

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІНСТИТУТ КОМП'ЮТЕРНИХ СИСТЕМ

Кваліфікаційна наукова
праця на правах рукопису

ТИМЧЕНКО Борис Ігорович

УДК 004.932.72'1

ДИСЕРТАЦІЯ

**МЕТОДИ ТА ІНСТРУМЕНТАЛЬНІ ЗАСОБИ ОБРОБКИ
ПЛАНАРНИХ ЗОБРАЖЕНЬ НА ОСНОВІ ШТУЧНИХ
НЕЙРОННИХ МЕРЕЖ В СИСТЕМАХ АВТОМАТИЗОВАНОГО
СКРИНІНГУ**

122 — Комп'ютерні науки
Інформаційні технології

Подається на здобуття наукового ступеня
доктора філософії

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

_____ Б. І. Тимченко

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

Одеса — 2021

АНОТАЦІЯ

Тимченко Б. І. Методи та інструментальні засоби обробки планарних зображень на основі штучних нейронних мереж в системах автоматизованого скринінгу. — Кваліфікаційна наукова праця на правах рукопису.

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

Ключові слова: .

ABSTRACT

Tymchenko B. I. Methods and tools for processing planar images based on artificial neural networks 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.

Key words: .

ЗМІСТ

Вступ	6
Розділ 1. Аналіз методів, алгоритмів і програмних засобів обробки планарних зображень	11
1.1. Аналіз задачі автоматизованого скринінгу	11
1.1.1. Задачі автоматизованого скринінгу в медицині	11
1.1.2. Задача автоматизованого скринінгу на виробництві .	15
1.1.3. Задачі автоматизованого скринінгу в метеорології . .	15
1.2. Аналіз методів та алгоритмів обробки зображень	17
1.2.1. Особливості планарних зображень	17
1.2.2. Feature detection	17
1.2.3. Задача класифікації зображень та відомі методи її розв’я- зання	17
1.2.4. Задача семантичної сегментації зображень та відомі методи її розв’язання	17
1.3. Методи глибинного машинного навчання	17
1.3.1. Загальні відомості про штучні нейронні мережі	18
1.3.2. Згорткові нейронні мережі	18
1.3.3. Сучасні архітектури нейронних мереж в задачах кла- сифікації та сегментації зображень	18
1.4. Трансферне навчання	19
1.5. Методи багатозадачного глибинного машинного навчання .	20
1.5.1. Визначення багатозадачного машинного навчання . .	20
1.5.2. Жорсткий розподіл параметрів	21
1.5.3. М’який розподіл параметрів	22

1.5.4.	Дистиляція прогнозів моделі	22
1.5.5.	Вивчення незалежних задач	22
1.6.	Аналіз методів оцінки якості класифікації та сегментації зображень в задачі автоматизованого скринінгу	22
1.6.1.	Помилки 1 і 2 роду та матриця невідповідностей . . .	22
1.6.2.	Влучність, чутливість та специфічність	23
1.6.3.	F1 та Dice score	23
1.6.4.	Cohen's Карра	23
1.6.5.	Криві характеристик класифікатора: ROC и PR . . .	23
Розділ 2. Метод		24
Розділ 3. Експерименти		26
3.1.	Класифікація стадії осередків діабетичної ретинопатії	26
3.1.1.	Опис наборів даних	26
3.1.2.	Детальний опис структури нейронної мережі	27
3.1.3.	Детальний опис процедури навчання	27
3.2.	Класифікація раку шкіри та сегментація родимок	29
3.2.1.	Опис набору даних SIIM-ISIC Melanoma Classification	29
3.2.2.	Детальний опис структури нейронної мережі	30
3.2.3.	Детальний опис процедури навчання	30
3.2.4.	Результати експерименту	31
3.3.	Класифікація та сегментація формацій хмар	31
3.3.1.	Опис набору даних Understanding Clouds from Satellite Images	31
3.3.2.	Детальний опис структури нейронної мережі	32
3.3.3.	Результати експерименту	32

