

ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ОДЕСЬКА ПОЛІТЕХНІКА»
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

Timchenko B. I. 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 ДУОП, Одеса, 2021.

У **вступі** обґрунтовано актуальність розробки моделей нейронних мереж та методів їх навчання в задачах автоматизованого скринінгу, а також моделей наборів даних і методів їх генерації; визначені мета, об'єкт, предмет, задачі і методи дослідження; показано зв'язок з науковими програмами, планами; наведено наукову новизну та практичне значення одержаних результатів; висвітлено особистий внесок здобувача.

В **першому розділі** наведено огляд літератури за тематикою даної роботи та спорідненими питаннями; висвітлено результати, які були отримані іншими дослідниками. Зокрема, наведено огляд досліджень, що стосуються задачі автоматизованого скринінгу в різних областях; проведено аналіз моделей та методів комп'ютероного зору, насамперед, моделей глибоких нейронних мереж та методів їх навчання. Проведено аналіз методів розпізнавання, класифікації та семантичної сегментації планарних зображень, використання для цього моделей та методів багатозадачного навчання. Додатково проаналізовано метрики, що використовуються для оцінки достовірності в задачах класифікації та сегментації зображень. Виявлено, що достовірність прогнозів може бути визначена як міра Дайса-Соренсена, або F1-міра. Наведено підходи використання багатозадачного навчання для класифікації та сегментації зображень в задачах автоматизованого скри-

нінгу.

Проведено аналіз зашумлення розмітки в наборах даних для різних 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 зайні маски на місцях, де немає об'єктів;
5. Зображенням присвоєні невірні класи.

Наведено огляд наявних моделей наборів даних, та методів їх генерації.

Виявлено, що відсутні моделі наборів даних, які б дозволяли змінювати рівень зашумлення розмітки відповідно до характеристик зашумлення в реальних наборах даних.

Виявлено, що 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*.

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*.

Таким чином, проведений огляд сучасної літератури на тему дисертації дозволяє аргументувати актуальність та практичну цінність проведених у роботі досліджень.

У другому розділі представлено параметричну модель наборів даних із зашумленою розміткою для задач сегментації та класифікації, що відповідає оцінкам моделей шуму в різних задачах автоматизованого скринінгу, а також метод контролюваної генерації як зображень, так і масок сегментації та міток класів що модифікуються відповідно до випадково обраних недоліків.

Запропоновано використання модельних наборів даних тестування достовірності моделей та методів навчання глибоких нейронних мереж в задачах класифікації та семантичної сегментації. Тестування виконується в контролюваних умовах за допомогою введення зашумлення в розмітку лише тренувального набору даних, в той час як тестувальних набір даних залишається з точною розміткою.

Модель набору даних має наступні параметри:

- Набір зображень фону \mathcal{X}_b
- Набір об'єктів \mathcal{X}_f
- Набір зображень текстур об'єктів \mathcal{X}_{tex}
- Кількість зображень в генерованому наборі даних N
- Розмір генерованих зображень 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}

На відміну від існуючих моделей, в запропоновану модель вводиться чотири додаткових параметри:

- Ймовірності зменшення та збільшення маски кожного з об'єктів P_e та P_d
- Допустимі масштаби збільшення та зменшення масок всіх об'єктів S_e та S_d

Введення цих параметрів дозволяє створити контролюване зашумлення розмітки відповідно до реальних характеристик.

На основі запропонованої параметричної моделі розроблено метод генерації наборів даних із зашумленою розміткою, який складається з двох етапів: етапу генерації фону та етапу розташування на ньому довільної кількості об'єктів. Для генерації фону можуть бути використані як звичайні натуруальні зображення, так і синтетичні текстури, чи заливка константним кольором. В якості об'єктів запропоновано використання зображень 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 під час навчання, 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.

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 зображені: $n_{obj} \sim \mathcal{U}(1, N_{obj})$
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 Вибрati координати розмiщення об'єкta:

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

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

4.3 Вибрati зображення об'єкta $x_{fg} \sim \mathcal{X}_f$ та вiдповiдний клас об'єкta $c_{fg} \sim \mathcal{Y}_f$

4.4 Змiнити розмiр зображення об'єкta за допомогою бiлiнiйної iнтерполяцiї:

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

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

4.6 Модiфiкувати зображення об'єкta за допомогою текстури:

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

4.7 Розмiстити зображення об'єкta на зображенi фонu:

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

4.8 Сформувати маску сегментацiї об'єкta:

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

де θ_{seg} - порiг бiнаризацiї вихiдного зображення об'єкta. Для набору даних MNIST $\theta_{seg} = 0.2$, для набору даних FashionMNIST $\theta_{seg} = 0.1$.

4.9 Модiфiкувати маску сегментацiї вiдповiдно до необхiдного рiвня 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 ерозiї.

4.10 Розмістити модифіковану маску сегментації об'єкта на загальному зображені 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.

Третій розділ роботи присвячено моделям глибоких нейронних мереж, а також методам багатозадачного навчання для одночасного підвищення достовірності класифікації та семантичної сегментації без зміни витрат часу.

Запропоновано модель глибокої нейронної мережі з використанням архітектури енкодер-декодер, що заснована на архітектурі нейронних мереж LinkNet [?]. Розроблена модель 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. Карті ознак після кожноого етапу просторового зменшення використовуються як входи для декодера сегментації, для декодера класифікації використовується карта ознак з найглибшого шару енкодера. Графічне представлення моделі зображено на рисунку 3

Також, запропоновану модель представлено математично. 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, \dots, 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):

$$v_1, v_2, \dots, v_n = F_{encoder}(x, \theta_{enc}) \quad (1)$$

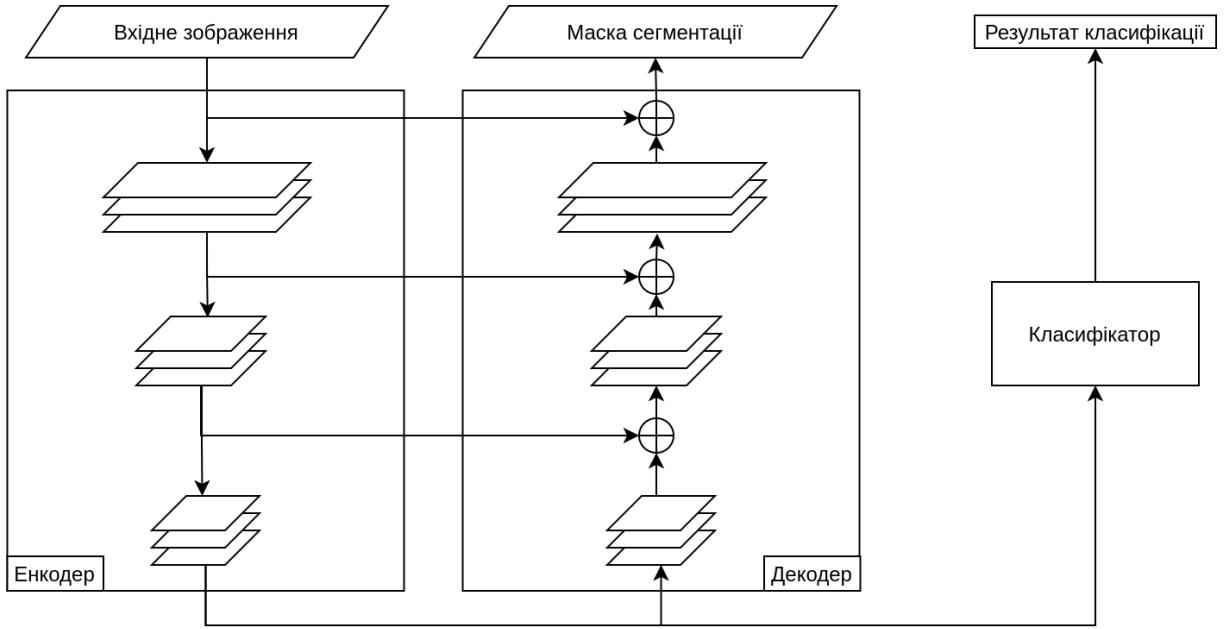


Рис. 1: Загальна архітектура нейронної мережі з декодером та класифікатором

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:

$$M_{seg} = F_{seg}((v_1, v_2 \dots v_n), \theta_{seg}) \quad (2)$$

$$C_{cls} = F_{cls}((v_n), \theta_{cls}) \quad (3)$$

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

Перевагою використання базової архітектури LinkNet є можливість застосовувати трансферне навчання для пришвидшення процесу тренування: набір параметрів енкодера θ_{enc} ініціалізується з використанням параметрів, отриманих після навчання енкодера на наборі даних Imagenet. Декодери ініціалізуються зв використанням методу Хе: $\theta \in \mathcal{U}(-b, b)$, де b - константа залежна від типу шару.

На основі розробленої моделі нейронної мережі запропоновано методи багатозадачного навчання глибоких нейронних мереж та прогнозування результатів. Метод тренування нейронних мереж в умовах частково-помилкової розмітки навчальних даних спирається на використання задач, похідних від оригінальної. Показано, що для задачі семантичної сегментації існує семантично-близька більш загальна задача, для якої розмітка навчальних даних є більш влучною, ніж для вихідної.

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-різномірних 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 зображені, зображення являє собою мішок з одним, чи декількома об'єктами та відповідним маркуванням, чи є об'єкти заданих класів на зображені.

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:

$$y_c = t < \sum_i^H \sum_j^W y_{s_{ij}} \quad (4)$$

де t - поріг мінімального розміру об'єкта в пікселях

Згенерована таким чином задача класифікації має меншу ймовірність хибної розмітки.

Для зменшення впливу помилкової частини розмітки, вперше 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 $(-\infty, \theta]$, тому для проміжку (θ, ∞) градієнт встановлено рівним нулю:

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

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:

$$L_{total} = \frac{L_{seg}] + L_{cls}}{2} \quad (5)$$

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

$$\nabla L_{total} = \frac{\nabla L_{seg}] + \nabla L_{cls}}{2} \quad (6)$$

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.

На основі запропонованих моделі нейронної мережі та методу її 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.

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 $(-\infty, +\infty)$ (логіти).

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

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

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

$$\begin{aligned}\hat{M}_{seg} &= \sigma(M_{seg}) \\ \hat{C}_{cls} &= \sigma(C_{cls})\end{aligned}$$

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{C}_{cls}

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

Для покращення можливостей інтерпретації прогнозів моделей, було вдосконалено метод локалізації 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 локалізації за допомогою направлена градієнтів від 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:

$$M_{unsup} = \hat{M}_{seg} \circ C_{cls} \quad (7)$$

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

$$C_{unsup} = \frac{\sum_{h=0}^H \sum_{w=0}^W M_{unsup(h,w)}}{\sum_{h=0}^H \sum_{w=0}^W M_{seg(h,w)} + c} \quad (8)$$

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 розглядати \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 $(\hat{M}_{seg} \rightarrow 1)$ для важливих ознак, що є спільними на зображеннях з навчального набору даних. де $\sigma(x)$ - логістична функція.

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

На етапі прогнозування використовується як задача класифікації, так і задача локалізації. Тільки у тому випадку, коли вихід класифікації перевищує заданий поріг T_c , виконується процедура декодування локалізації, в іншому випадку, карта локалізації вважається рівною нулю для всіх пікселів.

Оскільки виходи декодера сегментації M_{unsup} є неперервними, а їхній масштаб визначається процесом навчання, поріг бінаризації локалізації 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.

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 хибно-позитивних регіонів маски.

Наведені моделі та методи було апробовано на наступних предметних доменах: скринінг діабетичної ретинопатії, скринінг дефектів сталевих листів, скринінг меланоми, та скринінг хмарних утворень.

У четвертому розділі описано розроблені інструментальні засоби, що реалізують запропоновані методи та випробування розробленого методу в рамках експериментів як на синтетичних даних, що були згенеровані за допомогою запропонованої моделі, а також експерименти в реальних задачах скринінгу.

Інструментальні засоби розроблено мовою програмування *Python* з використанням фреймворку автоматичного диференціювання *PyTorch*. На основі розроблених інструментальних засобів створено ефективні програм-

ні модулі, які інтегровано з «хмарними» сервісами для вирішення ресурсомістких задач навчання нейронних мереж, що забезпечує високу обчислювальну потужність та швидкість прогнозування в задачах автоматизованого скринінгу.

За допомогою моделі протестовано запропоновані методи в різних умовах, виконаний аналіз внеску окремих компонентів та проведено аналіз стійкості запропонованого методу до різних рівнів зашумлення розмітки. В середньому, підвищення коефіцієнта Дайса відносно базової моделі склало 13%. Для високих рівнів зашумлення, підвищення склало 42%.

У рамках проекту «*Severstal: Steel Defect Detection*» на платформі для змагань з наук про дані Kaggle було використано запропоновані моделі нейронних мереж та запропоновані методи їх навчання та прогнозування результатів для задачі семантичної сегментації дефектів сталевих листів. Підвищення достовірності (міра Дайса) відносно базової моделі склало 4.2%.

У рамках проекту «*Understanding Clouds from Satellite Images*» на платформі для змагань з наук про дані Kaggle було використано запропоновані моделі нейронних мереж, а також методи їх навчання та прогнозування результатів для задачі семантичної сегментації патернів організації хмар на супутникових знімках. Підвищення достовірності (міра Дайса) відносно базової моделі склало 3.9%.

У рамках проекту «*APTOs 2019 Blindness Detection*» на платформі для змагань з наук про дані Kaggle було використано запропоновані методи багатозадачного навчання та прогнозування результатів в задачі класифікації фотографій очного дна для скринінгу діабетичної ретинопатії. Підвищення достовірності (F1-міра) відносно базової моделі склало 2.1%.

При виконанні проекту «*SIIM-ISIC Melanoma Classification*» на платформі для змагань з наук про дані Kaggle було використано запропонованій локалізації важливих для класифікації ознак зображення в задачі розпізнавання уражень шкіри для скринінгу меланоми. Використання

запропонованого методу дозволило спростити процес контролю навчання нейронних мереж, що допомогло попередити перенавчання і підвищити достовірність класифікації на 3.5%.

Розроблені в роботі методи та інструментальні засоби отримали **впровадження** в програмний продукт SafetyRadar компанії *ViTech Lab*, основним призначенням якого є скринінг наявності елементів засобів індивідуального захисту на людях в умовах будівельних майданчиків, або лікарень та лабораторій.

Ключові слова: аналіз зображень, глибокі нейронні мережі, багатозадачне машинне навчання, семантична сегментація, функції втрат.

ABSTRACT

Tymchenko B. I. Neural Network Methods for Planar Image Analysis in Automated Screening Systems. — Qualification scientific work in the form of manuscript.

Thesis for doctor of philosophy degree in speciality 122 — Computer science. — ICS ONPU, Odesa, 2021.

? ??????? the relevance of developing models of neural networks and methods of their training in automated screening tasks, as well as models of data sets and methods of their generation are substantiated; defined purpose, object, subject, tasks and research methods; the connection with scientific programs, plans is shown; the scientific novelty and practical significance of the obtained results are given; the personal contribution of the applicant is covered.

? ??????? ???? a review of the literature on the subject of this work and related issues; highlighted the results obtained by other researchers. In particular, an overview of research related to the problem of automated screening in various fields; the analysis of models and methods of computer vision, first of all, models of deep neural networks and methods of their training is carried out. The analysis of methods of recognition, classification and semantic segmentation of planar images, use of models and methods of multitasking learning for this purpose is carried out. Additionally, the metrics used to assess the reliability in the problems of classification and segmentation of images are analyzed. It was found that the reliability of predictions can be defined as a measure of Dice-Sorensen, or F1-measure. Approaches to the use of multitasking training for classification and segmentation of images in

automated screening tasks are presented.

An analysis of markup noise in data sets for various automated screening tasks was performed, which showed that the main problems with data markup in screening tasks are:

1. Segmentation masks that capture pixels adjacent to the object; item Segmentation masks that do not completely cover the object; item There are no segmentation masks for some objects; item There are extra masks in places where there are no objects; item Invalid classes assigned to image.

An overview of existing models of data sets and methods of their generation is given. It was found that there are no data set models that would allow to change the noise level of the markup according to the noise characteristics in real data sets. It was found that the inconsistency and imbalance of training data sets for automated screening tasks occurs due to the relatively small number of anomalous examples in populations and a significant level of markup error. This creates a problem when teaching deep neural networks by standard methods. Therefore, when building classification and segmentation models, it is necessary to solve the problem textit learning with partial error markup.

In automated screening tasks, it is important to reduce the number of false-positive diagnostic results, due to the relatively low prevalence of abnormal cases in populations and associated with false-positive results. Therefore, when building classification and segmentation systems, you need to focus on *reducing false-positive results*.

Thus, the review of modern literature on the topic of the dissertation allows to argue the relevance and practical value of the research conducted in the work.

The ??????? ??????? presents a parametric model of datasets with noisy markup for segmentation and classification tasks, which corresponds to

the estimates of noise models in various automated screening tasks, as well as a method of controlled generation of both images and segmentation masks and class labels modified accordingly to randomly selected shortcomings.

The use of model datasets for testing the reliability of models and methods of learning deep neural networks in the problems of classification and semantic segmentation is proposed. Testing is performed under controlled conditions by introducing noise into the markup of the training data set only, while the test data set remains with accurate markup.

The data set model has the following parameters:

- Set of background images \mathcal{X}_b
- Set of objects \mathcal{X}_f
- A set of image textures of objects \mathcal{X}_{tex}
- The number of images in the generated data set N
- Size of generated images S_{img} in pixels
- The average size of the object S_{obj} in pixels
- Maximum deviation of object size Δ_{max} as a percentage
- The maximum number of objects in the image N_{obj}

In contrast to existing models, the proposed model introduces four additional parameters:

- The probability of decreasing and increasing the mask of each of the objects P_e ?? P_d
- Values to increase and decrease the masks of all objects are allowed S_e ?? S_d

Entering these parameters allows you to create a controlled noise markup in accordance with the actual characteristics.

Based on the proposed parametric model, a method of generating data sets with noisy markup has been developed, which consists of two stages: the stage of background generation and the stage of placing an arbitrary number of objects on it. Both regular natural images and synthetic textures,

or a constant color fill can be used to generate the background. The use of images from MNIST or FashionMNIST datasets is proposed as objects.

The use of these data sets is due to three factors:

- Possibility of simple separation of the object from the original background;
- The presence of similar elements in different classes (for example, numbers 1 and 7, or classes *T-shirt* and *Dress*);
- High accuracy of modern neural networks on these data sets, which allows you to focus on the impact of noise in the markup during training, instead of recognizing objects directly.

Errors in the markup are artificially introduced by randomly applying morphological operations of erosion and dilatation with a square core to the segmentation masks of individual objects before adding them to the general mask.

The training and test datasets are generated based on the model, all images in the datasets are generated independently.

The algorithm for generating each of the images consists of the following steps:

1. Choosing a random background image: $x_{bg} \sim \mathcal{X}_b$
2. Select the number of objects in the image: $n_{obj} \sim \mathcal{U}(1, N_{obj})$
3. Initialization of the segmentation mask: $M = 0$??? ?? $M \in \mathcal{R}^{C \times S_{img} \times S_{img}}$
4. According to the number of images n_{obj} perform the following steps:
 - 4.1 Choose the size of the object: $s \sim \mathcal{U}(S_{obj} - \Delta_{max}, S_{obj} + \Delta_{max})$
 - 4.2 Select the coordinates of the object:

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

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

- 4.3 Select an image of the object $x_{fg} \sim \mathcal{X}_f$ and the corresponding object class $c_{fg} \sim \mathcal{Y}_f$

4.4 Resize the image of the object using bilinear interpolation:

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

4.5 Select a texture image $x_{tex} \sim \mathcal{X}_{tex}$

4.6 Modify the image of the object using the texture:

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

4.7 Place the image of the object on the background image:

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

4.8 Create an object segmentation mask:

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

?? θ_{seg} - the binarization threshold of the original image of the object.

For a data set MNIST $\theta_{seg} = 0.2$, for a data set FashionMNIST $\theta_{seg} = 0.1$.

4.9 Modify the segmentation mask according to the required error rate:

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

where $K^{S_d \times S_d}$ - kernel matrix $K^{S_e \times S_e}$ - erosion core matrix.

4.10 Place the modified object segmentation mask on the overall image of the segmentation mask:

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

Thanks to this generation option, it is possible to obtain multiple data sets with similar characteristics, which allows the use of non-parametric statistical methods, such as bootstrapping to evaluate models. In practice, to generate different data sets, it is sufficient to change the initialization of the random number generator.

?????? ?????? The work is devoted to models of deep neural networks, as well as methods of multitasking learning to simultaneously increase the reliability of classification and semantic segmentation without changing the time spent.

The proposed model of deep neural network architecture using encoder-decoder, based on the architecture of neural networks LinkNet [?]. The developed model consists of an encoder and two decoders: for segmentation and classification tasks, respectively. In the role of the encoder can be used by existing multi-architecture, such as VGGNet, ResNet, EfficientNet and others. Feature maps after each stage of spatial reduction are used as inputs for the segmentation decoder, for the classification decoder is used a feature map from the deepest layer of the encoder. Graphical representation of the model is shown in the figure3

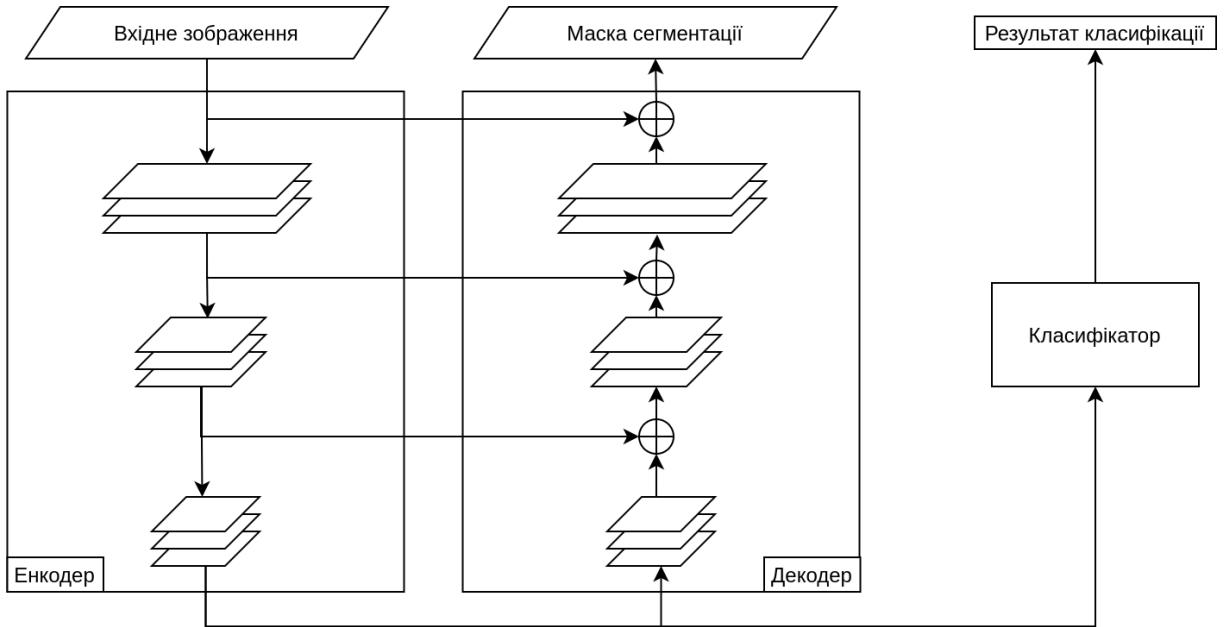


Figure 2: General neural network architecture with decoder and classifier

Also, the proposed model is presented mathematically. Let the neural network of the encoder be defined as $F_{encoder}$. Then, for the input image $x \in \mathcal{R}^{3 \times H \times W}$, $v_1, v_2, \dots, v_i, \dots, v_n$ - a set of sign cards, so $v_i \in \mathcal{R}^{c_i \times \frac{H}{i} \times \frac{W}{i}}$, ?? c_i -

the number of channels for each of the feature cards, ? n - number of stages encoder (depending on the architecture):

$$v_1, v_2 \dots v_n = F_{encoder}(x, \theta_{enc}) \quad (9)$$

?? θ_{enc} - encoder parameter set.

