфлікт градієнтів [**?**]: градієнти для різних задач мають протилежні напрямки. Для розв'язання цієї проблеми було запропоновано декілька підходів [**?**, **?**, **?**], найпростішим з яких є навчання на задачах, що не викликають конфлікту [**?**].

1.4.4. М'який розподіл параметрів. Одним з варіантів архітектур, що менш схильні до утворення конфліктних градієнтів є методи архітектури з м'яким розподілом параметрів.

В методах, що використовують м'який розподіл параметрів використовується декілька енкодерів для кожної з задач, між шарами яких відбувається обмін інформацією.

Наприклад, в роботі Cross-stitch networks (перехресні мережі)[**?**] представлена архітектура багатозадачної нейронної мережі, в якій кожній задачі відповідає окремий енкодер, але входом кожного шару кожного з енкодерів є лінійна комбінація виходів попереднього шару всіх енкодерів. Коефіцієнти лінійної комбінації також вивчаються за допомогою градієнтного спуску разом з іншими параметрами нейронної мережі. На рисунку 1.19 зображено архітектуру блоку лінійної комбінації для шарів двох нейронних мереж.

Узагальненням ідеї переносу інформації між послідовними шарами декількох специфічних до задачі нейронних мереж є шлюзові мережі (*англ. Sluice networks*). Вхід до кожного з шарів також є лінійної комбінацією виходів попередніх шарів для кожної з задач. На відміну від перехресних ме-

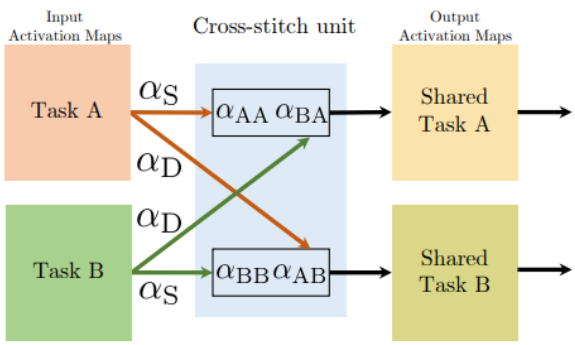

Рис. 1.19: Архітектура cross-stitch блоку [?]

реж, кожний шар нейронної мережі розділений на окремі підпростори: для конкретної задачі та спільний між всіма задачами. Також, додається обмеження на ортогональність підпросторів через додавання функції втрат, що мінімізує квадратовану норму Фробеніуса (матричну L^2 норму) між підпросторами в кожному з шарів.

Подальший розвиток шлюзових мереж було запропоновано в роботі «Нейромережеве дискримінтивне зменшення розмірності»(*англ. Neural Discriminative Dimensionality Reduction, NDDR*) [?]. Тут, замість використання лінійної комбінації шарів між різними задачами, виконується згортка конкатенованих виходів шарів всіх задач з ядром розміру 1x1 з виконанням подальшої нормалізації. Архітектуру NDDR зображено на рисунку 1.20.

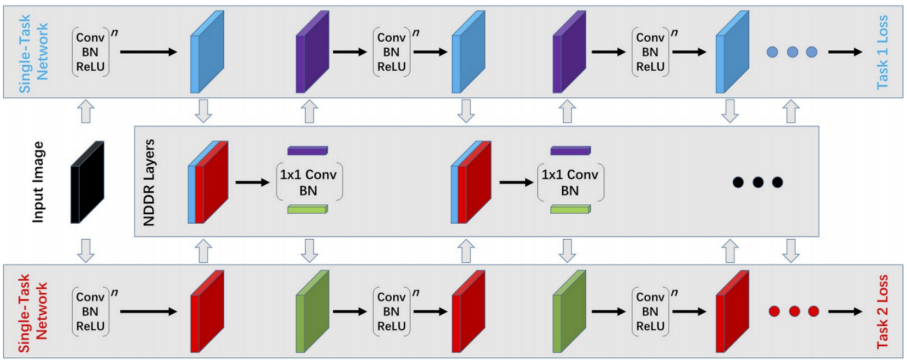

Рис. 1.20: Архітектура згорткової нейронної мережі NDDR [?]

Суттєвим недоліком методів, що використовують м'який розподіл параметрів є постійне збільшення кількості параметрів нейронної мережі, що,

в свою чергу, збільшує необхідну кількість ресурсів для роботи такої нейронної мережі зі збереженням продуктивності.

1.4.5. Дистиляція прогнозів моделі. Дистиляція прогнозів замість вивчення ознак, спільних до декількох задач, базується на використанні результатів для одних задач, як засобу вивчення ознак для інших. Наприклад, при одночасному вивченні задач класифікації та сегментації, карти сегментації можуть бути використані для подальшої класифікації зображення.

Першою роботою, що використала цей принцип була PAD-Net (Prediction-and-Distillation Network) [**?**]. В ній використовувалися результати задач семантичної сегментації, регресії карти глибини, регресії карти поверхневих нормалей та детекції контурів для отримання уточнених результатів в задачах регресії карти глибини та семантичної сегментації. Для цього автори запропонували три варіанти спеціальних модулів дистиляції:

— ПЕРЕВЕСТИ НОРМАЛЬНО
— Проста конкатенація результатів;
— Передача повідомлень між задачами
— Передача повідомлень з увагою

Подальшим розвитком цього методу стала робота PAP-Net, в якій було представлено модуль, що окремо вивчає попарні взаємозв'язки між задачами, та об'єднує вивчені карти ознак відповідно до попарних відношень.

1.4.6. Багатозадачне навчання в задачах класифікації та сегментації. Використання багатозадачного навчання в задачах класифікації та сегментації не є поширеним в літературі. Зазвичай, в роботах, що пропонують вирішення задачі сегментації, задача класифікації вирішується на етапі попередньої обробки зображень, щоб відсіяти хибнопозитивні

результати сегментації та зменшити час роботи системи в цілому через використання легких моделей класифікації. [?, ?, ?] Недоліком такого підходу є можливість поширення помилки від класифікатора, що в задачах скринінгу вимагає підвищення повноти класифікатора, з подальшою роботою сегментаційної мережі. Також, в задачах, що мають маленькі об'єкти на зображеннях, чи незбалансовані набори даних, такий підхід може зменшити влучність роботи системи порівняно з її частинами [?, ?]

На відміну від послідовної обробки, в роботі SegTHOR одночасно розв'язуються задачі класифікації та сегментації, доповнюючи одна одну. Так, автори пропонують нову архітектура нейронної мережі, що заснована на архітектурі UNet, але має додатковий вихід класифікації з останнього шару мережі сегментації. Отримання глобальних класів з виходу сегментації досягається через середнє арифметичне всіх пікселів для кожного з класів. Архітектура нейронної мережі SegTHOR зображена на рисунку 1.21.

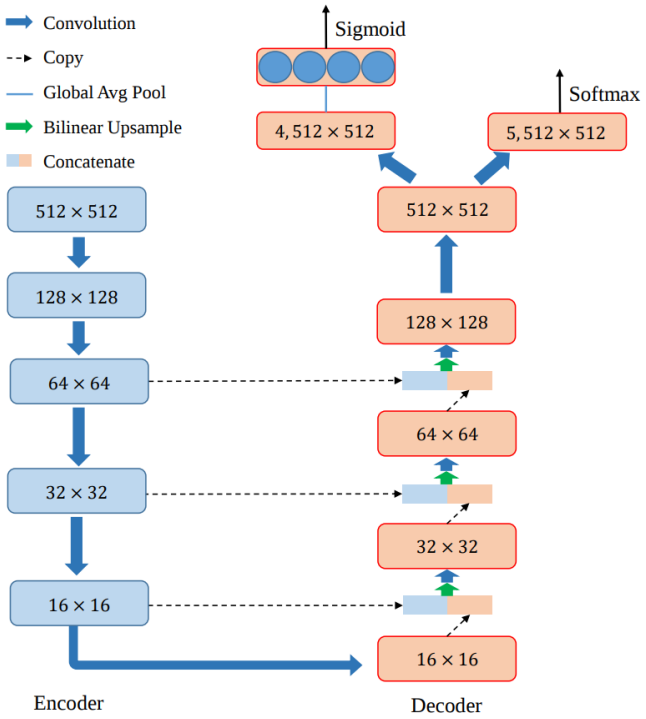

Рис. 1.21: Архітектура методу SegTHOR [?]

В нейронній мережі SegTHOR вихід класифікації використовується ли-

ше при тренуванні в якості додаткової задачі, та не використовується під час прогнозування масок сегментації.

В роботі [?] автори запропонували застосувати підхід багатозадачного машинного навчання до класифікації та сегментації новоутворень шкіри. На відміну від звичайних архітектур для сегментації зображень, в роботі запропоновано сегментацію будь-якого новоутворення, з паралельною класифікацією меланоми та себорейного кератозу як окремих задач. Автори показують, що така архітектура нейронної мережі незначно покращує достовірність як в задачах класифікації, так і сегментації. В даній архітектурі під час прогнозування результати різних задач використовуються окремо.

Автори [?] запропонували одночасно вивчати задачі класифікації та сегментації для локалізації ракових пухлин в мамографії. Як і [?], в даній роботі використовується нейронна мережа, що складається з одного енкодера і двох декодерів - для класифікації та сегментації відповідно. Архітектуру нейронної мережі зображено на рисунку ??.

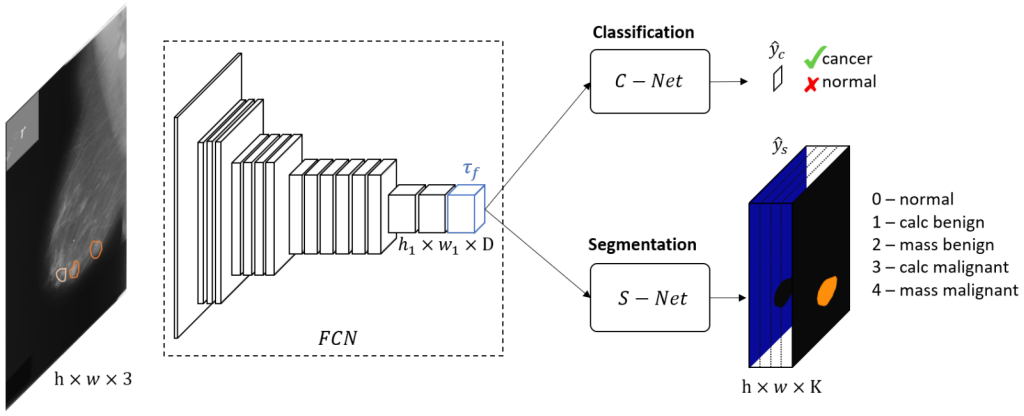

Рис. 1.22: Архітектура нейронної мережі класифікації та сегментації новоутворень на мамографії [?]

В роботі демонструється, що, на відміну від двох нейронних мереж, що вивчають задачі незалежно, запропонована архітектура має більшу точність як класифікації, так і сегментації. В даній архітектурі під час прогнозування результати різних задач використовуються окремо.

Окрім підходів з жорстким розподілом параметрів, в задачі сегментації може використовуватися каскадна схема. В роботі [?] запропоновано архітектуру каскадної нейронної мережі для розв'язання задачі сегментації будівель на аерофотознімках. Для покращення контурів сегментації, в роботі вводиться нова задача регресії піксельної відстані до найближчого контуру. Результат виконання задачі регресії використовується як частина вхідних даних для задачі сегментації. Архітектуру нейронної мережі зображено на рисунку 1.23. Автори показують, що таке використання декількох задач дозволяє підвищити достовірність сегментації в цілому та якість контурів окремо.