ВСТУП

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

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

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

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

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

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

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

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

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

Для досягнення поставленої мети необхідно виконати наступні задачі:

- Провести аналіз існуючих методів, алгоритмів та засобів обробки планарних зображень в системах автоматизованого скринінгу
- Розробити метод зниження дисперсії прогнозів нейронної мережі
- Розробити метод багатозадачного навчання штучних нейронних мереж для роботи в умовах частково помилкової розмітки навчального набору

- Розробити метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки семантичної сегментації в навчальному наборі даних
- Удосконалити архітектуру штучної нейронної мережі для багатозадачного навчання
- Розвинути нейромережевий метод обробки планарних зображень, який за рахунок навчання на декількох задачах одночасно підвищить точність класифікації та семантичної сегментації планарних зображень.
- Спроекувати та розробити інструментальні засоби, що виконують розроблений метод
- Провести випробування розробленого методу та інструментальних засобів в рамках експерименту

Об’єкт дослідження. Процес програмної обробки планарних зображень

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

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

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

Наукова новизна отриманих результатів. *вперше:*

- Розроблено метод багатозадачного навчання штучних нейронних мереж для роботи в умовах частково помилкової розмітки навчального набору даних в задачах автоматизованого скринінгу

- Розроблено метод зниження дисперсії прогнозів нейронної мережі за рахунок використання результатів семантично-близьких задач при багатозадачному навчанні
- Розроблено метод сегментації важливих для класифікації ознак зображення в умовах відсутності розмітки для сегментації в навчальному наборі даних

удосконалено:

- Архітектуру штучної нейронної мережі для багатозадачного навчання

отримав подальший розвиток:

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

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

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

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

Структура та обсяг дисертації. Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Робота містить X сторінок основного тексту, Y рисунків та Z таблиць. За-

гальний обсяг дисертації – N сторінок, список літературних джерел налічує K найменувань. Дисертація містить I додатків, розміщених на J сторінках.

РОЗДІЛ 1

АНАЛІЗ МЕТОДІВ, АЛГОРИТМІВ І ПРОГРАМНИХ ЗАСОБІВ ОБРОБКИ ПЛАНАРНИХ ЗОБРАЖЕНЬ

1.1. Аналіз задачі автоматизованого скринінгу

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

1.1.1.1. Скринінг раку шкіри. Рак шкіри є найбільш розповсюдженим видом злоякісних пухлин, і саме меланома є причиною більшості смертей від раку. Світова проблема захворюваності на меланому стрімко зростала за останні 50 років і стала проблемою, з якою намагаються боротися багато вчених з різних країн.

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

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

ознаки.

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

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

Коделла та ін. [10] запропонували систему сегментації та класифікації меланоми за дермоскопічними зображеннями шкіри. Для класифікації хвороб вони застосували ансамбль останніх методів машинного навчання, включаючи глибокі залишкові мережі, згорткові нейронні мережі тощо. Вони довели, що ансамблі здатні давати кращі результати, ніж моделі окремо.

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

CLASS значно покращила ефективність класифікації зображень та якість рекомендаційної частини системи.

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

1.1.1.2. Скринінг діабетичної ретинопатії. Діабетична ретинопатія є одним із найбільш загрозливих ускладнень діабету, при якому пошкодження сітківки викликає сліпоту. Вона пошкоджує кровоносні судини тканини сітківки, викликаючи витік рідини та погіршення зору. Поряд із захворюваннями, що призводять до сліпоти, такими як катаракта і 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 використовувати класичні методи комп'ютерного зору та машинного навчання, щоб забезпечити відповідне рішення цієї проблеми.

Наприклад, Priya et al. запропонували підхід на основі комп'ютерного зору для виявлення діабетичної ретинопатії за допомогою кольорових зображень очного дна. Автори створили набір ознак із вихідного 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 *condc* з найкращими результатами 73.4% точності та 68.4% для F-міри, використовуючи набір даних із 151 зображення з різною роздільною здатністю.