Then, the neural networks of the segmentation and classification decoder can be defined as F_{seg} and F_{cls} accordingly. For a set of feature maps generated by the encoder, we have:

$$M_{seg} = F_{seg}((v_1, v_2 \dots v_n), \theta_{seg}) \quad (10)$$

$$C_{cls} = F_{cls}((v_n), \theta_{cls}) \quad (11)$$

where ?????? ?????????? ?????????? ?????????? ?? ?????????? ???????????. θ_{seg} ?? θ_{cls} -

The advantage of using the basic LinkNet architecture is the ability to apply transfer training to speed up the training process: the set of encoder parameters θ_{enc} is initialized using the parameters obtained after training the encoder on the Imagenet data set. The decoder is initialized using the Xe method: $\theta \in \mathcal{U}(-b, b)$, ?? b - ?????????? ?????? ??? ??? ???.

Based on the developed neural network model, methods of multitasking learning of deep neural networks and prediction of results are proposed. The method of training neural networks in terms of partially erroneous markup of educational data is based on the use of tasks derived from the original. It is shown that for the problem of semantic segmentation there is a semantically closer more general problem for which the markup of educational data is more accurate than for the original.

Unlike previous methods based on the study of more detailed semantically similar problems, the developed method of using more accurate data for more general problems allows to improve the resolution of internal representations of the neural network, which, in turn, improves the results on the

original problem. Also, because the tasks are semantically similar, there is no conflict of gradients, which is typical for learning semantically heterogeneous problems. Thus, more general to the problem of semantic segmentation is the problem of classification. In this context, the task of classification is reduced to the task of learning by a set of samples: instead of marking each of the objects for all classes in the image, the image is a bag with one or more objects and appropriate labeling, whether objects of given classes in the image. Suppose that for the image $x \in \mathcal{R}^{3 \times H \times W}$ there is a segmentation mask $y_s \in \mathcal{R}^{C \times H \times W}$ where C - the number of classes, and, H and W the height and width of the image, respectively. If there is at least one marked object of class $?$, in the segmentation mask, the label of the corresponding class $y_c \in (0, 1)^C$ is set in the markup of the classification problem:

$$y_c = t < \sum_i^H \sum_j^W y_{s_{ij}} \quad (12)$$

where t - the threshold of the minimum size of the object in pixels

The classification problem generated in this way has a lower probability of erroneous markup.

To reduce the impact of the erroneous part of the markup, a change to the loss function when learning neural networks was proposed for the first time: the loss function for a problem with less accurate markup is limited at the top, so when learning multiple tasks, gradients from incorrect markup do not affect the learning process:

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

where θ - the threshold limit function loss. For the top constraint function, the gradient is defined only on the interval $(-\infty, \theta]$, ?? so for the period (θ, ∞) gradient is set to zero:

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

For each task separately calculated function loss. The top-down loss function is used to train the segmentation decoder, while the classification decoder uses the normal one. The total value of the loss function is defined as the arithmetic mean between individual values:

$$L_{total} = \frac{L_{seg} + L_{cls}}{2} \quad (13)$$

Accordingly, the total gradient function loss will amount gradients components:

$$\nabla L_{total} = \frac{\nabla L_{seg} + \nabla L_{cls}}{2} \quad (14)$$

The threshold for limiting the loss function for the segmentation problem is a parameter of the learning algorithm and should be chosen empirically depending on the level of errors in the markup. Thus, a gradient "pass" is provided to update the parameters from at least one loss function for each input example.

Based on the proposed neural network model and the method of its training, a method of combining semantically similar tasks at the forecasting stage has been developed in order to increase the reliability of classification and segmentation of planar images without increasing time. $C_{cls} \in \mathcal{R}^C$ $M_{seg} \in \mathcal{R}^{C \times H \times W}$ - Results decoders classification and segmentation, respectively, whose values are in the interval $(-\infty, +\infty)$ (logites).

For results on the interval $[0, 1]$ using logistic sigmoid activation function: $[0, 1]$

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

The proposed method is to weigh the segmentation map using normalized classifier logs. The first step is the transformation logit segmentation and classification evaluation period due to an uncalibrated $[0, 1]$:

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

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

These estimates have the same dimensions as the original mask and classes, for the convenience of representation of operations added additional dimensions to the vector of classes: $\hat{M}_{seg} \in \mathcal{R}^{C \times H \times W}$?? $\hat{C}_{cls} \in \mathcal{R}^{C \times 1 \times 1}$

The segmentation map is weighed using the Hadamard product between the matrices \hat{M}_{seg} ?? \hat{C}_{cls}

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

To improve the ability to interpret model predictions, the method of localization of important image features for classification was improved through the use of multitasking learning methods, which allowed its use in the absence of semantic segmentation markup in the training data set.

The basis of the proposed method is an iterative refinement of the map of localization features by directing the gradients from the classification problem:

The first step is to calculate the refined classification features. To do this, the Hadamard product is calculated between the output of the segmentation decoder normalized by the sigmoid function and the classification logs:

$$M_{unsup} = \hat{M}_{seg} \circ C_{cls} \tag{15}$$

Next, to obtain the classification result, the summation of the elements M_{unsup} is performed with normalization by the sum of the elements of the original non-normalized localization map:

$$C_{unsup} = \frac{\sum_{h=0}^H \sum_{w=0}^W M_{unsup(h,w)}}{\sum_{h=0}^H \sum_{w=0}^W M_{seg(h,w)} + c} \quad (16)$$

For numerical stability, added to the denominator constant low $c \approx 10^{-5}$

Since the scale of the normalized output of the segmentation decoder is in the range $[0, 1]$, the use of the Hadamard product allows us to consider \hat{M}_{seg} as a map of the importance of regions for the classification problem. During training the neural network, this structure encourages a combination of tasks neural network to the appointment of high value ($(\hat{M})_{seg} \rightarrow 1$) for important characteristics that are common in images of the training data set. where $\sigma(x)$ is a logistic function.

This iterative refinement prompts the neural network to assign high values ($\sigma(M) \rightarrow 1$) to important features that are common in the images from the training data set.

On stage forecasting is used as a classification problem and the problem of localization. Only when the classification output exceeds the specified threshold T_c , the localization decoding procedure is performed, otherwise, the localization map is considered equal to zero for all pixels.

Since the outputs of the segmentation decoder M_{unsup} are continuous, and their scale is determined by the learning process, the binarization threshold for localization T_{seg} may be different for different images. To select the optimal threshold T_{seg} on each of the input images, in the post-processing process, it is proposed to use adaptive binarization according to the Otsu method in order to avoid the need to calibrate the neural network predictions.

After performing binarization, mask received over $M_t = M > T_s$ operation performed morphological erosion with a square kernel size 1% of the image size to get rid of small possible false-positive regions of the mask.

These models and methods were tested in the following subject domains: screening for diabetic retinopathy, screening of steel sheet defects, screening

for melanoma, and screening of cloud formations.

? ?????????? ??????describes the developed tools that implement the proposed methods and tests of the developed method in the framework of experiments on synthetic data generated using the proposed model, as well as experiments in real screening problems.

The tools are developed in the *Python* programming language using the *PyTorch* automatic differentiation framework. Based on the developed tools, effective software modules have been created, which are integrated with "cloud" services for solving resource-intensive learning tasks of neural networks, which provides high computing power and speed of prediction in automated screening tasks.

Using the model, the proposed methods were tested in different conditions, the analysis of the contribution of individual components was performed and the analysis of the resistance of the proposed method to different levels of marking noise was performed. On average, the increase in the Dyce coefficient relative to the base model was 13%. For high noise levels, the increase was 42%.

As part of the *textit {ii} Severstal: Steel Defect Detection {ii}* project, the proposed neural network models were used on the Kaggle Data Science Competition Platform, and methods were proposed for their training and prediction of results for the problem of semantic segmentation of steel sheet defects. The increase in reliability (Dyce measure) relative to the base model was 4.2%.

As part of the project *iiUnderstanding Clouds from Satellite Images{ii}* on the platform for competitions in data science Kaggle used the proposed models of neural networks, as well as methods for training and predicting results for the problem of semantic segmentation of cloud patterns in satellite images. The increase in reliability (Dyce measure) relative to the base model was 3.9%.

As part of the *jiAPTOS 2019 Blindness Detection* project on the Kaggle Data Science Competition Platform, the proposed methods of multitasking learning and predicting results in the task of classifying fundus photographs for the screening of diabetic retinopathy were used. The increase in reliability (F1-measure) relative to the base model was 2.1%.

The project *jiSIIM-ISIC Melanoma Classification* on the Kaggle Data Science Competition Platform used the proposed localization of classification-relevant image features in the problem of recognizing skin lesions for melanoma screening. The use of the proposed method simplified the process of monitoring the learning of neural networks, which helped prevent retraining and increase the reliability of the classification by 3.5%.

The methods and tools developed in the work were ????????????? in the software product SafetyRadar of the company *VI Tech Lab*, the main purpose of which is to screen the presence of elements of personal protective equipment on people in construction sites, or hospitals and laboratories.

Key words: image analysis, deep neural networks, multitasking machine learning, semantic segmentation, loss functions.

ЗМІСТ

Вступ	49
--------------	-----------

Розділ 1. Аналіз моделей та методів обробки планарних зображень в задачах автоматизованого скринінгу	53
1.1. Задача автоматизованого скринінгу	53
1.1.1. Задачі автоматизованого скринінгу в медицині	53
1.1.2. Задачі автоматизованого скринінгу в метеорології . .	58
1.2. Проблема зашумлення розмітки в наборах даних	59
1.2.1. Природне зашумлення в наборах даних	60
1.2.2. Штучне зашумлення в наборах даних	61
1.3. Моделі глибинного машинного навчання	63
1.3.1. Згорткові нейронні мережі	64
1.3.2. Визначні сучасні моделі класифікації зображень на основі ШНМ	66
1.3.3. Визначні моделі ШНМ для сегментації зображень .	75
1.4. Методи багатозадачного глибинного машинного навчання .	82
1.4.1. Визначення багатозадачного машинного навчання .	83
1.4.2. Трансферне навчання	84
1.4.3. Жорсткий розподіл параметрів	84
1.4.4. М'який розподіл параметрів	86
1.4.5. Дистиляція прогнозів моделі	88
1.4.6. Багатозадачне навчання в задачах класифікації та сегментації	88
1.5. Методи оцінки якості класифікації та сегментації зображень в задачі автоматизованого скринінгу	92

1.5.1. Помилки 1 і 2 роду та матриця невідповідностей	92
1.5.2. Влучність, чутливість та специфічність	93
1.5.3. Mіра F1 та індекс подібності Дайса-Соренсена	93
1.5.4. Каппа Коена	94
1.6. Висновки та постановка задачі дослідження	95
Розділ 2. Параметрична модель наборів даних із зашумленою розміткою і метод їх генерації	98
2.1. Модель набору даних із зашумленою розміткою	99
2.2. Метод генерації наборів даних	99
2.3. Види згенерованих зображень	101
2.4. Висновки до другого розділу	107
Розділ 3. Нейромережева модель аналізу планарних зображень та методи її навчання	108
3.1. Модель багатозадачної штучної нейронної мережі	108
3.1.1. Загальна структура нейронної мережі	109
3.1.2. Структура декодера семантичної сегментації	110
3.1.3. Структура декодера класифікації	111
3.2. Методи навчання та прогнозування багатозадачних ШНМ в умовах частково помилкової розмітки семантичної сегментації	113
3.2.1. Автоматична фільтрація помилкової розмітки	114
3.2.2. Генерація похідної задачі до задачі семантичної сегментації	116
3.2.3. Агрегація функцій втрат під час навчання	118
3.2.4. Багатозадачне прогнозування результатів семантичної сегментації	119
3.3. Метод локалізації важливих для класифікації ознак зображення в умовах відсутності розмітки для сегментації в навчальному наборі даних	122

3.3.1. Напівавтоматичне навчання ШНМ в задачі локалізації	123
3.3.2. Прогнозування результатів сегментації	124
3.4. Висновки до третього розділу	125
Розділ 4. Інструментальні засоби та експериментальні дослідження	127
4.1. Експерименти на синтетичних даних	127
4.1.1. Оцінка прогнозів нейронної мережі	127
4.1.2. Структура нейронних мереж	128
4.1.3. Параметри експериментів	129
4.2. Експерименти на реальних даних	145
4.3. Класифікація та сегментація формаций хмар	145
4.3.1. Опис набору даних Understanding Clouds from Satellite Images	145
4.3.2. Процедура навчання	148
4.3.3. Результати експерименту	149
4.3.4. Класифікація стадії осередків діабетичної ретинопатії	149
4.3.5. Процедура навчання	151
4.3.6. Результати експерименту	153
4.4. Класифікація раку шкіри та локалізація родимок	153
4.4.1. Опис набору даних SIIM-ISIC Melanoma Classification	153
4.4.2. Процедура навчання	155
4.4.3. Результати експерименту	156
4.5. Інструментальні засоби	157
4.5.1. Узагальнена структура системи навчання глибоких нейронних мереж	158
4.5.2. Узагальнена структура прогнозування за допомогою глибоких нейронних мереж	159
4.6. Висновки до четвертого розділу	159

ВСТУП

Актуальність теми зумовлюється стрімким розвитком різних областей технологій і процесів, які потребують швидкого реагування на нетипові зміни для подальшого коригування, поки ціна реакції є низькою. Задача виявлення таких нетипових змін називається задачею скринінгу. Для підвищення продуктивності, створюються інтелектуальні системи, що дозволяють вирішувати задачу скринінгу автоматично без участі людини. В таких системах постає задача автоматизованого скринінгу.

Актуальним методом скринінгу є скринінг заснований на аналізі зображень, що дозволяє спростити забір проб, а також використати інструментальні засоби віддаленого спостереження, такі як рентгенографія, або супутникове зондування. В задачах скринінгу, що заснований на зображеннях можуть використовуватися як планарні (рентгенограми, двомірні дефектоскопи, та ін.), так і об'ємні зображення (томографії, тривимірні ультразвукові дослідження та ін.). Особливістю планарних зображень є відсутність перспективних спотворень та можливість знахтувати глибиною об'єктів.

Важливими складовими інтелектуальних систем автоматизованого скринінгу є класифікація зображень та локалізація регіонів інтересу.

Ефективність роботи систем автоматизованого скринінгу напряму залежить від достовірності класифікації та локалізації, оскільки хибно-позитивні результати можуть привести до передчасного реагування, а хибно-негативні результати - до пропуску наявних проблем та пізнього реагування. Традиційно, в таких системах використовуються класичні методи комп'ютерного зору, однак достовірність їхніх прогнозів суттєво обмежена через різноманіття можливих проблем та складність їх формалізації за допомогою правил.

Все частіше для аналізу зображень впроваджуються методи на основі машинного навчання, зокрема глибинного, через значне підвищення достовірності прогнозів. Однак, в наявні підходи на основі глибоких нейронних мереж можуть бути застосовані виключно за умови наявності великих тренувальних наборів розмічених даних. Часто, в задачах скринінгу, отримання великої кількості даних, а також якісної розмітки від професіоналів є дуже дорогим, та іноді не є можливим в принципі через малу кількість прикладів, іх різноманіття та неузгодженість розмітників під час формування наборів даних. Достовірність прогнозів визначається як міра Дайса-Соренсена, або F1-міра на валідаційному наборі з точною розміткою.

Таким чином, має місце **протиріччя** між підвищенням вимог щодо достовірності прогнозів нейромережевих методів з одного боку, та необхідністю зниження витрат на отримання розмітки навчальних даних.

В дисертаційній роботі пропонується новий підхід до розв'язання цього протиріччя шляхом розробки моделей нейронних мереж та методів їх навчання для аналізу планарних зображень, що здатні використовувати неточно розмічені тренувальні набори даних.

Запропонований метод використовує багатозадачне навчання для підвищення достовірності класифікації та семантичної сегментації за рахунок навчання нейронних мереж на більш загальних та семантично-близьких задачах одночасно з основною задачею з подальшою комбінацією результатів виконання цих задач.

Тобто, підвищення достовірності прогнозів нейромережевих методів в задачах автоматизованого скринінгу через розробку та удосконалення методів багатозадачного навчання є **актуальним**.

Робота пов'язана з дослідними проектами: № 700-145 «Моделі, методи та інструментальні засоби підтримки прийняття рішень з підвищення ефективності гідроаеродинамічних процесів в діючому енергетичному обладнанні» (№ ДР 0115U000413); № 448-68 «Дослідження інформаційних сх-

вищ як моделей предметних областей в системах підтримки прийняття рішень» (№ ДР 0104U002401); № 37-62 «Методи моделювання та робочого діагностування складних цифрових систем і мереж» (номер ДР 0110U008194)

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

Для досягнення мети дослідження поставлено і розв'язано такі **задачі**:

- Здійснено аналіз зашумлення розмітки в наборах даних для задач автоматизованого скринінгу, проведено аналіз існуючих моделей, методів, алгоритмів та засобів аналізу планарних зображень в системах автоматизованого скринінгу і обґрунтовано вибір напрямку досліджень.
- На основі проведеного аналізу розроблено параметричну модель набору даних, що складається з планарних зображень із зашумленою розміткою для задач класифікації та семантичної сегментації, а також метод його генерації.
- Розроблено модель нейронної мережі та методи навчання й прогнозування для аналізу планарних зображень, що за рахунок використання декількох семантично-близьких задачах одночасно дозволяють підвищити точність класифікації та семантичної сегментації в умовах зашумлення розмітки даних.
- Розроблено інструментальні засоби, що виконують розроблені моделі та методи, і провести їх випробування в рамках експерименту

Об'єктом дослідження є процес аналізу планарних зображень.

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

Методи дослідження:

Наукова новизна отриманих результатів.

- *Удосконалено* параметричну модель набору даних та метод його генерації, що за допомогою додавання контролюваного зашумлення розмітки, яке відповідає характеристикам зашумлення в реальних задачах автоматизованого скринінгу дає можливість виконувати тестиування нейромережевих методів семантичної сегментації та класифікації через використання синтетичних наборів даних.
- *Удосконалено* модель згорткової нейронної мережі та метод багатозадачного навчання штучних нейронних мереж для роботи в умовах частково помилкової розмітки навчального набору даних в задачах автоматизованого скринінгу що за рахунок навчання на декількох задачах дозволяє підвищити достовірність семантичної сегментації на 8%.
- *Вперше* розроблено метод об'єднання семантично-близьких задач на етапі прогнозування, що за допомогою використання результатів декількох задач дозволило підвищити достовірність класифікації та сегментації планарних зображень на 12% без підвищення витрат часу.
- *Отримав подальший розвиток* метод локалізації важливих для класифікації ознак зображення в умовах відсутності розмітки для сегментації в навчальному наборі даних, що за рахунок

Практичне значення одержаних результатів полягає у створенні інструментальних засобів генерації наборів даних відповідно до розробленої параметричної моделі, а також навчання та прогнозування нейронних мереж для задач автоматизованого скринінгу.

На основі розроблених інструментальних засобів створено ефективні програмні модулі, які інтегровано з «хмарними» сервісами для вирішення ресурсомістких задач навчання нейронних мереж, що забезпечує високу обчислювальну потужність та швидкість прогнозування в задачах автома-

тизованого скринінгу.

Розроблені в роботі методи отримали **впровадження** в програмний продукт SafetyRadar компанії *VI Tech Lab*.

РОЗДІЛ 1

**АНАЛІЗ МОДЕЛЕЙ ТА МЕТОДІВ ОБРОБКИ ПЛАНАРНИХ
ЗОБРАЖЕНЬ В ЗАДАЧАХ АВТОМАТИЗОВАНОГО
СКРИНІНГУ**

1.1. Задача 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. [?]

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

1.1.1. Задачі автоматизованого скринінгу в медицині. Серед багатьох застосувань скринінгу в медицині, особливо нагальною є потреба в методах скринінгу для агресивних прихованіх 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 розрізнати. 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 спільноті 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 глаукома, 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 вико-

ристовувати 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 зображення, 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 вивчені ефективних та надійних алгоритмів глибокого навчання для різних проблем діагностики діабетичної ретинопатії, та звернули увагу на напрямки подальших дослід-

дженъ.

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 90,9%. Сарки 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. 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 не-професіоналами, 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 види-

ному спектрі, так і в інших діапазонах: мультиспектральні та інфрачервоні.

Один із методів виявлення хмар був розроблений Чжу та співавт. [2], вони 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. Проблема зашумлення розмітки в наборах даних

В задачах машинного навчання, поняття зашумлення розмітки (англ. *label noise*) є досить розмитим і його нелегко визначити точно. В літературі цей термін використовується для позначення невідповідності розмітки

істинним даним, її пошкодження, або неідеальність.

Треба відрізняти зашумлення розмітки від зашумлення вихідних даних. Так, для множини зображень X та відповідної її множини міток Y утворюються пари (x_i, y_i) . Зазвичай, в задачах класифікації та семантичної сегментації, множина Y містить дискретні мітки y_i рівня зображення, або рівня окремих пікселей відповідно. Ці мітки часто відповідають справжньому класу елемента x_i , але вони можуть бути піддані впливу шумового процесу перед тим, як будуть представлені алгоритму навчання.

1.2.1. Природне зашумлення в наборах даних. Відповідно до [?], ідентифікація джерел зашумлення розмітки не обов'язково важлива, коли фокус аналізу зосереджений на наслідках зашумлення. Однак, коли модель зашумлення має бути вбудована безпосередньо в алгоритм навчання, може бути важливим обрати модель, яка точно пояснює джерела та вид шуму.

Зазвичай, природне зашумлення розмітки відбувається, коли до процесу розмітки залучені люди-анотатори. Тоді, причинами некоректної розмітки можуть бути схожі структури, недостатній досвід роботи анотаторів, випадкові помилки, а також людські упередженості.

Автори [?] пропонують наступну класифікацію видів зашумлення розмітки:

- Рівномірне зашумлення - ймовірність зміни мітки класу на мітку будь-якого іншого класу розподілена рівномірно.
- Залежне від класу зашумлення - ймовірність зміни мітки класу на мітку іншого класу залежить від цих класів.
- Залежне від ознак зашумлення - ймовірність зміни мітки класу на мітку іншого класу залежить від ознак об'єкта. На відміну від інших двох видів, цей майже не зустрічається в роботах через складність виконання.

Наприклад, набір даних Imagenet, що використовується як основа для трансферного навчання в інженерній практиці, має значний рівень зашумлення розмітки, через те, що в ньому на одному зображенні може знаходитися декілька об'єктів різних класів, але мітка є тільки для одного. Автори [?] оцінюють рівень зашумлення та пропонують варіант для його корекції через розбиття зображень з декількома об'єктами на декілька зображень.

Також, в задачах семантичної сегментації, зашумлення розмітки може бути упередженим [?] відповідно до людей-анотаторів, кожний з яких розмічає зображення певною мірою суб'єктивно. Особливо частою ця проблема є в задачах медичної сегментації - в деяких предметних областях навіть один і той самий анотатор може давати різну розмітку одних і тих самих зображень, в залежності від власного стану [?].

Для того, щоб оцінити приблизний рівень шуму в складних умовах для людей-анотаторів, автори [?] створили спеціальний набір даних, що складається з пар класів зі схожими ознаками, наприклад, кіт та рись, або вовк та койот. Після розмітки прикладів людьми, автори використали перехресну перевірку для оцінки ймовірності зміни мітки класу в парах, яка вийшла рівною 6.44%. Одночасно з набором даних, автори також запропонували метод відновлення правильних міток класів SELFIE.

1.2.2. Штучне зашумлення в наборах даних. В процесі тестування різних методів машинного навчання в умовах зашумлення розмітки виникає потреба в тестових наборах даних, що мають незашумлену розмітку, в той час, як відповідні тренувальні набори даних мають зашумлену розмітку. Оскільки збір таких наборів даних є практично неможливим, основним варіантом отримання є додавання штучного шуму в оригінальну, якісну розмітку тільки для тренувального набору даних.

