

Харківський національний університет радіоелектроніки

Факультет навчально-науковий центр заочної форми навчання

Кафедра електронних обчислювальних машин

Рівень вищої освіти другий (магістерський)

Спеціальність 123 – Комп'ютерна інженерія
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Комп'ютерні системи та мережі
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ _____ ” _____ 20__ р.

ЗАВДАННЯ

НА АТЕСТАЦІЙНУ РОБОТУ

студентові Мазуру Федору Леонідовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Моделі семантичної сегментації об'єктів міського середовища на цифрових аерофотзнімках

затверджена наказом по університету від “ 23 ” жовтня 2020 р. № 168 Стз

2. Термін подання студентом роботи до екзаменаційної комісії 14 грудня 2020 р.

3. Вхідні дані до роботи _____

1. Глибокі нейронні мережі в CV

2 Вимоги до якості роботи глибоких нейронних мереж при вирішенні задач CV

3 База даних цифрових аерофотознімків

4. Перелік питань, що потрібно опрацювати в роботі _____

Аналіз предметної області

Формальне представлення процесу сегментації цифрових зображень

Обчислювальний експеримент

Розробка практичної рекомендації

Висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 12 слайдів

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз сучасного стану проблеми.	27.10.20 – 05.11.20	
2	Постановка задачі дослідження	06.11.20 – 10.11.20	
3	Аналіз літератури	11.11.20 – 18.11.20	
4	Формальне представлення процесу сегментації цифрових зображень	19.11.20 – 03.12.20	
5	Експеримент і аналіз отриманих результатів	04.12.20 – 09.12.20	
6	Оформлення пояснювальної записки	10.12.20 – 13.12.20	

Дата видачі завдання 26 жовтня 2020 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

проф. Олізаренко С.А.
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка атестаційної роботи: 70 с., 20 рис., 2 табл., 2 дод., 36 джерел.

МОДЕЛЬ, СЕГМЕНТАЦІЯ, НЕЙРОННА МЕРЕЖА, ГЛИБОКЕ НАВЧАННЯ, ЦИФРОВИЙ ЗНІМОК, ЗОБРАЖЕННЯ.

Метою атестаційної роботи є дослідження та модифікація моделей семантичної сегментації об'єктів міського середовища з використанням методології глибокого навчання.

В ході дослідження проведено аналіз відомих методів для автоматизації виявлення так класифікації об'єктів на цифрових знімках. Встановлено, що методи, які реалізують піксельно-орієнтований підхід доцільно застосовувати для сегментації об'єктів на цифрових знімках через їх більшу точність. В роботі виконано модифікацію моделі U-Net. У модифікованій моделі U-Net була випробувана технологія "battlenack" в поєднанні з технологією «transfer learning» що дозволило зменшити валідаційну помилку до 7 – 8%.

ABSTRACT

Master's thesis: 70 pages, 20 figures, 2 table, 2 appendixes, 36 sources.

MODEL, SEGMENTATION, NEURAL NETWORK, DEEP LEARNING,
DIGITAL IMAGE, IMAGE.

The purpose of the attestation work is to study and modify the models of semantic segmentation of urban objects using the methodology of deep learning.

The study analyzed known methods for automating the detection of classification of objects in digital images. It has been found that methods that implement a pixel-oriented approach should be used to segment objects in digital images due to their greater accuracy. The work modifies the U-Net model. In the modified U-Net model, the "battlenack" technology was tested in combination with the "transfer learning" technology, which allowed to reduce the validation error to 7 – 8%.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	8
ВСТУП	9
1 АНАЛІЗ АВТОМАТИЗАЦІЇ ПРОЦЕСІВ ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЇ ОБ'ЄКТІВ МІСЬКОГО СЕРЕДОВИЩА НА ЦИФРОВИХ АЕРОФОТОЗНІМКАХ.....	11
1.1 Глибоке навчання. Терміни та поняття.....	11
1.1.1 Переваги використання методів глибокого навчання.....	17
1.2 Аналіз змісту процесів виявлення так класифікації об'єктів на цифрових знімках.....	20
1.3 Аналіз основних підходів щодо виявлення так класифікації об'єктів на цифрових знімках	24
1.4 Дослідження топології згорткової нейронної мережі для обробки цифрових зображень.....	27
2 ДОСЛІДЖЕННЯ ПІДХОДІВ ЩОДО СЕГМЕНТАЦІЇ ОБ'ЄКТІВ НА ЦИФРОВИХ ЗНІМКАХ З ВИКОРИСТАННЯМ МЕТОДОЛОГІЇ ГЛИБОКОГО НАВЧАННЯ.....	36
2.1 Дослідження основних топологій глибоких нейрон мереж для сегментації об'єктів на цифрових знімках	36
2.2 Оцінка необхідної кількості навчальних прикладів вибірки зображень для навчання ЗНМ.....	45
2.3 Визначення загальних положень процедури навчання ЗНМ	46
3 МОДЕЛЮВАННЯ СЕГМЕНТАЦІЇ ОБ'ЄКТІВ МІСЬКОГО СЕРЕДОВИЩА НА ЦИФРОВИХ АЕРОФОТЗНІМКАХ	50
3.1 Розробка інформаційної технології автоматизованого виявлення та класифікації простих об'єктів оптико-електронної повітряної розвідки	50

3.2 Обґрунтування і формування репрезентативної навчальної вибірки при малому обсязі даних.....	52
3.3 Опис експерименту	55
ВИСНОВКИ.....	58
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	60
ДОДАТОК А Графічний матеріал атестаційної роботи	63
ДОДАТОК Б Топологія модифікованої нейромережевої моделі U-Net для сегментації об'єктів на цифрових знімках.....	70

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ
І ТЕРМІНІВ

АРМ – автоматизоване робоче місце

ЗНМ – згорткова нейронна мережа

ІТ – інформаційна технологія

CUDA – Compute Unified Device Architecture

DIGITS – Deep Learning GPU Training System

IMU – Inertial Measurement Unit

JSIPS – Joint Service Imagery Processing System

GPS – Global Positioning System

GPU – процесор графічного відеоадаптера

ВСТУП

Аналіз способів обробки цифрових аерофотознімків міського середовища показує, що на даний час виявлення та класифікація (розпізнавання) об'єктів міського середовища на аерофотознімках виконується неавтоматизовано, а саме шляхом залучення операторів-дешифрувальників. При цьому оперативність та якість результатів виявлення та класифікації об'єктів значною мірою залежить від знань та умінь оператора-дешифрувальника [1-5, 14-23]. Наявність великої інтенсивності вхідних даних, фізична та психічна завантаженість оператора-дешифрувальника призводять до неможливості своєчасної та якісної обробки цифрових аерофотознімків міського середовища.

Таким чином, в теперішній час виникає протиріччя. З однієї сторони необхідно забезпечити своєчасну обробку даних з заданою якістю. А з іншої відсутні ефективні методи вирішення задач виявлення та класифікації об'єктів на цифрових знімках земної поверхні, що забезпечують якість та оперативність видачі не нижчу ніж у операторів-дешифрувальників. Одним з основних етапів виявлення та класифікації об'єктів на цифрових знімках земної поверхні є сегментація зображень.

Для вирішення вказаного протиріччя у роботі сформульовано та вирішено завдання, яке полягає в дослідженні моделей семантичної сегментації об'єктів міського середовища для підвищення якості та оперативності обробки даних на цифрових аерофотознімках міста.

На даний час найкращі результати з сегментації цифрових знімків показують згорткові нейронні мережі (ЗНМ), які мають властивості часткової стійкості до зміни розмірів, зміщення, повороту об'єкта на знімках та паралельного обчислення для обробки зображень в режимі часу, наближеного до реального. Тому застосування інтелектуальних знання-орієнтованих методів обробки зображень на основі об'єктно-орієнтованого

підходу, які дозволяють підвищити оперативність обробки даних з заданою якістю, є доцільним.

