

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

КОСТИРЯ МИХАЙЛО АНДРІЙОВИЧ

УДК 004.8

**СЕМАНТИЧНА СЕГМЕНТАЦІЯ З ВИКОРИСТАННЯМ ЗГОРТКОВОЇ
НЕЙРОННОЇ МЕРЕЖІ**

Галузь знань 12 «Інформаційні технології» за спеціальністю

122 «Комп'ютерні науки та інформаційні технології»

122 - ДР.А - 401.21610111

Автореферат

дипломної роботи на здобуття освітньої кваліфікації

«бакалавр комп'ютерних наук та інформаційних технологій»

Миколаїв – 2020

Дипломна робота є рукопис.

Робота виконана в Чорноморському національному університеті імені Петра Могили Міністерства освіти і науки України на кафедрі Інтелектуальних інформаційних систем

Науковий керівник:

д.т.н., професор, зав. кафедри ІС
Ю.П. Кондратенко

Рецензент:

к.т.н., доцент, доцент кафедри ІПЗ
А.В. Швед

Захист відбудеться «_22_» червня 2020 р. о 9³⁰ год. на засіданні екзаменаційної комісії (ауд. 2-403) у Чорноморському національному університеті імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

З дипломною роботою можна ознайомитися в бібліотеці Чорноморського національного університету імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

Автореферат представлений «___» червня 2020 р.

Секретар

екзаменаційної комісії,

викладач кафедри ІС

М.О. Таранов

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Дипломна робота присвячена обробці зображень, сфері, що користується попитом у сучасному світі як для охорони, так і для інформаційних цілей. Це, в першу чергу, обумовлено важливістю обробки даних комп'ютером з високою точністю, та подальшим використанням отриманої інформації. Правильно обрані методи та технології обробки інформації дозволять значно підвищити рівень контроль візуальної дати у сучасному світі.

Метою дипломної роботи є семантична сегментація зображень з використанням згорткової нейронної мережі.

Практичне значення отриманих результатів полягає у можливості застосування отриманої системи для обробки і аналізу дати для розпізнавання об'єктів на зображенні для подальшої обробки у практичних цілях.

Структура дипломної роботи. Пояснювальна записка до дипломної роботи складається із вступу, ____ розділів, висновків, додатків. Загальний обсяг роботи складає ____ сторінки, ____ рисунків, ____ таблиць та ____ посилань на літературні джерела.

ОСНОВНИЙ ЗМІСТ РОБОТИ

Одним із найбільш потужних і актуальних типів штучного інтелекту є комп'ютерний зір. Ця сфера фокусується на реплікації частин візуальної системи людини, надаючи комп'ютерам можливість ідентифікувати та обробляти об'єкти на зображеннях та відео. До недавніх часів, комп'ютерний зір мав обмежені можливості.

Завдяки нещодавнім відкриттям в області штучного інтелекту і машинного навчання, сфера комп'ютерного зору здобула значного розвитку і навіть перевершила людину у деяких завданнях, що стосуються детекції та класифікації об'єктів. Одним із основних факторів, що сприяли стрімкому розвитку комп'ютерного зору, є значна кількість дати, що ми генеруємо на сьогоднішній день.

Результат роботи такої системи є надзвичайно актуальним у наш час. Сфера сегментації зображень дуже широка і використовується у значній кількості напрямів. Одним із таких напрямів є створення візуальної репрезентації медичних зображень для подальшої обробки. За рахунок процесу сегментації, на зображенні можна виділити конкретні клітини, або області, що мають потенційний інтерес. Таким чином можна попередити рак, за рахунок класифікації ракових клітин. Майбутнє класифікації у сфері медицини лежить за комп'ютерами і використанням нейронних мереж.

Іншими словами - система, що здатна сегментувати зображення, є основним елементом для подальшого використання у інших напрямках. Це фундамент, на основі якого створюються подальші системи, що обробляють результати сегментації. Широка сфера застосування і актуальність на сьогоднішній день - причини такої створення системи, що описана у цій роботі.

Перший розділ.