Наприклад, в роботі [?] запропоновані методи додавання шуму в розмітку на основі зазначененої вище класифікації. Для всіх класів в наборі даних

складається матриця зашумлення N , так що:

Для *рівномірного зашумлення*:

$$N_{ij} = \begin{cases} p & \text{якщо } i = j \\ \frac{1-p}{M-1} & \text{якщо } i \neq j \end{cases}$$

де M - кількість класів.

Для *залежного від класу зашумлення* автори пропонують використовувати матрицю невідповідностей, що створюється попередньо навченою нейронною мережею.

Для *зашумлення, залежного від ознак*, в роботі запропоновано використання претренованої нейронної мережі в якості екстрактора ознак, з подальшою її кластеризацією та побудовою матриці невідповідностей.

Так само, можна використовувати зазначені методи і для задачі семантичної сегментації, змінюючи мітки класу для всього об'єкту в цілому, однак, це не відповідає більшості реальних зашумлень масок сегментації.

Окрім зміни міток класів для об'єктів в цілому, маски семантичної сегментації можуть не повністю покривати об'єкти, або частково не відповідати їх контурам. Для моделювання цих недоліків розмітки, в роботі [?] запропоновано зашумлення упередженого та неупередженого видів: неупереджене зашумлення складається з еластичної трансформації масок для кожного з об'єктів, в той час як упереджене передбачає константний зсув маски відносно об'єкта для всіх зображень. Такі модифікації можна розглядати як окремий випадок зашумлення, залежного від ознак для задачі класифікації, якщо прийняти семантичну сегментацію як задачу попіксельної класифікації.

Також, окремим випадком рівномірного зашумлення є відсутність масок сегментації для певних об'єктів, або їх присутність в місцях, де немає об'єктів. В такому разі рівномірне зашумлення відбувається між класами

о'єктів та класом фону.

В деяких випадках, немає можливості отримати набір даних з реальної предметної області для оцінки методу. В такому разі, дослідники використовують генерацію синтетичних даних, відповідно до характеристик, притаманних предметній області.

Так, автори [?] запропонували генерацію набору коротких відео-послідовностей з цифр, зо рухаються для перевірки можливостей рекурентної моделі до передбачення наступної позиції в послідовності.

Автори [?] розширили попередній набір даних на задачу семантичної сегментації з додаванням зашумлення вхідних даних.

В інженерній роботі [?] запропоновано параметричну модель генерації наборів даних для оцінки моделей та методів семантичної сегментації. Об'єкти даного набору даних є рукописними цифрами MNIST, а фон генерується з випадкових текстур, при цьому, будь-яке зашумлення розмітки відсутнє.

1.3. Моделі глибинного машинного навчання

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

Штучні нейронні мережі — це обчислювальні системи, натхнені біологічними нейронними мережами. ШНМ ґрунтуються 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 сигна-

лізувати 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 відомими 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 зображені на рис. 1.1.

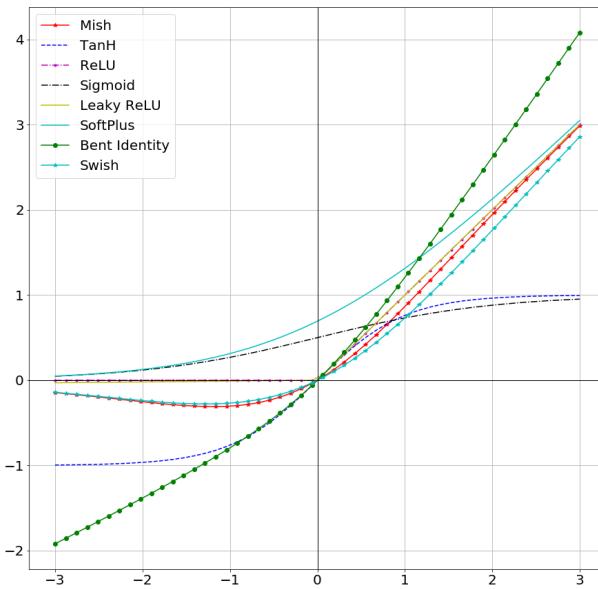


Рис. 1.1: Графіки популярних функцій активації [?]

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 [?, ?, ?].

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.

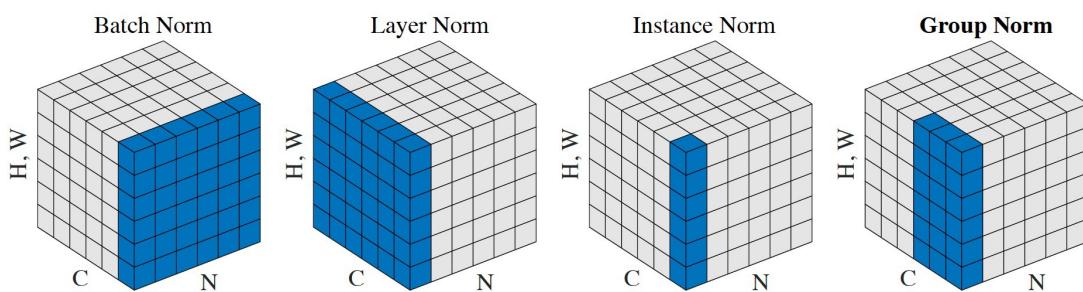


Рис. 1.2: Візуалізація різних типів нормалізації [?]

1.3.2. Визначні сучасні моделі класифікації зображень на основі ШНМ.

Модель нейронної мережі 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 з 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 великої кількості зображень порівняно високої роздільної здатності, автори розробили алгоритм паралельного навчання на декількох графічних процесорах (GPU).

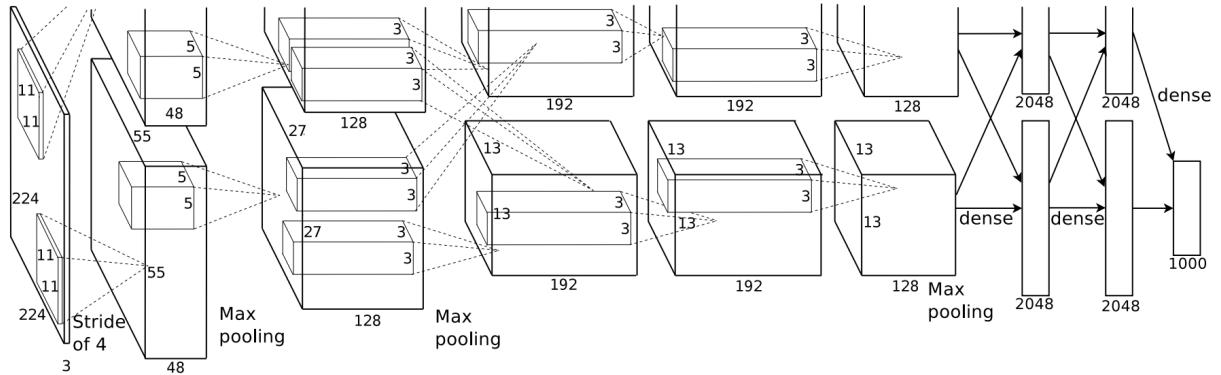


Рис. 1.3: Графічна репрезентація моделі нейронної мережі AlexNet [?]

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

Модель нейронної мережі 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 є одна згортка з розміром ядра 7×7 , але кількість задіяних параметрів становить $3 \cdot (3^2 C^2)$ порівняно з $7^2 C^2$ параметрів для ядер розміром 7×7 . 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.

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 моделей VGGNet з різною кількістю шарів [?]

В даній роботі було продемонстровано, що глибина нейронної мережі вигідна для точності класифікації, а також того, що можна досягти високої точності за допомогою звичайної архітектури згорткової нейронної мережі.

Сімейство моделей 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 1×1 . Один такий шар називається *блоком Inception*, архітектура якого зображена на рис. 1.5

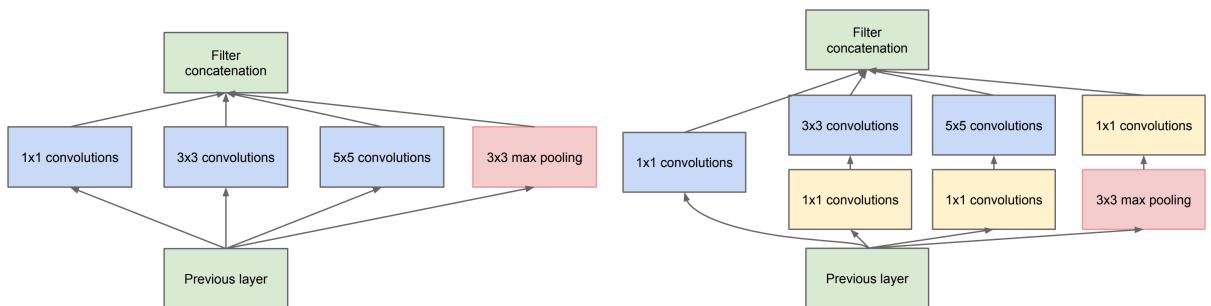


Рис. 1.5: Блок-схема Inception v1 до та після розкладу[?]

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 5×5 на дві згортки з розміром ядра 3×3 , а згортки з розміром ядра 3×3 на дві згортки розміру 1×3 та 3×1 відповідно. Це дозволяє зменшити кількість параметрів в нейронній мережі при збереженні глибини, зменшуючи необхідні ресурси для обчислення.

Схема блоків Inception v2 та v3 зображене на рис. 1.6

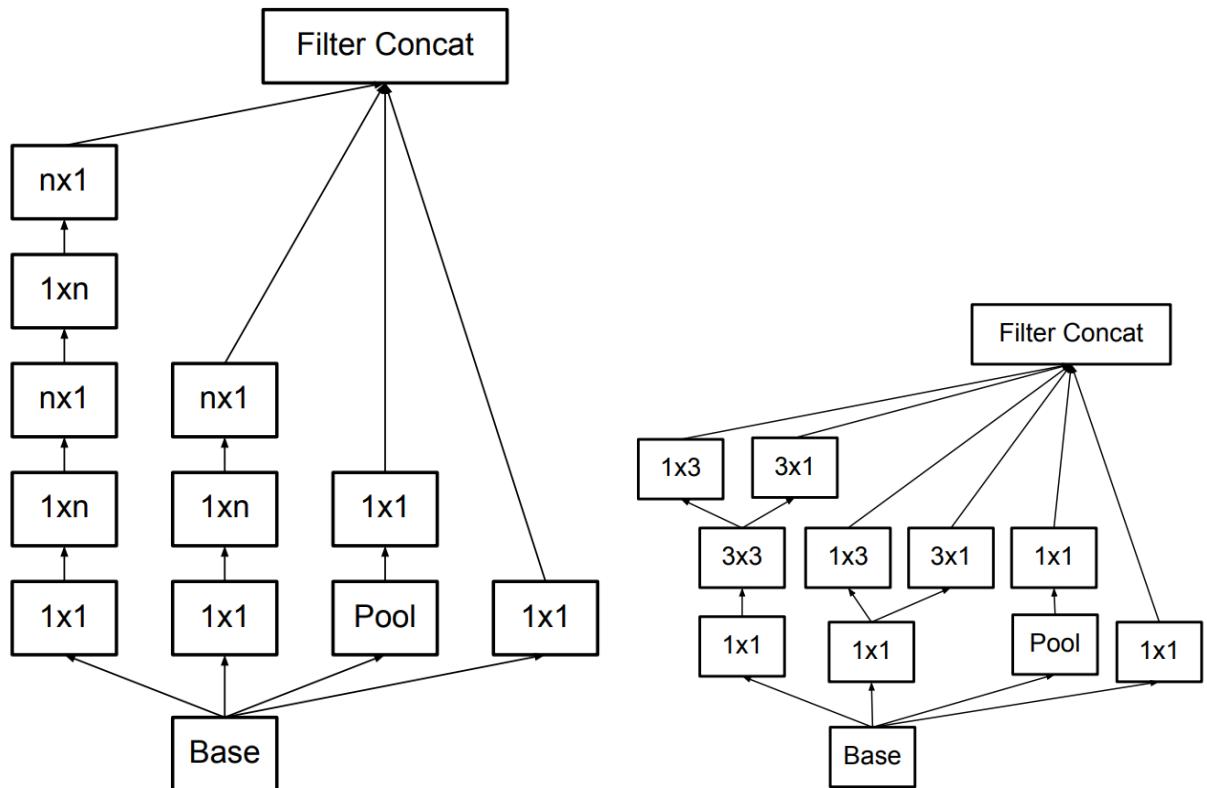


Рис. 1.6: Блок-схема Inception v2 до та після розкладу[?]

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 затухан-

ня градієнтів, але мають властивості регуляризації, особливо на останніх стадіях навчання.

Сімейство моделей 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

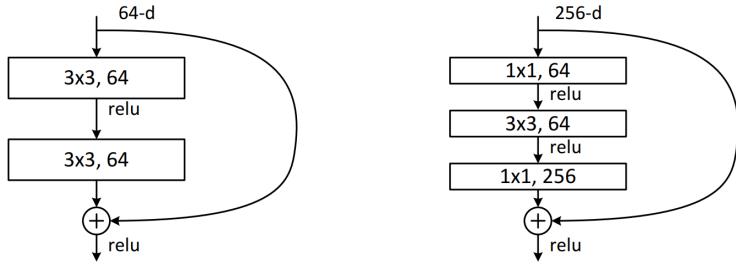


Рис. 1.7: Схема залишкових блоків нейронних мереж ResNet18 та ResNet50 [?]

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

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

де $[x_0, x_1, x_3 \dots x_{l-1}]$ - 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 змінити кількість параметрів в нейронній мережі та підвищити точність.

Граф нейронної мережі DenseNet зображене на рис. 1.8.

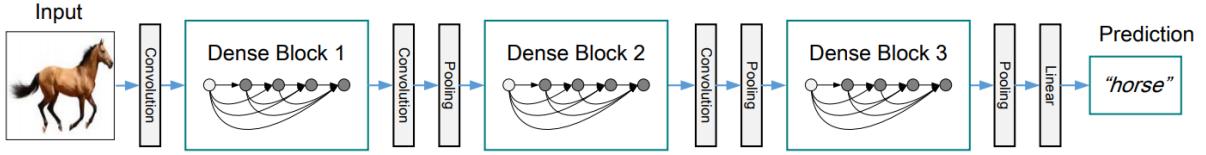


Рис. 1.8: Графічна репрезентація моделі нейронної мережі DenseNet [?]

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

Сімейство моделей 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:

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

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

Графічну репрезентацію блоку ResNeXt зображенено на рис. 1.9

В роботі автори проводять експерименти відносно важливості параметрів

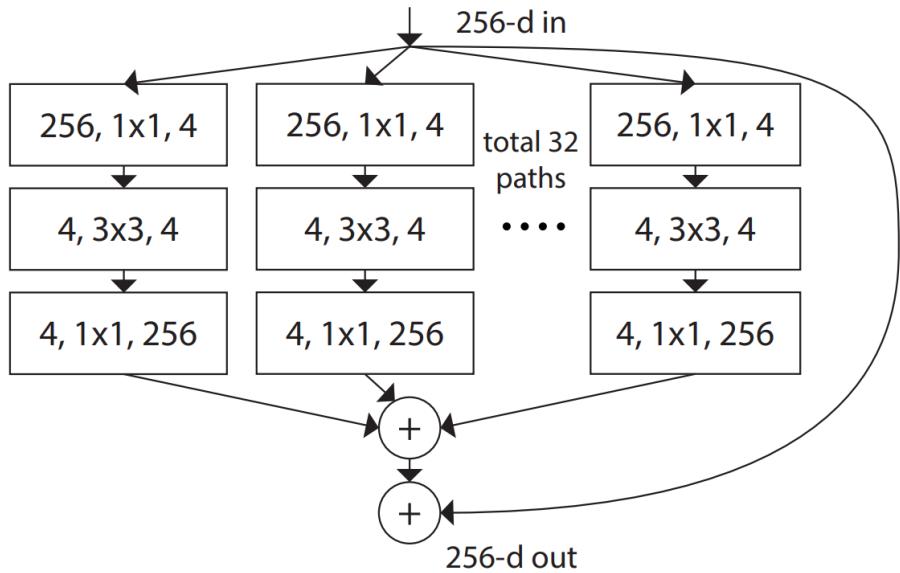


Рис. 1.9: Графічна репрезентація блоку ResNeXt [?]

тру 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 репрезентації вхідних даних.

Сімейство моделей нейронних мереж EfficientNet. Сімейство нейронних мереж EfficientNet також є логічним продовженням архітектури ResNet. В даній роботі вперше розглядається параметрична модель та узагальнений метод масштабування згорткових нейронних мереж для збільшення точності під заданий обчислювальні можливості.

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

Інтуїтивно зрозуміло, що для зображень із більш високою роздільною здатністю необхідно збільшити глибину мережі, щоб більше збільшити поле зору та захопити більше пікселів у більшому зображені. 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, модель нейронної мережі формулюється як задача максимізації точності в залежності від ширини, глибини та роздільної здатності:

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

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

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

та $\hat{\mathcal{F}}_i$ - функція, що задається моделлю 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 :

$$d = \alpha^\phi \quad (1.5)$$

$$w = \beta^\phi \quad (1.6)$$

$$r = \gamma^\phi \quad (1.7)$$

так, що виконується наступне:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (1.8)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \quad (1.9)$$

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 моделі 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 метод трансферного навчання 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 зображенний на рис. 1.10.

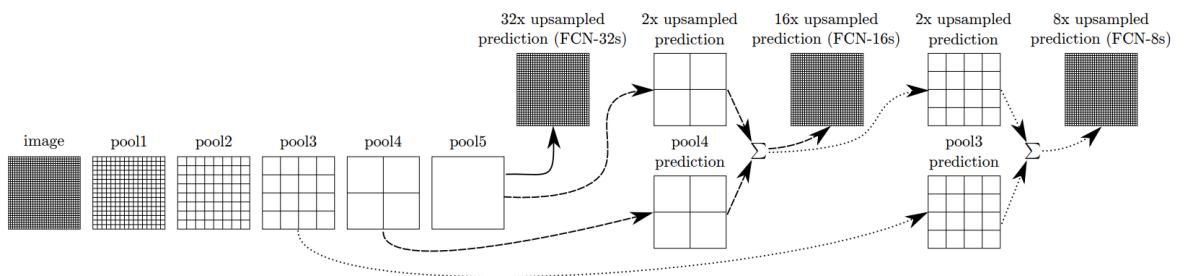


Рис. 1.10: Графічна репрезентація моделі нейронної мережі FCN [?]

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

Модель нейронної мережі 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 відображення з

компактного простору з низькою роздільною здатністю у карту активації 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 активації низького розширення. Так, декодер використовує індекси операцій підвибірки, обчислені на відповідному шарі енкодера, для нелінійного підвищення розширення. Карти активації підвищеного розширення виявляються розрідженими, для того, щоб зробити їх щільними, виконується декілька операцій згортки. Структура SegNet зображена на рисунку ??.

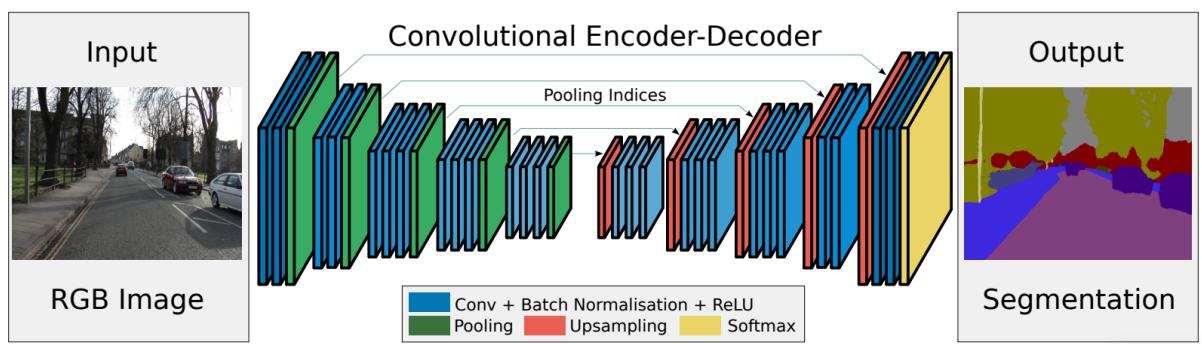


Рис. 1.11: Граф нейронної мережі SegNet [?]

Модель нейронної мережі 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 ??.

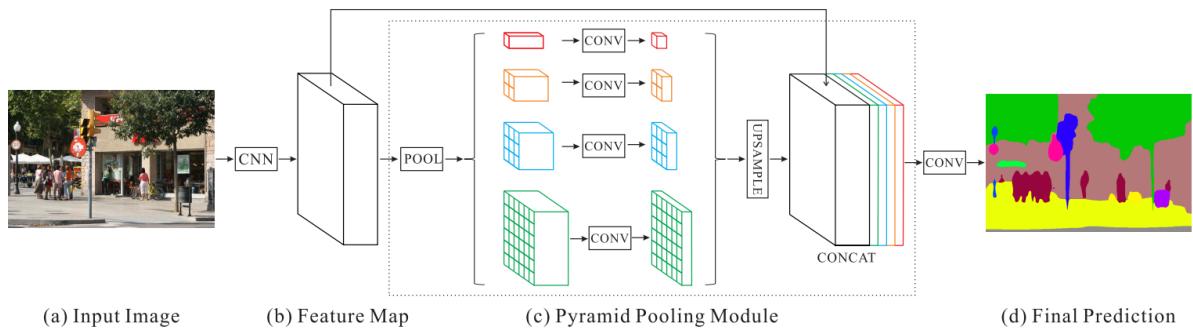


Рис. 1.12: Граф нейронної мережі PSPNet [?]

Модель нейронної мережі 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 різноманітних 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.

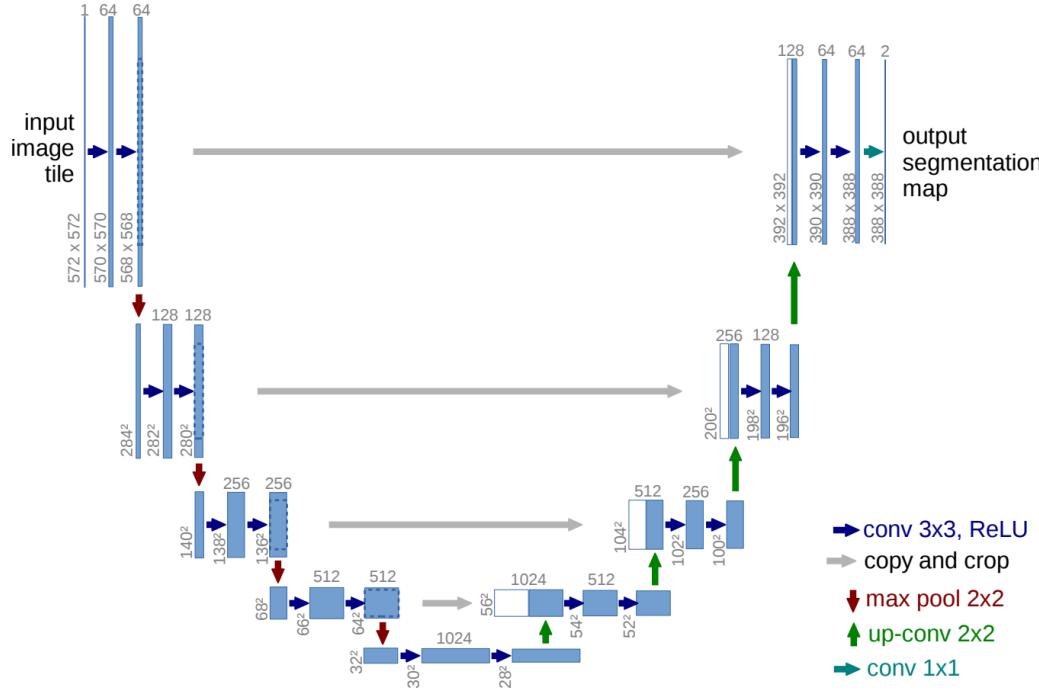


Рис. 1.13: Граф нейронної мережі U-Net [?]