Мета і завдання дослідження. Метою роботи є підвищення якості та оперативності виявлення та класифікації об'єктів міського середовища на цифрових аерофотзнімках на основі застосування сучасних моделей семантичної сегментації.

Для досягнення мети в роботі поставлені і вирішені наступні задачі:

- проведено аналіз напрямків автоматизації процесів виявлення та класифікації простих об'єктів міського середовища на цифрових аерофотзнімках;

- визначено підхід щодо сегментації об'єктів міського середовища на цифрових аерофотзнімках;

- визначено підхід щодо формування навчальної вибірки зображень для навчання ЗНМ сегментації об'єктів міського середовища;

- проведено моделювання процесів виявлення та класифікації об'єктів міського середовища з використанням ЗНМ для семантичної сегментації.

Об'єкт дослідження – процеси автоматизованого виявлення та класифікації об'єктів міського середовища з використанням ЗНМ для семантичної сегментації.

Предмет дослідження – моделі та методи виявлення та класифікації об'єктів міського середовища з використанням ЗНМ для семантичної сегментації.

1 АНАЛІЗ АВТОМАТИЗАЦІЇ ПРОЦЕСІВ ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЇ ОБ'ЄКТІВ МІСЬКОГО СЕРЕДОВИЩА НА ЦИФРОВИХ АЕРОФОТОЗНІМКАХ

1.1 Глибоке навчання. Терміни та поняття

Нейронна мережа у якомусь сенсі є спробою відтворення роботи людського мозку на комп'ютері за допомогою шарів нейронів.

Штучний інтелект - здатність машини чи програми знаходити рішення за допомогою обчислень. Під час перших досліджень в області штучного інтелекту вчені намагалися відтворити людський інтелект для вирішення конкретних завдань: наприклад, гри з людиною. Було введено велику кількість правил, яким має слідувати комп'ютер. На основі цих правил комп'ютер приймав рішення погоджено з конкретним списком можливих дій.

Машинне навчання - це спроба навчити комп'ютери самостійно навчатися на великій кількості даних замість жорстко постульованих правил. Машинне навчання дозволяє комп'ютерам самостійно навчатися. Це можливо завдяки обчислювальній потужності сучасних комп'ютерів, які можуть легко обробляти великі набори даних.

Контрольоване навчання (навчання з учителем, supervised learning) має на увазі використання помічених наборів даних, що містять вхідні дані і очікувані вихідні результати. Під час тренування нейронної мережі за допомогою контрольованого навчання, подаються як вхідні дані, так і очікувані вихідні результати. Якщо результат, що генерується нейронною мережею, є помилковим, вона скоректує свої обчислення.

Це ітераційний процес, що закінчується тоді, коли мережа перестає робити помилки. Прикладом завдання з контрольованим навчанням є передбачення погоди. Нейромережа вчиться робити прогноз погоди з використанням історичних даних. Навчальні дані включають в себе вхідні

дані (тиск, вологість, швидкість вітру) та вихідні результати (температура).

Неконтрольоване навчання (навчання без учителя, *unsupervised learning*) - це машинне навчання з використанням наборів даних без певної структури. При тренуванні нейромережі неконтрольовано, вона самостійно проводить логічну класифікацію даних. Прикладом завдання з неконтрольованим навчанням є передбачення поведінки відвідувачів інтернет-магазинів. У цьому випадку мережа не навчається на розмічених даних. Замість цього вона самостійно класифікує вхідні дані і відповідає на питання, які користувачі найчастіше купують різні товари.

Глибоке навчання - це один з методів машинного навчання. Глибоке навчання (рисунок 1.1) дозволяє навчати модель передбачати результат по набору вхідних даних. Для навчання мережі можна використовувати як контрольоване, так і неконтрольоване навчання.

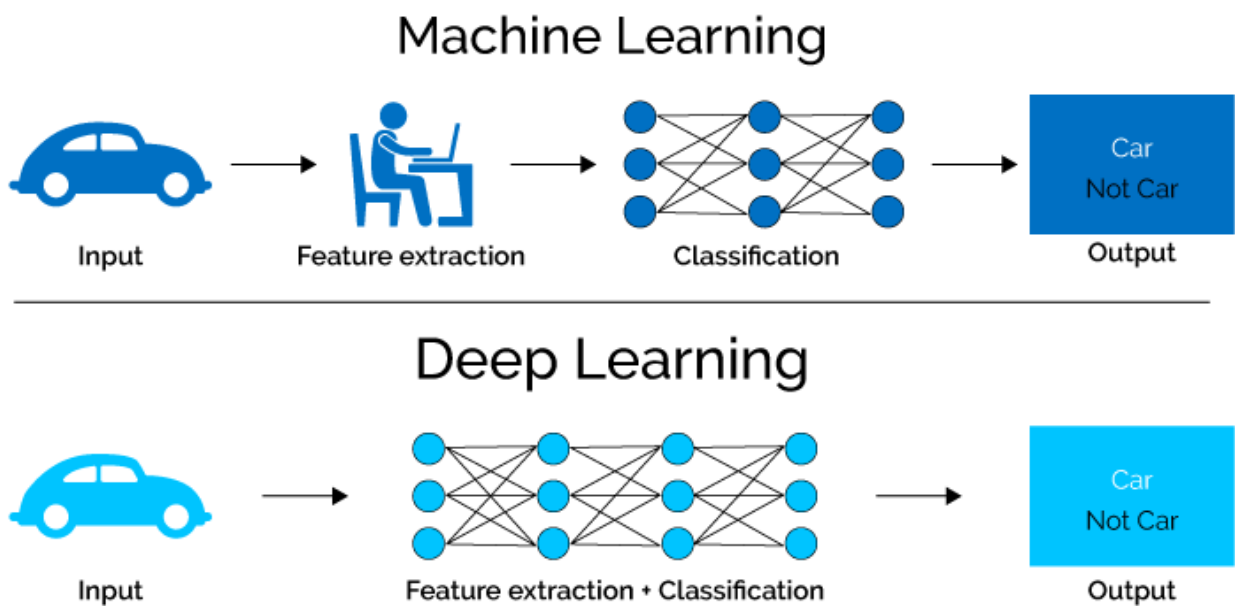


Рисунок 1.1 – Порівняння машиного навчання та глибокого

Розглянемо роботи глибокого навчання на прикладі роботи сервісу з оцінки вартості авіаквитка. Покажемо процес з використанням контрольованого навчання. Слід зауважити, що по завданню сервіс

передбачав ціну на авіаквиток за такими вхідними даними:

- аеропорт вильоту;
- аеропорт призначення;
- дата відбуття;
- авіакомпанія.

Нейронна мережа для вирішення такого завдання представлена на рисунку 1.2.