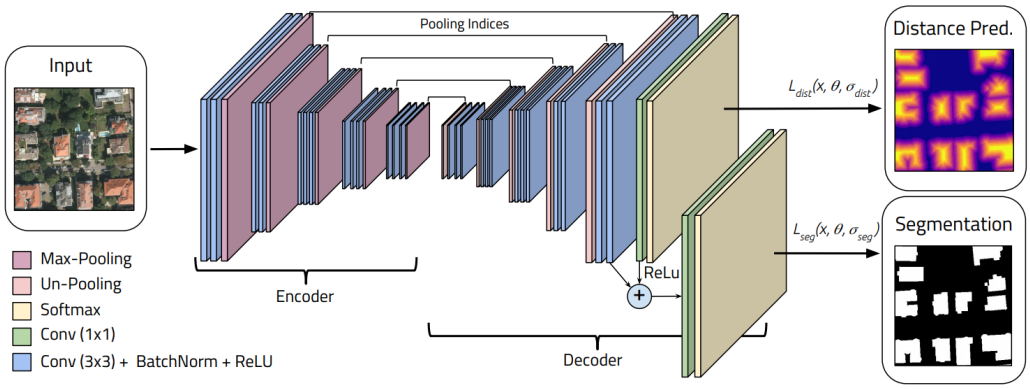

Рис. 1.23: Архітектура нейронної мережі для сегментації будівель [?]

Деякі методи машинного навчання з частковим залученням вчителя використовують архітектури, натхнені багатозадачним машинним навчанням. Так, популярною задачею машинного навчання з частковим залученням вчителя є семантична сегментація зображень з використанням лише міток рівня зображення, точок, або обмежувальних коробок навколо об'єктів. Наприклад, в роботі [?] запропоновано EM-алгоритм сумісно з глибокими нейронними мережами для розв'язання задачі сегментації через тренування нейронної мережі з мітками рівня зображення, або обмежувальних коробок. Задача класифікації в даному випадку використовується на етапі оцінювання очікування в EM алгоритмі як апріорна інформація.

Альтернативний метод сегментації з частковим залученням вчителя базується на функції втрат, що дозволяє одночасно оптимізувати наступні задачі:

— регресію локалізації об'єктів

— розширення локацій об'єктів відповідно до класів, присутніх на зображеннях

— обмеження границь масок сегментації об'єктів до границь об'єктів

Архітектура поєднання задач зображена на рисунку 1.24.

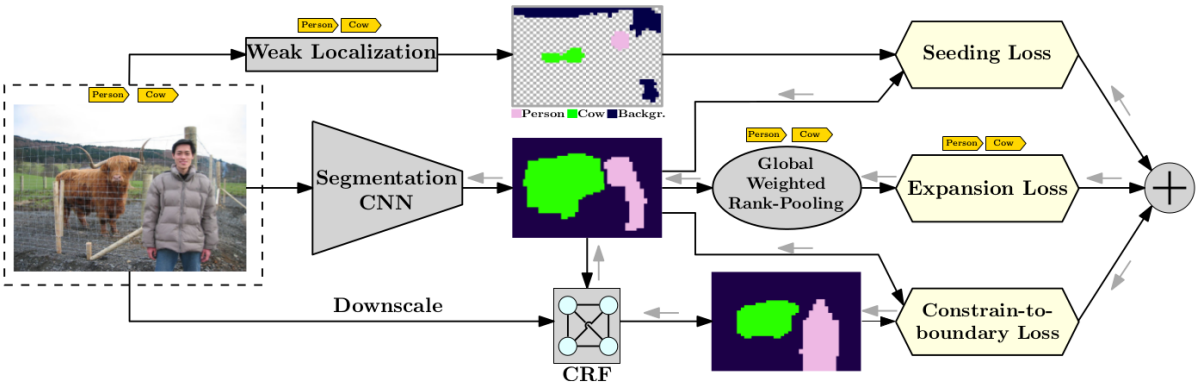

Рис. 1.24: Архітектура методу SEC-CNN [?]

1.5. Аналіз методів оцінки якості класифікації та сегментації зображень в задачі автоматизованого скринінгу

1.5.1. Помилки 1 і 2 роду та матриця невідповідностей. В задачах математичної статистики, помилка першого роду - це хибне відхилення правильної гіпотези (хибно-позитивний результат, FP), тоді як помилка другого роду - це прийняття хибної гіпотези (хибно-негативний результат, FN). Дані поняття використовуються, коли потрібно прийняття бінарного рішення на основі деякого критерію, що може мати похибку.

В задачах контрольованого навчання, зокрема бінарної класифікації, за за нульову гіпотезу зазвичай приймається приналежність елемента ви-

бірки до класу, що зустрічається в виборці частіше. Також, в задачах багатокласової класифікації використовується розширення поняття помилок першого і другого роду на задачу з більшою кількістю гіпотез - матриця невідповідностей. Матриця невідповідностей - це матриця, в строках якої зазначені зразки прогнозованого класу, а кожен із стовпців представляє зразки справжнього класу. Така матриця дозволяє оцінити, чи допускає система невідповідності між класами.

1.5.2. Влучність, чутливість та специфічність. Влучність, чутливість та специфічність є похідними мірами якості бінарного класифікатора та розраховуються на основі значень в матриці невідповідностей. Влучність (англ. precision) - вимірює частку істинно-позитивних зразків серед знайдених.

$$Precision = \frac{TP}{TP + FP} \tag{1.10}$$

Чутливість, або повнота (англ. sensitivity, recall) - вимірює частку істинно-позитивних зразків серед усіх позитивних зразків.

$$Recall = \frac{TP}{TP + FN} \tag{1.11}$$

Специфічність (англ. specificity) - вимірює частку істинно-негативних зразків серед усіх негативних зразків.

$$Specificity = \frac{TN}{TN + FP} \tag{1.12}$$

Між влучністю і повнотою, та чутливістю і специфічністю існує обернена залежність, коли можливо підвищити одну ціною зниження іншої. Через наявність такої залежноси, на практиці, ці міри не використовуються окремо. Замість цього використовують агреговані метрики, що дозволяють оцінити якість класифікатора в цілому.

1.5.3. Міра F1 та індекс подібності Соренсена. F1 та індекс подібності Соренсена (*англ. Dice score*) є еквівалентними метриками, що, зазвичай використовуються в задачах класифікації (F1) та сегментації (Dice score).

Міра F1 визначається як середнє гармонічне між влучністю та повнотою:

$$F1 = \frac{2}{precision^{-1} + recall^{-1}} = 2 \cdot \frac{precision \cdot recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \tag{1.13}$$

Індекс подібності Соренсена визначається на множинах, і, в задачах сегментації, визначається на піксельних масках сегментації:

$$D_c = 2 \cdot \frac{A \cap B}{|A| + |B|} \tag{1.14}$$

Також, для бінарних даних $A \cap B$ відповідає TP, а $|A| + |B|$ відповідає всій множині об'єктів, тобто еквівалентно $TP + FP + FN$. Таким чином, коефіцієнт Соренсена є еквівалентним мірі F1.

1.5.4. Каппа Коена. Оскільки використання точності неможливе в задачах з сильним дисбалансом класів, значення точності можна перенормувати. Перенормування виконується за допомогою статистики chance adjusted index: точність класифікації нормується з використанням точності, яку можна було б отримати випадково. Під випадковою тут розуміється точність рішення, яке отримано з нашого випадковою перестановкою класів:

$$\kappa = 1 - \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} o_{ij}}{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} e_{ij}}, \tag{1.15}$$

Де k - це кількість класів, o_{ij}, та e_{ij} це елементи предсказуваних та

істинних матриць. w_{ij} розраховується наступним чином для квадратично-зваженої метрики:

$$w_{ij} = \frac{(i-j)^2}{(k-1)^2}, \tag{1.16}$$

Через чутливість каппи Коена, дослідники повинні ретельно інтерпретувати цю метрику. Наприклад, якщо розглянути дві пари оцінювачів з однаковим відсотком угоди, але різними пропорціями по класах, це суттєво вплине на значення метрики.

Іншою проблемою є кількість класів: із збільшенням кількості класів, метрика підвищується. Крім того, каппа Коена може бути низькою, хоча рівень згоди високий, і окремі оцінки є точними.

Це робить Каппу Коена мінливим коефіцієнтом для аналізу.

1.6. Висновки та постановка задачі дослідження

В першому розділі дисертаційної роботи проведено аналіз методів розпізнавання, класифікації та сегментації планарних зображень, використання для цього методів глибокого навчання, та методів багатозадачного навчання.

Проаналізовано метрики, що використовуються в задачах класифікації та сегментації зображень. Наведено підходи використання багатозадачного навчання для класифікації та сегментації зображень в задачах скринінгу.

Показано, що складність розробки систем автоматизованого скринінгу пов'язана з необхідністю вирішення наступних завдань:

1. Неузгодженість та незбалансованість навчальних наборів даних для задач автоматизованого скринінгу виникає через відносно малу кількість аномальних прикладів в популяціях та значний рівень похибки розмітників. Це створює проблему при навчанні глибоких нейронних мереж стандартними методами. Тому при побудові моделей

класифікації та сегментації потрібно вирішувати задачу навчання з урахуванням частково-помилкової розмітки.

2. В задачах автоматизованого скринінгу важливо зменшення кількості хибно-позитивних результатів діагностики, через відносно малу розповсюдженість аномальних прикладів в популяціях та пов'язані з хибно-позитивними результатами подальші дії. Тому, при побудові систем класифікації та сегментації потрібно зосередитися на зменшенні хибно-позитивних результатів.
3. Попередні дослідження показали, що системи автоматизованого скринінгу, засновані на методах багатозадачного глибинного навчання, дозволяють підвищити достовірність класифікації та сегментації за рахунок спільного навчання на цих задачах. Однак, попередні дослідження не враховують можливої комбінації результатів різних задач для підвищення достовірності під час прогнозування. Ці недоліки запропоновано подолати за рахунок комбінування результатів під час фази прогнозування.

У зв'язку з перерахованим вище, сформульовано мету і задачі дослідження.

Метою дисертаційної роботи є підвищення достовірності класифікації та семантичної сегментації планарних зображень без підвищення витрат часу за рахунок розроблення методу об'єднання задач багатозадачного навчання з використанням штучних нейронних мереж.

Для досягнення мети, в роботі виконано наступні задачі:

— Проведено аналіз існуючих методів, алгоритмів та засобів обробки планарних зображень в системах автоматизованого скринінгу
— Удосконалено архітектуру штучної нейронної мережі для багатозадачного навчання
— Розвинуто нейромережевий метод обробки планарних зображень, який за рахунок навчання на декількох задачах одночасно підви-

щить точність класифікації та семантичної сегментації планарних зображень

— Розроблено метод об'єднання задач класифікації та сегментації на етапі прогнозування, що дозволяє підвищити достовірність класифікації та сегментації без підвищення витрат часу
— Вдосконалено метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки семантичної сегментації в навчальному наборі даних
— Проведено випробування розробленого методу та інструментальних засобів в рамках експерименту
— Спроектовано та розроблено інструментальні засоби, що виконують розроблений метод

РОЗДІЛ 2
МОДЕЛЬ ШУМНИХ ДАНИХ В ЗАДАЧІ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ

Для систематизованого вивчення впливу рівня шуму в даних, та відповідного впливу запропонованих рішень, розроблено модель зашумлених даних задачі сегментації, що відповідає оцінкам моделей шуму в різних задачах автоматизованого скринінгу.

2.1. Аналіз зашумлення розмітки сегментації в задачі автоматизованого скринінгу

Основними проблемами з розміткою даних в задачах скринінгу є:

— Маски сегментації, що захоплюють сусідні з об'єктом пікселі [?]
— Маски сегментації, що покривають об'єкт не повністю [?]
— Відсутні маски сегментації для деяких об'єктів [?]
— Присутні зайві маски на місцях, де немає об'єктів [?]

2.2. Моделювання зашумлення розмітки сегментації в задачі автоматизованого скринінгу

Для моделювання цих дефектів розмітки запропоновано метод контрольованої генерації як зображень, так і масок сегментації що модифікуються відповідно до випадково обраних недоліків.

2.2.1. Алгоритм генерації наборів даних. Контрольована генерація зображень виконується в два етапи: генерація фону та розташування на ньому довільної кількості об'єктів. Для генерації фону можуть бути ви-

користані як звичайні натуральні зображення, так і синтетичні текстури, чи заливка константним кольором. В якості об'єктів можуть бути використані зображення з наборів даних MNIST [**?**], або FashionMNIST [**?**].