Комп'ютерний зір - науковий розділ, що базується навколо надання комп'ютерам розуміння на основі цифрових зображень, чи відео [1]. З перспективи науковця, комп'ютерний зір має автоматизувати виконання задач, на які здатні візуальна система людини. Семантична сегментація зображень - один із ключових розділів комп'ютерного зору, процес поділу цифрового зображення на ряд сегментів (об'єктів зображення). Сегментацію можна описати як класифікацію зображень на піксельному рівні [2]. Основною метою сегментації є спрощення, або зміна представлення зображення для подальшого аналізу. В основному, сегментація зображень використовується для відокремлення об'єктів та границь (ліній, кривих) на зображенні. Іншими словами, це процес присвоєння ярликів до кожного пікселю зображення так, що пікселі під одним ярликом мають схожі характеристики.

Результатом сегментації зображення є набір сегментів, що покриває всю площу зображення, або набір контурів. Кожний піксель у регіоні має схожі характеристики на основі одного з критеріїв поділу, таких як колір, щільність, або текстура. Сусідні регіони значно відрізняються один від одного.

Семантична сегментація вивчалась на протязі багатьох років, була створена значна кількість дата сетів для верифікації алгоритмів. Сегментація зображень відіграла важливу роль у багатьох розділах комп'ютерного зору, таких як репрезентація сцени, слідкування за об'єктами у відео, та ряді медичних проблем.

На абстрактному рівні, задачу можна сформулювати наступним чином: “Маючи RGB кольорове (висота \times ширина \times 3), або чорно-біле (висота \times ширина \times 1) зображення, створити сегментаційну карту, що показує відносність кожного пікселю до певного класу (наприклад, людина, дерево, автомобіль)”.

Другий розділ.

Перш за все, варто пояснити вибір саме нейронної мережі для досягнення поставленої задачі, адже це далеко не єдиний спосіб аналізу даних. Нижче буде детально наведено ряд потенційних алгоритмів, що можуть бути використані для обробки зображення.

Розглянемо варіант використання звичайної нейронної мережі, у якій всі вузли будь-якого шару, з'єднані з усіма вузлами наступного. Можливість такого вибору зумовлена тим, що за останні 10 років нейронні мережі показали найкращі результати в області аналізу даних [16, 17]. Проте, якщо використати таку нейронну мережу для семантичної сегментації зображення, то для RGB кольорового зображення, на вхід будуть поступати сотні тисяч індивідуальних змінних. У комбінації з повністю з'єднаними наступними шарами, навіть найпростіша мережа буде складатись із мільйонів ребер. Більше того, швидше за все, така нейронна мережа не буде показувати відмінних результатів з точки зору точності і варіативності зображень. Це пов'язано з тим, що така значна кількість вузлів призводить не лише до обчислювальної складності, але і до перетренування на основі навчальних зображень. Результатом є нездатність мережі до адекватного сприйняття нових зображень, попередньо не оброблених даних, у чому нейронні мережі загалом показують відмінні результати. Крім цього, кількість вузлів при такій архітектурі є статичною, іншими словами, мережа не може приймати на вхід зображення різних розмірів, тому їх довелось би заздалегідь підганяти під необхідні простори, потенційно втрачаючи важливу інформацію.

Проте, невідповідність такої архітектури до вимог завдання не говорить про те, що використання нейронної мережі є недоречним в принципі. Насправді, існує багато типів нейронних мереж, кожен з яких краще підходить для вирішення певної задачі. Один із типів, що є більш доречним для розв'язання задач комп'ютерного зору, розглянуто нижче.

Згорткові мережі дуже схожі на звичайні мережі, вони також складаються з нейронів, ваг, та зміщення (bias). Мережа у цілому так само виражається оцінювальною функцією, на вхід приймається зображення, а на виході - ймовірності належності до класу. Значною різницею є те, що згорткова мережа (у випадку семантичної сегментації), заздалегідь передбачає використання зображення у якості вхідних даних. Це дозволяє надати архітектурі мережі певних особливостей, що значно покращують її загальну продуктивність. Нейрони згорткової мережі розташовані у трьох просторах (ширина, висота, глибина). Кожен шар трансформує 3D вхідні дані у 3D вихідні нейронні активації.

Надалі припускається що вхідними даними є RGB зображення. У такому разі, нейрони у рамках одного шару організовані у трьох просторах: ширина, висота, та глибина. Завдання кожного шару полягає у тому, щоб взяти на вході 3D дані, виконати певну функцію (у якій є похідна), та вивести нові 3D дані на виході.