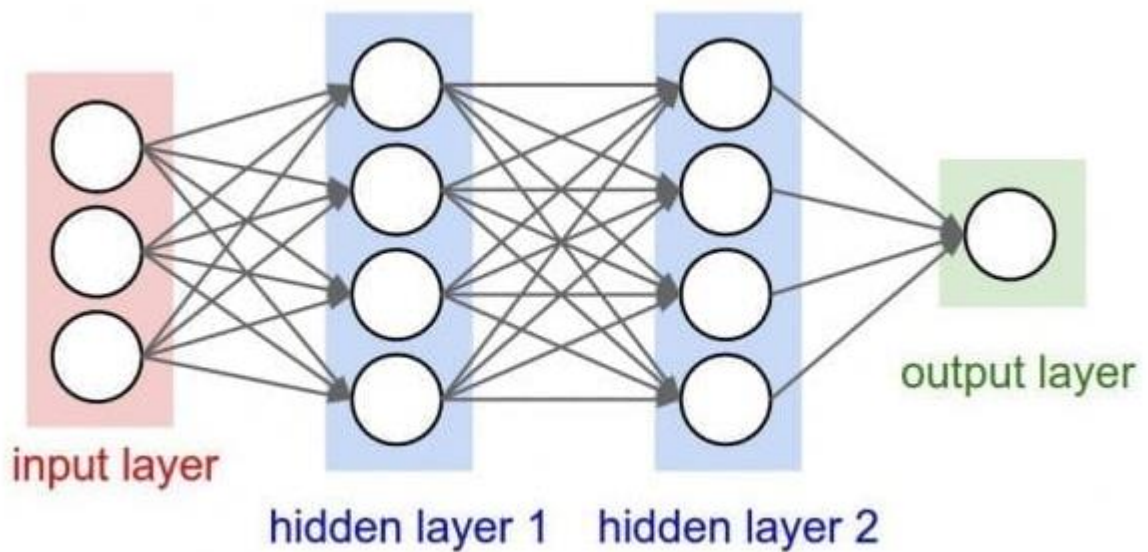


Рисунок 1.2 – Глибока нейронна мережа з двома прихованими шарами

Нейрони згруповані в три різних типи шарів:

- вхідний шар;
- прихований шар (шари);
- вихідний шар.

Вхідний шар приймає вхідні дані. У нашому випадку є чотири нейрони на вхідному шарі: аеропорт вильоту, аеропорт призначення, дата вильоту і авіакомпанія. Вхідний рівень передає ці дані в перший прихований шар. Приховані шари виконують математичні обчислення з вхідними даними.

Одне із завдань при створенні нейронних мереж - визначення кількості прихованих шарів і нейронів у кожному шарі. Слово «глибина» в терміні

«глибоке навчання» означає наявність більш ніж одного прихованого шару.

Вихідний шар видає результат. У нашому випадку це прогноз ціни на квиток.

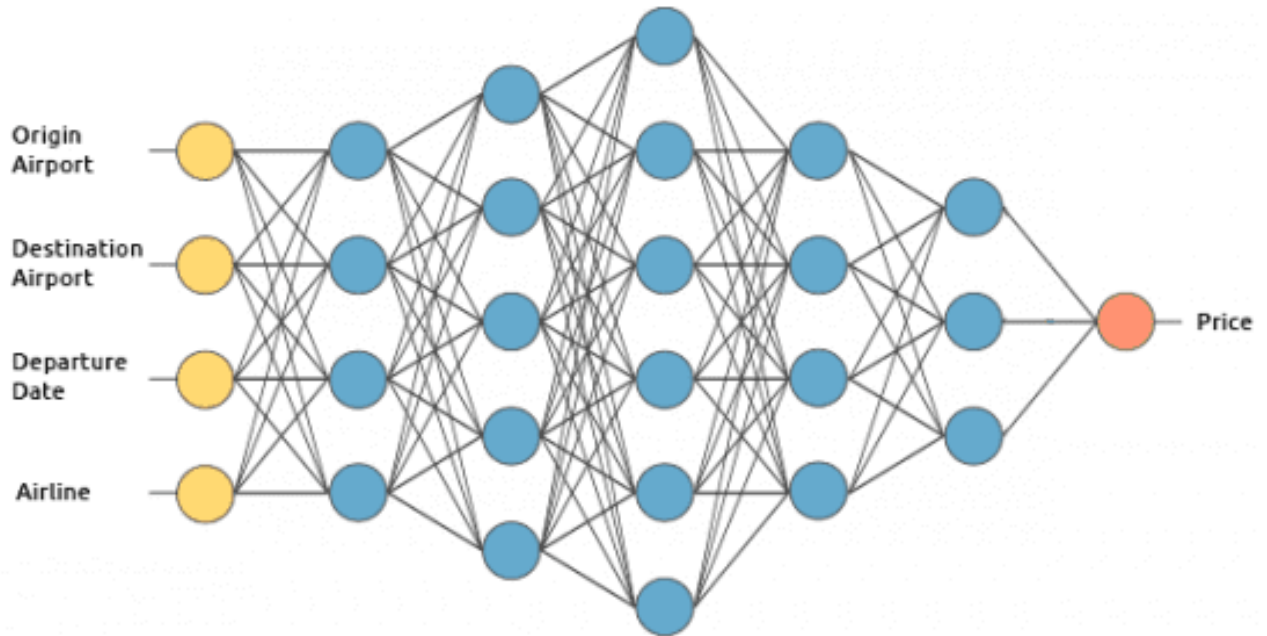


Рисунок 1.3 – Вирішення завдання з використанням глибокої нейронної мережі з двома прихованими шарами

Нейрони пов'язані між собою з певною вагою. Вага визначає важливість елемента вхідних даних. Вихідні ваги задаються випадковим чином. При прогнозуванні ціни на квиток «дата вильоту» є одним з найбільш важливих факторів.

Отже, зв'язки нейрона «часу вильоту» матимуть велику вагу.

Кожен нейрон має функцію активації. Її сенс важко зрозуміти без залучення математичних міркувань. Однією з її цілей є «стандартизація» даних на виході з нейрона.

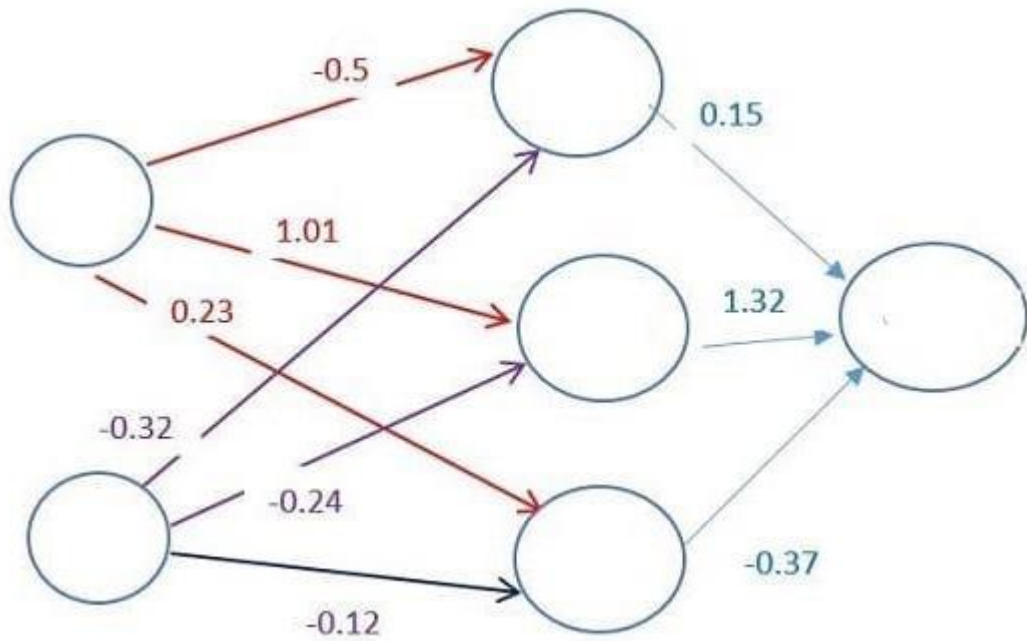


Рисунок 1.3 – Зв'язки нейронів з ваговими коефіцієнтами

Після того, як набір вхідних даних пройшов через всі шари нейронної мережі, функція активації повертає вихідні результати через вихідний рівень.