Модель нейронної мережі 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 перенесі 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 близчими до LinkNet. [?, ?, ?]

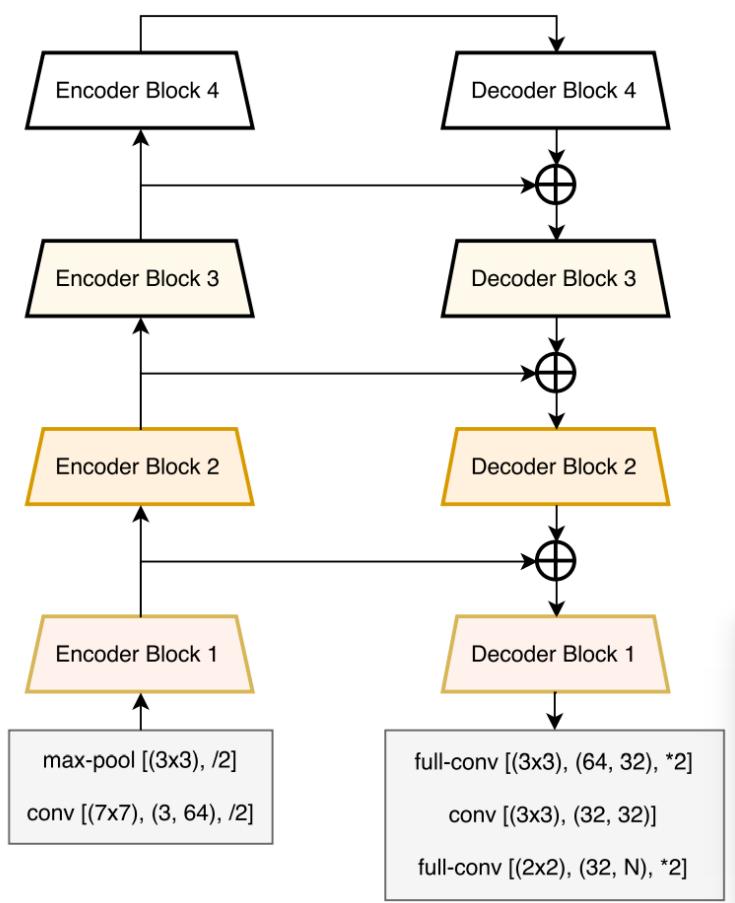


Рис. 1.14: Граф нейронної мережі LinkNet [?]

Модель нейронної мережі 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 сегментації за допомогою операції згортки.

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

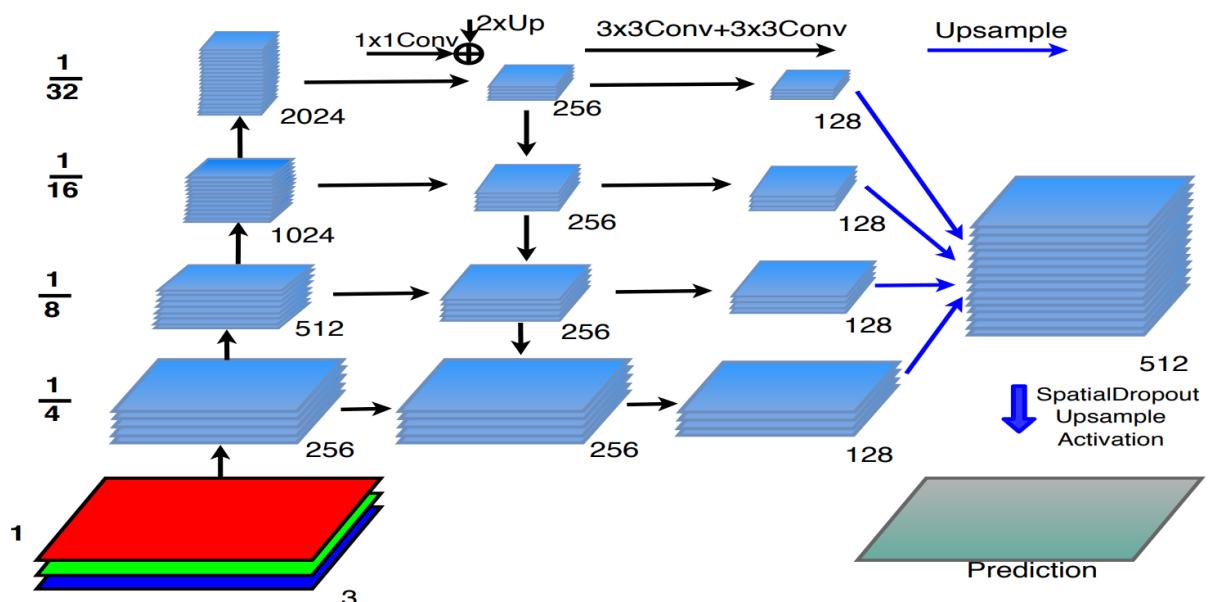


Рис. 1.15: Граф нейронної мережі FPN [?]

Сімейство моделей 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.

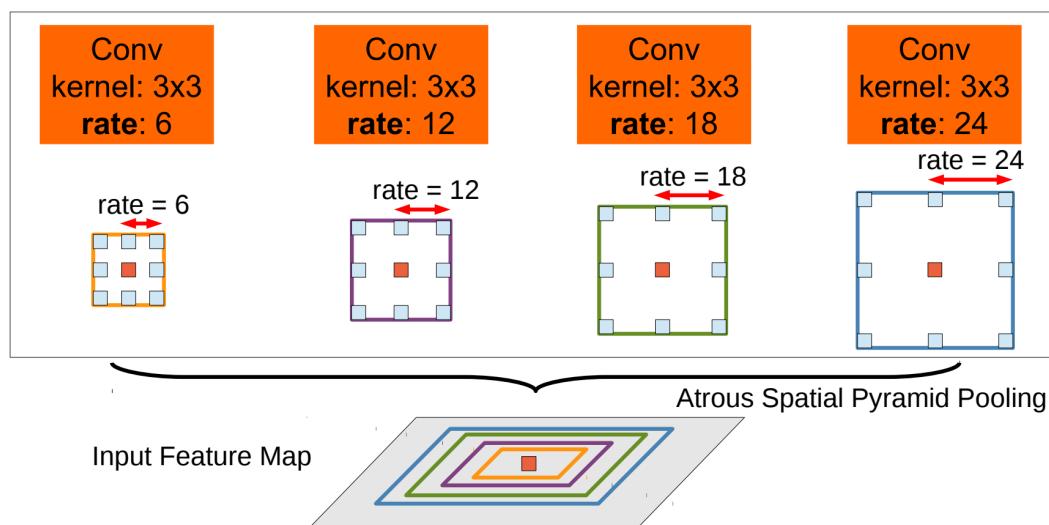


Рис. 1.16: Структура блоку ASPP в нейронних мережах DeepLab [?]

1.4. Методи багатозадачного глибинного машинного навчання

Багатозадачне машинне навчання - це один з підходів машинного навчання, в якому одночасно вирішуються кілька навчальних задач, використовуючи спільноти 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 привести 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 контролюваного 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 п навчальних задач T_1, T_2, \dots, 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 п або деяких з них задачах.

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 зображені та розмірів (задача регресії). Для задач контролюваного навчання, кожна задача T_i містить в собі набор даних D_i , що складається з t 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 [1] 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 зображені (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 зображене на рис. 1.17.

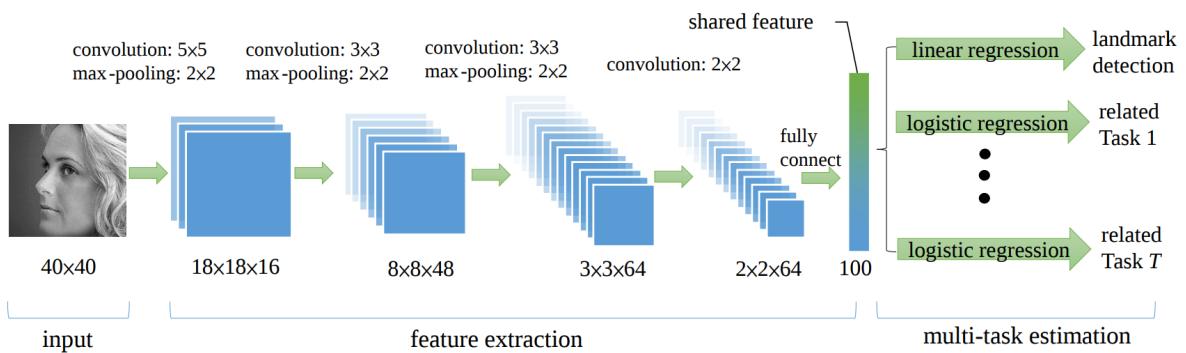


Рис. 1.17: Архітектура нейронної мережі TCDCN [?]

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 зображене на рис. 1.18. Схожий підхід також використовується в деяких архітектурах нейронних мереж для розв'язання задачі детекції об'єктів [?, ?, ?]

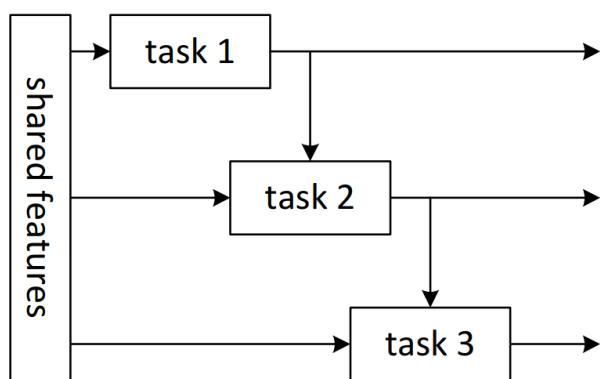


Рис. 1.18: Архітектура багатозадачної каскадної мережі [?]

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 склонні до утворення конфліктних градієнтів є методи архітектури з м'яким розподілом параметрів.

В методах, що використовують м'який розподіл параметрів використовується декілька енкодерів дляожної 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. Коєфіцієнти лінійної комбінації також вивчаються за допомогою градієнтного спуску разом з іншими параметрами нейронної мережі. На рисунку 1.19 зображені 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 дляожної з задач. На відміну від перехресних ме-

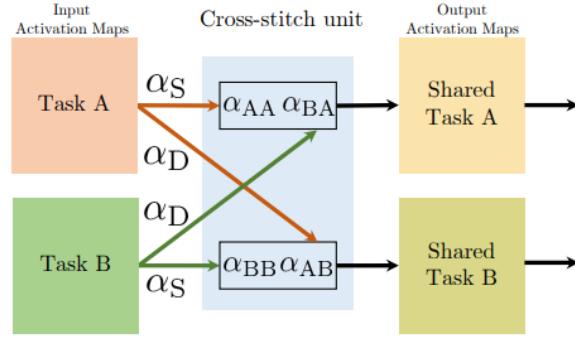


Рис. 1.19: Архітектура cross-stitch блоку [?]

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 шлюзовых 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 зображену на рисунку 1.20.

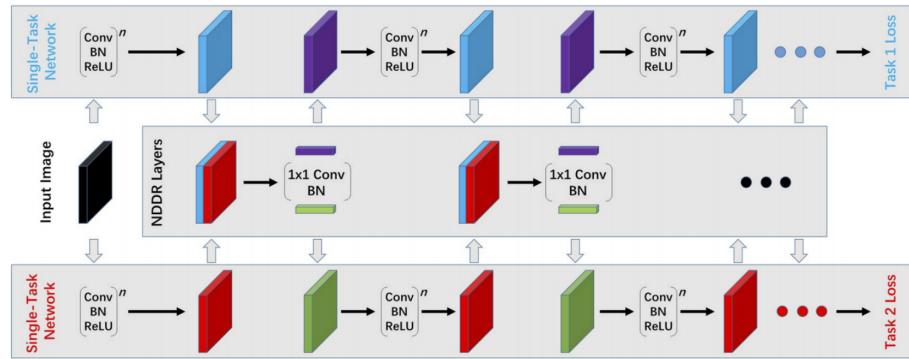


Рис. 1.20: Архітектура згорткової нейронної мережі NDDR [?]

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:

- Проста конкатенація результатів;
- Передача повідомлень між задачами
- Передача повідомлень з вагами

Подальшим розвитком цього методу стала робота PAP-Net, в якій було представлено модуль, що окрім 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 скрінінгу 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 архітектуру 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.

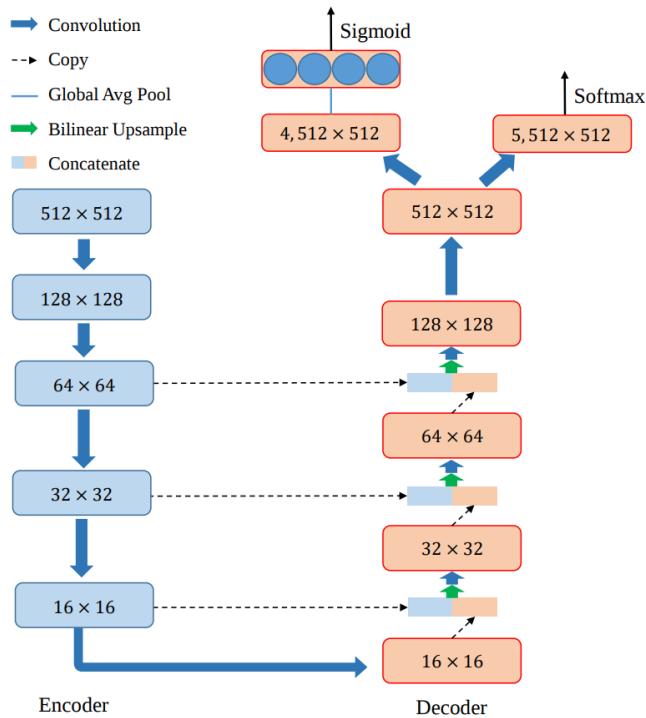


Рис. 1.21: Архітектура методу SegTHOR [?]

В нейронній мережі SegTHOR вихід класифікації використовується лише 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 зображену на рисунку ??.

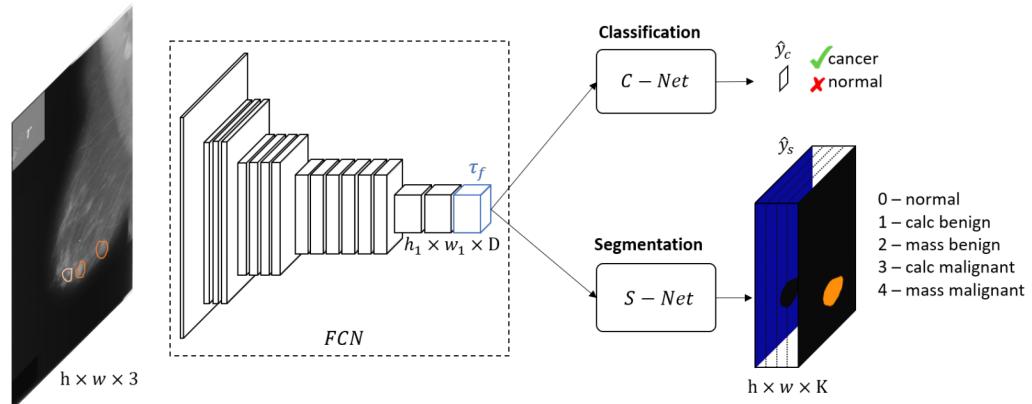


Рис. 1.22: Архітектура нейронної мережі класифікації та сегментації новоутворень на мамографії [?]

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 архі-

тектурою 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 зображенено на рисунку 1.23. Автори показують, що таке використання декількох задач дозволяє підвищити достовірність сегментації в цілому та якість контурів окремо.

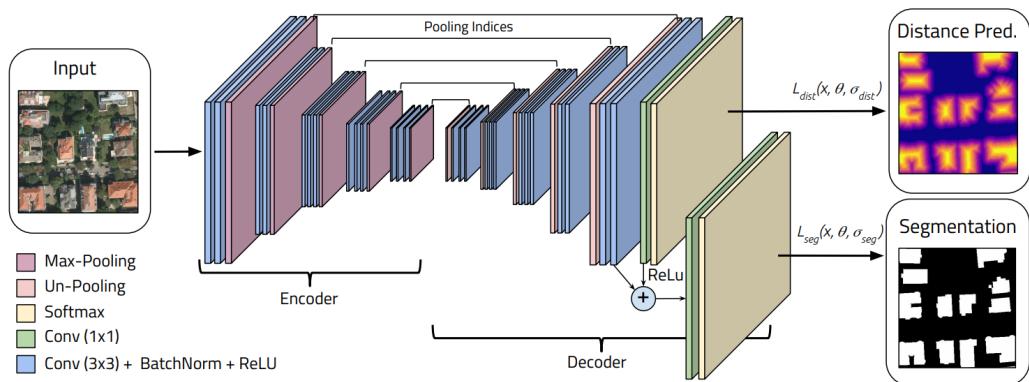


Рис. 1.23: Архітектура нейронної мережі для сегментації будівель [?]

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 ЕМ-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 ЕМ 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.

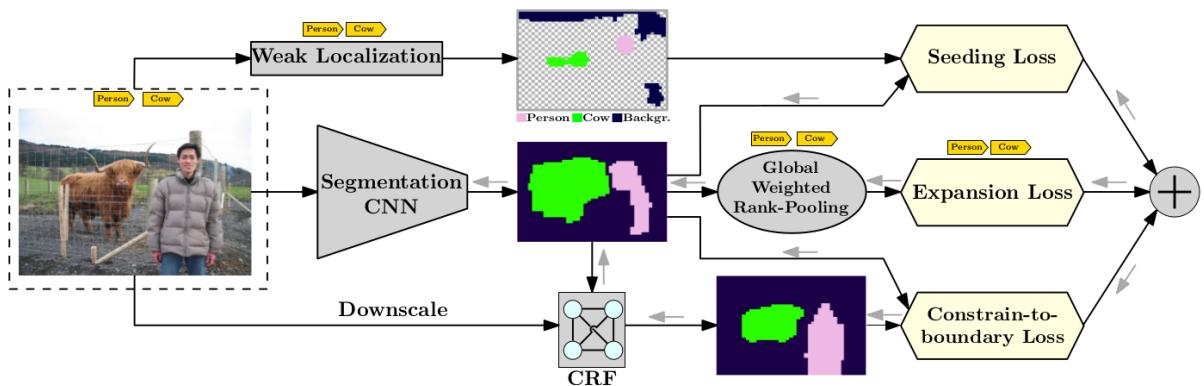


Рис. 1.24: Архітектура методу SEC-CNN [?]

1.5. Методи 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 контролюваного навчання, зокрема бінарної класифікації, за нульову гіпотезу зазвичай приймається принадлежність елемента вибірки 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.

$$Precision = \frac{TP}{TP + FP} \quad (1.10)$$

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

$$Recall = \frac{TP}{TP + FN} \quad (1.11)$$

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

$$Specificity = \frac{TN}{TN + FP} \quad (1.12)$$

Між влучністю і повнотою, та чутливістю і специфічністю існує обернена залежність, коли можливо підвищити одну ціною зниження іншої. Чезрьом наявність такої залежності, 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 Дайса-Соренсена. F1 та індекс подібності Дайса-Соренсена (англ. Dice score) є еквівалентними ме-

триками, що, зазвичай використовуються в задачах класифікації (F1) та сегментації (Dice score).

Міра F1 визначається як середнє гармонічне між влучністю та повнотою:

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

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

$$D_c = 2 \cdot \frac{A \cap B}{|A| + |B|} \quad (1.14)$$

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:

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

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:

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

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** наведено огляд літератури за тематикою даної роботи та спорідненими питаннями; висвітлено результати, які були отримані іншими дослідниками. Зокрема, наведено огляд досліджень, що стосуються задачі автоматизованого скринінгу в різних областях; проведено аналіз моделей та методів комп'ютероного зору, насамперед, моделей глибоких нейронних мереж та методів їх навчання. Проведено аналіз методів розпізнавання, класифікації та семантичної сегментації планарних зображень, використання для цього моделей та методів багатозадачного навчання. Додатково проаналізовано метрики, що використовуються для оцінки достовірності в задачах класифікації та сегментації зображень. Виявлено, що достовірність прогнозів може бути визначена як міра Дайса-Соренсена, або F1-міра. Наведено підходи використання багатозадачного навчання для класифікації та сегментації зображень в задачах автоматизованого скринінгу.

Проведено аналіз зашумлення розмітки в наборах даних для різних 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;
5. Зображенням присвоєні невірні класи.

Наведено огляд наявних моделей наборів даних, та методів їх генерації.

Виявлено, що відсутні моделі наборів даних, які б дозволяли змінювати рівень зашумлення розмітки відповідно до характеристик зашумлення в реальних наборах даних.

Виявлено, що 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*.

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*.

Таким чином, проведений огляд сучасної літератури на тему дисертації дозволяє аргументувати актуальність та практичну цінність проведених у роботі досліджень.

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

Метою дослідження є підвищення достовірності класифікації та се-

мантичної сегментації планарних зображень в системах автоматизованого скринінгу через розробку моделі та методів аналізу планарних зображень на основі штучних нейронних мереж.

Для досягнення мети дослідження поставлено і розв'язано такі **задачі**:

- Здійснено аналіз зашумлення розмітки в наборах даних для задач автоматизованого скринінгу, проведено аналіз існуючих моделей, методів, алгоритмів та засобів аналізу планарних зображень в системах автоматизованого скринінгу і обґрунтовано вибір напрямку досліджень.
- На основі проведеного аналізу розроблено параметричну модель набору даних, що складається з планарних зображень із зашумленою розміткою для задач класифікації та семантичної сегментації, а також метод його генерації.
- Розроблено модель нейронної мережі та методи навчання й прогнозування для аналізу планарних зображень, що за рахунок використання декількох семантично-близьких задачах одночасно дозволяють підвищити точність класифікації та семантичної сегментації в умовах зашумлення розмітки даних.
- Розроблено інструментальні засоби, що виконують розроблені моделі та методи, і провести їх випробування в рамках експерименту

Об'єктом дослідження є процес аналізу планарних зображень.

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

РОЗДІЛ 2

ПАРАМЕТРИЧНА МОДЕЛЬ НАБОРІВ ДАНИХ ІЗ ЗАШУМЛЕНОЮ РОЗМІТКОЮ І МЕТОД ЇХ ГЕНЕРАЦІЇ

Для систематизованого вивчення впливу рівня шуму в даних, та оцінки відповідного впливу запропонованих рішень, розроблено модель набору даних із зашумленням розмітки семантичної сегментації, що відповідає оцінкам моделей шуму в різних задачах автоматизованого скринінгу.

На основі проведеного аналізу виявлено, що основними проблемами з розміткою масок сегментації планарних зображень 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.

Для відображення означених недоліків розмітки, запропоновано штучне введення помилок за допомогою випадкового застосування морфологічних операцій ерозії та дилатації з квадратним ядром до масок сегментації окремих об'єктів перед додаванням їх до загальної маски.

Операції ерозії та дилатації відповідають неупередженному зашумленню, оскільки за їхнім результатом не можна відновити вихідну конфігурацію маски [?]. Також, означені операції реалізують зашумлення, що залежать від ознак, оскільки нерівномірно модифікують тонкі та більш товсті лінії та порожнини.

На відміну від геометричних перетворень маски, включення, або невключення пікселів з граничних регіонів маски є частішим випадком зашумлення розмітки [?].

2.1. Модель набору даних із зашумленою розміткою

Запропонована модель має наступні параметри:

- Набір зображень фону \mathcal{X}_b
- Набір зображень об'єктів \mathcal{X}_f
- Набір зображень текстур об'єктів \mathcal{X}_{tex}
- 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}

На відміну від інших моделей наборів даних, запропонована модель має параметри, що відповідають безпосередньо за зашумлення масок сегмен-тації:

- 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

2.2. Метод генерації наборів даних

Запропоновано метод контрольованої генерації як зображень, так і масок сегмен-тації що модифікуються відповідно до випадково обраних недоліків.