Використання зазначених наборів даних зумовлено трьома факторами:

— Можливість простого відділення об'єкта від вихідного фону;

— Наявність схожих елементів в різних класів (наприклад, цифри 1 та 7, або класи *T-shirt* та *Dress*);

— Висока точність роботи сучасних нейронних мереж на цих наборах даних, що дозволяє сконцентруватися на впливі шуму в розмітці, замість розпізнавання безпосередньо об'єктів

Помилки в розмітку штучно вводяться за допомогою випадкового застосування морфологічних операцій ерозії та дилатації з квадратним ядром до масок сегментації окремих об'єктів перед додаванням їх до загальної маски.

Окрім наборів даних зображень фону $\mathcal{X}_b$, об'єктів $\mathcal{X}_f$ та текстур об'єктів $\mathcal{X}_{tex}$, модель має наступні параметри:

— Кількість зображень в наборі даних N

— Розмір зображення S_{img} в пікселях

— Середній розмір об'єкта S_{obj} в пікселях

— Максимальне відхилення розміру об'єкта Δ_{max} в відсотках

— Максимальна кількість об'єктів на зображенні N_{obj}

— Ймовірності ерозії та дилатації маски кожного з об'єктів P_e та P_d

— Розміри ядер ерозії та дилатації масок всіх об'єктів S_e та S_d

На основі моделі генеруються навчальний та тестовий набори даних, всі зображення в наборах генеруються незалежно.

Алгоритм генерації кожного з зображень складається з наступних кроків:

1. Вибір випадкового зображення фону: $x_{bg} \sim \mathcal{X}_b$

2. Вибір кількості об'єктів на зображенні: $n_{obj} \sim \mathcal{U}(1, N_{obj})$

3. Ініціалізація маски сегментації: $M = 0$ так що $M \in \mathcal{R}^{C \times S_{img} \times S_{img}}$

4. Відповідно до кількості зображень n_{obj} виконати наступні кроки:

4.1 Вибрати розміри об'єкта: $s \sim \mathcal{U}(S_{obj} - \Delta_{max}, S_{obj} + \Delta_{max})$

4.2 Вибрати координати розміщення об'єкта:

$$i_f \sim \mathcal{U}(0, S_{img} - s)$$
$$j_f \sim \mathcal{U}(0, S_{img} - s)$$

4.3 Вибрати зображення об'єкта $x_{fg} \sim \mathcal{X}_f$ та відповідний клас об'єкта $c_{fg} \sim \mathcal{Y}_f$

4.4 Змінити розмір зображення об'єкта за допомогою білінійної інтерполяції:

$$\hat{x}_{fg} = R_{bilinear}(x_{fg})$$

4.5 Вибрати зображення текстури $x_{tex} \sim \mathcal{X}_{tex}$

4.6 Модифікувати зображення об'єкта за допомогою текстури:

$$\hat{x}_{fg} = x_{fg} \circ x_{tex}[i_f : i_f + s, j_f : j_f + s]$$

4.7 Розмістити зображення об'єкта на зображенні фону:

$$x_{bg}[i_f : i_f + s, j_f : j_f + s] = (1 - x_{fg}) \circ x_{bg} + \hat{x}_{fg}$$

4.8 Сформувати маску сегментації об'єкта:

$$M_{seg} = x_{fg} > \theta_{seg}$$

де θ_{seg} - поріг бінаризації вихідного зображення об'єкта. Для набору даних MNIST $\theta_{seg} = 0.2$, для набору даних FashionMNIST $\theta_{seg} = 0.1$.

4.9 Модифікувати маску сегментації відповідно до необхідного рівня помилок:

$$M_{seg} = \begin{cases} M_{seg} \oplus K^{S_d \times S_d} & \text{якщо } p_d \sim \mathcal{U}(0,1) < P_d \\ M_{seg} \ominus K^{S_e \times S_e} & \text{якщо } p_e \sim \mathcal{U}(0,1) < P_e \end{cases}$$

де $K^{S_d \times S_d}$ - матриця ядра $K^{S_e \times S_e}$ - матриця ядра ерозії.

4.10 Розмістити модифіковану маску сегментації об'єкта на загальному зображенні маски сегментації:

$$M[c_{fg}, i_f : i_f + s, j_f : j_f + s] = max\ \{M[c_{fg}, i_f : i_f + s, j_f : j_f + s], M_{seg}\}$$

Завдяки такому варіанту генерації, можливо отримання безлічі наборів даних із схожими характеристиками, що дозволяє використовувати непараметричні статистичні методи, такі як бутстрепінг для оцінки моделей. На практиці, для генерації різних наборів даних, достатньо зміни ініціалізації генератора випадкових чисел.

2.2.2. Види згенерованих зображень. В даному розділі продемонстровано згенеровані зображення та відповідні маски сегментації для різних налаштувань моделі.

За допомогою моделі можна генерувати набори даних з досить різноманітними характеристиками. На рисунках 2.1 та 2.2 зображено найпростіший випадок для генерації за допомогою моделі: заповнення фону та об'єктів константними значеннями (0 і 1) відповідно.

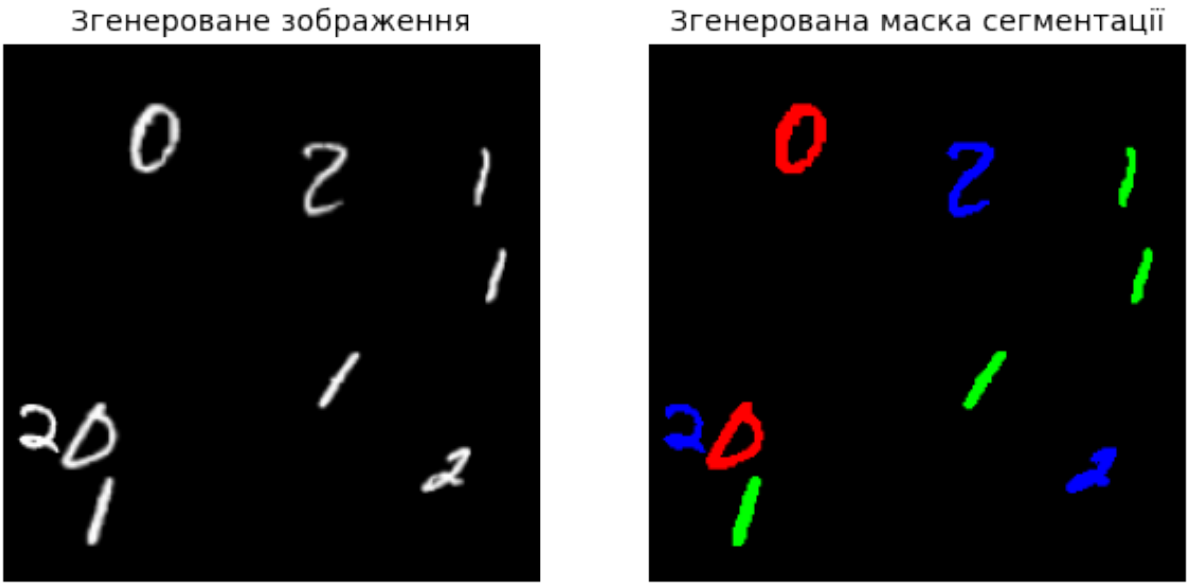

Рис. 2.1: Об'єкти MNIST, фон: константний, колір об'єктів: білий

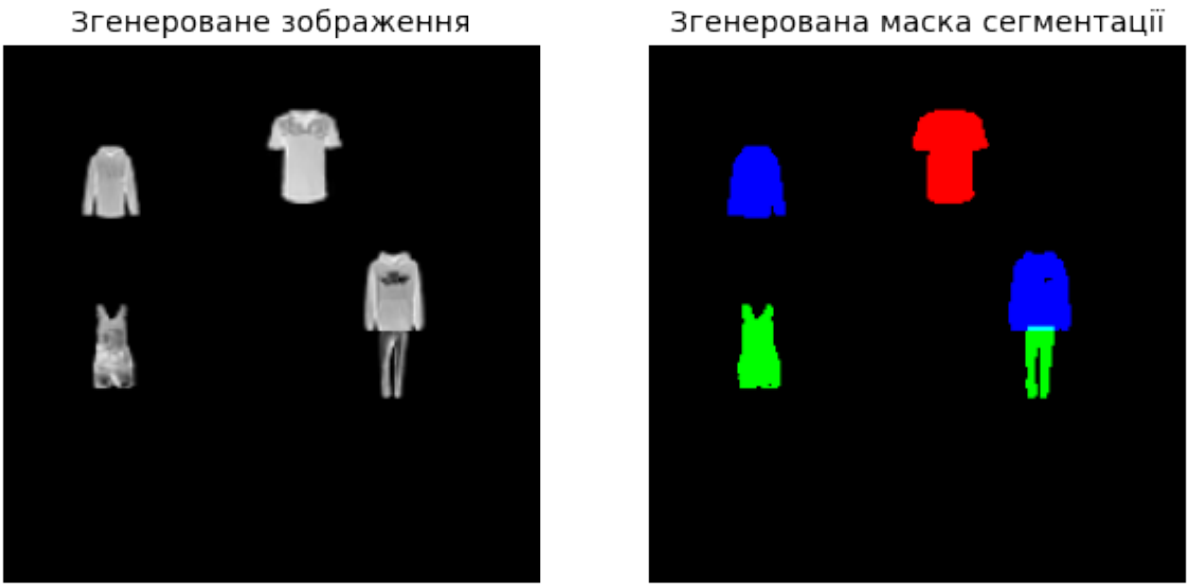

Рис. 2.2: Об'єкти FashionMNIST, фон: константний, колір об'єктів: білий

На рисунках 2.3 та 2.4 зображено фон з використанням зображення з набору даних Imagenette та заповнення об'єктів константним значенням 1.

Рис. 2.3: Об'єкти MNIST, фон: Imagenette, колір об'єктів: білий

Рис. 2.4: Об'єкти FashionMNIST, фон: Imagenette, колір об'єктів: білий

На рисунках 2.5 та 2.6 зображено фон з використанням зображення з набору даних Imagenette та заповнення об'єктів константним значенням 0.