Практика показує що власні архітектурні рішення стосовно комбінацій шарів не мають значного впливу на результат у 90% проблем. Замість створення власної архітектури для вирішення одиничної проблеми, рекомендується знайти найефективнішу на даний момент, завантажити натреновану модель, та підігнати її під свою дату. Необхідність у створенні власної згорткової мережі з нуля, або її тренуванні, постає дуже рідко.

Третій розділ.

Машинне навчання - комплексна дисципліна. Створення власних моделей було дуже трудомістким процесом у минулі часи. Проте, завдяки фреймворкам для машинного навчання, процеси отримання дати, тренування моделі, та отримання результатів, стали значно простішими. Одним із таких фреймворків є TensorFlow від Google.

TensorFlow поєднує у собі ряд алгоритмів, та моделей нейронних мереж. Для забезпечення зручного інтерфейсу використовується Python, в той час як виконання операцій проводиться на C++, що сприяє високій швидкості.

Написання та тренування глибоких нейронних мереж для класифікації написаних від руки чисел, упізнавання зображень, обробки людської мови, та багатьох інших, дуже полегшено завдяки TensorFlow. Більше того, TensorFlow надає можливість створення графів для демонстрації потоку даних. Кожний вузол графу репрезентує математичну операцію, в той час як кожне з'єднання - тензор.

Найбільшою вигодою використання TensorFlow є абстракція. Замість необхідності у імплементації власних алгоритмів низького рівня, або вирішення правильного шляху передачі вихідних даних однієї функції на вхід до іншої, розробник може повністю сконцентруватись на загальній логіці системи, у той час як TensorFlow бере на себе деталі реалізації.

Датасет є важливою складовою машинного навчання, але їх отримання, підключення, та адаптація до системи все ще є надто складним. Кожний дослідник має проходити через стадію написання скриптів для завантаження та підготовки даних для роботи, кожен з яких має різні формати та складності.

TensorFlow Datasets спрямований на полегшення роботи з даними, даючи доступ до відомих публічних датасетів, використовуючи зручний інтерфейс. Фреймворк бере на себе всю роботу по завантаженню даних, та їх збереженню на дисці у зручному форматі. Кожен датасет має власну версію, що забезпечує незмінність даних, та можливість репродукції результатів. Зі зміною дати, версія датасету інкрементується.

При роботі з будь-якою проблемою машинного навчання, завжди рекомендується провести певний час, намагаючись зрозуміти проблему. Це не тільки допомагає з пошуком необхідного забезпечення, але і сприяє мотивації розробника.

Для демонстраційних цілей буде використано датасет Oxford-IIIT Pet Dataset, та фреймворк TensorFlow. Датасет складається з зображень, відповідних підписів, та піксельних масок. Кожний піксель відноситься до однієї з 3-х категорій: клас

тварини, клас границі, та клас заднього фону. Дана сет заздалегіть належить до TensorFlow Datasets, так що все що необхідно - завантажити його. У якості архітектури нейронної мережі буде використано модифіковану U-Net. Модель складається з двох шляхів. Перший шлях - це шлях скорочення (також його називають кодером), який використовується для отримання контексту зображення. Кодер представляє собою звичайну CNN. Другий шлях - це симетричний шлях розширення (також його називають декодером), який використовується для забезпечення точної локалізації за допомогою транспонованих конволюцій.

Першим кроком проводиться аугментація даних, у даному випадку - звичайне горизонтальне віддзеркалення зображення. Це робиться для більш повного використання тренувальної дати.

Тренування буде виконуватись на протязі 20 епох. У результаті, з епохальним інкрементом результати передбачення стають більш точними.

Потенційно, збільшення кількості епох сприятиме покращенню точності результатів. Проте, до таких рішень варто відноситись з обережністю, адже це може сприяти перетренування мережі до конкретних тестових даних, що зробить її не ефективною при роботі з новими даними.

Розділ з охорони праці.