Зі зростанням популярності підходів, заснованих на глибокому навчанні, з'явилися методи, які застосовують глибокі згорткові нейронні мережі

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

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

Інші дослідники також намагалися здійснити трансферне навчання за допомогою згорткових нейронних мереж. Хагош та ін. hagos спробував навчити InceptionNet V3 для класифікації 5 класів з нейронною мережею, натренованою на наборі даних ImageNet і досяг точності 90,9%. 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. Задача автоматизованого скринінгу на виробництві.

1.1.3. Задачі автоматизованого скринінгу в метеорології. Однією з найцінніших особливостей визначення кліматичної моделі Землі є поведінка хмар. Однак дослідження їх поведінки є однією з найскладніших частин, оскільки вимагає досконалого розуміння всіх процесів в атмосфері.

Класифікація різних типів організації хмар допомагає покращити розуміння цих хмар, що, в свою чергу, допоможе нам побудувати кращі кліматичні моделі.

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

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

Один із методів виявлення хмар був розроблений Чжу та співавторами [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 Enhanced Thematic 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 мо-

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

1.2. Аналіз методів та алгоритмів обробки зображень

1.2.1. Особливості планарних зображень. Глибина сцени незначна відносно відстані до камери. Масштаб об'єктів в сцені є постійним.

1.2.2. Feature detection.

1.2.3. Задача класифікації зображень та відомі методи її розв'язання.

1.2.4. Задача семантичної сегментації зображень та відомі методи її розв'язання. Before 2000, we used several methods in digital image processing: threshold segmentation, region segmentation, edge segmentation, texture features, clustering and so on. From 2000 to 2010, there are four main methods: graph theory, clustering, classification and combination of clustering and classification.

1.3. Методи глибинного машинного навчання

Глибинне машинне навчання - це частина більш широкого сімейства методів машинного навчання, заснованих на штучних нейронних мережах.

(ШНМ). Алгоритми глибинного машинного навчання моделюють високо-рівневі абстракції за допомогою глибинного графу з декількох шарів, що побудовані з лінійних, чи нелінійних перетворень.

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

1.3.2. 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 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.

Згортки Активації Підвибірки Нормалізації

1.3.3. Сучасні архітектури нейронних мереж в задачах класифікації та сегментації зображень.

1.3.3.1. Визначні сучасні архітектури ШНМ класифікації зображень. AlexNet VGGNet Inception ResNet DenseNet EfficientNet

1.3.3.2. Визначні архітектури ШНМ для сегментації зображень. FCN Перше застосування повністю згорткової нейронної мережі в задачі сегментації Перше застосування індуктивного переносу з задачі класифікації на задачу сегментації

SegNet Перше застосування архітектури енкодер-декодер для задачі семантичної сегментації Передача проміжної інформації від енкодера до декодера через індекси

UNet Передача проміжної інформації від енкодера до декодера через проміжні з'єднання Використання різних стратегій аугментації даних для навчання на малій вибірці зображень Використання зваженої функції втрат для розділення близьких об'єктів

LinkNet Перше використання індуктивного переносу в архітектурі енкодер-декодер виключно для енкодера Перше використання архітектури ResNet в якості енкодера

Feature pyramid network Перше використання піраміди ознак для сегментації об'єктів різного розміру Перше використання агрегації

DeerLab Перша архітектура сегментації, що використовує мапи ознак для різних масштабів Перше використання atrous spatial pyramid pooling

1.4. Трансферне навчання

Трансферне навчання – це проблема машинного навчання, яка фокусується на збереженні знань, отриманих під час вирішення однієї задачі з подальшим застосуванням її результатів до іншої, але близької до неї. З практичної точки зору, повторне використання або передача інформації з раніше засвоєних завдань для вивчення нових може значно підвищити

ефективність їх розв'язання [savchenko]. Деякі дослідники [] вважають багатозадачне машинне навчання підмножиною трансферного навчання.