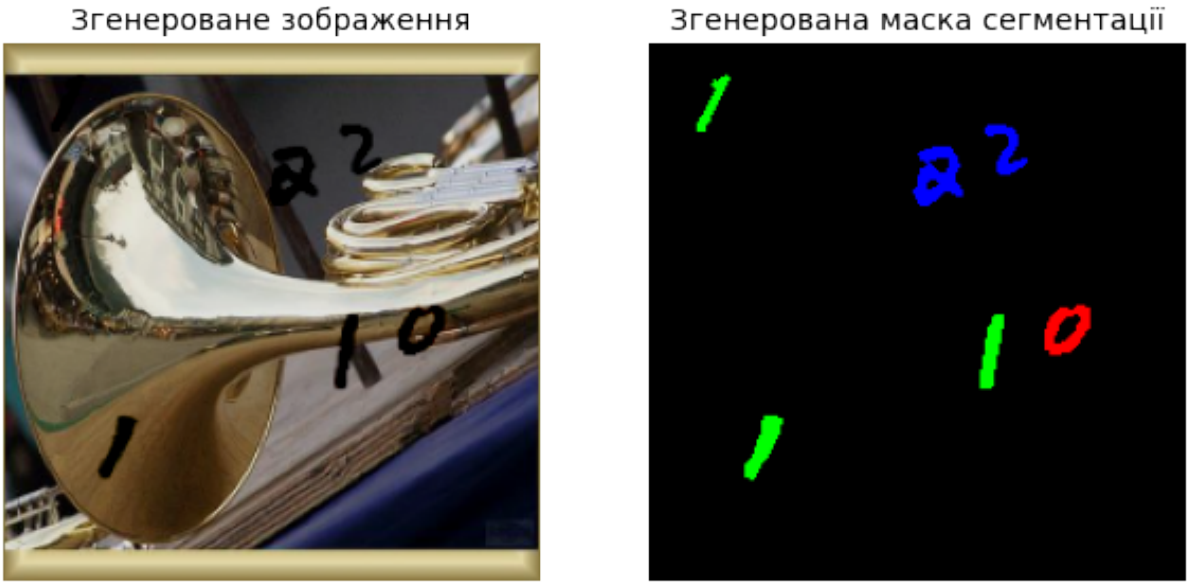

Рис. 2.5: Об'єкти MNIST, фон: Imagenette, колір об'єктів: чорний

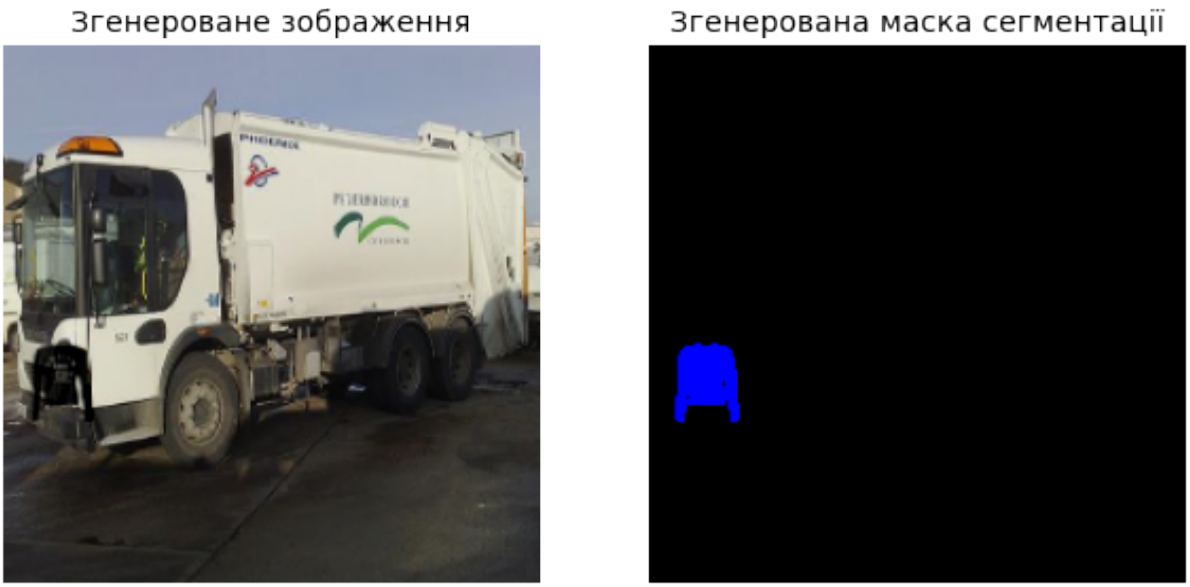

Рис. 2.6: Об'єкти FashionMNIST, фон: Imagenette, колір об'єктів: чорний

На рисунках 2.7 зображено найскладніший варіант набору даних: як фон, так і текстури об'єктів взято з набору даних Imagenette.

Рис. 2.7: Об'єкти MNIST, фон: Imagenette, текстури об'єктів: Imagenette

Рис. 2.8: Об'єкти FashionMNIST, фон: Imagenette, текстури об'єктів: Imagenette

Для наочності, модифікації масок сегментації показано на простих зображеннях з константним кольором фону та об'єктів.

На рисунку 2.9 зображено вихідне зображення, маску сегментації з ерозією ядром 3×3 та їх комбінацію для наочності.

Рис. 2.9: Об'єкти MNIST, ерозія маски ядром 3×3

На рисунку 2.10 зображено вихідне зображення, маску сегментації з дилатацією ядром 3×3 та їх комбінацію для наочності.

Рис. 2.10: Об'єкти MNIST, дилатація маски ядром 3×3

На рисунку 2.11 зображено вихідне зображення, маску сегментації з дилатацією та ерозією ядром 3×3 та їх комбінацію для наочності.

Рис. 2.11: Об'єкти MNIST, ерозія та дилатація маски ядром 3×3

Також, при виборі розмірів ядер ерозії потрібно бути уважним до вихідних масок - дуже великі ядра ерозії можуть знищити маску сегментації для об'єкта. На рисунку 2.12 зображено вихідне зображення, маску сегментації з ерозією ядром 7×7 та їх комбінацію для наочності.

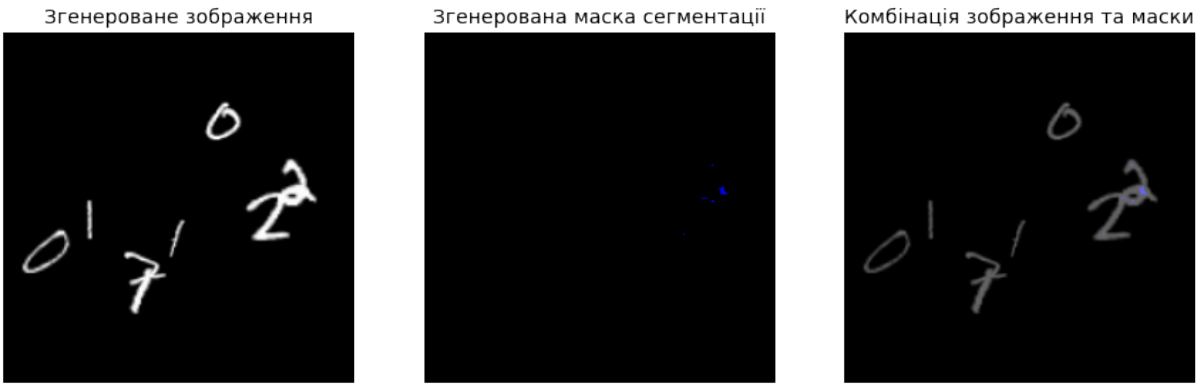

Рис. 2.12: Об'єкти MNIST, ерозія маски ядром 7×7

2.2.3. Висновки до другого розділу.

РОЗДІЛ 3
МЕТОДИ ПОЄДНАННЯ ЗАДАЧ ГЛИБИННОГО БАГАТОЗАДАЧНОГО НАВЧАННЯ

Використання комбінації декількох задач машинного навчання може сприяти вивченню кращих наборів ознак, та підвищити точність роботи нейронних мереж. Також, можливе сумісне використання прогнозів декількох задач для уточнення кожного з окремих прогнозів, так і побудови нових.

3.1. Метод багатозадачного навчання ШНМ в умовах частково помилкової розмітки семантичної сегментації

В умовах частково-помилкової розмітки даних виникає потреба зменшення впливу неправильно розмічених прикладів на процес навчання, та їх ефективна фільтрація. Часто на одному зображенні можуть бути розташовані як коректно-розмічені, так і некоректно-розмічені об'єкти. Для підвищення точності навчання нейронних мереж в таких умовах, запропоновано використати метод багатозадачного навчання з використанням похідної більш загальної задачі.

3.1.1. Генерація похідної задачі до задачі семантичної сегментації. ВОДУ СЮДА

Семантично-близькими називають задачі, що мають однакові вхідні дані, та пов'язані вихідні дані.

На відміну від [?], що спирається на вивчення більш детальних семантично-близьких задач, в розробленому методі використання більш точних даних, для загальніших задач дозволяє покращити роздільність внутрішніх

представлень нейронної мережі, що, в свою чергу, покращує результати на вихідній задачі. Також, оскільки задачі є семантично близькими, не відбувається конфлікту градієнтів, що є типовим при навчанні семантично-різнорідних задач [?].

Оскільки задача семантичної сегментації може бути розглянута як задача піксельної класифікації, можна розглядати задачу класифікації всього зображення як більш загальну до неї. В такому контексті, класифікація буде окремим випадком навчання за набором зразків [?]: замість розмітки кожного з об'єктів для всіх класів на зображенні, зображення являє собою мішок з одним, чи декількома об'єктами та відповідним маркуванням, чи є об'єкти заданих класів на ньому.

Попередній аналіз показав, що неточна розмітка в задачах сегментації полягає в наявності зайвих, або у відсутності деяких розмічених пікселів, або об'єктів. Наявність такої неточної розмітки дозволяє створити на її основі точну розмітку для класифікації.

Нехай для зображення $x \in \mathcal{R}^{3\times H\times W}$ існує маска сегментації $y_s \in \mathcal{R}^{C\times H\times W}$ де C - кількість класів, а H та W висота та ширина зображення відповідно. Якщо в масці сегментації є хоча б один розмічений об'єкт класу , встановлюється мітка відповідного класу $y_c \in (0,1)^C$ в розмітці задачі класифікації:

$$y_c = t < \sum_{i}^{H}\sum_{j}^{W} y_{s_{ij}} \tag{3.1}$$

де t - поріг мінімального розміру об'єкта в пікселях

Згенерована таким чином задача класифікації має меншу ймовірність хибної розмітки.

3.1.2. Автоматична фільтрація помилкової розмітки. Маючи велику кількість параметрів, сучасні нейронні мережі здатні до запам'я-

товування тренувального набору даних замість вивчення корисних ознак для узагальнення на інші набори даних (перенавчання). В умовах наявності помилкової розмітки в тренувальному наборі даних, нейронні мережі схильні до вивчення і помилок розмітки на протязі навчання.

Традиційні методи запобігання перенавчанню, такі як аугментація даних, регуляризація ваг мережі, та завчасна зупинка навчання хоча і показують покращені результати, але можуть виявитися недостатніми в умовах великої ймовірності помилок в навчальних даних. Також, вони можуть погіршити якість навчання через зменшення ємності нейронної мережі, або, при надмірному використанні, можуть спонукати нейронну мережу акцентувати увагу на текстурах, чо часто є недопустимим в задачах семантичної сегментації.

Для того, щоб ефективно зменшити вплив помилкової розмітки, необхідно вилучити її з процесу навчання нейронної мережі. Якщо неможливо вилучити неправильно розмічені приклади з набору даних до початку навчання, це можна робити ітеративно в процесі навчання. Такі функції втрат, як перехресна ентропія, або фокальна функція втрат призначають експоненційно високе значення для неправильних прогнозів нейронних мереж. Це має шкідливий ефект при наявності помилкової розмітки в тренувальному наборі даних - правильні прогнози на неправильно-розмічених даних створюють конфліктні градієнти та знижують впевненість прогнозів нейронної мережі.

Для виключення неправильної розмітки в процесі навчання запропоновано обмеження функції втрат зверху:

$$\mathcal{L}\rceil = min(L, \theta) \tag{3.2}$$

де θ - поріг обмеження функції втрат.

Графіки обмежених зверху функцій втрат зображено на рисунку 3.1.

3.1

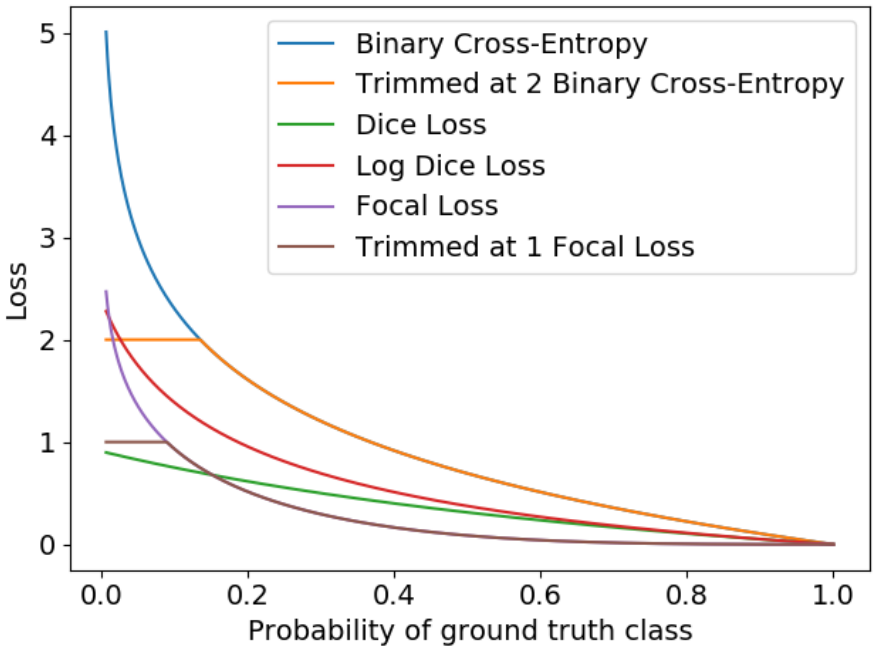

Рис. 3.1: Приклади обмежених зверху функцій втрат

Для функції обмеження зверху градієнт визначено лише на проміжку $(-\inf, \theta]$, тому для проміжку $(\theta, \inf)$ градієнт встановлено рівним нулю:

$$\nabla min(L, \theta) = \begin{cases} 1 & L \in (-\inf, \theta] \\ 0 & L \in (\theta, \inf) \end{cases} \tag{3.3}$$

Таким чином, приклади з занадто великою похибкою (переважно з помилковою розміткою) ефективно виключаються з навчання.