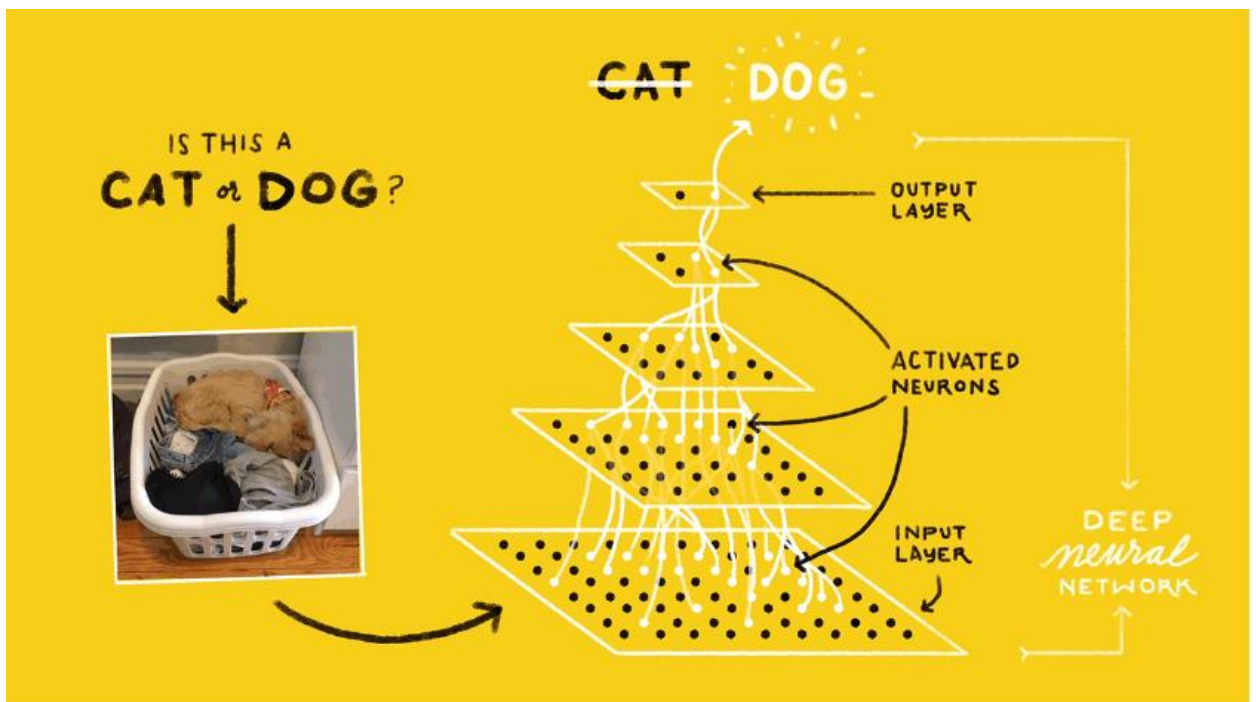


Рисунок 1.4 – Принцип роботи методу глибокого навчання

Для оцінки вартості квитка потрібно знайти історичні дані про ціни на квитки. Через велику кількість можливих комбінацій аеропортів і дат вильоту потрібен дуже великий список цін на квитки. Для навчання мережі потрібно подати в ній підготовлені дані і порівняти згенеровані їй вихідні результати з результатами з нашого тестового набору даних. Оскільки мережа ще не навчена, результати будуть невірними. Після пропуску всіх даних можна визначити функцію, яка буде показувати нам, наскільки результати роботи алгоритму відрізняються від реальних даних. Ця функція називається функцією втрат. В ідеалі має бути так, щоб функція втрат дорівнювала нулю. У цьому випадку вихідні результати роботи мережі повністю збігаються з результатами тестового набору даних.

Для того щоб зменшити значення функції втрат потрібно міняти ваги між нейронами. Можна робити це випадковим чином до тих пір, поки функція втрат не стане рівною нулю, але це не дуже ефективно. Замість цього ми пропонується використовувати метод градієнтного спуску (рисунок 1.5).

Метод найшвидшого спуску - це метод, який дозволяє знайти мінімум функції.

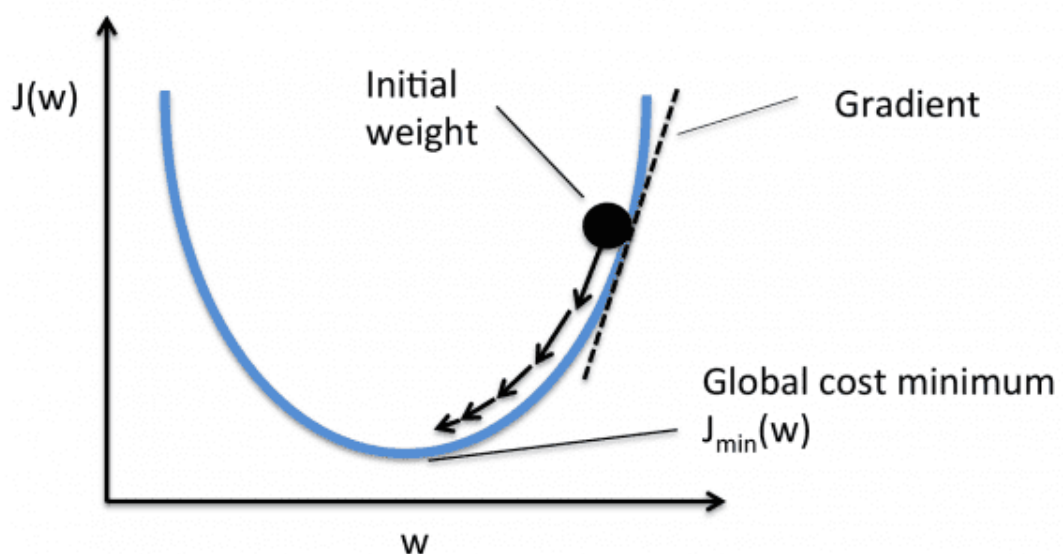


Рисунок 1.4 – Метод градієнтного спуску

У випадку поточного завдання іде пошук мінімуму функції втрат. Суть методу полягає в тому, щоб трохи зменшувати ваги після кожної ітерації. Обчислюючи похідну (або градієнт) функції втрат при певному наборі ваг, можна визначити, в якому напрямку знаходиться мінімум.

Для мінімізації функції втрат потрібно багаторазово перебирати дані. Саме тому потрібна велика обчислювальна потужність. Уточнення ваг за допомогою градієнтного спуску виконується автоматично. В цьому і полягає метод глибокого навчання.

Після навчання можна використовувати сервіс для прогнозування цін на авіаквитки.

Таким чином, глибоке навчання використовує нейронну мережу для відтворення інтелекту людей/тварин. Існує три типи шарів нейронів в нейронній мережі: вхідний шар, прихований шар (шари) і вихідний шар. Зв'язки між нейронами мають ваги, які визначають важливість елемента вхідних даних. Для навчання глибокої нейронної мережі необхідний реально великий набір даних. Ітераційно порівнюючи вихідні результати з включеними в набір даними, можна обчислити функцію втрат, що вказує, наскільки сильно помиляється алгоритм. Після кожної ітерації (epoch) ваги між нейронами перерозподіляються за допомогою методу градієнтного спуску для мінімізації функції втрат.

1.1.1 Переваги використання методів глибокого навчання

Незважаючи на те, що глибоке навчання демонструє високу продуктивність, для ряду конкретних ситуацій використання класичного машинного навчання все ще краще. Тут маються на увазі завдання, для вирішення яких необхідно щось на зразок лінійної регресії або дерева рішень, а не велика глибока мережа. Отже, перевагами є:

- краща в своєму класі продуктивність. Глибокі мережі досягли точності, яка сильно перевершує класичні методи машинного навчання в

багатьох областях, включаючи мову, природню мову, комп'ютерне бачення та інші. У багатьох задачах класичне машинне навчання навіть не може скласти конкуренцію. На наведеному нижче графіку (рисунок 1.6) показана точність класифікації зображень різними методами на вибірці даних ImageNet. Синій колір вказує на кількість помилок класичних методів машинного навчання, а фіолетовий колір - на кількість помилок методу глибокої згорткової нейронної мережі (CNN). З рисунку видно, що глибоке навчання в рази перевершує класичне машинне. А в 2015 році нейронна мережа зробила навіть менше помилок, ніж людина (відзначимо, що в групі випробовуваних було всього 5 осіб).