В спеціальному розділі з охорони праці було вивчено проблеми, пов'язані з забезпеченням здорових і безпечних умов, у яких відбувається праця людини – одне з найбільш важливих завдань у розробці нових технологій і систем виробництва. Дослідження й виявлення можливих причин виробничих нещасних випадків, професійних захворювань, аварій, вибухів, пожеж, і розробка заходів і вимог, спрямованих на усунення цих причин дозволяють створити безпечні й сприятливі умови для праці людини. Комфортні й безпечні умови праці – один з основних факторів, який впливає на продуктивність і безпеку праці, здоров'я працівників.

ЗАГАЛЬНІ ВИСНОВКИ

Під час написання дипломної роботи було проаналізовано існуючі методи, алгоритми й реалізації систем семантичної сегментації. Після чого було проведено порівняльну характеристику кожного й визначено їх переваги і недоліки.

Було вивчено формулювання загальної постанови і проведення порівняльного аналізу використовуваних підходів і методів вирішення завдання класифікації об'єктів на зображенні.

Розглянуто в роботі:

- опис декількох методів вирішення задачі, огляд їх переваг і недоліків;
- опис формалізації обраного методу у вигляді алгоритму і реалізації у вигляді програми;
- опис методики і результатів проведення комп'ютерного експерименту, що дозволяє виявити недоліки і оцінити можливості застосування реалізованого методу в процесі аналізу і синтезу інформаційних систем.

В результаті виконання дипломної роботи вирішено наступні завдання:

- проаналізовано сучасний стан задачі комп'ютерного зору та сегментації зображень;
- проведено огляд існуючих технологій для вирішення поставленої задачі;
- реалізовано згорткову нейронну мережу з архітектурою U-Net для семантичної сегментації зображень;
- протестовано розроблений програмний застосунок.

Було засвоєно та розібрано класифікацію зображення на піксельному рівні, як завдання комп'ютерного зору. Ця задача може бути сформульована як створення сегментаційної карти, де кожен піксель відповідає певному класу. У реальній системі пікселю може відповідати клас людини, дороги, тощо.

АНОТАЦІЯ

до бакалаврської роботи

Тема: «Семантична сегментація з використанням згорткової нейронної мережі»

Студент: Костира Михайло Андрійович

Керівник: д.т.н, професор Кондратенко Юрій Пантелійович

Дипломна робота присвячена розробці та здійсненню програмної реалізації системи для семантичної сегментації зображень.

Об'єктом дослідження є процес семантичної сегментації зображення.

Предметом дослідження є згорткові нейронні мережі для семантичної сегментації зображень.

Метою дослідження є удосконалення процесу семантичної сегментації, шляхом використання згорткової нейронної мережі.

Дипломна робота складається з фахової частини і спеціальної частини з охорони праці. Пояснювальна записка дипломної роботи складається зі вступу, трьох розділів та висновків.

У першому розділі розкрито важливість систем комп'ютерного зору у сучасному світі.

У другому розділі описуються існуючі технології та алгоритми, також виконано порівняння методів з метою виявлення недоліків.

У третьому розділі описано проектування та програмну реалізацію розробленої системи. Описано створений додаток і результати його тестування.

У четвертому розділі описана частина з охорони праці.

Дипломна робота містить ___ сторінок, ___ рисунків, ___ джерел, ___ додатків.

Ключові слова: комп'ютерний зір, штучний інтелект, семантична сегментація, нейронна мережа.

ABSTRACT
for bachelor's work

Subject: “Semantic segmentation using a convolutional neural network”

Student: Kostyria Mykhailo Andriyovich

Leader: D.Sc., Professor Kondratenko Yuri Panteliyovych

Thesis is devoted to the development of semantic segmentation system using artificial methods.

The object of research is the process of semantic segmentation of an image.

The subject of research is convolutional neural networks for semantic segmentation.

The purpose of the thesis is improving the process of semantic segmentation using a convolutional neural network.

Thesis consists of a professional part and a special part on labor protection. The explanatory note of the thesis consists of an introduction, three sections and conclusions.

The first section reveals the importance of semantic segmentation systems in the modern world.

The second section describes existing technologies and algorithms, and compares methods to identify shortcomings.

The third section describes the created application and the results of its testing.

The fourth section describes the part of labor protection.

Diploma contains ___ pages, ___ drawings, ___ sources, ___ applications.

Key words: computer vision, artificial intelligence, semantic segmentation, neural network.