Окрім прикладів з неправильною розміткою, з навчання випадково будуть виключені приклади, що занадто складні для вивчення нейронною мережею на ранніх етапах, що може сповільнити процес, або взагалі зупинити його. Щоб запобігти сповільненню процесу, запропоновано використання багатозадачного навчання з використанням додаткової більш загальної задачі.

В такому випадку, обмеження функції втрат використовується лише для оригінальної задачі, а більш загальна додаткова задача використовує

оригінальну функцію втрат без обмеження. Таким чином, завжди існують градієнти від більш загальної розмітки, що спонукають навчання на прикладах, для яких немає градієнтів через обмеження функції втрат. Архітектуру нейронної мережі детальніше розглянуто в розділі 3.1.3

3.1.3. Архітектура багатозадачної нейронної мережі. Для того, щоб основна і додаткова задачі оновлювали параметри нейронної мережі, що відповідають за виділення ознак, використано схему багатозадачного навчання з жорстким розподілом параметрів.

Загальна архітектура нейронної мережі. Нейронна мережа виконана з використанням архітектури енкодер-декодер, заснована на архітектурі нейронних мереж LinkNet [**?**]. Нейронна мережа складається з енкодера та двох декодерів: для задач сегментації та класифікації відповідно. В ролі енкодера можуть бути використані наявні багатостадійні архітектури, такі як VGGNet, ResNet, EfficientNet та ін. Карти ознак після кожного етапу просторового зменшення використовуються як входи для декодера сегментації, для декодера класифікації використовується карта ознак з найглибшого шару енкодера. Узагальнену архітектуру зображено на рисунку 3.2

Нехай нейронну мережу енкодера визначено як $F_{encoder}$. Тоді для вхідного зображення $x \in \mathcal{R}^{3\times H\times W}$, $v_1, v_2, ...v_i, ...v_n$ - набір карт ознак, так що $v_i \in \mathcal{R}^{c_i\times\frac{H}{i}\times\frac{W}{i}}$, де c_i - кількість каналів для кожної з карт ознак, а n - кількість стадій енкодера (залежить від архітектури):

$$v_1, v_2...v_n = F_{encoder}(x, \theta_{enc}) \tag{3.4}$$

де θ_{enc} - набір параметрів енкодера.

Тоді, нейронні мережі декодера сегментації та класифікації можна визначити як F_{seg} та F_{cls} відповідно. Для набору карт ознак, що генерує енкодер, маємо:

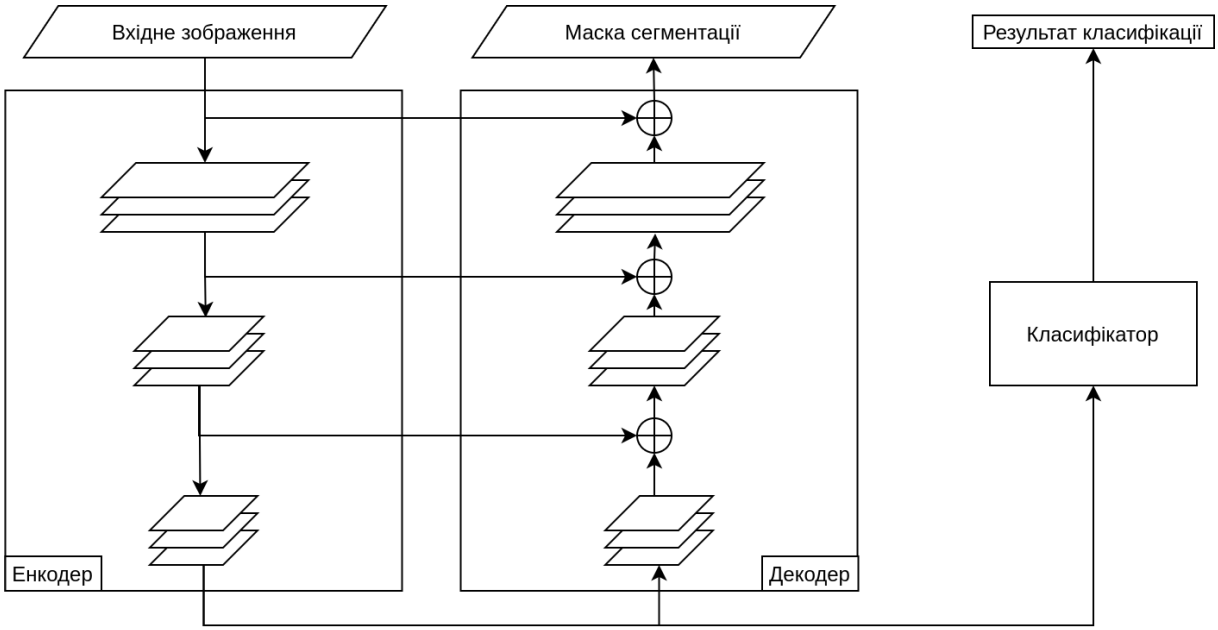

Рис. 3.2: Загальна архітектура нейронної мережі з декодером та класифікатором

$$M_{seg} = F_{seg}((v_1, v_2 ... v_n), \theta_{seg}) \tag{3.5}$$

$$C_{cls} = F_{cls}((v_n), \theta_{cls}) \tag{3.6}$$

де θ_{seg} та θ_{cls} - набори параметрів декодерів сегментації та класифікації відповідно.

Для пришвидшення процесу навчання, використано трансферне навчання: набір параметрів енкодера θ_{enc} ініціалізується з використанням параметрів, отриманих після навчання енкодера на наборі даних Imagenet [**?**]. Декодери ініціалізується зв використанням методу Xe: $\theta \in \mathcal{U}(-b, b)$, де b - константа залежна від типу шару [**?**]

Структура декодера семантичної сегментації. Структура декодера сегментації ідентична декодеру архітектури UNet [**?**] за виключенням останнього шару. Декодер складається з декількох стадій, кожна з яких має розміри карт ознак відповідно до карт ознак енкодера.

Оскільки в задачі сегментації з частково-помилковою розміткою одному пікселю може бути призначено декілька класів одночасно, необхідно це враховувати при використанні обмежених функцій втрат. Для формування фінальної карти сегментації використовується спеціальна версія блоку декодера. Структуру окремих стадій показано на рисунку 3.3.

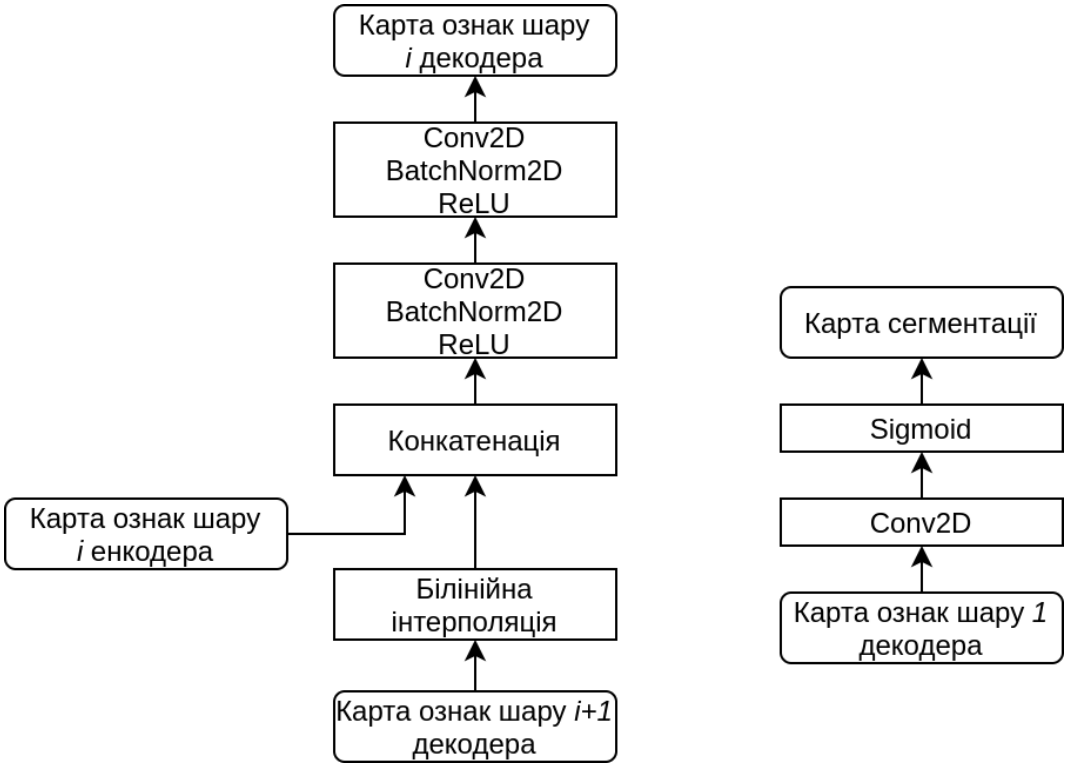

Рис. 3.3: Структура стадій декодеру сегментації

В ролі вхідної карти ознак для найпершого шару декодера використовується останній шар енкодера.

Структура декодера класифікації. Структура декодера класифікації є важливою складовою успішного навчання при використанні обмеженої функції втрат в задачі класифікації. Основними задачами декодеру класифікації є провадження градієнтів до енкодера у випадку фільтрації неправильної розмітки в задачі сегментації, а також покращення роздільності ознак найглибшого шару енкодера.

Оскільки, в даному випадку, на зображенні може бути декілька об'-

єктів, а задача класифікації є задачею навчання за набором зразків, декодер класифікації має враховувати як глобальний просторовий контекст, так і локальні ознаки, притаманні об'єктам. Для цього, першим шаром класифікатора є конкатенація результатів двох операцій глобальної підвибірки: з операцією усереднення (GlobalAvgPooling), та вибору максимума (GlobalMaxPooling) для кожного з каналів карти ознак енкодера.

Для підвищення роздільності ознак, класифікатор являє собою лінійну функцію, що спонукає енкодер до формування лінійно-роздільних представлень. Також, для зменшення впливу відносного масштабу карт ознак, використовується шар пакетної нормалізації перед лінійним шаром.

Структура декодера класифікації зображена на рисунку 3.4.

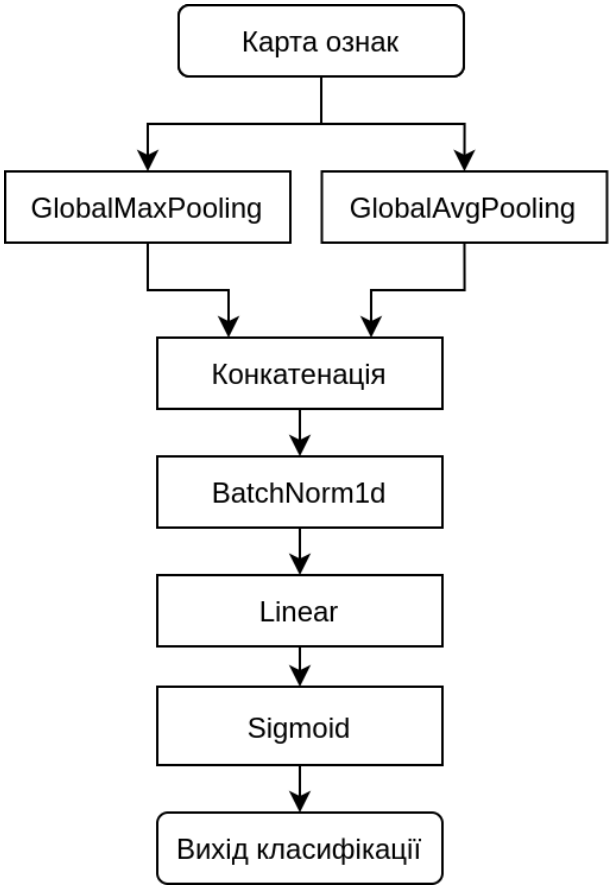

Рис. 3.4: Структура декодера класифікації

Агрегація функцій втрат під час навчання. Під час навчання нейронної мережі використовується стандартний алгоритм зворотного поширення помилки з оптимізатором Adam [**?**].