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

1.5. Методи багатозадачного глибинного машинного навчання

Багатозадачне машинне навчання - це один з підходів машинного навчання, в якому одночасно вирішуються кілька навчальних задач, використовуючи спільності та відмінності між ними. Кожна з таких задач може бути загальною навчальною задачею, такою як навчання з учителем (наприклад, задача класифікації, або регресії), навчання без вчителя (кластеризація), напіваавтоматичне 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 [caguna], що спільне навчання на цих задачах може призвести до значного підвищення продуктивності в порівнянні з навчанням на кожній задачі окремо. Отже, багатозадачне навчання направлено на підвищення якості узагальнення кількох пов'язаних задач. Далі будуть розглядатися виключно задачі контрольованого навчання. Подібно до людського навчання, корисно вивчати кілька навчальних задач разом, оскільки знання, що містяться в одній задачі, можуть бути використані іншими. Наприклад, людина, що вивчає математику та математичну статистику, може використати цей досвід для вивчення інших схожих галузей, наприклад, машинного навчання.

1.5.1. Визначення багатозадачного машинного навчання. Для n 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 p або деяких з них задчах. В літературі визначається декілька видів БЗМН: Однорідне БЗМН: Для кожної з задач прогнозується лише один вихід. Наприклад, розпізнавання цифр MNIST зазвичай використовується для оцінки алгоритмів БЗМН. В цьому разі, воно розглядається як 10 завдань бінарної класифікації. Неоднорідне БЗМН: Для кожної з задач прогнозується свій набір виходів. Наприклад, сучасні задачі детектування об'єктів потребують одночасного прогнозування класу об'єкта (задача класифікації), його положення на зображенні та розмірів (задача регресії). Для задач контрольованого навчання, кожна задача T_i містить в собі набір даних D_i , що складається з m елементів. . .

Дослідники вважають, що коли різні задачі використовують однакові набори вхідних даних, БЗМН зводиться до задачі класифікації або регресії з кількома виходами. Однак, новіші дослідження визначають БЗМН на одному наборі вхідних даних, або незалежно від кількості наборів вхідних даних. В методах БЗМН, заснованих на глибоких нейронних мережах, виділяють дві групи відповідно до способу розподілу параметрів між різними задачами.

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

В найпершій роботі, що використовує цей принцип, було запропоновано для детекції ключових точок лиця використати додаткові задачі регресії пози голови та класифікації атрибутів. Модифікацією попереднього підходу є багатозадачні каскадні мережі, в яких результати для одних задач використовуються як доповнення до вхідної інформації для наступних. Така організація шарів дозволяє підвищити точність на більш складних завда-

ннях.

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.

1.5.3. М'який розподіл параметрів.

1.5.4. Дистиляція прогнозів моделі.

1.5.5. Вивчення незалежних задач.

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

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

В задачах контрольованого навчання, зокрема бінарної класифікації, за нульову гіпотезу зазвичай приймається приналежність 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.

1.6.2. 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. Чутливість, або повнота (англ. sensitivity, recall) - вимірює частку істинно-позитивних зразків серед усіх позитивних зразків. Специфічність (англ. specificity) - вимірює частку істинно-негативних зразків серед усіх негативних зразків. Між влучністю і повнотою, та чутливістю і специфічністю існує обернена залежність, коли можливо підвищити одну ціною зниження іншої. Через наявність такої залежності, на практиці, ці міри не використовуються окремо. Замість цього використовують агреговані метрики, що дозволяють оцінити якість класифікатора в цілому.

1.6.3. F1 та Dice score.

1.6.4. Cohen's Kappa.

1.6.5. Криві характеристик класифікатора: ROC и PR.