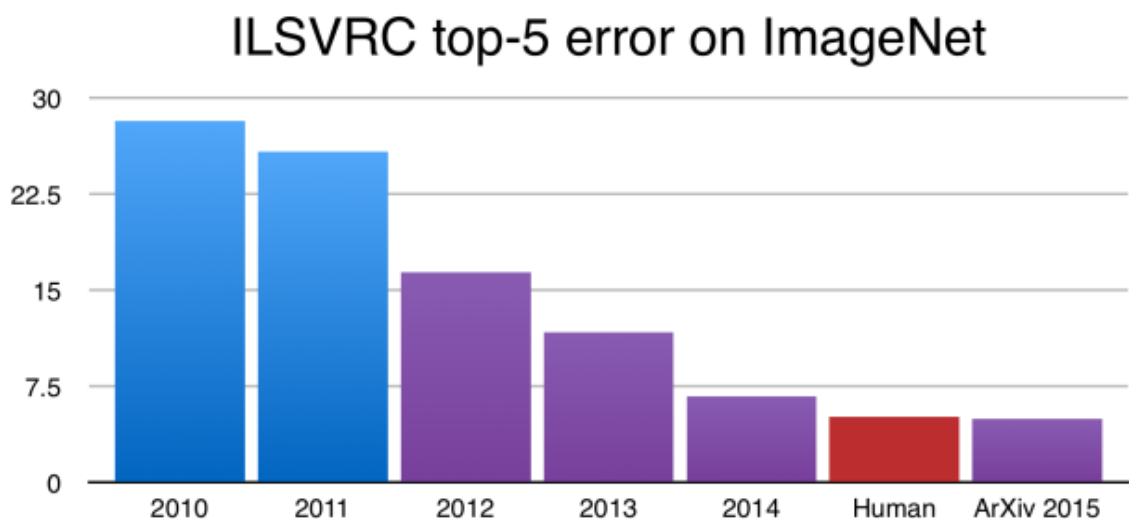


Рисунок 1.6 – Точність класифікації зображень різними методами

- **Маштабованість.** Глибокі мережі краще справляються з ростом кількості даних, ніж класичні алгоритми. Представлений нижче графік (рисунок 1.7) наочно це ілюструє. В даному випадку для підвищення точності необхідні більш складні методи.

- **Розробка функцій не потрібна.** Класичні алгоритми часто вимагають складної розробки функцій. Зазвичай глибокий аналіз даних спочатку виконується на наборі даних. Потім для спрощення процесу обробки можна

зменшити розмірність. Нарешті, для переходу до алгоритму відбираються найкращі функції. У глибокій мережі такого немає, оскільки дані передаються безпосередньо в мережу і, як правило, хороші результати виходять з самого початку. Таким чином, повністю виключається великий і складний етап розробки функцій для всього процесу.

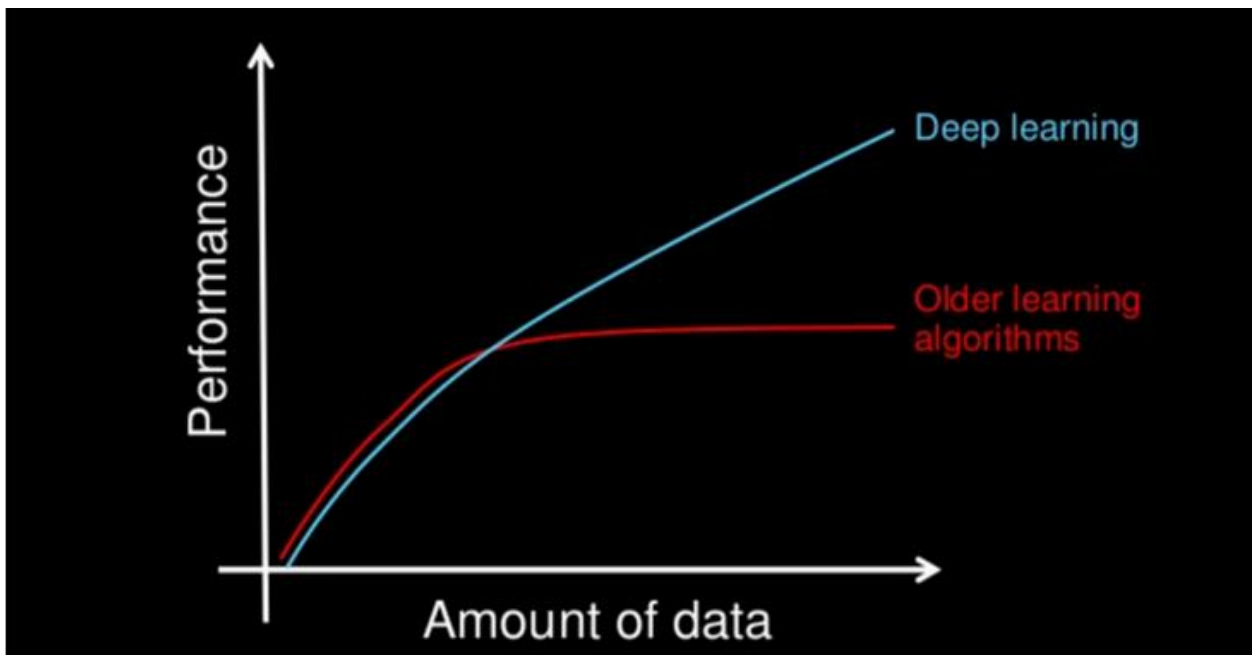


Рисунок 1.7 – Переваги глибокого навчання

- Адаптація та застосування у суміжних областях. Методи глибокого навчання набагато легше адаптувати до різних областей, ніж класичні алгоритми машинного навчання. По-перше, передача напрацьованих навичок навчання (transfer learning) дозволяє використовувати заздалегідь підготовлені глибокі мережі для різних завдань в межах однієї області. Наприклад, в комп'ютерному зорі, заздалегідь підготовлені мережі для класифікації зображень часто використовуються в якості попереднього етапу в процесі виявлення об'єктів і проведенні сегментації. Використання попередньо підготовлених мереж полегшує навчання всієї моделі і часто допомагає досягти більш високої продуктивності за більш короткий період часу. Крім того, основні ідеї та методи глибокого навчання часто легко

переносяться в іншу область. Наприклад, як тільки хтось створює теорію глибокого навчання для розпізнавання мови, вивчати, як застосовувати глибокі мережі для обробки природної мови, стає простіше, оскільки базові знання досить схожі. У класичному машинному навчанні все інакше: для створення високопродуктивних моделей потрібні специфічні для конкретної області методи і розробка функцій. База знань класичного машинного навчання для різних областей і додатків абсолютно різна і часто вимагає великого спеціалізованого вивчення в кожній конкретній області.

1.2 Аналіз змісту процесів виявлення так класифікації об'єктів на цифрових знімках

Застосування автоматизованих систем для моніторингу міського середовища на цифрових аерофотзнімках дозволяє досягти якісно нового рівня, підвищити ефективність за рахунок впровадження сучасних ІТ (методів, систем і засобів) отримання, передавання, збирання, обробки, зберігання і використання інформації.