Для кожної з задач окремо обчислюється функція втрат. Для навчання декодера сегментації використовується обмежена зверху функція втрат, в той час як для декодера класифікації - звичайна

Загальне значення функції втрат визначається як арифметичне середнє між індивідуальними значеннями:

$$L_{total} = \frac{L_{seg}\rceil + L_{cls}}{2} \tag{3.7}$$

Відповідно, загальний градієнт функції втрат буде сумою градієнтів складових частин:

$$\nabla L_{total} = \frac{\nabla L_{seg}\rceil + \nabla L_{cls}}{2} \tag{3.8}$$

Поріг обмеження функції втрат для задачі сегментації є параметром алгоритму навчання та має обиратися емпірично в залежності від рівня помилок в розмітці. Таким чином, забезпечується "прохід"градієнтів для оновлення параметрів від хоча б однієї функції втрат для кожного вхідного прикладу.

3.1.4. Метод зниження хибно-позитивних результатів семантичної сегментації за рахунок використання задачі класифікації. Оскільки одна нейронна мережа одночасно вивчає як задачу сегментації, так і задачу класифікації, стає можливою імплементація фільтрування хибно-позитивних ознак без підвищення витрат часу на етапі прогнозування результатів. В такому випадку можливі два варіанти комбінації результатів двох задач: послідовний і паралельний.

Нехай $C_{cls} \in \mathcal{R}^{C}$ та $M_{seg} \in \mathcal{R}^{C \times H \times W}$- результати декодерів класифікації та сегментації відповідно, значення яких знаходяться на проміжку

$(-\inf, +\inf)$ (логіти).

Для отримання результатів на проміжку $[0, 1]$ використовується логістична сигмоїдна функція активації:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.9}$$

Послідовний варіант полягає у відкиданні результатів сегментації, якщо результат класифікації менший, за деякий поріг t_{cls}:

$$M_{refined} = \begin{cases} \sigma(M_{seg}) & \text{якщо } \sigma(C_{cls}) > t_{cls} \\ 0^{C \times H \times W} & \text{якщо } \sigma(C_{cls}) \leqslant t_{cls} \end{cases} \tag{3.10}$$

Такий варіант має декілька суттєвих недоліків:

— Необхідність вибору додаткового параметра θ_{cls} вносить додаткову можливість помилки; неправильний підбір цього параметра може призвести до значного збільшення хибно-негативних результатів.

— Можливість розповсюдження помилки класифікатора виникає через значно меншу кількість параметрів порівняно з декодером сегментації.

Паралельний варіант полягає у зважуванні карти сегментації за допомогою нормованих логітів класифікатора. Першим кроком є трансформація логітів сегментації та класифікації в некалібровані оцінки на проміжку $[0, 1]$:

$$\hat{M}_{seg} = \sigma(M_{seg}) \tag{3.11}$$

$$\hat{C}_{cls} = \sigma(C_{cls}) \tag{3.12}$$

Ці оцінки мають ті самі розмірності, що й оригінальні маска та класи, для зручності репрезентації операцій додано додаткові розмірності до вектору класів: $\hat{M}_{seg} \in \mathcal{R}^{C \times H \times W}$ та $\hat{C}_{cls} \in \mathcal{R}^{C \times 1 \times 1}$

Зважування карти сегментації відбувається за допомогою добутку Адамара між матрицями $\hat{M}_{seg}$ та $\hat{}_{cls}$

$$M_{refined} = \hat{M}_{seg} \circ \hat{}_{cls} \tag{3.13}$$

Графічну репрезентацію структури зображено на рисунку 3.5.

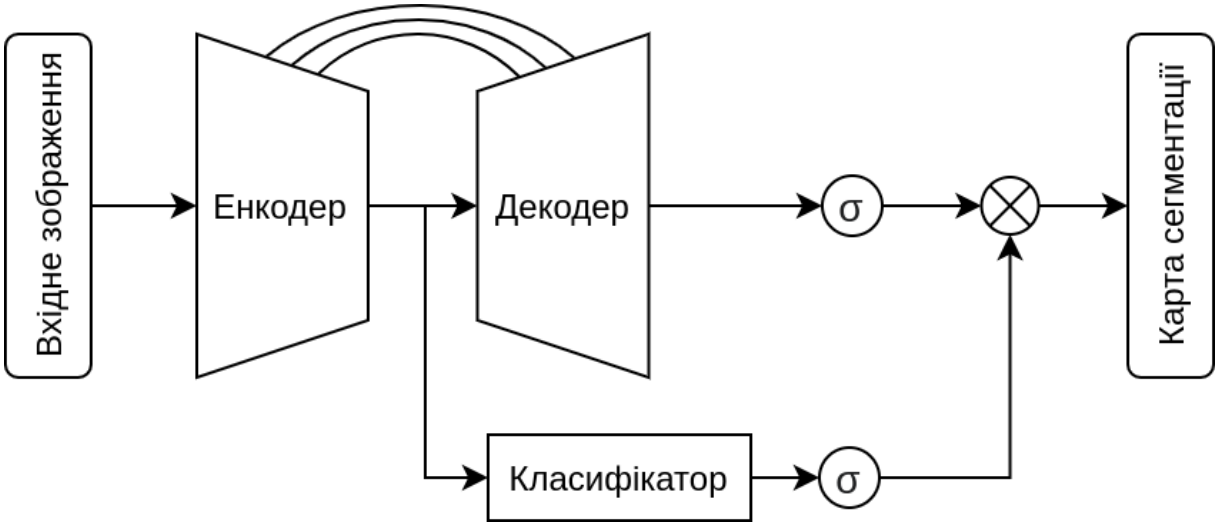

Рис. 3.5: Структура об'єднання прогнозів класифікації та сегментації

Паралельний варіант має перевагу над послідовним у відсутності додаткового параметру навчання. Також, зважені карти сегментації допомагають збереженню більшої кількості інформації для етапу пост-обробки масок сегментації.

Однак, в разі використання добутку двох сигмоїдних функцій, потрібно зважати, що змінюється центральна точка, і, відповідно поріг бінаризації в подальшій обробці. Навіть у випадку співпадання значень класифікації та сегментації, фінальний прогноз буде відрізнятися від послідовного підходу.

Графік сигмоїдної функції, та добутку двох ідентичних сигмоїдних функцій зображено на рисунку 3.6.

При використанні того самого порогу бінаризації для карт сегментації, зменшиться кількість позитивних пікселів в результаті.

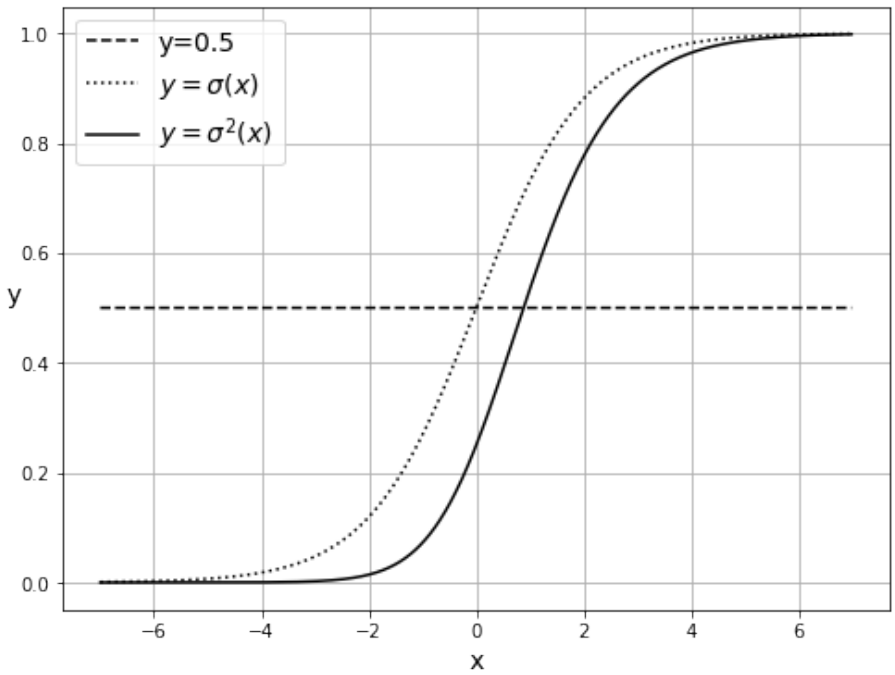

Рис. 3.6: Структура декодера класифікації

3.2. Метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки для сегментації в навчальному наборі даних

На основі двох представлених методів, запропоновано метод навчання з частковим залученням вчителя та метод пост-обробки, що дозволяє виконувати сегментацію важливих для класифікації ознак. Даний метод складається з двох етапів: етапу навчання та етапу прогнозування.

Відповідно до попереднього розділу використані позначення:

x - вхідне зображення енкодера розмірами $3 \times H \times W$

$F_{encoder}(x, \theta_{enc})$ - функція енкодера зображень,

θ_{enc} - набір параметрів нейронної мережі енкодера,

$v_1, v_2, ...v_i, ...v_n$ - набір карт ознак енкодера, відповідно до рівняння 3.4,

F_{seg} та F_{cls} - нейронні мережі декодера сегментації та класифікації,

M_{seg} та C_{cls} - результати сегментації та класифікації відповідно до рівнянь 3.5 та 3.6

$\hat{M}_{seg}$ та $\hat{C}_{cls}$ - нормовані результати сегментації та класифікації відповідно до рівнянь 3.11 та 3.12

3.2.1. Напівавтоматичне навчання ШНМ в задачі сегментації.

Для покращення можливостей інтерпретації прогнозів моделей, було вдосконалено метод сегментації важливих для класифікації ознак зображення за рахунок використання методів багатозадачного навчання, що дозволило його використання в умовах відсутності розмітки семантичної сегментації в навчальному наборі даних.

В основі запропонованого методу лежить ітеративне уточнення карти ознак сегментації за допомогою направлення градієнтів від задачі класифікації. В даному випадку, під час навчання декодер сегментації використовується в ролі механізму уваги, що навчається автоматично за рахунок градієнтів до задачі класифікації.

Першим етапом, обчислюється уточнена ознак класифікації. Для цього, обчислюється добуток Адамара між нормованим за допомогою сигмоїдної функції виходом декодера сегментації та логітами класифікації:

$$M_{unsup} = \hat{M}_{seg} \circ C_{cls} \tag{3.14}$$

Далі, для отримання результату класифікації виконується сумація елементів M_{unsup} з нормалізацією за сумою елементів оригінальної ненормалізованої карти сегментації:

$$C_{unsup} = \frac{\sum_{h=0}^{H} \sum_{w=0}^{W} M_{unsup(h,w)}}{\sum_{h=0}^{H} \sum_{w=0}^{W} M_{seg(h,w)} + c} \tag{3.15}$$

Для чисельної стабільності, до знаменника додано малу константу $c \approx 10^{-5}$

Оскільки масштаб нормованого виходу декодера сегментації знаходиться на проміжку $[0, 1]$, використання добутку Адамара дозволяє розгля-

дати $\hat{M}_{seg}$ як карту важливості регіонів для задачі класифікації. В процесі навчання нейронної мережі, така структура поєднання задач спонукає нейронну мережу до призначення високих значень ($(\hat{(M)}_{seg} \to 1)$ для важливих ознак, що є спільними на зображеннях з навчального набору даних.

Для запобігання вивченню нейронною мережею карт сегментації, що складаються виключно з високих значень, необхідна така ініціалізація параметру зсуву останнього шару декодера сегментації F_{seg}, щоб активації, за замовчуванням, були близькими до нуля. Для сигмоїдної функції активації таке значення дорівнює -4.6

3.2.2. Прогнозування результатів сегментації. На етапі прогнозування використовується як декодер класифікації, так і декодер, що відповідає за задачу сегментації. Тільки у тому випадку, коли вихід декодера класифікації перевищує заданий поріг T_c, виконується процедура декодування сегментації.

Оскільки виходи декодера сегментації M_{unsup} є неперервними, а їхній масштаб визначається процесом навчання, поріг бінаризації сегментації T_{seg} може бути різним для різних зображень. Для вибору оптимального порогу T_{seg} на кожному з вхідних зображень, в процесі пост-обробки запропоновано використати адаптивну бінарізацію за методом Оцу [**?**], щоб уникнути необхідності калібрації прогнозів нейронної мережі.