РОЗДІЛ 2

МЕТОД

2.1 Метод багатозадачного навчання штучних нейронних мереж для роботи в умовах частково помилкової розмітки навчального набору даних в задачах автоматизованого скринінгу (10 страниц) Использовать задачи, в каждой из которых разный уровень неуверенности Например, в задаче с шумной разметкой сегментации, наличие классов на картинке будет более точным, а сеть - более уверенной в предсказаниях наличия классов на изображении. В задаче классификации с шумной разметкой как классификации, так и сегментации, возможно обучение сети на участках с корректной разметкой и последующим дообучением на неразмеченных участках. Если задача может быть решена несколькими способами, каждый из этих способов может быть организован как отдельная задача, результаты решения которых будут объединены для получения совместного результата. Использование нескольких задач снижает чувствительность к неправильно размеченным примерам и повышает способность сети к генерализации

2.1.1 Архітектура штучної нейронної мережі для багатозадачного навчання Архитектура нейросети основана на Shared trunk подходе, так как используется один и тот же набор данных, и разніе задачи нужно решать для всех даних в єтом наборе. При необходимости, интеграция разных видов данных достигается через объединение представлений с нескольких энкодеров.

2.1.2 Задача класифікації зображень Использование задачи сегментации как структурированного внимания для регионов, содержащих необходимые для классификации признаки.

2.1.2 Задача семантичної сегментації зображень Задача сегментации мо-

жет быть рассмотрена как задача попиксельной классификации. /*Вывод условной вероятности при использовании двух классификаторов*/

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

2.3.1 Комбінування задачі класифікації та сегментації Использование классификации как маски для отсеечения малых ложноположительных регионов в сегментации

2.2 Метод зниження дисперсії прогнозів нейронної мережі за рахунок використання результатів семантично-близьких задач при багатозадачному навчанні 2.2.1 Вимірювання дисперсії прогнозів нейронної мережі 2.2.2 Зниження дисперсії в задачі класифікації 2.2.3 Зниження дисперсії в задачі семантичної сегментації

2.3.2 Напівавтоматичне навчання ШНМ в задачі сегментації Использование пикселей внутри маски как региона, по которому считаются логиты классификации При этом полезные для классификации регионы получают высокое значение в маске, а бесполезные - близкое к 0 Важно - использовать L2 регуляризацию, чтобы вся маска на стала единицами

2.3.3 Пост-обробка результатів сегментації Вихід маски від CNN безперервним в діапазоні $[0, 1)$. Оскільки декодер сегментації навчається без нагляду, істинний діапазон результатів сегментації неможливо передбачити заздалегідь. Також, в залежності від загального рівня активацій нейронної мережі, він може бути різним для різних вхідних зображень. Щоб полегшити калібрування передбачень моделі, використовується пост-обробка, що дозволяє бінарізувати результати сегментації нейронною мережею. Після того, як маска була бінаризована, до неї застосовується операція морфологічного розкриття квадратним ядром, щоб зменшити кількість малих хибнопозитивних областей.

2.4 Висновки до розділу 2

РОЗДІЛ 3

ЕКСПЕРИМЕНТИ

3.1. Класифікація стадії осередків діабетичної ретинопатії

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

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

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

Крім того, були використані менші набори даних: набір зображень індійської діабетичної ретинопатії (IDRiD) idrid, з якого використано 413 фотографій очного дна, та MESSIDOR (Методи оцінки методів сегментації та індексації в області офтальмології сітківки) messidor, з якого використано 1200 фотографій очного дна. Оскільки розмітка оригінального набору даних MESSIDOR відрізняється від інших наборів даних, ми використали версію, яку група офтальмологів messidor_relabeled помаркувала відповідно до стандартного протоколу.

Усі зазначені набори даних мають однаковий розподіл міток класів, що

є фундаментальною властивістю для цієї задачі.

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

3.1.2. Детальний опис структури нейронної мережі.

3.1.3. Детальний опис процедури навчання.

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

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

Для попереднього навчання, ініціалізована нейронна мережа навчається на протязі 20 епох на наборі даних DRDC2015 за допомогою стохастичного градієнтного спуску. Основна мета попереднього навчання - створити ініціалізацію параметрів на розподілі даних, що є близьким до цільового. Після попереднього навчання, параметри нейронної мережі використовуються як ініціалізація для основного навчання.