Всі цифрові методи обробки даних в інтересах виявлення та класифікації об'єктів міського середовища на цифрових аерофотзнімках можливо розділити на дві групи [34]:

- попередня обробка – направлена на покращення зображень для їх дешифрування за рахунок подавлення шумів, балансування кольорів, конвертацію у потрібні формати та ін.;

- тематична обробка (дешифрування) – направлена на отримання отримання інформації про об'єкти місцевості по їх зображеннях на знімках.

Цифрові методи першої групи в роботі не розглядаються. Основна увага в роботі буде приділена тематичній обробці (дешифруванню). Дешифрування базується на знаннях закономірностей відтворення властивостей об'єктів на зображеннях, особливостей їх спектральних характеристик (сигнатур), геометричної форми, розміщення на місцевості.

Процедурно дешифрування включає до себе виявлення, класифікацію та інтерпретацію об'єктів і місцевості.

Під виявленням розуміється процес або результат пошуку та сприйняття об'єктів на початковому етапі дешифрування знімків (зображень). При цьому на зображенні локалізуються ділянки, де найбільш імовірно розташовані об'єкти.

Під класифікацією розуміється процес машинного аналізу об'єктів (образів) та їх ознак на зображеннях з метою визначення належності об'єкта до певного класу.

Етап інтерпретації, це заключний етап дешифрування, в процесі якого проводяться аналіз і узагальнення кількісних та якісних характеристик (ознак) об'єкта з метою встановлення його стану, значущості та можливостей в конкретній обстановці.

Етапи виявлення та класифікації цифрових знімків здійснюють сортування пікселів зображення на класи, визначені у алфавіті класів, за значеннями їх ознак. Виявлення та класифікація виконується за рахунок статистичного аналізу просторових, спектральних та часових характеристиках знімку.

Результатом виконання етапів виявлення та класифікації в процесі тематичної обробки даних є дані про клас об'єкта та його місцезнаходження.

Розрізняють піксельно-орієнтовані та об'єктно-орієнтовані підходи до автоматизації процесу виявлення та класифікації об'єктів міського середовища на цифрових аерофотзнімках.

Піксельно-орієнтований підхід передбачає групування пікселів знімку у класи за інформацією про спектральну яскравість в кожному пікселі знімку. Об'єктно-орієнтований підхід передбачає групування пікселів знімку у класи з урахуванням спектральних особливостей, форми та структури об'єктів виявлення та класифікація яких здійснюється.

Піксельно-орієнтований підхід є більш доцільним з точки зору виявлення та класифікації об'єктів міського середовища на цифрових аерофотзнімках.

Під класифікацією простих об'єктів мається на увазі процедура віднесення об'єкта, зображеного на цифровому знімку, до певного класу. Класифікація можлива лише у випадках, коли кількість класів обмежена скінченним числом N_{class} , а кожний об'єкт може бути віднесений тільки до одного класу з усієї сукупності класів N_{class} , та існує апіорна інформація про об'єкти в класах N_{class} , за якою однозначно можливо розділити об'єкти на класи. За таких умов та вихідних даних класифікація об'єктів на цифрових аерофотознімках зводиться до віднесення об'єкта з деякою ймовірністю до деякого з класів N_{class} . Тобто вихідними даними процедури класифікації є вектор значень ймовірностей того, що об'єкт належить до конкретного класу. Кількість елементів вектора ймовірностей завжди дорівнює кількості класів N_{class} .

Завдання класифікації об'єктів на зображеннях входить в клас завдань з розпізнавання образів, які вирішуються в рамках розробки інтелектуальних ІТ. Всі завдання класифікації відповідають моделі класифікації. Вона складається з трьох частин – датчика, блока виділення ознак та класифікатора (рисунки 1.8).

Датчик приймає дані з навколишнього середовища і перетворює їх до вигляду, зручного для обробки.

Блок виділення ознак перетворює вхідні дані на вектор ознак. Під ознакою розуміється кількісна оцінка об'єкта довільної природи. Сукупність ознак, які відносяться до одного об'єкта, утворюють вектор ознак. Вектори ознак зазвичай приймають значення у багатовимірному просторі ознак. Кожному об'єкту відповідає лише один унікальний вектор ознак, і, навпаки, кожному вектору ознак відповідає лише один об'єкт. Класифікатор інколи називають вирішальним правилом. Він на основі вектору ознак відносить об'єкт до одного з декількох класів N_{class} .

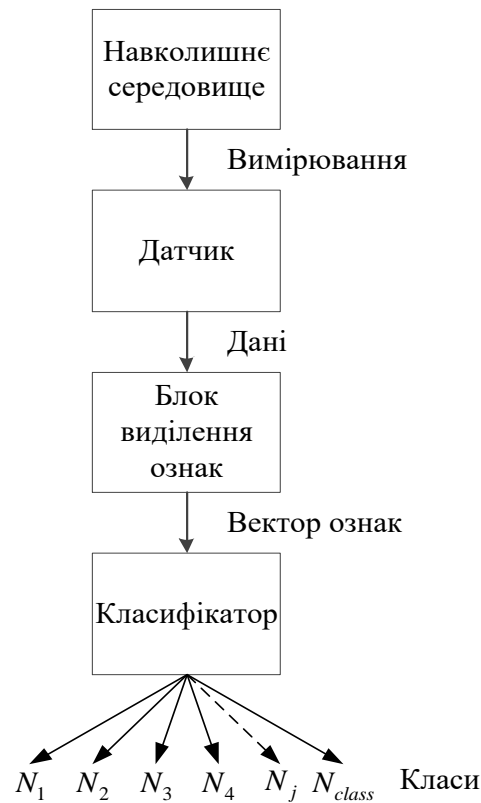


Рисунок 1.8 – Модель класифікації

Формально постановка завдання дослідження здійснюється наступним чином.

Нехай задано множину об'єктів $\Omega = \{\omega_1, \dots, \omega_h\}$, де h – кількість об'єктів, які зображені на аерофотознімку.

$g(\omega) : \Omega \rightarrow N$, $N = \{1, 2, \dots, n\}$ – наперед невідома функція (класифікатор), що розбиває простір образів $\Omega^1, \Omega^2, \dots, \Omega^{N_{class}}$ на N_{class} класів, які не перетинаються;

X – простір ознак;

$x(\omega) : \Omega \rightarrow X$ – функція, яка ставить у відповідність кожному об'єкту ω точку $x(\omega)$ в просторі ознак X ;

$x(\omega)$ – це вектор ознак, який вказує на місце об'єкта в багатовимірному просторі ознак X ;

$g(x): X \rightarrow N$ – класифікатор або оцінка $g(\omega)$ на основі $x(\omega)$,
 $g(x) = g(x(\omega))$.

Необхідно побудувати такий класифікатор $g(x)$, щоб класифікація проходила з мінімальною кількістю помилок на всій множині об'єктів.

Множина об'єктів $\Omega = \{\omega_1, \dots, \omega_h\}$ у даній постановці завдання становить навчальну вибірку розміром h . Для задач класифікації зображень зазвичай по вибірці Ω неможливо за допомогою правил описати класифікатор $g(x)$. Але він може бути отриманий для конкретної множини об'єктів за допомогою використання методів, що реалізують процеси його навчання. Класифікація об'єктів на цифрових знімках за допомогою методів з навчанням активно використовується розпізнавання у ряді прикладних задач і не суперечить даній постановці завдання дослідження.

1.3 Аналіз основних підходів щодо виявлення так класифікації об'єктів на цифрових знімках

Існуючі методи виявлення так класифікації об'єктів на цифрових знімках реалізуються за рахунок логічних і арифметичних операції та класифікації пікселів зображення за їх просторовими, спектральними та часовими характеристиками. Виділяють два основних підходи до процедури класифікації, а саме піксельно-орієнтований та об'єктно-орієнтований.