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

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 натуруальні зображення, так і синтетичні текстири, чи заливка константним кольором. В якості об'єктів можуть бути використані зображення з наборів даних MNIST [?], або Fashi-

onMNIST [?].

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

Запропонований метод генерації описується за допомогою алгоритму.

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 зміні ініціалізації генератора випадкових чисел.

2.3. 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.

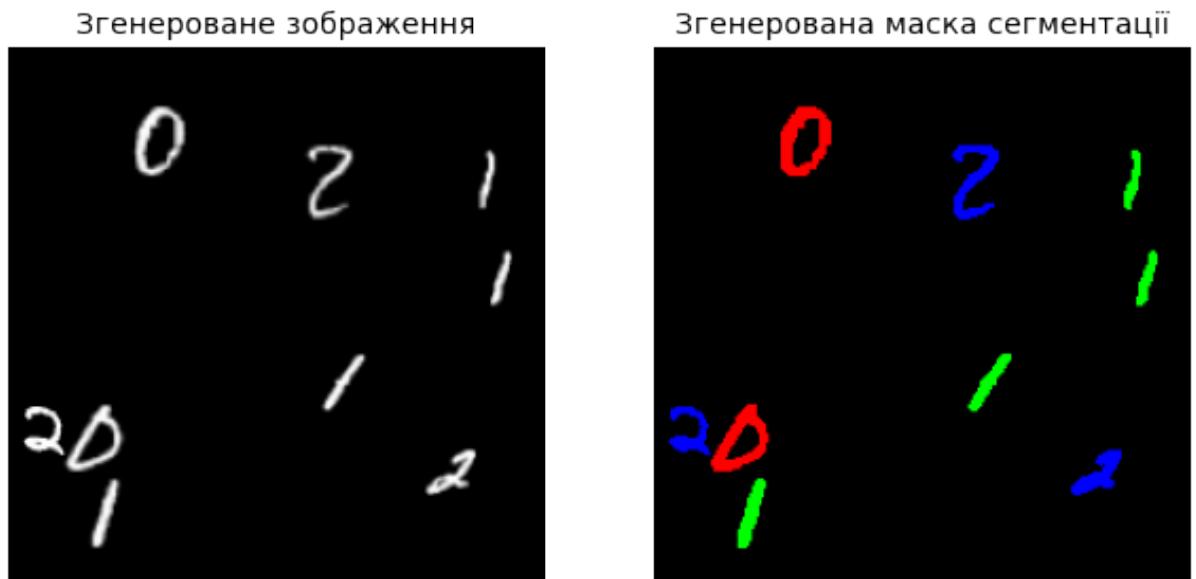


Рис. 2.1: Об'єкти MNIST, фон: константний, колір об'єктів: білий

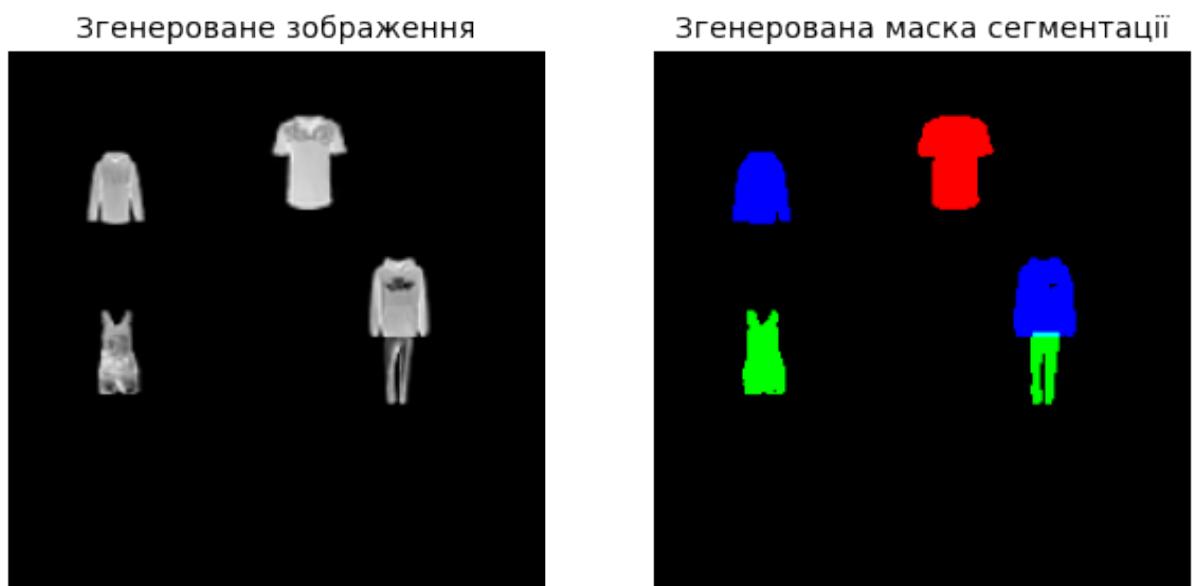


Рис. 2.2: Об'єкти FashionMNIST, фон: константний, колір об'єктів: білий

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

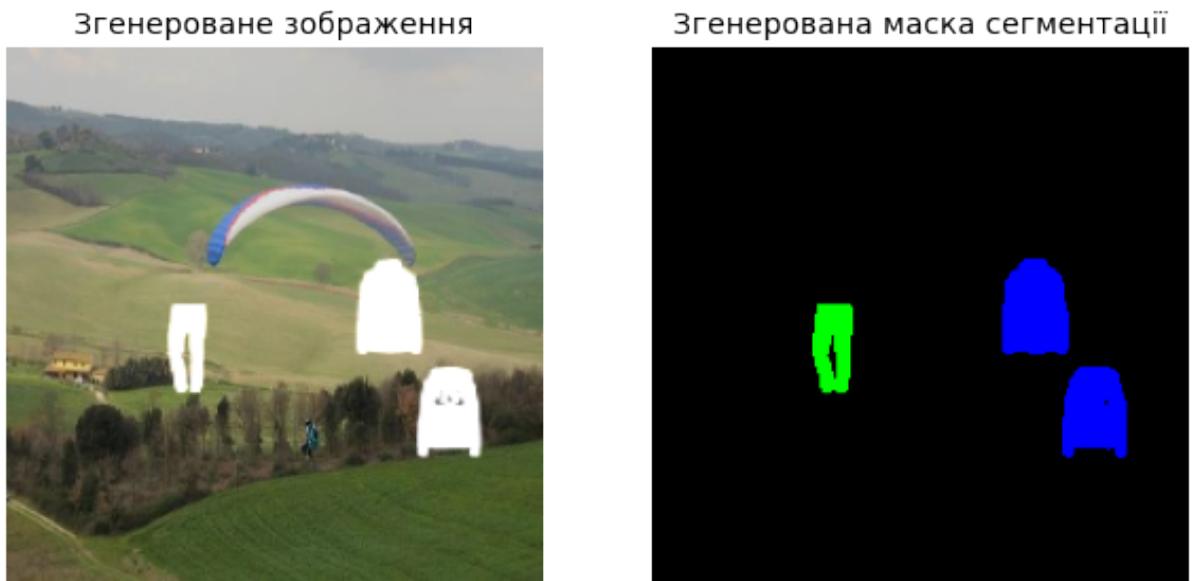


Рис. 2.4: Об'єкти FashionMNIST, фон: Imagenette, колір об'єктів: білий

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

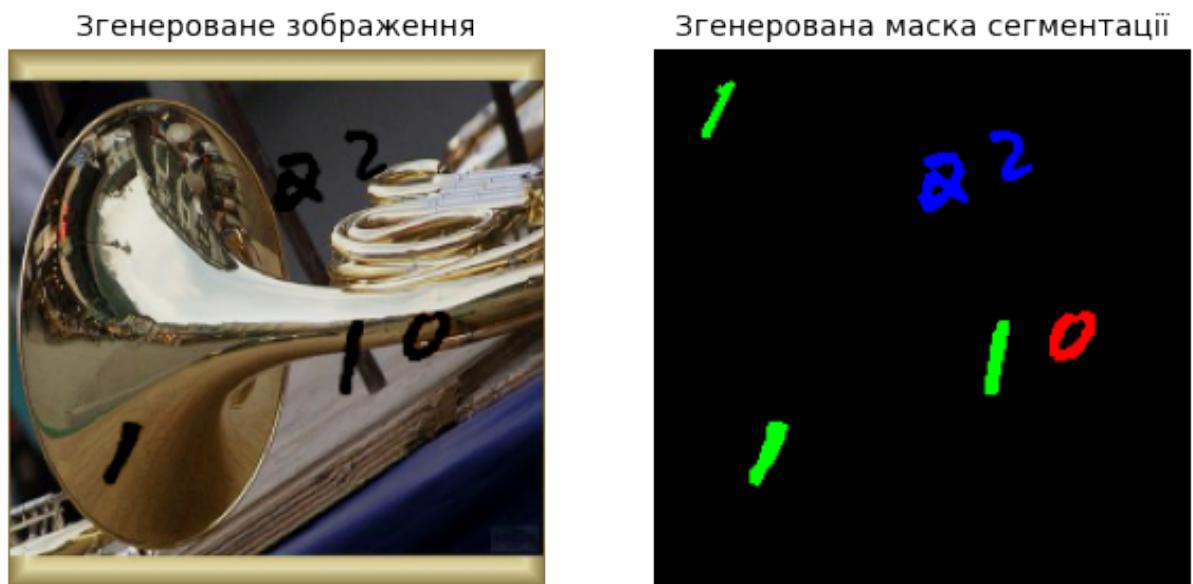


Рис. 2.5: Об'єкти MNIST, фон: Imagenette, колір об'єктів: чорний



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 еrozією ядром 3×3 та їх комбінацію для наочності.



Рис. 2.11: Об'єкти MNIST, еrozія 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 зображене вихідне зображення, маску сегментації з ерозією ядром 7×7 та їх комбінацію для наочності.

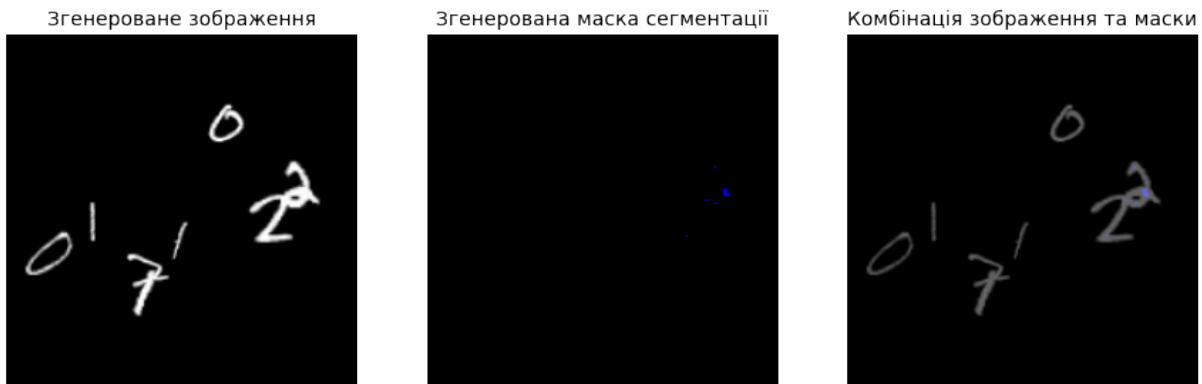


Рис. 2.12: Об'єкти MNIST, ерозія маски ядром 7×7

2.4. Висновки до другого розділу

У другому розділі представлено параметричну модель наборів даних із зашумленою розміткою для задач сегментації та класифікації, що відповідає оцінкам моделей шуму в різних задачах автоматизованого скринінгу, а також метод контролюваної генерації як зображень, так і масок сегментації та міток класів що модифікуються відповідно до випадково обраних недоліків.

Запропоновано використання модельних наборів даних тестування довірності моделей та методів навчання глибоких нейронних мереж в задачах класифікації та семантичної сегментації. Тестування виконується в контролюваних умовах за допомогою введення зашумлення в розмітку лише тренувального набору даних, в той час як тестувальних набір даних залишається з точною розміткою.

РОЗДІЛ 3

НЕЙРОМЕРЕЖЕВА МОДЕЛЬ АНАЛІЗУ ПЛАНАРНИХ ЗОБРАЖЕНЬ ТА МЕТОДИ ЇЇ НАВЧАННЯ

Запропоновано модель багатозадачної нейронної мережі, що дозволяє вирішувати задачі семантичної сегментації та класифікації для аналізу планарних зображень, а також метод її навчання, що заснований на методах багатозадачного навчання.

На відміну від однозадачного машинного навчання, багатозадачне машинне навчання дозволяє використати спільноті та відмінності між різними задачами для підвищення узагальнювальної здатності на кожній з окремих задач. Використання комбінації декількох задач машинного навчання може сприяти вивченю кращих наборів ознак, та підвищити точність роботи нейронних мереж. Також, сумісне використання прогнозів декількох задач для уточнення кожного з окремих прогнозів, так і побудови нових.

Основним припущенням, на яке спирається розроблений метод є можливість створення більш точної розмітки тренувального набору даних для задачі класифікації на основі розмітки семантичної сегментації для того ж набору вхідних зображень. В даному контексті, задача семантичної сегментації називається основною задачею, а класифікації - додатковою.

3.1. Модель багатозадачної штучної 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.

3.1.1. Загальна структура нейронної мережі. Нейронна мережа виконана з використанням архітектури енкодер-декодер, заснована на архітектурі нейронних мереж LinkNet [?]. На відміну від оригінальної структури LinkNet, запропонована модель нейронної мережі складається з одного 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. Карті 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 3.1

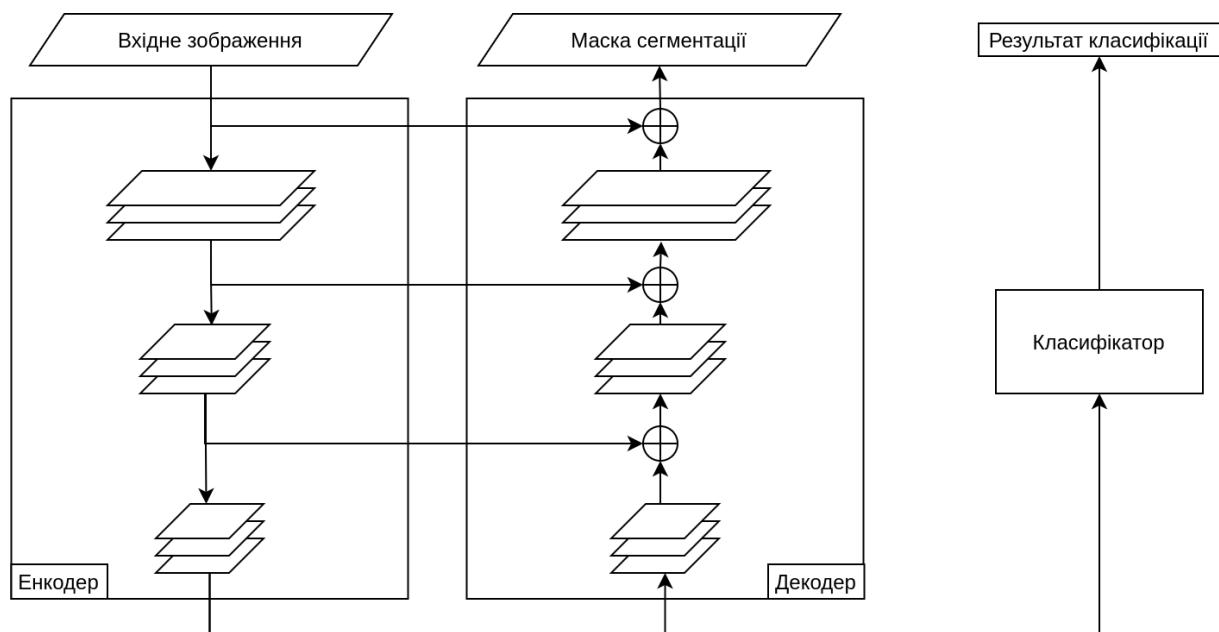


Рис. 3.1: Загальна архітектура нейронної мережі з декодером та класифікатором

Таким чином, виконується жорсткий розподіл параметрів, оскільки параметри енкодера відповідають одночасно за обидві задачі. Така структура дозволяє використати методи трансферного навчання (індуктивного пере-

носу) для підвищення роздільноті внутрішніх представлень і пришвидшення процесу навчання: набір параметрів енкодера θ_{enc} ініціалізується з використанням параметрів, отриманих після навчання енкодера на наборі даних Imagenet [?]. Декодери ініціалізується з використанням методу Хе: $\theta \in \mathcal{U}(-b, b)$, де b - константа залежна від типу шару [?]

Окрім графічного представлення, запропонована модель може бути представлена алгебраїчно: нехай нейронну мережу енкодера визначено як $F_{encoder}$. Тоді для вхідного зображення $x \in \mathcal{R}^{3 \times H \times W}$, $v_1, v_2, \dots, v_i, \dots, v_n$ - 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 дляожної з карт ознак, а n - кількість стадій енкодера (залежить від архітектури):

$$v_1, v_2, \dots, v_n = F_{encoder}(x, \theta_{enc}) \quad (3.1)$$

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:

$$M_{seg} = F_{seg}((v_1, v_2, \dots, v_n), \theta_{seg}) \quad (3.2)$$

$$C_{cls} = F_{cls}((v_n), \theta_{cls}) \quad (3.3)$$

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

Окрім загальної архітектури моделі, критичними для коректної роботи запропонованих методів є структури декодерів дляожної з задач, зокрема, нормалізація їх внутрішніх представлень.

3.1.2. Структура декодера семантичної сегментації. Декодер складається з декількох стадій, кожна з яких має розміри карт ознак відповідно до карт ознак енкодера.

Структура декодера сегментації повторює декодер архітектури UNet [?]. На відміну від архітектури UNet, в даному декодері застосовується пакетна нормалізація після операцій конкатенації з шарами енкодера та перед згортковими шарами. Ця нормалізація необхідна для забезпечення незалежності масштабів значень карт ознак на різних шарах декодера сегментації та енкодера, які також спільно використовуються декодером класифікації.

Також, замість зворотних згорток, в декодері використовується білінійна інтерполяція для підвищення просторового розширення карт ознак, що дозволяє уникнути "шахового патерну" в прогнозованих масках [?]/

Оскільки в задачі сегментації з частково-помилковою розміткою одному пікселю може бути призначено декілька класів одночасно, необхідно це враховувати при використанні функцій втрат. Для формування фінальної карти сегментації використовується спеціальна версія блоку декодера, що має сигмоподібну функцію активації. Це дозволяє нейронній мережі прогнозувати одночасно декілька класів для одного пікселя, замість спроб виділення одного конкретного класу при використанні нормованої експоненційної функції (Softmax).

Структуру окремих стадій показано на рисунку 3.2.

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

3.1.3. Структура декодера класифікації. 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 максимуму (GlobalMaxPooling) для кожного з каналів карти ознак енкодера.

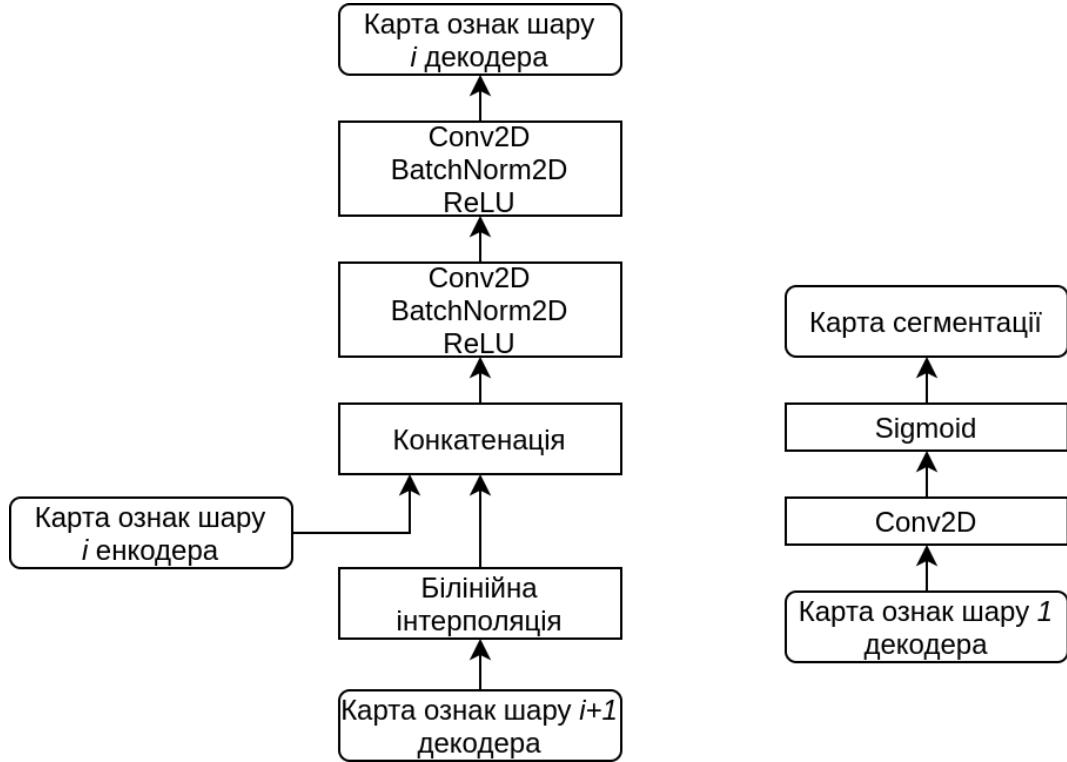


Рис. 3.2: Структура стадій декодеру сегментації

Так само, як і в декодері сегментації, для забезпечення незалежності масштабу останнього шару енкодера, першим шаром є шар пакетної нормалізації.

Для підвищення роздільності ознак останнього шару енкодера, класифікатор являє собою лінійну функцію, що спонукає енкодер до формування лінійно-роздільних представлень v_n , які також використовуються декодером сегментації. Ця структура знижує кількість нелінійних перетворень, потрібних декодеру.

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.

Структура декодера класифікації зображена на рисунку 3.3.

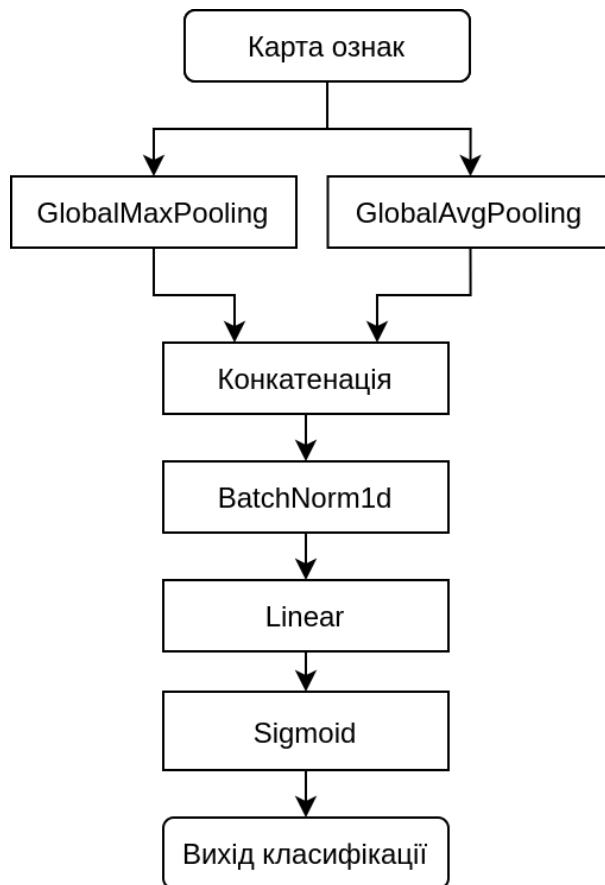


Рис. 3.3: Структура декодера класифікації

3.2. Методи навчання та прогнозування багатозадачних ШНМ в умовах частково помилкової розмітки семантичної сегментації

На основі запропонованої моделі багатозадачної нейронної мережі розроблено метод багатозадачного навчання, що в умовах частково-помилкової розмітки дозволяє зменшити вплив неправильно розмічених прикладів на процес навчання за рахунок їх ефективної фільтрації безпосередньо під час навчання. Так, часто на одному зображені 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.