Для цього, вихід декодера сегментації M_{unsup} квантизується до 256 значень M_q, після чого ітеративно знаходиться поріг T_{seg}, що мінімізує дисперсію всередині каналів маски для кожного з класів, яка визначається як зважена сума дисперсій класів переднього та заднього плану:

$$\sigma_w^2(T_{seg}) = min: \ \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \tag{3.16}$$

Тут ω_0, ω_1 - ймовірності класів при розділенні порогом t, а σ_0^2 та σ_1^2 - дисперсії класів.

Після виконання бінаризації, над отриманою маскою $M_t = M_q > T_{seg}$ виконується морфологічна операція ерозії [**?**] з квадратним ядром розміру 1% від розміру зображення для того, щоб позбутися малих можливо хибно-позитивних регіонів маски:

$$M_e = M_t \ominus k \tag{3.17}$$

де k - ядро ерозії.

Результатом такого перетворення э бінаризована карта сегментації, в якій розмічено найбільш вірогідні ознаки, що вплинули на результат класифікації.

3.2.3. Висновки до третього розділу.

РОЗДІЛ 4
ІНСТРУМЕНТАЛЬНІ ЗАСОБИ

4.1. Узагальнена структура системи автоматизованого скринінгу

4.2. Узагальнена структура системи навчання глибоких нейронних мереж

4.3. Узагальнена структура прогнозування за допомогою глибоких нейронних мереж

РОЗДІЛ 5
ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ В РІЗНИХ ЗАДАЧАХ СКРИНІНГУ

В даному розділі розглянуто експерименти на синтетичних даних, що були згенеровані за допомогою моделі, а також експерименти в реальних задачах скринінгу.

5.1. Експерименти на синтетичних даних

5.1.1. Метод оцінки прогнозів нейронної мережі.

Основні метрики для оцінки задач. В наведених нижче експериментах використовуються наступні метрики:

— Метрика F1 для задачі класифікації

— Коефіцієнт Дайса для задачі семантичної сегментації

— Коефіцієнт роздільності внутрішніх представлень K_{sep}

В експериментах поставлена задача багатокласової класифікації та семантичної сегментації, для отримання фінального значення метрик, вони усереднені по класах.

Метод кількісної оцінки роздільності внутрішніх представлень нейронної мережі. Для оцінки роздільності внутрішніх представлень, використовується метод K-середніх. Оскільки в задачах класифікації та семантичної сегментації відома кількість класів, можна заздалегідь визначити необхідну кількість кластерів ознак у внутрішньому представленні нейронної мережі.

Для набору даних над ознаками з внутрішнього представлення обчислюється кількість кластерів, що відповідає кількості класів в основній

задачі. Після чого, обчислюється середня Евклідова відстань між центрами кластерів. Таким чином, більшій відстані між кластерами відповідає краща роздільність внутрішніх представлень.

Також, для наочної демонстрації складу внутрішніх представлень, ознаки візуалізовано за допомогою алгоритму нелінійного відображення Uniform Manifold Approximation and Projection (UMAP) [**?**] на двомірну площину.

5.1.2. Структура нейронних мереж. У всіх експериментах використана однакова структура нейронних мереж якщо не зазначено інакше.

Структура базової моделі. Базова модель являє собою екземпляр архітектури UNet. В ролі енкодера використана архітектура нейронної мережі ResNet34, що ініціалізована вагами претренованої на наборі даних Imagenet нейронної мережі.

Декодер виконано відповідно до класичної архітектури UNet, що складається з п'яти стадій (рисунок 3.3). Кількість каналів в згортках стадій від найглибшої: 256, 128, 64, 32, 16. Функція активації декодера - ReLU. На кожній стадії використовується пакетна нормалізація ознак. Додаткові з'єднання від енкодера передаються за допомогою операції конкатенації. Функція активації останнього шару декодера - логістична сигмоїда.

Структура запропонованої моделі. Енкодер та декодер сегментації запропонованої моделі ідентичні до базової моделі. Додано декодер класифікації, архітектура якого зображена на рисунку. Входом декодера класифікації є вихід останнього шару енкодера.

При навчанні запропонованої моделі використовується обмежена функція втрат для задачі сегментації, а при прогнозуванні використовується метод об єднання задач класифікації та сегментації.

5.1.3. Параметри експериментів.

Загальна процедура навчання. Складається з одного циклу навчання та валідації.

Використано наступні параметри навчання:

— Розмір зображення: 224×224 пікселя

— Розмір пакету навчання: 32

— Кількість тренувальних зображень: 15000

— Кількість тестових зображень: 5000

— Кількість епох навчання: 20

— Оптимізатор: RAdam [**?**]

— Темп навчання: 10^{-3}

— Коефіцієнт L2 регуляризації параметрів мережі: 10^{-4}

— Закон зміни темпу навчання: косинус

Експеримент 1. Порівняння базового та запропонованого методів на простому наборі даних без зашумлення розмітки. Параметри генерації набору даних зазначено в таблиці 5.1.

Таблиця 5.1: Параметри набору даних експеримента 1

Об'єкти	MNIST
Текстура фону	константна
Текстура об'єктів	константна
P_e	0
P_d	0
N_{obj}	10
S_{obj}	36
Δ_{max}	0.2

Результати експерименту наведено в таблиці 5.2.

Таблиця 5.2: Результати експеримента 1

Модель	Міра Дайса	F1-міра	K_{sep}
Базова модель	0.86	-	14.33
Запропонований метод	0.87	0.98	17.51

На рисунку 5.1 зображено приклади прогнозів нейронної мережі для тестових зображень.

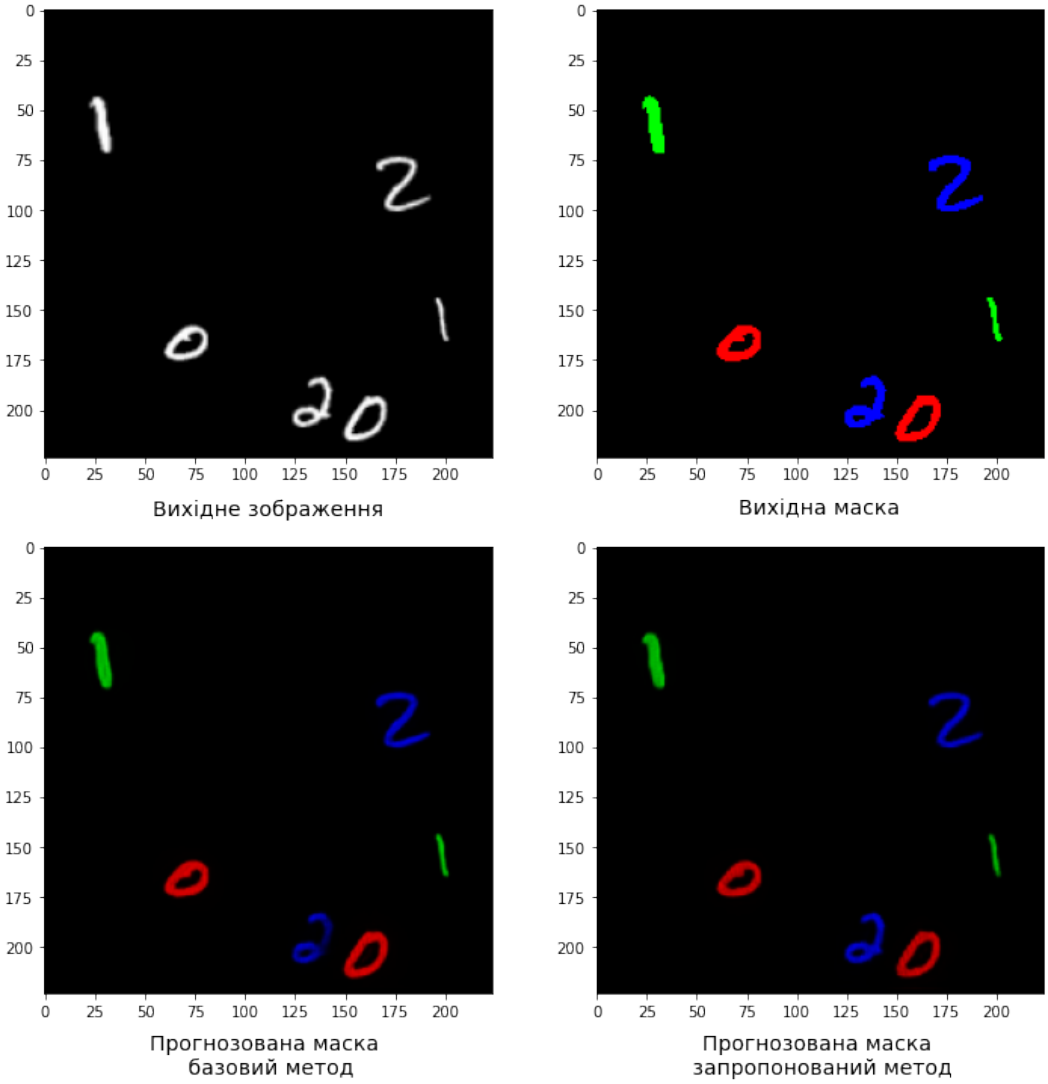

Рис. 5.1: Структура декодера класифікації

На рисунках 5.2 та 5.3 зображено UMAP-проекцію ознак останнього шару енкодера для набору тестових зображень.

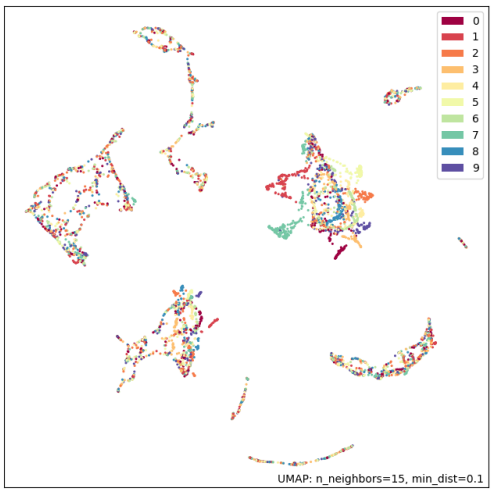

Рис. 5.2: first figure

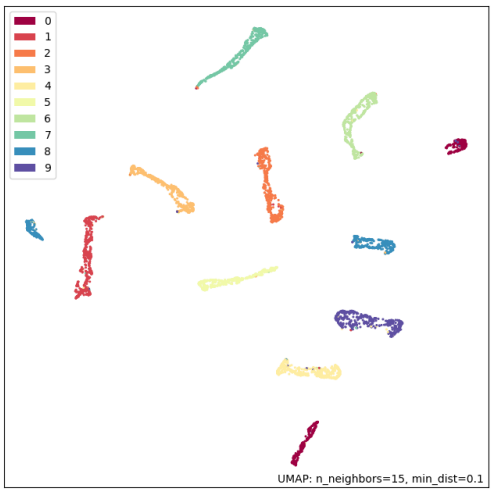

Рис. 5.3: second figure

Експеримент 2. Порівняння базового та запропонованого методів на простому наборі даних із зашумленням розмітки. Параметри генерації набору даних зазначено в таблиці 5.3.

Таблиця 5.3: Параметри набору даних експеримента

Об'єкти	
Текстура фону	
Текстура об'єктів	
P_e	
P_d	
N_{obj}	10
S_{obj}	36
Δ_{max}	0.2

Результати експерименту наведено в таблиці 5.4.

Таблиця 5.4: Результати експеримента

Модель	Міра Дайса	F1-міра
Базова модель		
Запропонований метод		

На рисунку **??** зображено приклади прогнозів нейронної мережі для тестових зображень.

На рисунку **??** зображено UMAP-проекцію ознак останнього шару енкодера для набору тестових зображень.

Експеримент 3.

Експеримент 4.

Експеримент 5.

Експеримент 6. Стійкість методів до зашумлення розмітки на простих даних.

Експеримент 7. Стійкість методів до зашумлення розмітки на складних даних.

Дослідження внеску окремих компонентів. Параметри генерації набору даних зазначено в таблиці **??**.