Піксельно-орієнтований підхід передбачає класифікацію кожного пікселя зображення окремо за інформацією про значення його спектральної яскравості. Відбувається класифікація кожного пікселя окремо, а один піксель зображення може відповідати тільки одному класу.

Основні методи які реалізують піксельно-орієнтований підхід включають методи без навчання: ISODATA, k-means, метод контурного аналізу та методи з навчанням: random forest, boosting, нейронні мережі (персептрон), вейвлети.

Об'єктно-орієнтовані методи направлені на роботу не з окремим пік селем зображення, а з групою пікселів, якою можливо описати об'єкт класифікації. Група пікселів оцінюється за ректальними особливостями, формою, структурою, а також просторовим положенням окремих груп між собою.

Основні методи які реалізують об'єктно-орієнтований підхід включають support vector machine, згорткові нейронні мережі, детектори HOG, Haar, LBP, SIFT.

Аналіз можливостей перелічених методів для процесу класифікації здійснювався з використанням вибірки з 648 тисяч розмічених на 200 класів зображень Imagenet 2012 за результатами конкурсу ImageNet Challenge з розпізнавання зображень.

Застосування методу контурного аналізу дозволяє класифікувати антропогенні та природні об'єкти на цифрових аерофотознімках.

Метод ISODATA призначений для поділу заданої множини образів на підмножини (кластери), пов'язані певним властивістю, шляхом оцінки геометричної відстані між точок у просторі ознак. Він спирається на евристичний алгоритм, а результат його роботи багато в чому залежить від заданих початкових параметрів. Зазвичай метод ISODATA використовується у системах комп'ютерного аналізу багатозональних і спектрозональних аерокосмічних зображень (класифікація за спектральним ознаками) та реалізований у прикладних пакетах для роботи з космічними знімками, наприклад, ERDAS Imagine, ER Mapper (Австралія), EASY/PAGE (PCI) (Канада), ILVIS GIS (Голландія).

Він використовується для класифікації антропогенних та природних об'єктів, водних ресурсів, та земель сільськогосподарського призначення.

В таблиці 1.1 наведено основні характеристики вказаних методів.

Таблиця 1.1 – Характеристики основних методів виявлення та класифікації об'єктів на зображення

Назва методу	Точність класифікації (кількість правильно класифікованих об'єктів вибірки)	Можливість реалізації в режимі часу, наближеного до реального	Можливість одночасного виявлення та класифікації об'єктів на зображеннях
ISODATA	53-58%	Ні	Ні
K-means	61-66%	Ні	Ні
Метод контурного аналізу	50-56%	Так	Ні
Random forest	70-73%	Ні	Ні
Boosting	60-64%	Так	Ні
Нейронні мережі (персептрон)	59-67%	Ні	Так
Вейвлети	68-71%	Ні	Так
Support vector machine	75-81%	Ні	Так
Згорткові нейронні мережі	80-89%	Так	Так
Детектори HOG, Haar, LBP, SIFT.	–	Так	Ні

Перспективним методом також є метод Support vector machine, який зазвичай використовується для задач класифікації та регресійного аналізу.

Метод дозволяє здійснювати перехід від вектору вхідних даних до нового вектору з набагато більшою розмірністю, а далі здійснити лінійне розбиття простору ознак на класи. Але основним недоліком даного методу є велика кількість обчислень при його реалізації та відсутність існуючих апаратних для виявлення та класифікації об'єктів на зображеннях в режимі часу, наближеному до реального.

Як видно з таблиці 1.1 тільки метод згорткових нейронних мереж функціонує в режимі часу наближеному до реального та забезпечує одночасне виявлення та класифікацію об'єктів на зображенні.

Особливістю згорткових нейронних мереж є властивості часткової стійкості до зміни розмірів, зміщення, повороту об'єкта на знімках та паралельного обчислення, що дозволяє здійснювати обробку зображень в режимі часу, наближеного до реального.

1.4 Дослідження топології згорткової нейронної мережі для обробки цифрових зображень

Штучна нейронна мережа – це паралельно розподілена система обробки інформації, утворена з однорідних елементів (нейронів). Вона має властивість накопичувати експериментальні знання, узагальнювати їх і робити доступними для користувача в формі, зручній для інтерпретації та прийняття рішень. У задачах розпізнавання образів нейронна мережа є універсальним апроксиматором, який ділить простір ознак на компактні області, які відповідають класам об'єктів. Фактично нейронна мережа реалізує відображення n -вимірному вектору $x^{net} = (x_1^{net}, x_2^{net}, \dots, x_n^{net})^T$ в m -вимірний вектор $y^{net} = (y_1^{net}, y_2^{net}, \dots, y_m^{net})^T$, тобто моделює деякий оператор $F^{net} : R^{n^{net}} \rightarrow R^{m^{net}}$.

На кожний вхід нейрона подається сигнал x_i^{node} . З кожним входом зв'язана синоптична вага w_j^{node} цього входу. В тілі нейрона обчислюється

функція $\psi^{node}(\sum_{i=0}^{n^{node}} w_i^{node} x_i^{node})$. Таким чином реалізується нелінійне відображення багатовимірного простору входів $R^{n^{node}}$ у скалярний вихід $R^{1^{node}}$. Такий підхід було запропоновано W. S. McCulloch та W. A. Pitts в роботі. В подальшому описаний нейрон отримав назву нейрон Маккалоха-Піттса. На даний час він застосовується для опису нелінійного перетворювача з багатьма входними та вихідними сигналами. Штучний нейрон Маккалоха-Піттса схематично зображено на рисунку 1.2.

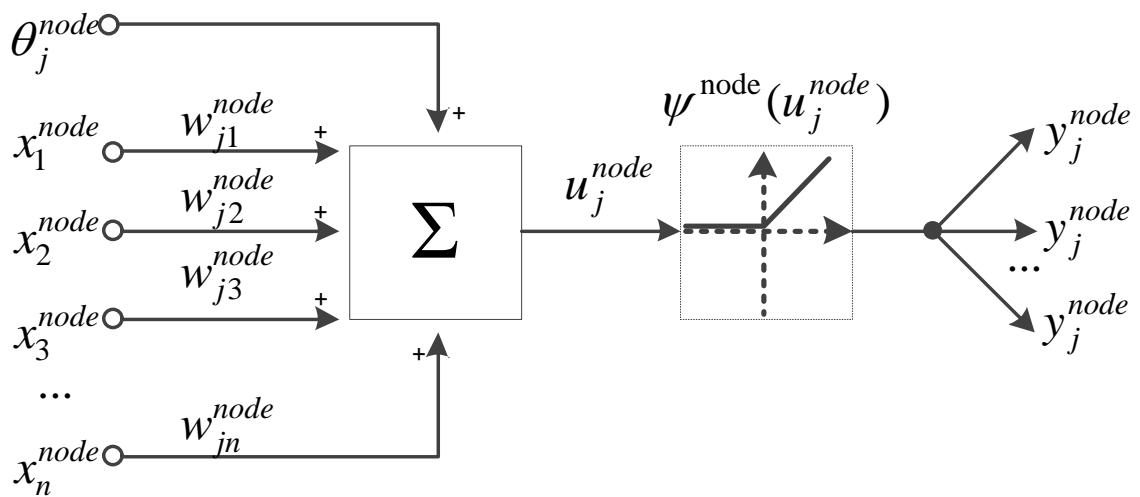


Рисунок 1.9 – Штучний нейрон Маккалоха-Піттса

На вхід j -го нейрона поступає n сигналів $x_1^{node}, x_2^{node}, \dots, x_n^{node}$, які корегуються відповідними ваговими коефіцієнтами w_j^{node} . Після цього кореговані сигнали $w_{j1}^{node} x_1^{node}, w_{j2}^{node} x_2^{node}, \dots, w_{jn}^{node} x_n^{node}$ разом зі зміщенням (зсувом) θ_j^{node} подаються на суматор Σ . Він формує внутрішній сигнал u_j^{node} . Далі внутрішній сигнал нейрона поступає на передаточну функцію або функцію активації нейрона $\psi^{node}(u_j^{node})$.