3.2.1. Автоматична фільтрація помилкової розмітки. Маючи велику кількість параметрів, сучасні нейронні мережі здатні до запам'ятовування 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 по-гіршити якість навчання через зменшення ємності нейронної мережі, або, при надмірному використанні, можуть спонукати нейронну мережу акцен-тувати увагу на текстурах, чо часто є недопустимим в задачах семантичної сегментації.

Для того, щоб ефективно зменшити вплив помилкової розмітки, необ-хідно 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, що має шкідливий ефект при наявності помилкової розмітки в тре-нуальному наборі даних - правильні прогнози на неправильно-розмічених даних створюють конфліктні градієнти та знижують впевненість прогнозів нейронної мережі.

Для виключення неправильної розмітки в процесі навчання запропоно-вано обмеження функції втрат, з відповідним визначенням градієнтів для нової функції:

$$\mathcal{L}^{\top} = \min(L, \theta) \quad (3.4)$$

де θ - поріг обмеження функції втрат.

Графіки обмежених зверху функцій втрат зображені на рисунку 3.4.

3.4

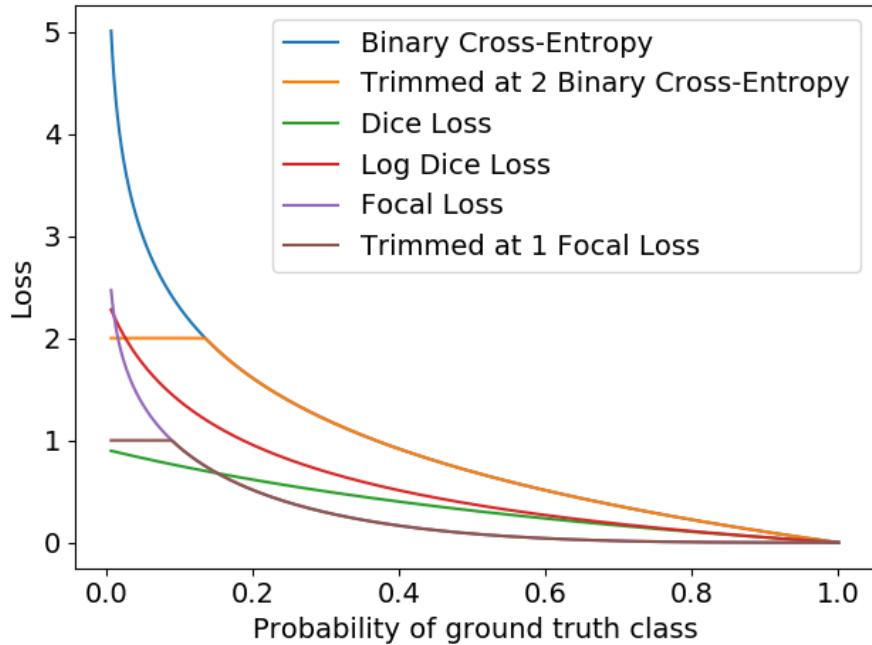


Рис. 3.4: Приклади обмежених зверху функцій втрат

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

$$\nabla \min(L, \theta) = \begin{cases} 1 & L \in (-\infty, \theta] \\ 0 & L \in (\theta, \infty) \end{cases} \quad (3.5)$$

Всі загальновживані методи оптимізації параметрів нейронних мереж засновані на методі оновлення параметрів першого порядку, що, в загальному випадку, є модифікацією стохастичного градієнтного спуску:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} L \quad (3.6)$$

Де θ - параметри нейронної мережі, η - темп навчання, а ∇_{θ} - градієнт функції втрат відносно параметрів, таким чином, приклади з занадто вели-

кою похибкою (переважно з помилковою розміткою) ефективно виключаються з навчання, оскільки мають нульовий градієнт відносно параметрів.

Недоліком цього методу фільтрації помилкової розмітки є можливість виключення прикладів, що занадто складні для вивчення нейронною мережею на ранніх етапах навчання. Це може сповільнити процес навчання, або взагалі зупинити його. Щоб запобігти сповільненню процесу, запропоновано використання багатозадачного навчання з використанням додаткової більш загальної задачі.

3.2.2. Генерація похідної задачі до задачі семантичної сегментації. Для того, щоб забезпечити успішне навчання нейронних мереж на ранніх етапах, та врахувати занадто складні приклади, необхідно ввести інформацію про ці об'єкти в обхід відфільтрованої розмітки.

На основі розмітки для задачі семантичної сегментації може бути побудовано розмітку для інших більш загальних задач, таких як задача детекції (наприклад, обмежувальні прямокутники навколо об'єктів), регресії (наприклад, координати центрів об'єктів), або класифікації (класи об'єктів). Такі задачі є більш зашальними, оскільки мають менше незалежних змінних в виходах нейронної мережі. Наприклад, замість класифікації кожного з пікселів зображення в задачі семантичної сегментації, в задачі детекції потрібно лише прогнозування класів та координат обмежувальних прямокутників.

Також, з точки зору багатозадачного машинного навчання, ці задачі є семантично-близькими: такі задачі, що мають одинакові 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 неточності 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:

$$y_c = t < \sum_i^H \sum_j^W y_{s_{ij}} \quad (3.7)$$

де t - поріг мінімального розміру об'єкта в пікселях

Згенерована таким чином задача класифікації має меншу ймовірність хибної розмітки.

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 при-

кладах, для яких немає градієнтів через обмеження функції втрат.

3.2.3. Агрегація функцій втрат під час навчання. Оскільки модель багатозадачної нейронної мережі має два виходи дляожної з задач, для одночасного оновлення всіх параметрів необхідне одночасне використання двох функцій втрат. Дляожної з задач окремо обчислюється функція втрат. Для навчання декодера сегментації використовується обмежена зверху, в той час як для декодера класифікації - звичайна. Кожна з функцій втрат нормалізується відносно кількості прикладів в одному пакеті навчання для зниження впливу цього гіперпараметру.

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

$$L_{total} = \frac{L_{seg} + L_{cls}}{2} \quad (3.8)$$

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

$$\nabla L_{total} = \frac{\nabla L_{seg} + \nabla L_{cls}}{2} \quad (3.9)$$

Під час навчання нейронної мережі використовується стандартний алгоритм зворотного поширення помилки з оптимізатором Adam [?], що відрізняється від методу стохастичного градієнтного спуску наявністю аддитивної нормалізації моментів градієнтів.

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.

3.2.4. Багатозадачне прогнозування результатів семантичної сегментації. Оскільки запропонована модель нейронної мережі може одночасно вивчати як задачу сегментації, так і задачу класифікації, стає можливою імплементація фільтрування хибно-позитивних ознак без підвищення витрат часу на етапі прогнозування результатів.

Для фільтрації хибно-позитивних результатів запропоновано метод комбінації задач класифікації та сегментації, що дозволяє використовувати задачу класифікації, що попередньо була навчена на менш зашумленій розмітці як фільтр для задачі семантичної сегментації. В такому випадку можливі два варіанти комбінації 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 $(-\infty, +\infty)$ (логіти).

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

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.10)$$

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

$$M_{refined} = \begin{cases} \sigma(M_{seg}) & \text{якщо } \sigma(C_{cls}) > t_{cls} \\ 0^{C \times H \times W} & \text{якщо } \sigma(C_{cls}) \leq t_{cls} \end{cases} \quad (3.11)$$

Такий варіант має декілька суттєвих недоліків:

- Необхідність вибору додаткового параметра θ_{cls} вносить додаткову можливість помилки; неправильний підбір цього параметра може привести до значного збільшення хибно-негативних результатів.
- Можливість розповсюдження помилки класифікатора виникає через значно меншу кількість параметрів порівняно з декодером се-

гментації.

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 $[0, 1]$:

$$\hat{M}_{seg} = \sigma(M_{seg}) \quad (3.12)$$

$$\hat{C}_{cls} = \sigma(C_{cls}) \quad (3.13)$$

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{C}_{cls}

$$M_{refined} = \hat{M}_{seg} \circ \hat{C}_{cls} \quad (3.14)$$

Графічну репрезентацію структури зображено на рисунку 3.5.

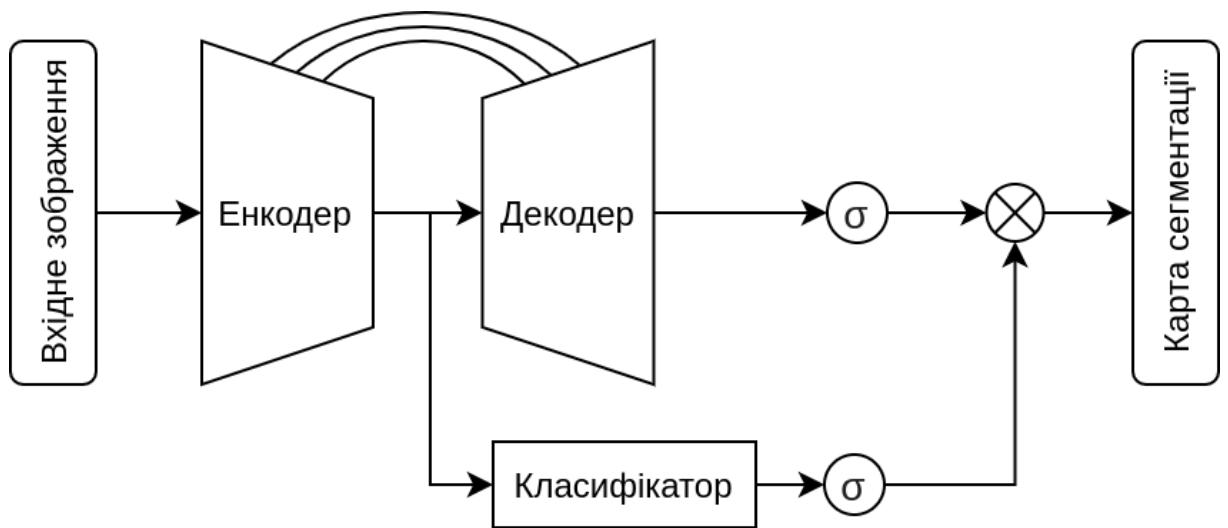


Рис. 3.5: Структура об'єднання прогнозів класифікації та сегментації

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 співпадіння значень класифікації та сегментації, фінальний прогноз буде відрізнятися від послідовного підходу.

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

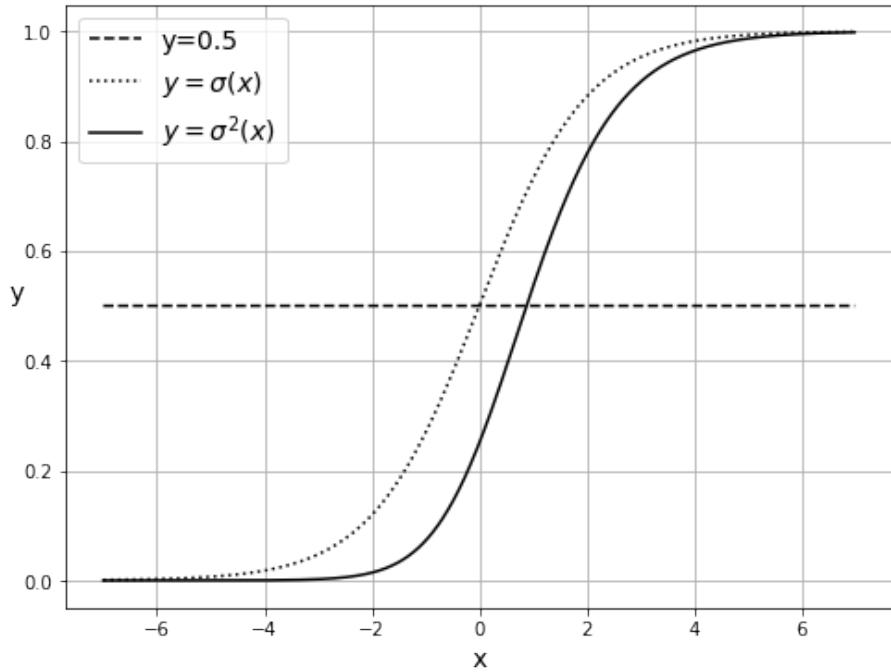


Рис. 3.6: Структура декодера класифікації

При використанні того самого порогу бінаризації для карт сегментації, зменшиться кількість позитивних пікселів в результаті, тому при прямому переході між однозадачним методом прогнозування, потрібно використовувати квадрат t_{cls} :

$$\hat{t}_{cls} = t_{cls}^2 \quad (3.15)$$

Таким чином, кількість позитивних пікселів не заміниться за рахунок зміни в нормалізації сигмоїд.

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

На основі двох представлених методів, розроблено метод навчання з частковим залученням вчителя та метод пост-обробки, що дозволяє використовувати локалізацію важливих для класифікації ознак на вихідному зображенні, що є корисним під час інтерпретації прогнозів моделей. За рахунок використання методів багатозадачного навчання досягається можливість застосування методу в умовах відсутності розмітки семантичної сегментації в навчальному наборі даних.

Даний метод застосовується для навчання тієї самої моделі, але за умови наявності лише розмітки для задачі класифікації. Так само, як і в розділі 3.1 використані наступні 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,

$v_1, v_2, \dots, v_i, \dots, v_n$ - набір карт ознак енкодера, відповідно до рівняння 3.1,

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

M_{seg} та C_{cls} - результати сегментації та класифікації відповідно до рівнянь 3.2 та 3.3

\hat{M}_{seg} та \hat{C}_{cls} - нормовані результати сегментації та класифікації відповідно до рівнянь 3.12 та 3.13

В основі запропонованого методу лежить ітеративне уточнення карти ознак сегментації за допомогою направлених градієнтів від задачі класифікації. В даному випадку, під час навчання декодер локалізації використо-

вується в ролі механізму уваги [?], що навчається автоматично за рахунок градієнтів до задачі класифікації. Хоча, при використанні даного методу, якість семантичної сегментації є низькою в районі контурів об'єктів, отриманих результатів достатньо для локалізації цих об'єктів на зображеннях.

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

3.3.1. Напівавтоматичне навчання ШНМ в задачі локалізації.

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

$$M_{unsup} = \hat{M}_{seg} \circ C_{cls} \quad (3.16)$$

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

$$C_{unsup} = \frac{\sum_{h=0}^H \sum_{w=0}^W M_{unsup(h,w)}}{\sum_{h=0}^H \sum_{w=0}^W M_{seg(h,w)} + c} \quad (3.17)$$

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 розглядати \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 $(\hat{M}_{seg} \rightarrow 1)$ для важливих ознак, що є спільними на зображеннях з навчального набору даних.

Для запобігання вивченю нейронною мережею карт сегментації, що складаються виключно з високих значень, необхідна така ініціалізація па-

раметру зсуву останнього шару декодера сегментації F_{seg} , щоб активації, за замовчуванням, були близькими до нуля. Для сигмоїдної функції активації таке значення дорівнює -4.6

3.3.2. Прогнозування результатів сегментації. На етапі прогнозування використовується як декодер класифікації, так і декодер, що відповідає за задачу локалізації. 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} квантується 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:

$$\sigma_w^2(T_{seg}) = \min : \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad (3.18)$$

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:

$$M_e = M_t \ominus k \quad (3.19)$$

де k - ядро еrozії.

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

3.4. Висновки до третього розділу

1. Запропоновано модель багатозадачної нейронної мережі, що дозволяє вирішувати задачі семантичної сегментації та класифікації для аналізу планарних зображень.
2. Виявлено, що окрім загальної архітектури моделі, критичними для коректної роботи запропонованих методів є структури декодерів для кожної з задач, зокрема, нормалізація їх внутрішніх представлень.
3. На основі запропонованої моделі багатозадачної нейронної мережі розроблено метод багатозадачного навчання, що в умовах частково-помилкової розмітки дозволяє зменшити вплив неправильно розмічених прикладів на процес навчання за рахунок їх ефективної фільтрації безпосередньо під час навчання.
4. Запропоновано метод автоматичної фільтрації помилкової розмітки, що спирається на відсіювання градієнтів в точках, що відповідають занадто високим значенням функції втрат.
5. Запропоновано використання задачі класифікації як більш загальної до задачі семантичної сегментації при багатозадачному навчанні для оновлення параметрів енкодера в точках, що були помилково відфільтровані як зашумлені.
6. Удосконалено метод об'єднання задач сегментації та класифікації на етапі прогнозування для зменшення кількості хибно-позитивних результатів.

7. На основі двох представлених методів, розроблено метод навчання з частковим залученням вчителя та метод пост-обробки, що дозволяє виконувати локалізацію важливих для класифікації ознак на вихідному зображенні, що є корисним під час інтерпретації прогнозів моделей.

РОЗДІЛ 4

ІНСТРУМЕНТАЛЬНІ ЗАСОБИ ТА ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

В даному розділі розглянуто експерименти із розробленими моделями нейронних мереж та методами їх навчання (розділ 3.1) на синтетичних даних, що були згенеровані за допомогою запропонованої в розділі 2 моделі, а також експерименти, поставлені на наборах даних з реальних задач скринінгу.

4.1. Експерименти на синтетичних даних

Для оцінки запропонованих методів в контрольованих умовах зашумлення розмітки, проведено експерименти на згенерованих наборах даних, що були синтезовані за допомогою розробленої моделі наборів даних із зашумленою розміткою.

4.1.1. Оцінка прогнозів нейронної мережі.

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

- Метрика F1 для оцінки задачі класифікації
- Коефіцієнт Дайса для оцінки 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.

Кількісна оцінка роздільноті внутрішніх представлень нейронної мережі. Для оцінки роздільноті внутрішніх представлень, використовується метод К-середніх. Оскільки в задачах класифікації та семантичної сегментації заздалегідь відома кількість класів, можна 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 проектуються за допомогою алгоритму нелінійного відображення Uniform Manifold Approximation and Projection (UMAP) [?] на двомірну площину.

4.1.2. Структура нейронних мереж. У всіх експериментах використана однаакова 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 3.2). 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 виконані відповідно до структури, описаної в розділі 3.1. При навчанні запропонованої моделі використовується обмежена функція втрат для задачі сегментації, а при прогнозуванні використовується метод об'єднання задач класифікації та сегментації.

4.1.3. Параметри експериментів.

Загальна процедура навчання. Складається з одного циклу навчання та валідації, якщо не вказано інакше.

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 4.1.

Таблиця 4.1: 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 4.2.

Таблиця 4.2: Результати експеримента 1

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

На рисунку 4.1 зображено приклади прогнозів нейронної мережі для тестових зображень.

На рисунках 4.2 та 4.3 зображено УМАР-проекцію ознак останнього шару енкодера для набору тестових зображень.

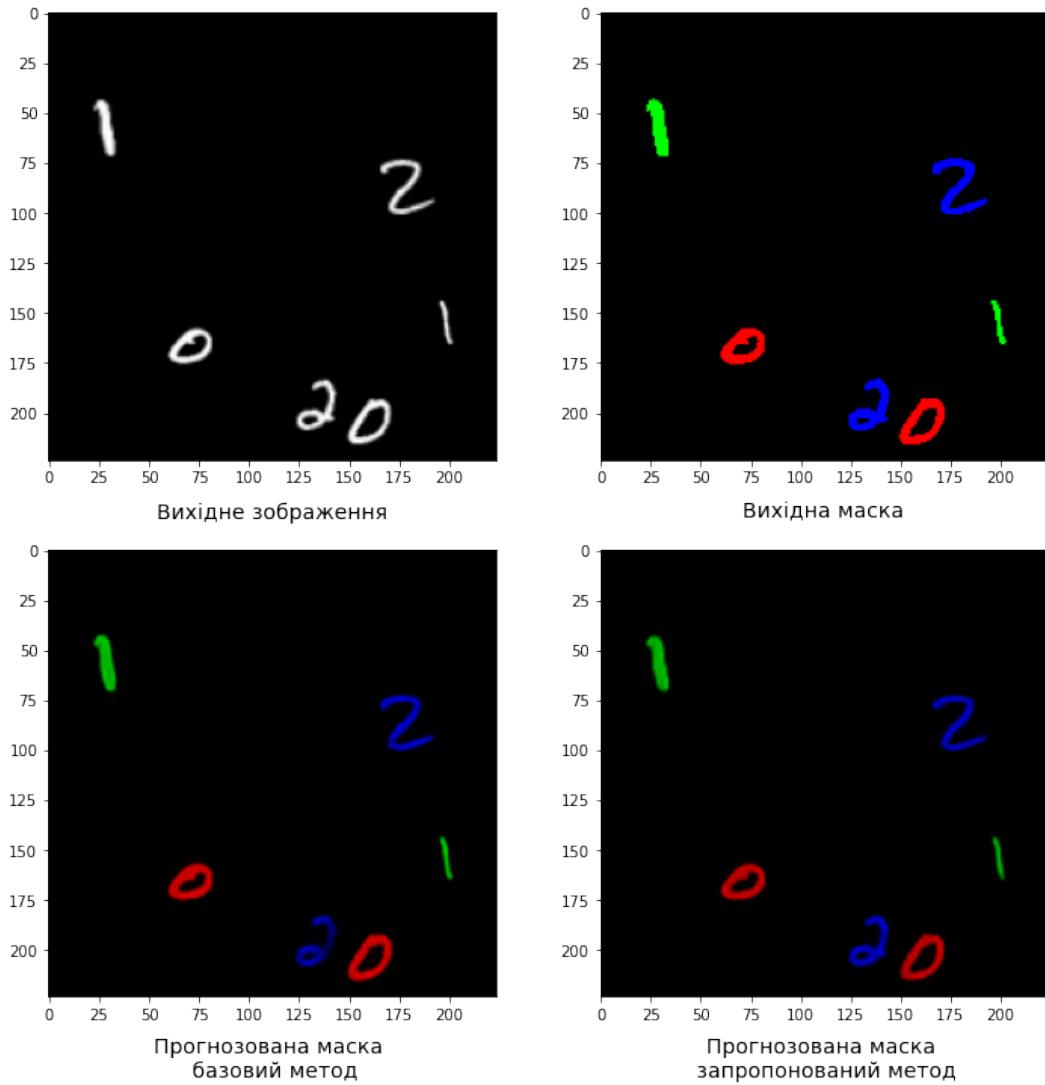


Рис. 4.1: Структура декодера класифікації

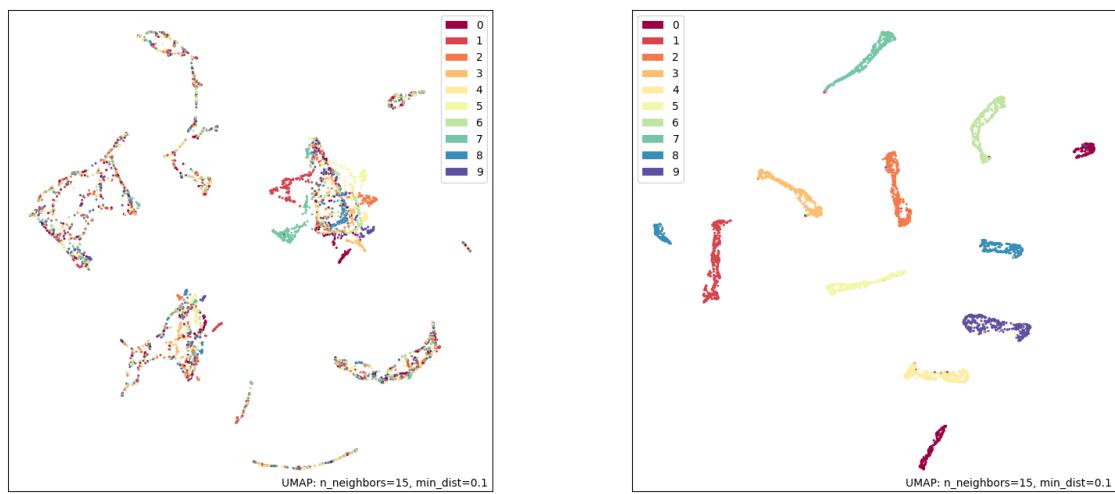


Рис. 4.2: УМАР ознак
базової моделі

Рис. 4.3: УМАР ознак
запропонованої моделі

Експеримент 2. Порівняння базового та запропонованого методів на простому наборі даних із зашумленням розмітки. Параметри генерації набору даних зазначено в таблиці 4.3.