Таблиця 5.5: Параметри набору даних для проведення дослідження абляції

Текстура фону	Imagentte
Текстура об'єктів	Imagenette
P_e	0.2
P_d	0.2
N_{obj}	10
S_{obj}	36
Δ_{max}	0.2

В таблиці 5.6 наведено значення метрики при послідовному додаванні складових частин запропонованого метода до базової моделі.

Таблиця 5.6: Дослідження абляції

Змінна	Коефіцієнт Дайса	F1-міра
Базова модель		-
+ декодер класифікації		
+ обмеження функції втрат		
+ комбінація прогнозів		

З зазначеної таблиці видно, що найбільший приріст дають обмеження функції втрат та комбінація прогнозів, хоча для отримання повного ефекту необхідні всі складові частини запропонованого методу.

5.2. Експерименти на реальних даних

5.2.1. Класифікація стадії осередків діабетичної ретинопатії.

5.2.1.1. Опис наборів даних. Використані в цьому експерименті зображень, були взяті з декількох наборів даних. Для претренування нейронних мереж було використано відкритий набір даних від Kaggle: “Diabetic Retinopathy Detection Challenge 2015” dr2015.

Цей набір даних є найбільшим із загальнодоступних. Він складається з 35126 фотографій очного дна для лівого та правого ока американських громадян, з розміченими стадіями діабетичної ретинопатії за стандартним протоколом:

— Відсутність діабетичної ретинопатії (мітка 0)
— Легка діабетична ретинопатія (мітка 1)
— Помірна діабетична ретинопатія (мітка 2)
— Важка діабетична ретинопатія (мітка 3)
— Проліферативна діабетична ретинопатія (мітка 4)

Крім того, були використані менші набори даних: набір зображень індійської діабетичної ретинопатії (IDRiD) idrid, з якого використано 413 фотографій очного дна, та MESSIDOR (Методи оцінки методів сегментації та індексації в області офтальмології сітківки) messidor, з якого використа-

но 1200 фотографій очного дна. Оскільки розмітка оригінального набору даних MESSIDOR відрізняється від інших наборів даних, ми використали версію, яку група офтальмологів messidor_relabeled помаркувала відповідно до стандартного протоколу.

Усі зазначені набори даних мають однаковий розподіл міток класів, що є фундаментальною властивістю для цієї задачі.

Оцінка проводиться на наборі даних Kaggle APTOS2019 aptos2019, дослідники мають доступ лише до тренувальної та валідаційної частин. Повний набір даних складається з 18590 фотографій очного дна, які розділені на набори з 3662 тренувальних, 1928 валідаційних та 13000 тестових зображень, що були розділені організаторами змагань Kaggle.

5.2.1.2. Детальний опис структури нейронної мережі.

5.2.2. Детальний опис процедури навчання.

5.2.2.1. Ініціалізація нейронної мережі. Відпочатку, енкодер ініціалізується параметрами нейронної мережі, що була натренована на наборі даних ImageNet. Параметри декодерів ініціалізуються випадково (ініціалізація Xe).

5.2.2.2. Трансферне навчання. Використання трансферного навчання можливо, оскільки природні особливості діабетичної ретинопатії узгоджуються між різними людьми і не залежать від набору даних. Крім того, різні набори даних збираються на різному обладнанні. Включення цих знань у модель підвищує її здатність до узагальнення та підвищує важливість природних ознак за рахунок зменшення чутливості до особливостей обладнання.

Для попереднього навчання, ініціалізована нейронна мережа навчається на протязі 20 епох на наборі даних DRDC2015 за допомогою стохастичного градієнтного спуску. Основна мета попереднього навчання - створити

ініціалізацію параметрів на розподілі даних, що є близьким до цільового. Після попереднього навчання, параметри нейронної мережі використовуються як ініціалізація для основного навчання.

Під час попереднього навчання, кожен декодер мінімізує свою функцію втрат: перехресну ентропію для класифікаційного декодера, бінарну перехресну ентропію для декодера порядкової регресії та середню абсолютну похибку для декодера регресії.

Основне тренування проводиться на наборах даних APTOS2019, IDRID та MESSIDOR разом. Починаючи з ваг, отриманих на етапі попереднього тренування, виконується 5-кратна перехресна перевірка та оцінка моделі на відкладеному наборі даних. На цьому етапі функції втрат для декодерів змінено: Focal loss focalloss для класифікаційного декодера, бінарна Focal loss focalloss для декодера порядкової регресії та середньоквадратична помилка для декодера регресії.

Зміна функцій втрат на даному етапі пов'язана з необхідністю знизити значення втрат для зразків, що вже коректно класифіковані, щоб зменшити вплив ознак, що не відносяться до захворювання.

Для кожної з підвиборок було натреновано одну модель, протягом 75 епох, використовуючи алгоритм оптимізації Rectified Adam, radam, зі зміною темпу навчання за косинусним законом. Щоб зберегти параметри енкодера, поки декодери перебувають у випадковому стані, оновлення параметрів енкодера не відбувається на протязі перших п'яти епох навчання.

5.2.2.3. Комбінування результатів декількох задач. На етапі після основного навчання, для кожної з моделей, навчається модель лінійної регресії з виходів декодерів в єдине значення.

Лінійна регресія навчається після основного навчання, оскільки в іншому випадку, сходиться до неоптимальних локальних мінімумів з вагами двох декодерів, близьких до нуля. Ці нульові ваги запобігають оновленню

відповідних ваг декодерів і, відповідно, запобігають навчанню. Початкові ваги для кожного з виходів декодерів були встановлені рівними 1/3, а потім тренувались протягом п'яти епох, щоб мінімізувати середньоквадратичну функцію помилки.

5.2.2.4. Результати експерименту.

5.3. Класифікація раку шкіри та сегментація родимок

5.3.1. Опис набору даних SIIM-ISIC Melanoma Classification. Дані зображень, використані в цьому дослідженні, були взяті з декількох наборів даних з однаковою структурою: SIIM&ISIC з 2017, 2018, 2019 та 2020 років. Ці набори даних були створені Міжнародною співпрацею з обробки зображень шкіри (англ. International Skin Imaging Collaboration - ISIC), а зображення отримані з наступних джерел:

— лікарня Клінік де Барселона
— Віденський медичний університет
— Центр раку Меморіал Слоун Кеттерінг
— Австралійський інститут меланоми
— Квінслендський університет
— Афінська медична школа

Загалом, ці набори даних складаються із приблизно 50000 RGB-зображень, з яких близько 3000 мають зображення злоякісних уражень. Набір даних містить 434 повторюваних зображення. Окрім даних про зображення, були надані метадані про пацієнтів. Один пацієнт має декілька зображень різних родинок.

Зображення та метадані надані у форматі DICOM, який є загальновживаним форматом даних медичних зображень. Крім того, набір даних доступний у форматі JPEG із розмірами зображень, змінених до 1024x1024.

Метадані також надаються за межами формату DICOM, у файлах CSV [12].

Метадані містять наступну інформацію:

— image_name - унікальний ідентифікатор, вказує на ім'я файлу відповідного зображення DICOM;
— patient_id - унікальний ідентифікатор пацієнта;
— sex - стать пацієнта;
— age_approx - приблизний вік пацієнта на момент фотографії (ціле число);
— anatom_site_general_challenge - розташування родинки на тілі;
— diagnosis - детальна інформація про діагностику (рядок);
— benign_malignant - показник злоякісності зображеного ураження (cnhjrf, одна з "benign"і "malignant");
— target - бінарізована версія цільової змінної.

Значення для anatom_site_general_challenge та diagnosis беруться із заздалегідь визначеного кінцевого набору.

Метадані доступні для кожного пацієнта, тому різні зображення можуть мати однаковий набір функцій рівня пацієнта. Ми використовуємо всі доступні метадані, крім ідентифікатора пацієнта та діагностики, оскільки вони доступні лише в навчальних наборах даних.

Набір даних має дисбаланс високого класу. Розподіл діагнозів показано на рис. 1. Для значень unknown в полі diagnosis, автори набору даних гарантують, що новоутворення не є злоякісним [13].

Через те, що зображення в наборі даних з різних джерел та мають різні стандарти візуалізації вони мають структурований шум у вигляді лінійних смуг, областей, позначених ручкою, центруючих ліній тощо. Залежно від місця новоутворення та статі, також спостерігається волосся. Всі ці додаткові ознаки можуть значно вплинути на тренувальний процес і можуть призвести до перенавчання.

5.3.2. Детальний опис структури нейронної мережі.

5.3.3. Детальний опис процедури навчання. В даному експерименті використано одноступеневий навчальний процес, який включає одночасно трансферне навчання та багатозадачне навчання.

Навчання проводиться у 3-кратній схемі перехресної перевірки. Енкодер відпочатку ініціалізується за допомогою вагів, отриманих за допомогою алгоритму [25]. У наших експериментах ми спостерігали, що така ініціалізація призводить до стабільно кращих результатів, ніж ініціалізація Imagenet.

5.3.4. Результати експерименту.

5.4. Класифікація та сегментація формацій хмар

5.4.1. Опис набору даних Understanding Clouds from Satellite Images. Набір даних Understanding Clouds from Satellite Images (UCSID) складається з 10000 RGB зображень, зроблених з двох супутників TERRA та AQUA, що знаходяться на полярній орбіті [6]. Кожен з цих супутників проходить над певну областю один раз на день. Через обмежене поле зору камер, встановлених на цих супутниках, кожне зображення зшито з двох супутників, що знаходяться над однією й тою самою областю одночано, але на різних орбітах. Решта зображення, для якої не було знято даних під час прольоту супутників (між орбітами) заповнена чорним кольором. Оскільки зображення підлягають компресії, є артефакти стиснення, та чорний колір має деякі незначні аномалії.

На знімках є регіони, які містять певні хмарні утворення, та помічені дослідниками відповідно: Риба, Квітка, Гравій, Цукор (англ. Fish, Flower, Gravel, Sugar). Кожне зображення має принаймні одне хмарне утворення і може містити до всіх чотирьох одночасно.

Розмітка для цих регіонів була створена під час краудсорсингу в Інституті метеорології імені Макса Планка в Гамбурзі, Німеччина, та Laboratoire de m?t?orologie Dynamique у Парижі, Франція. Команда з 68 вчених виявила ділянки з хмарами на кожному зображенні, і кожне зображення було розмічене, в середньому, трьома різними вченими. Кожен вчений мав виділити хмари за допомогою прямокутних областей на власний розсуд. Основна розмітка була створена об'єднанням областей, що були розічені всіма вченими для цього зображення, після видалення чорної смуги з цих областей [7].

Через те, як був зібраний UCSID, він має значну кількість пікселів шуму масках для хмар. Оскільки маски складаються з прямокутників, які повністю перекривають хмару, існує багато пікселів, які позначені як хмари, але, насправді, відповідають фону. Також, не всі хмари позначені масками. Через об'єднання масок від різних анотаторів, класи можуть суттєво перекриватися, включаючи випадки, коли всі чотири класи присвоюються одним і тим самим пікселям.

Усі зображення мають вихідну роздільну здатність 2100x1400 пікселів. Більшість зображень містять більше ніж один клас хмар. Немає зображень без хмар. Розподіл класів хмар на зображеннях показано на рисунку.

Різні типи хмар зазвичай зустрічаються разом, розподіл комбінацій різних типів хмар зазначено на рисунку. UCSID розділено на 5546 тренувальних, та 3698 валідаційних зображень. Крім того, немає доступу до міток валідаційного набору даних. Значення метрики можливо отримати через систему валідації.

До розподілу міток класів в наборі даних не було виконано змін (недодискретизація, передискретизація тощо).

5.4.2. Детальний опис структури нейронної мережі.

5.4.3. Результати експерименту.

ВИСНОВКИ

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Fu J., Zhu X., Li Y. Recognition of surface defects on steel sheet using transfer learning // CoRR. — 2019. — Vol. abs/1909.03258. — 1909.03258.

[2] Noergaard M., Grauslund J. Automated screening for diabetic retinopathy – a systematic review // Ophthalmic Research. — 2018. — 01. — Vol. 60.

[3] Tymchenko B., Marchenko P., Spodarets D. Deep learning approach to diabetic retinopathy detection. — 2020. — 2003.02261.