Під час попереднього навчання, кожен декодер мінімізує свою функцію втрат: перехресну ентропію для класифікаційного декодера, бінарну пере-

хресну ентропію для декодера порядкової регресії та середню абсолютну похибку для декодера регресії.

Основне тренування проводиться на наборах даних APTOS2019, IDR1D та MESSIDOR разом. Починаючи з ваг, отриманих на етапі попереднього тренування, виконується 5-кратна перехресна перевірка та оцінка моделі на відкладеному наборі даних. На цьому етапі функції втрат для декодерів змінено: Focal loss focalloss для класифікаційного декодера, бінарна Focal loss focalloss для декодера порядкової регресії та середньоквадратична помилка для декодера регресії.

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

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

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

Лінійна регресія навчається після основного навчання, оскільки в іншому випадку, сходиться до неоптимальних локальних мінімумів з вагами двох декодерів, близьких до нуля. Ці нульові ваги запобігають оновленню відповідних ваг декодерів і, відповідно, запобігають навчанню. Початкові ваги для кожного з виходів декодерів були встановлені рівними $1/3$, а потім тренувались протягом п'яти епох, щоб мінімізувати середньоквадратичну функцію помилки.

3.1.3.4. Результати експерименту.

3.2. Класифікація раку шкіри та сегментація родимок

3.2.1. Опис набору даних SIIM-ISIC Melanoma Classification.

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

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

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

Зображення та метадані надані у форматі DICOM, який є загальнозживаним форматом даних медичних зображень. Крім того, набір даних доступний у форматі JPEG із розмірами зображень, змінених до 1024x1024. Метадані також надаються за межами формату DICOM, у файлах CSV [12].

Метадані містять наступну інформацію:

- `image_name` - унікальний ідентифікатор, вказує на ім'я файлу відповідного зображення DICOM;
- `patient_id` - унікальний ідентифікатор пацієнта;
- `sex` - стать пацієнта;

- `age_approx` - приблизний вік пацієнта на момент фотографії (ціле число);
- `anatom_site_general_challenge` - розташування родинки на тілі;
- `diagnosis` - детальна інформація про діагностику (рядок);
- `benign_malignant` - показник злоякісності зображеного ураження (`cnhjrf`, одна з "benign" і "malignant");
- `target` - бінарізована версія цільової змінної.

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

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

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

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

3.2.2. Детальний опис структури нейронної мережі.

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

Навчання проводиться у 3-кратній схемі перехресної перевірки. Енкодер

відпочатку ініціалізується за допомогою вагів, отриманих за допомогою алгоритму [25]. У наших експериментах ми спостерігали, що така ініціалізація призводить до стабільно кращих результатів, ніж ініціалізація Imagenet.

3.2.4. Результати експерименту.

3.3. Класифікація та сегментація формацій хмар

3.3.1. 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 область один раз на день. Через обмежене поле зору камер, встановлених на цих супутниках, кожне зображення зшито з двох супутників, що знаходяться над однією й тою самою областю одночасно, але на різних орбітах. Решта зображення, для якої не було знято даних під час прольоту супутників (між орбітами) заповнена чорним кольором. Оскільки зображення підлягають компресії, є артефакти стиснення, та чорний колір має деякі незначні аномалії.

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

Розмітка для цих регіонів була створена під час краудсорсингу в Інституті метеорології імені Макса Планка в Гамбурзі, Німеччина, та Laboratoire de météorologie Dynamique у Парижі, Франція. Команда з 68 вчених виявила ділянки з хмарами на кожному зображенні, і кожне зображення було розмічене, в середньому, трьома різними вченими. Кожен вчений мав виділити хмари за допомогою прямокутних областей на власний розсуд. Основна розмітка була створена об'єднанням областей, що були розічені

всіма вченими для цього зображення, після видалення чорної смуги з цих областей [7].

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

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

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

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

3.3.2. Детальний опис структури нейронної мережі.

3.3.3. Результати експерименту.