Таблиця 4.3: Параметри набору даних експеримента

Об'єкти	MNIST
Текстура фону	константна
Текстура об'єктів	константна
P_e	0.1
P_d	0.25
N_{obj}	10
S_{obj}	36
Δ_{max}	0.2

Результати експерименту наведено в таблиці 4.4.

Таблиця 4.4: Результати експеримента

Модель	Міра Дайса	F1-міра	K_{sep}
Базова модель	0.81	-	12.10
Запропонований метод	0.91	0.93	15.73

На рисунку 4.4 зображені приклади прогнозів нейронної мережі для тестових зображень.

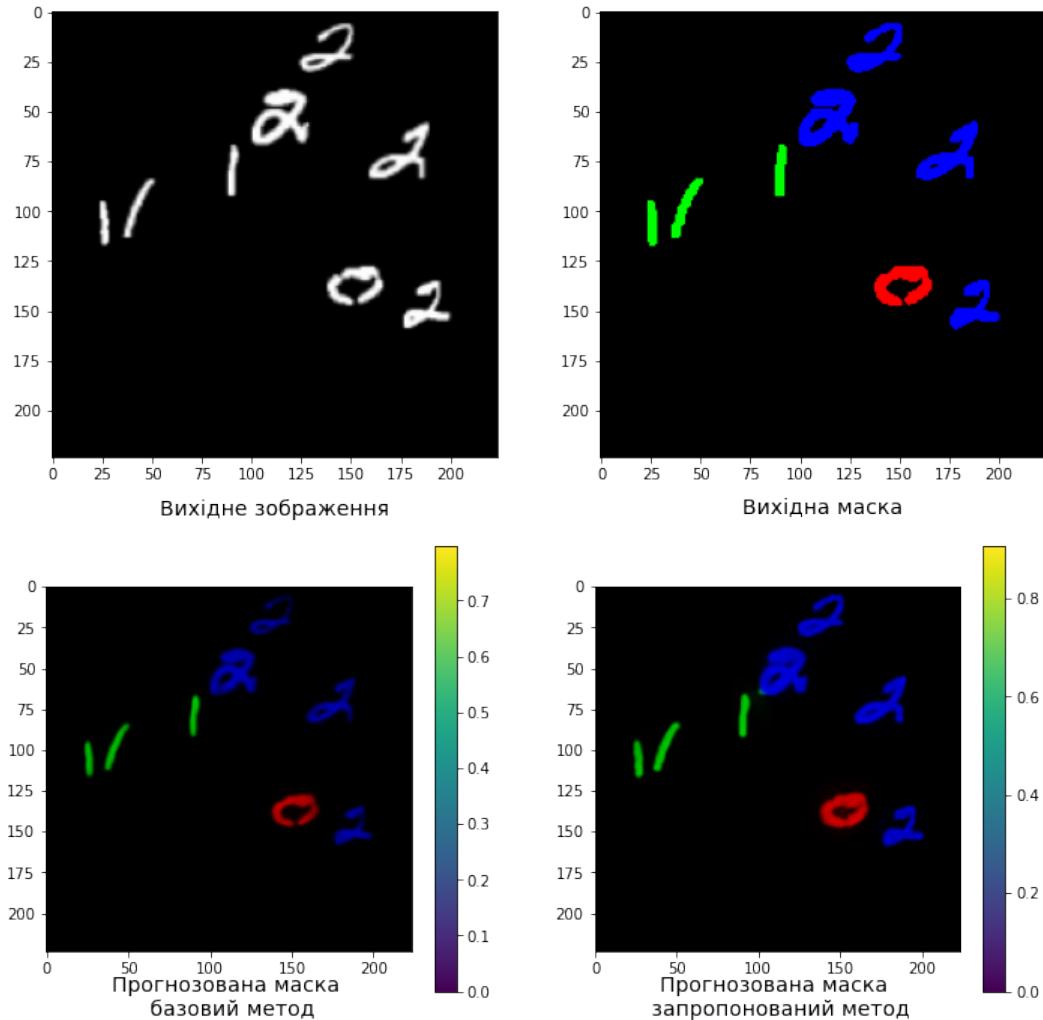


Рис. 4.4

На рисунках 4.5 та 4.6 зображені УМАР-проекцію ознак останнього шару енкодера для набору тестових зображень.

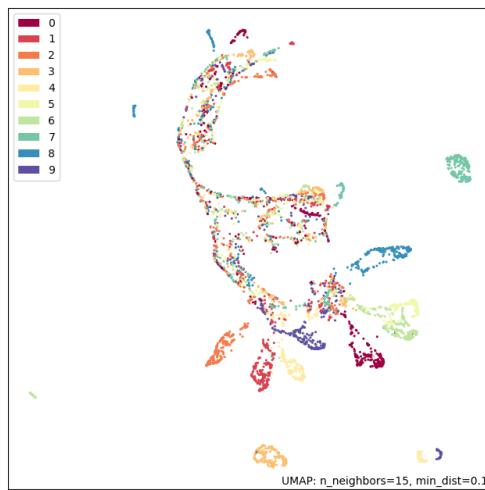


Рис. 4.5: УМАР ознак
базової моделі

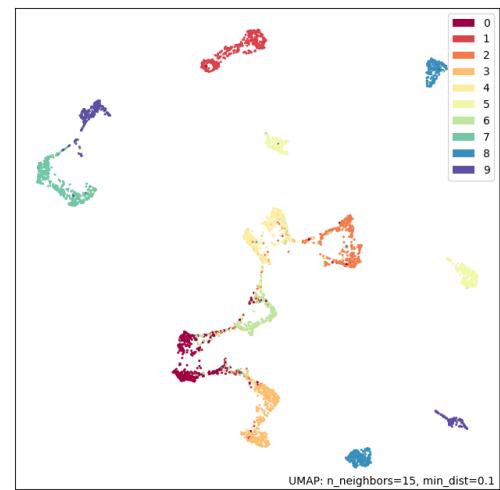


Рис. 4.6: УМАР ознак
запропонованої моделі

В умовах зашумленої розмітки, запропонований метод дає більш впевнені прогнози, також зберігаючи тонкі лінії, що порушуються при виконанні операції ерозії над розміткою.

Експеримент 3. Порівняння базового та запропонованого методів на простому наборі даних з ускладненими об'єктами та зашумленням розмітки. Параметри генерації набору даних зазначено в таблиці 4.5.

Таблиця 4.5: Параметри набору даних експеримента

Об'єкти	FashionMNIST
Текстура фону	константна
Текстура об'єктів	константна
P_e	0.1
P_d	0.25
N_{obj}	10
S_{obj}	36
Δ_{max}	0.2

Результати експерименту наведено в таблиці 4.6.

Таблиця 4.6: Результати експеримента

Модель	Mіра Дайса	F1-міра	K_{sep}
Базова модель	0.83	-	27.49
Запропонований метод	0.87	0.88	32.12

На рисунку 4.7 зображені приклади прогнозів нейронної мережі для тестових зображень.

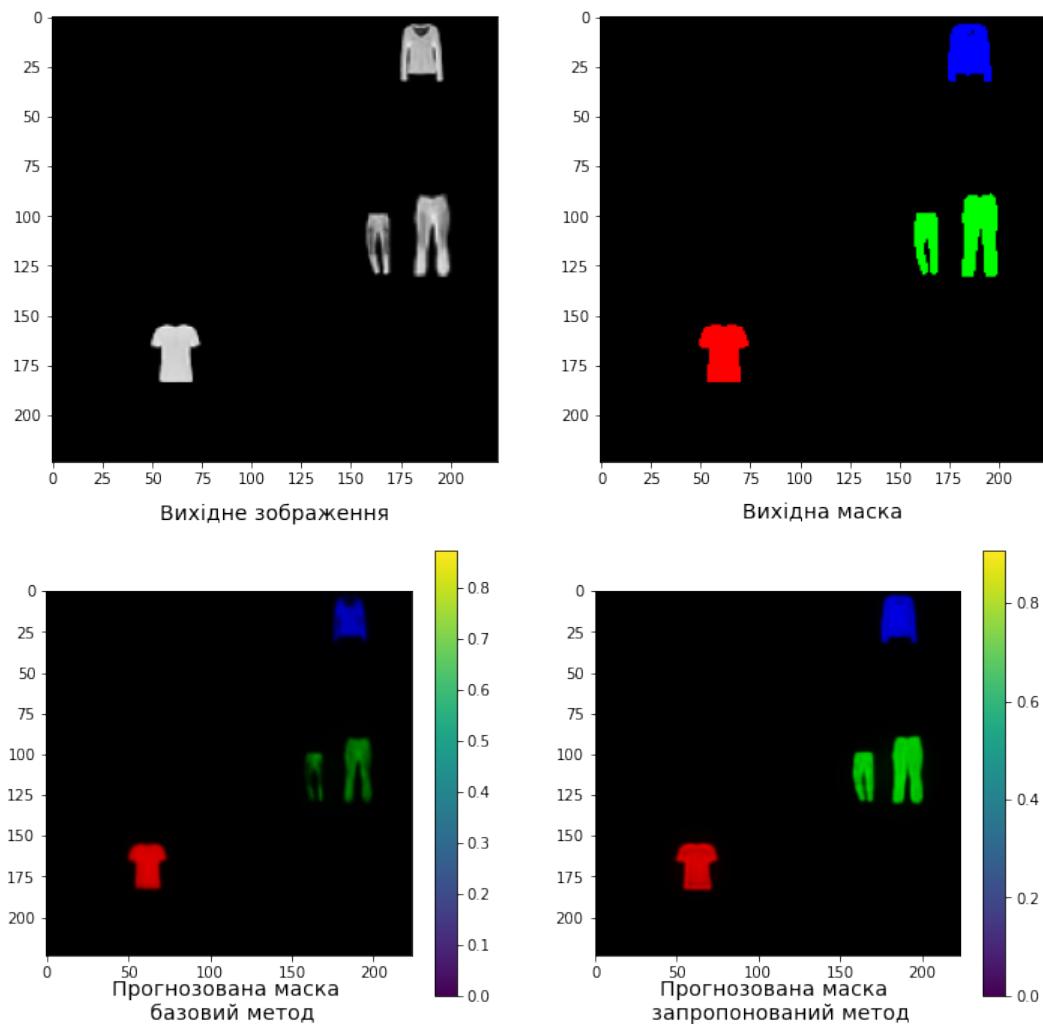


Рис. 4.7

На рисунках 4.8 та 4.9 зображені УМАР-проекцію ознак останнього

шару енкодера для набору тестових зображень.

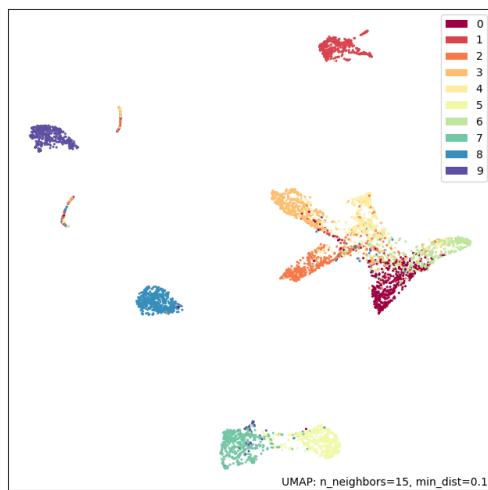


Рис. 4.8: УМАР ознак
базової моделі

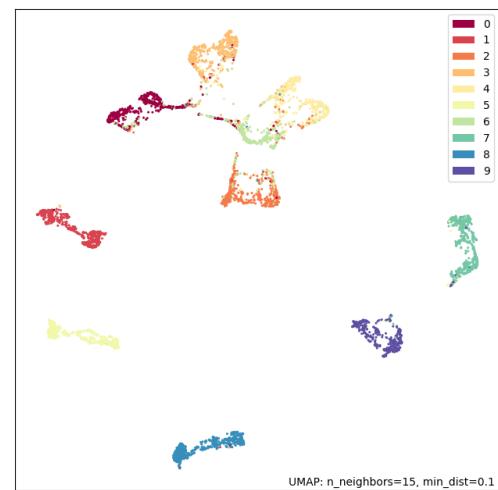


Рис. 4.9: УМАР ознак
запропонованої моделі

Як і в попередньому експерименті, в умовах зашумленої розмітки, запропонований метод дає більш впевнені прогнози, також зберігаючи більше малих деталей.

Експеримент 4. Порівняння базового та запропонованого методів на складному наборі даних з простими об'єктами та зашумленням розмітки. Параметри генерації набору даних зазначено в таблиці 4.7.

Таблиця 4.7: Параметри набору даних експеримента

Об'єкти	MNIST
Текстура фону	Imagenette
Текстура об'єктів	Imagenette
P_e	0.1
P_d	0.25
N_{obj}	10
S_{obj}	36
Δ_{max}	0.2

Результати експерименту наведено в таблиці 4.8.

Таблиця 4.8: Результати експеримента

Модель	Mира Дайса	F1-міра	K_{sep}
Базова модель	0.69	-	15.26
Запропонований метод	0.72	0.75	18.90

На рисунку 4.10 зображено приклади прогнозів нейронної мережі для тестових зображень.

На рисунках 4.11 та 4.12 зображено UMAP-проекцію ознак останнього шару енкодера для набору тестових зображень.

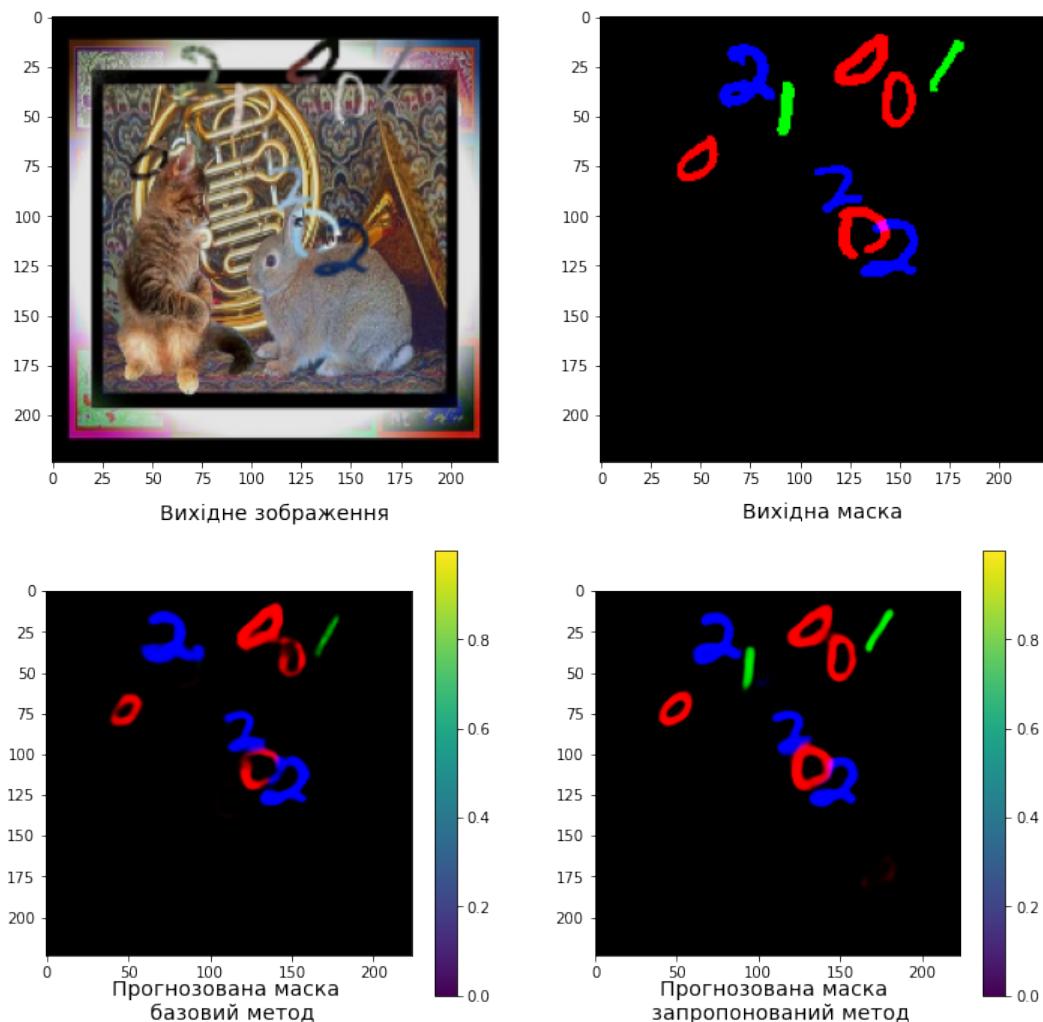


Рис. 4.10

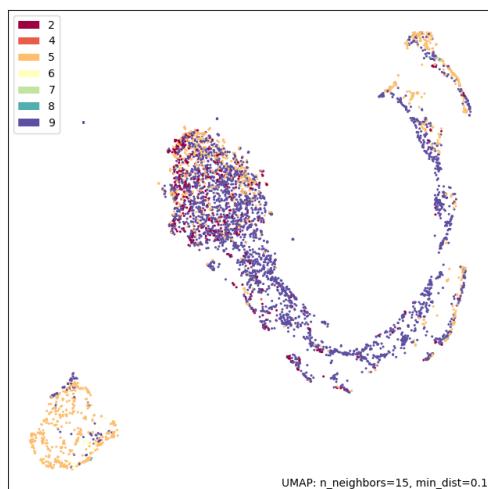


Рис. 4.11: УМАР ознак
базової моделі

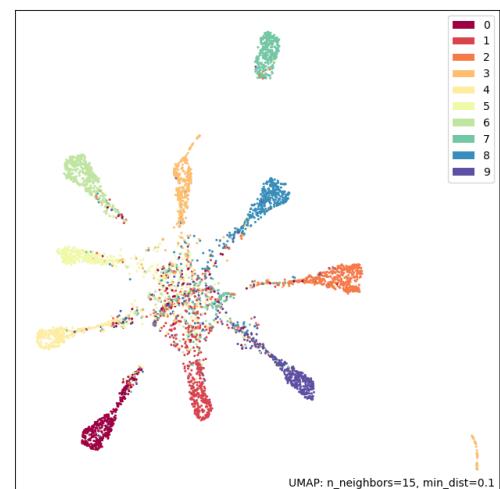


Рис. 4.12: УМАР ознак
запропонованої моделі

Як і в попередньому експерименті, в умовах зашумленої розмітки, запропонований метод дає більш впевнені прогнози, також зберігаючи більше малих деталей. Однак, внутрішні представлення мають гіршу роздільність через більш складний тип вхідних даних.

Експеримент 5. Порівняння базового та запропонованого методів на складному наборі даних з ускладненими об'єктами та зашумленням розмітки. Параметри генерації набору даних зазначено в таблиці 4.9.

Таблиця 4.9: Параметри набору даних експеримента

Об'єкти	FashionMNIST
Текстура фону	Imagenette
Текстура об'єктів	Imagenette
P_e	0.1
P_d	0.25
N_{obj}	10
S_{obj}	36
Δ_{max}	0.2

Результати експерименту наведено в таблиці 4.10.

Таблиця 4.10: Результати експеримента

Модель	Міра Дайса	F1-міра	K_{sep}
Базова модель	0.70	-	7.02
Запропонований метод	0.75	0.78	10.6

На рисунку 4.13 зображено приклади прогнозів нейронної мережі для тестових зображень.

На рисунках 4.14 та 4.15 зображено UMAP-проекцію ознак останнього шару енкодера для набору тестових зображень.

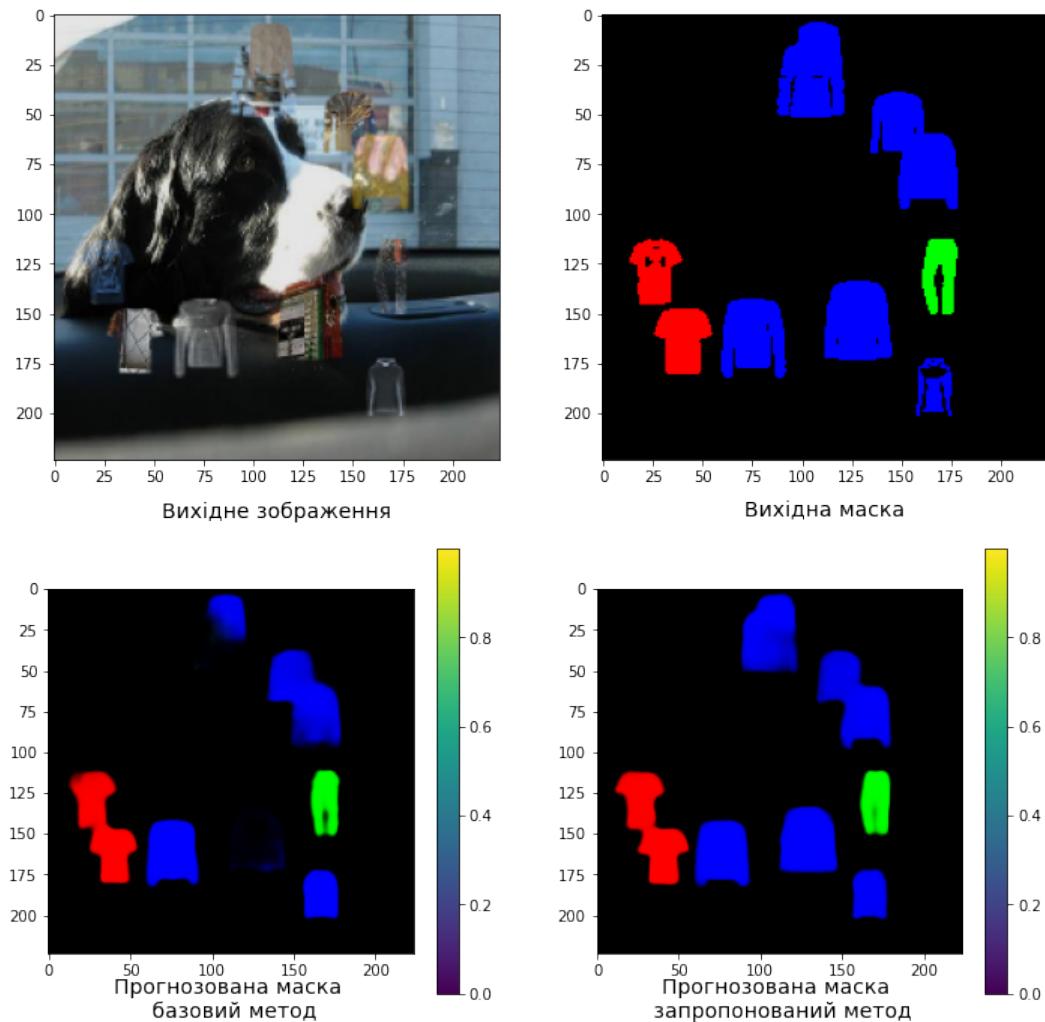


Рис. 4.13

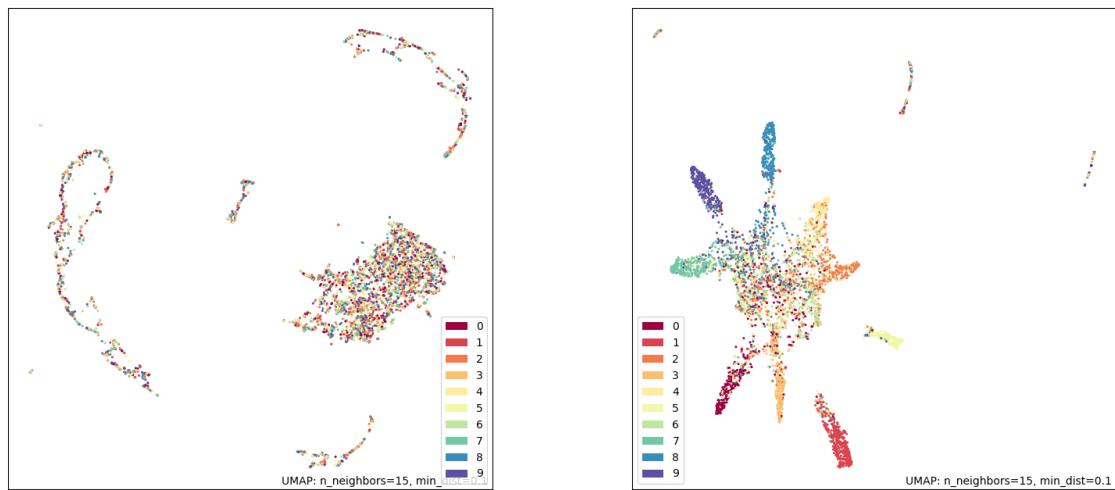


Рис. 4.14: UMAP ознак
базової моделі

Рис. 4.15: UMAP ознак
запропонованої моделі

Як і в попередньому експерименті, в умовах зашумленої розмітки, запропонований метод дає більш впевнені прогнози, також зберігаючи більше малих деталей. Однак, внутрішні представлення мають гіршу роздільність через більш складний тип вхідних даних.

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

Для оцінки стійкості базового та запропонованого методів до сили зашумлення розмітки, проведено серію експериментів зі зміною вірогідності зашумлення для кожного з об'єктів в наборі даних.

Результати експерименту наведено в таблиці 4.11

Таблиця 4.11: Порівняння стійкості моделей до шуму

P_e	P_d	Міра Дайса Базовий метод	Міра Дайса Запропонований метод
0.1	0.1	0.84	0.96
0.1	0.25	0.81	0.91
0.25	0.1	0.65	0.83
0.5	0.25	0.51	0.62
0.25	0.5	0.57	0.69
0.5	0.5	0.3	0.47
0.75	0.75	0.12	0.23

З таблиці видно, що запропонований метод є більш стійким до більшого рівня зашумлення розмітки. Також, як запропонований, так і базовий методи менш стійкі до підвищеної ймовірності ерозії.

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

Для перевірки стійкості до шуму розмітки в умовах наблизених до реальних, попередній експеримент проведено з використанням фону та текстур об'єктів з набору даних Imagenette. Результати експерименту наведено в таблиці 4.12.

Таблиця 4.12: Порівняння стійкості моделей до шуму

P_e	P_d	Міра Дайса Базовий метод	Міра Дайса Запропонований метод
0.1	0.1	0.72	0.78
0.1	0.25	0.70	0.75
0.25	0.1	0.61	0.69
0.5	0.25	0.48	0.53
0.25	0.5	0.55	0.60
0.5	0.5	0.21	0.41
0.75	0.75	0.08	0.15

Результати експерименту збігаються з результатами попереднього експерименту, хоча мають тенденцію до зменшення середнього значення метрики через підвищенну складність даних.

Дослідження внеску окремих компонентів. Параметри генерації набору даних зазначено в таблиці 4.13.

Таблиця 4.13: Параметри набору даних для проведення дослідження внеску окремих компонентів

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

В таблиці 4.14 наведено значення міри Дайса в задачі семантичної сегментації при послідовному додаванні складових частин запропонованого метода до базової моделі. Використані скорочення:

- MN - MNIST
- FM - FashionMNIST
- IN - ImageNette
- С - константний колір

Використано наступний порядок кодів в шифрі:

1. Набір даних, з якого взято об'єкти
2. Набір даних, з якого взято текстури фону
3. Набір даних, з якого взято текстури об'єктів

Таблиця 4.14: Дослідження внеску окремих компонентів

Змінна	MN/C/C	FM/C/C	MN/IN/IN	FM/IN/IN
Базова модель	0.81	0.83	0.69	0.70
+ декодер класифікації	0.83	0.84	0.69	0.71
+ обмеження функції втрат	0.88	0.86	0.71	0.74
+ комбінація прогнозів	0.91	0.87	0.72	0.75

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. Також, більш складні набори даних зменшують відносний приріст метрик, що є очікуваним через потребу збільшення кількості параметрів в нейронній мережі для можливості вивчення ускладнених представлень.

4.2. Експерименти на реальних даних

4.3. Класифікація та сегментація формаций хмар

4.3.1. 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 областью 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 областью одночасно, але на різних орбітах. Решта зображень, 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 météorologie Dynamique у Парижі, Франція. Команда з 68 вчених виявила ділянки з хмарами на кожному зображення, 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 розмічені 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 перетинатися - один і той самий піксель може бути помічений декількома класами, 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 4.16.

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 (недо-

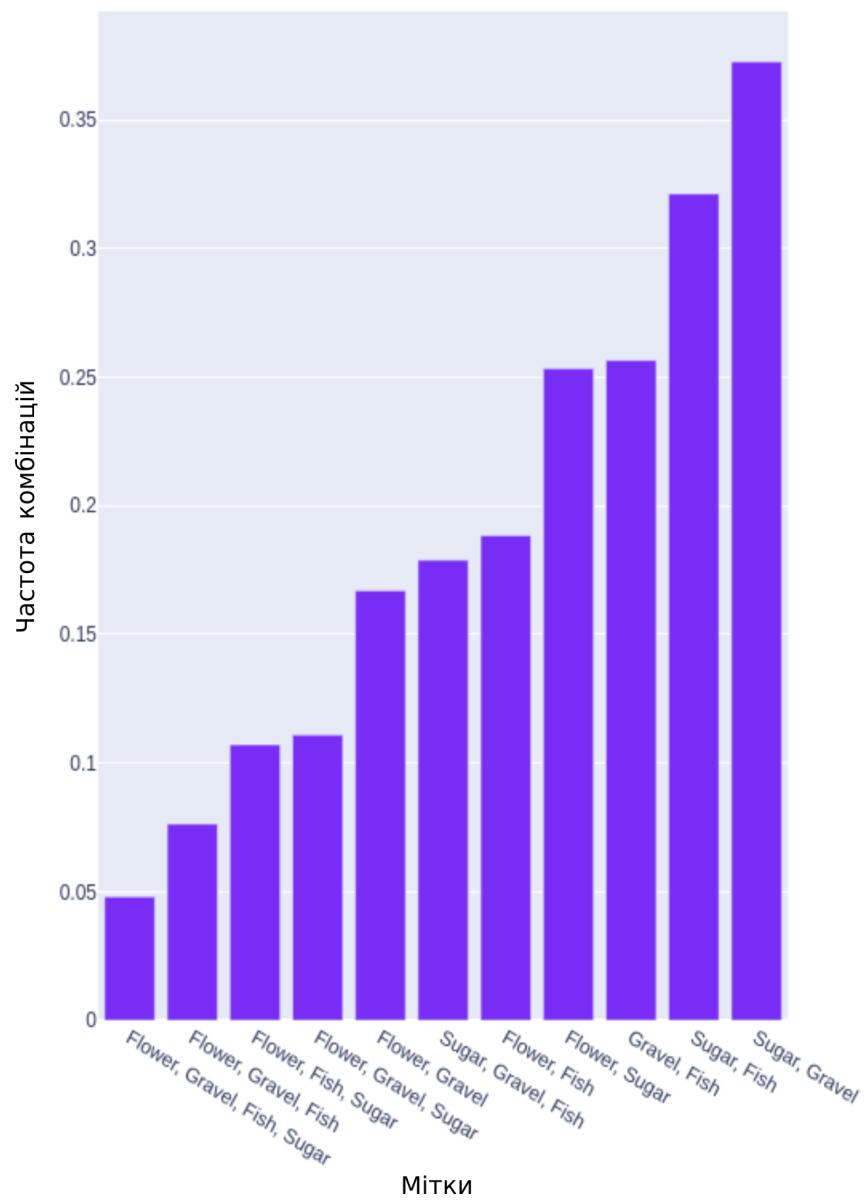


Рис. 4.16: Частота зустрічі декількох типів хмар на одному зображенні

дискретизація, передискретизація тощо). Для отримання більш стабільного значення метрик в умовах зашумлення тестової розмітки, оцінювання якості моделей виконано за допомогою k -кратної перехресної перевірки [?].

4.3.2. Процедура навчання. Тренування та прогнозування проводиться на зменшених версіях оригінальних зображень. Загальні параметри навчання:

- Розмір зображення: 350×525 пікселя
- Розмір пакету навчання: 64
- Кількість тренувальних зображень: 5546
- Кількість тестових зображень: 3698
- Кількість епох навчання: 50
- Оптимізатор: Adam
- Темп навчання: 10^{-4}
- Коефіцієнт L2 регуляризації параметрів мережі: 10^{-4}
- Закон зміни темпу навчання: косинус
- $k : 5$

Структура базової моделі. Базова модель являє собою екземпляр архітектури UNet. В ролі енкодера використана архітектура нейронної мережі ResNet50, що ініціалізована вагами претренованої на наборі даних Imagenet нейронної мережі.

Декодер виконано відповідно до класичної архітектури UNet, що складається з п'яти стадій. 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 виконані відповідно до структури, описаної в роз-

ділі 3.1. При навчанні запропонованої моделі використовується обмежена функція втрат для задачі сегментації, а при прогнозуванні використовується метод об'єднання задач класифікації та сегментації.

4.3.3. Результати експерименту. Чисельні результати експерименту зазначено в таблиці ??.

Таблиця 4.15: Результати експеримента

Модель	Міра Дайса	F1-міра
Базова модель	0.56 ± 0.03	-
Запропонований метод	0.63 ± 0.02	0.70 ± 0.01

Приклад зображення з тестового набору даних з вихідною розміткою та прогнозом нейронної мережі зображене на рисунку 4.17.

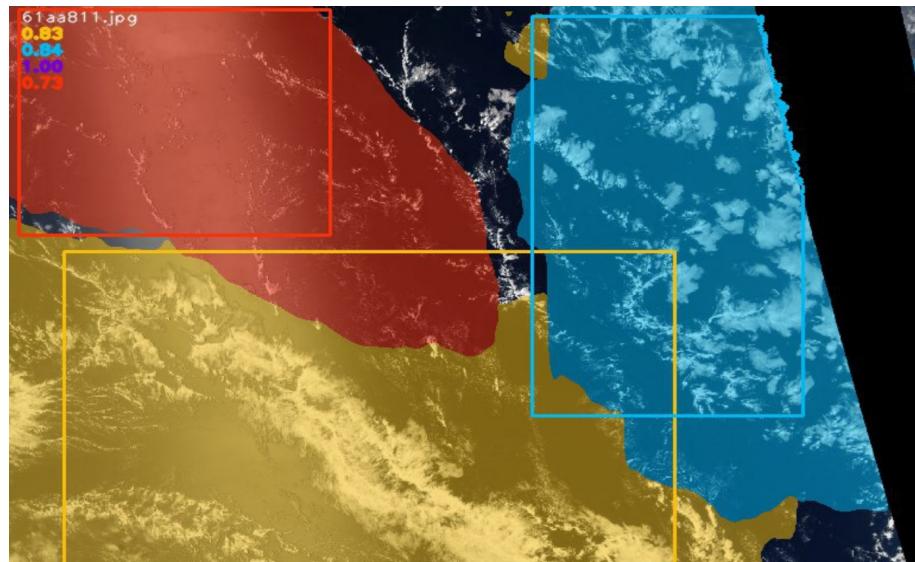


Рис. 4.17: Оригінальна розмітка (контури) та прогнози нейронної мережі (заливка)

4.3.4. Класифікація стадії осередків діабетичної ретинопатії.

В даному експерименті розглянуто можливість переносу підходу із задачі сегментації на задачу класифікації. Так, в ролі похідних задач до зада-

чі класифікації використані задачі лінійної регресії мітки класу, а також порядкової регресії, що відповідає специфіці задачі [?].

4.3.4.1. Опис наборів даних. Використані в цьому експерименті зображення, 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) [?], з якого використано 413 фотографій очного дна, та MESSIDOR (Методи оцінки методів сегментації та індексації в області офтальмології сітківки) [?], з якого використано 1200 фотографій очного дна. Оскільки розмітка оригінального набору даних MESSIDOR відрізняється від інших наборів даних, ми використали версію, яку група офтальмологів [?] помаркувала відповідно до стандартного протоколу.

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

Оцінка проводиться на наборі даних Kaggle 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.

4.3.5. Процедура навчання. Тренування та прогнозування проводиться на зменшених версіях оригінальних зображень. Загальні параметри навчання:

- Розмір зображення: 380×380 пікселя
- Розмір пакету навчання: 32
- Кількість тренувальних зображень: 5275
- Кількість тестових зображень: 1928
- Кількість епох навчання: 75
- Оптимізатор: RAdam
- Темп навчання: 10^{-4}
- Коефіцієнт L2 регуляризації параметрів мережі: 10^{-4}
- Закон зміни темпу навчання: косинус
- $k : 5$

Структура базової моделі. Базова модель являє собою екземпляр архітектури ResNet. В ролі енкодера використана архітектура нейронної мережі ResNeXt50, що ініціалізована вагами претренованої на наборі даних Imagenet нейронної мережі.

Декодер класифікації являє собою два лінійних шари, функція активації - ReLU. Функція активації останнього шару декодера - логістична сигмоїда.

Структура запропонованої моделі. Енкодер та декодери запропонованої моделі виконані відповідно до структури, описаної в розділі 3.1. В означеній задачі використано ідентичні декодери для різних задач, структуру яких зазначено на рисунку ???. При навчанні запропонованої моделі використовується обмежена функція втрат для задачі класифікації, а при прогнозуванні використовується запропонований метод об'єднання задач.

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.

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

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.

Процедура навчання. 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: фокальна функція втрат [?] для класифікаційного декодера, бінарна фокальна функція втрат [?] для декодера порядкової регресії та середньоквадратична помилка для декодера регресії.

4.3.6. Результати експерименту. Чисельні результати експерименту зазначено в таблиці 4.16.

Таблиця 4.16: Результати експеримента

Модель	F1-міра
Базова модель	0.68 ± 0.05
Запропонований метод	0.79 ± 0.06

4.4. Класифікація раку шкіри та локалізація родимок

Метою даного експерименту є тестування запропонованого методу локалізації важливих для класифікації ознак в умовах відсутності розмітки сегментації.

4.4.1. 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. Okрім даних про зображення, були надані метадані про пацієнтів. Один пацієнт має декілька зображень різних родинок.

Зображення та метадані надані у форматі DICOM, який є загально-вживаним 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 [?].

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 за- здалегідь визначеного кінцевого набору.

Метадані доступні для кожного пацієнта, тому різні зображення можуть мати одинаковий 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 рисунку 4.18.

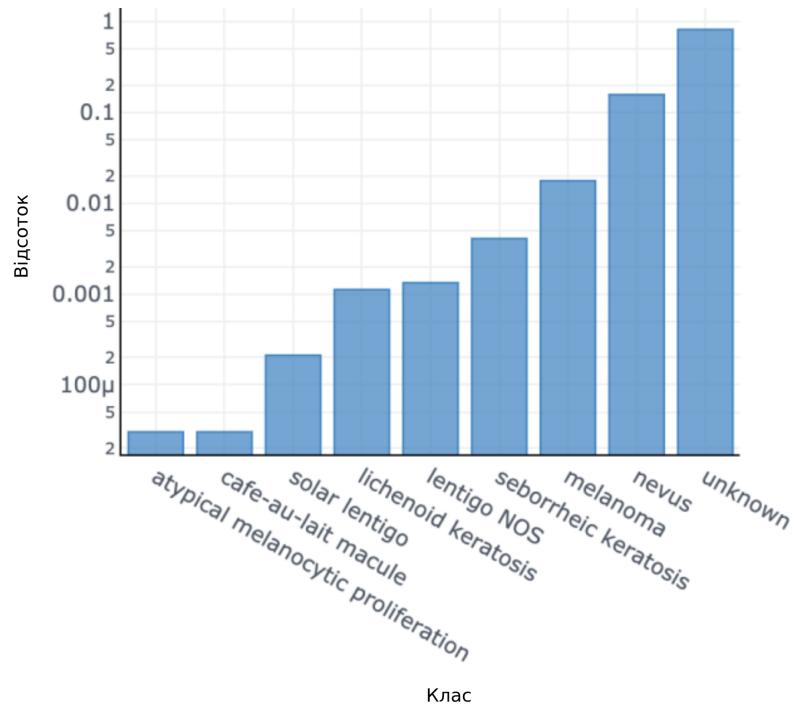


Рис. 4.18: Розподіл класів в наборі даних SIIM-ISIC Melanoma Classification

Для значень *unknown* в полі *diagnosis*, автори набору даних гарантують, що новоутворення не є злоякісним [?].

4.4.2. Процедура навчання. Тренування та прогнозування проводиться на зменшених версіях оригінальних зображень. Загальні параметри навчання:

- Розмір зображення: 512×512 пікселів
- Розмір пакету навчання: 64
- Кількість тренувальних зображень: 25000

- Кількість тестових зображень: 5000
- Кількість епох навчання: 50
- Оптимізатор: Adam
- Темп навчання: 10^{-4}
- Коефіцієнт L2 регуляризації параметрів мережі: 10^{-4}
- Закон зміни темпу навчання: косинус
- $k : 5$

Структура базової моделі. Базова модель являє собою екземпляр архітектури ResNet. В ролі енкодера використана архітектура нейронної мережі ResNet50, що ініціалізована вагами претренованої на наборі даних Imagenet нейронної мережі.

Декодер класифікації являє собою два лінійних шари, функція активації - ReLU. Функція активації останнього шару декодера - логістична сигмоїда.

Структура запропонованої моделі. Енкодер та декодери запропонованої моделі виконані відповідно до структури, описаної в розділі 3.1. Для спрощення задачі локалізації, декодер сегментації має менше стадій та виводить карту уваги в меншій роздільній здатності, ніж вхідне зображення.

Процедура навчання. Процедури навчання і прогнозування відповідають запропонованому методу напівавтоматичного навчання задачі локалізації, що описаний в розділі 3.3.1.

4.4.3. Результати експерименту. Чисельні результати експерименту зазначено в таблиці 4.17.

Таблиця 4.17: Результати експеримента

Модель	F1-міра
Базова модель	0.83 ± 0.01
Запропонований метод	0.86 ± 0.01

Приклад локалізації нейронною мережею зображенено на рисунку 4.19.

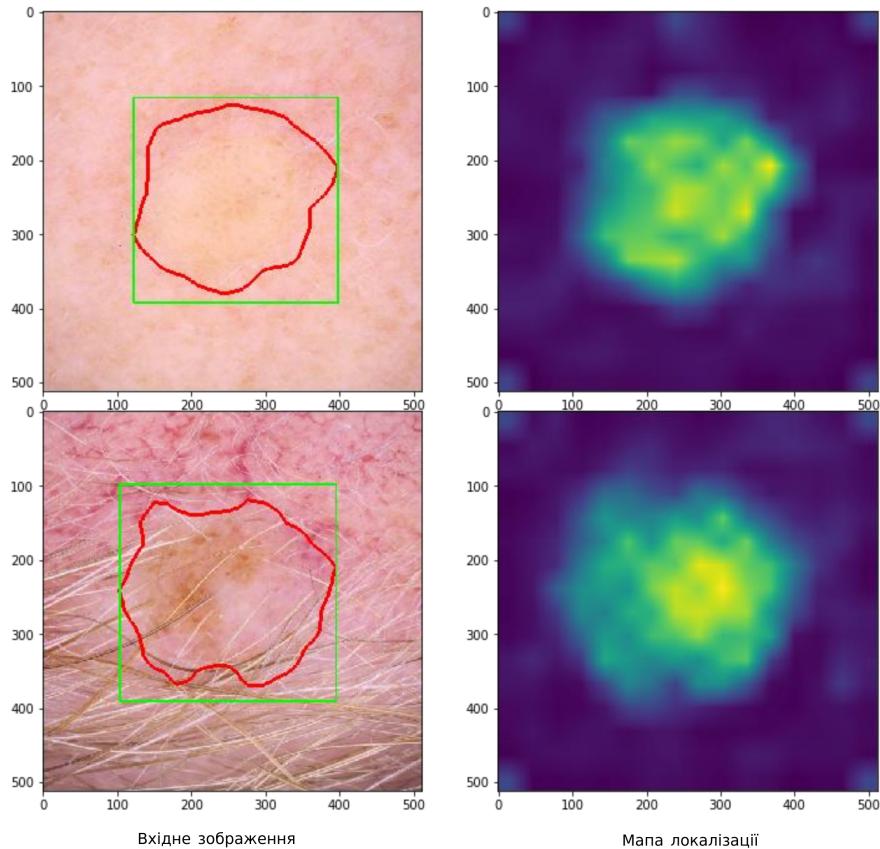


Рис. 4.19: Приклад напівавтоматичної локалізації родинок

4.5. Інструментальні засоби

Застосування розроблених моделей та методів значно ускладнюється через відсутність інструментальних програмних засобів, що забезпечували б їх роботу. Розробка інструментальних засобів, що реалізують моделі та методи є актуальною задачею.

Задача навчання глибоких нейронних мереж є обчислювально-складною

задачею, тому для прискорення процесів навчання та прогнозування необхідно використовувати хмарні технології обчислень для задач машинного навчання.

На основі запропонованих у дисертації моделей, розроблених методів навчання нейронних мереж та прогнозування результатів, реалізовано інструментальні засоби у вигляді програмних модулів для навчання та прогнозування на основі хмарних сервісів.

Всі інструментальні засоби реалізовано мовою програмування Python. Провайдером хмарних сервісів є Amazon Web Services [?]. Моделі глибоких нейронних мереж та методи іх навчання реалізовано на базі фреймворку автоматичного диференціювання Pytorch [?]. Відтворюваність результатів експериментів забезпечується використанням програмної високорівневої обгортки Catalyst [?], яка дозволяє зберігання конфігурації експериментів у широко анотованих конфігураційних файлах.

Інструментальні засоби навчання моделей розгорнуті за допомогою сервісу AWS EC2 [?], в той час як прогнозування результатів розгорнуто за допомогою AWS Sagemaker [?].

Зберігання даних та артефактів навчання моделей (файли конфігурації та бінарні файли) зберігаються за допомогою сервісу AWS S3 [?].

4.5.1. Узагальнена структура системи навчання глибоких нейронних мереж. Сучасні нейронні мережі в цілому, та запропоновані їх моделі потребують високої кількості обчислювальних ресурсів в процесі навчання. Необхідно складовою обладнання для їх навчання є відеоприскорювачі загального призначення (англ. General-Purpose Graphic Processing Unit - GPGPU). Економічно-вигідним є використання арендних відеоприскорювачів в «хмарних сервісах», що потребує відповідної зміни парадигм проектування систем навчання.

4.5.2. Узагальнена структура прогнозування за допомогою глибоких нейронних мереж. В даний час швидко розвивається складність програмного забезпечення, що виконує наукові та інженерні задачі. Сучасні системи автоматизованого скринінгу являють собою складні багатокомпонентні системи програмних модулів, які можуть бути фізично розподілені один від одного. Так, наприклад, обладнання для скринінгу може знаходитися безпосередньо в лікарні, в той час як сервіси для обробки даних можуть бути пов'язані з обладнанням, або знаходитися на сторонніх сервісах. Також, все частіше нагальною є потреба зміни модулів в процесі роботи системи, наприклад, при оновленні нейронних мереж, що є їхніми складовими частинами.

Також, протягом останніх років виник інтерес до запуску нейронних мереж на пограничних пристроях, наприклад, безпосередньо на камерах відеоспостереження, для зниження витрат на підтримку інфраструктури.

4.6. Висновки до четвертого розділу

ВИСНОВКИ