Аналітично нейрон Маккалоха-Піттса записується виразом:

$$y_j^{node} = \psi^{node} \left(\sum_{i=1}^n w_{jn}^{node} x_i^{node} + \theta_j^{node} \right).$$

Група нейронів, яка отримує сигнал безпосередньо від джерела сигналу, утворює вхідний шар нейронної мережі. Наступні нейрони по лінії проходження сигналу, які отримують сигнали від вхідного шару, утворюють скриті шари. Останній шар по лінії проходження сигналу утворює вихідний шар. Він містить стільки виходів N_{class} , на скільки класів здійснюється класифікація. Зв'язок між нейронами різних шарів утворює топологію штучної нейронної мережі.

Широке застосування штучних нейронних мереж для обробки інформації обумовлено універсальними апроксимаційними можливостями штучних нейронних мереж. Згідно теореми Колмогорова-Арнольда про представлення безперервних функцій кількох змінних у вигляді суперпозиції безперервних функцій одної змінної, в роботі [16] було доведено, що функція кількох змінних може бути представлена за допомогою нейронної мережі з двома скритими шарами та фіксованим числом нейронів у кожному шарі. Так, скінченна лінійна комбінація фіксованих одновимірних функцій може однозначно апроксимувати будь-яку безперервну функцію n дійсних змінних на заданому гіперкубі.

З наведеного вище зрозуміло, що застосування штучних нейронних мереж для здійснення класифікації є доцільним у разі, коли вхідними даними є n -вимірний вектор, вихідними даними – m -вимірний вектор, а операцію відображення $F^{net} : R^{n^{net}} \rightarrow R^{m^{net}}$ в аналітичному вигляді записати неможливо.

Багатошарова повнозв'язна штучна нейронна мережа, виконана з однакових нейронів, називається перцептрон. Такий тип мережі має широку сферу застосування. Але для роботи з зображенням вимагає настільки великих обчислювальних потужностей, що існуючі апаратні засоби не задовольняють вимогам. Тому для виявлення та класифікації простих об'єктів ОЕПР на цифрових знімках пропонується використати тип нейронних мереж, який здійснює попередню обробку зображення перед подачею його на

перцептрон, а саме виділення ознак об'єкта на зображенні. Такий підхід пропонується реалізувати за допомогою використання ЗНМ.

Приводом до розробки топології згорткових нейронних мереж стали дослідження зорового апарату кішок, проведені D. H. Hubel та T. N. Wiesel у 1960-х роках. За результатами їх роботи було відкрито два типи клітин головного мозку, які виконують зорову функцію. Першому типу клітин (прості нейрони) притаманна властивість локальної чутливості. Вони здійснюють виділення елементарних характеристик образів, зображених на рисунку 1.3, а саме: прямих ліній, кінцевих точок, кутів та ін.

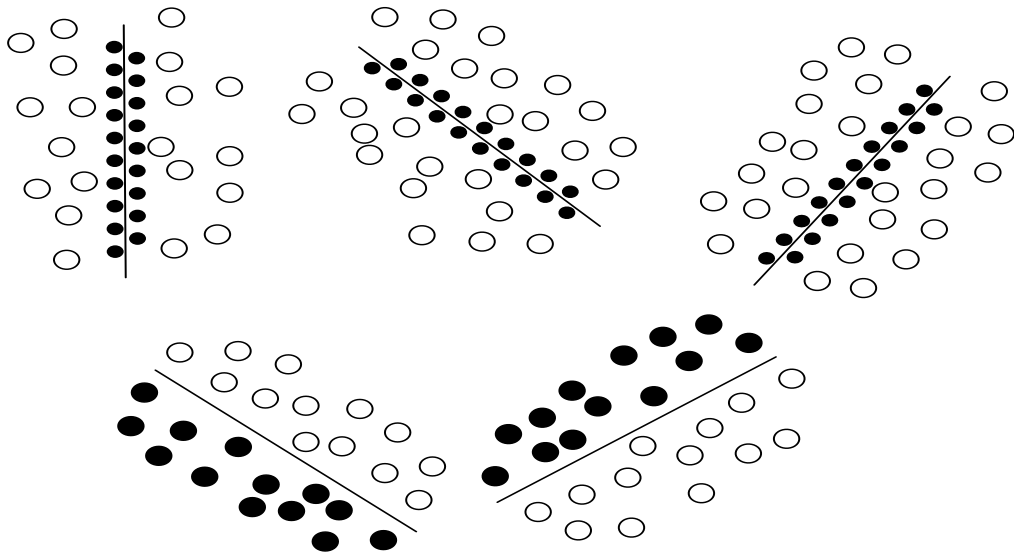


Рисунок 1.10 – Рецептивне поле простих нейронів

Клітини другого типу (складні нейрони) здійснюють комбінування елементарних характеристик образів у ознаки, а саме: перетин ліній, їх взаємне положення. Діяльність описаних простих та складних нейронів можливо змодельовати, застосувавши підходи щодо побудови нейронів Маккалоха-Піттса.

Описану топологію вперше було застосовано в неокогнітроні, запропонованому в роботі. Але налаштування параметрів запропонованої топології здійснювалося вручну. Це обмежувало використання її тільки для

задач класифікації елементарних образів, наприклад, друкованих символів або простих знаків.

Наступним етапом у розвитку описаних підходів стала розробка топології ЗНМ LeNet-5. Топологія включала три концептуальні підходи, а саме, локальні рецептивні поля, колективні вагові коефіцієнти, просторову субдискретизацію. Вона реалізовувалася за рахунок введення в топологію нейронної мережі згорткових шарів (Convolutional layer) та шарів субдискретизації (Pooling layer). Такі концептуальні підходи породили нові властивості штучних нейронних мереж, а саме, часткову стійкість до змін масштабу об'єктів на знімку, їх зсуву, повороту, зміни ракурсу та інших спотворень, можливість здійснювати паралельні розрахунки.

Локальне рецептивне поле нейрона – це набір рецепторів (сигналів x^{net}), який впливає на функціонування нейрона. Під рецепторами розуміються нейрони, які безпосередньо сприймають зовнішні сигнали. Перший шар (рецепторів) складається з простих нейронів, наступні за ним шари – зі складних нейронів. Для кожного нейрона наступного шару виходить попереднього шару і є рецептивним полем. На рисунку 1.4 для наглядності зображено рецептивне поле шару складних нейронів у формі букви “х”. На практиці рецептивні поля складних нейронів квадратні. Таким чином здійснюється подача на вхід одного нейрона не всього зображення, а лише його квадратної області. Такий підхід дозволяє зберігати просторову залежність між ознаками об'єктів.

Підхід колективних вагових коефіцієнтів передбачає, що вагові коефіцієнти нейронів згорткового та субдискретизаційного шару застосовуються одразу до всього вхідного зображення. Вагові коефіцієнти шарів формуються в процесі навчання ЗНМ. Підхід колективних вагових коефіцієнтів значно зменшує загальну кількість вагових коефіцієнтів у топології нейронної мережі та дозволяє виявляти ознаки, на які навчені вагові коефіцієнти по всій області зображення. В одному шарі ЗНМ використовується одразу кілька колективних вагових коефіцієнтів. Чим більша кількість

колективних вагових коефіцієнтів, тим більша кількість ознак може бути виявлена на зображенні.

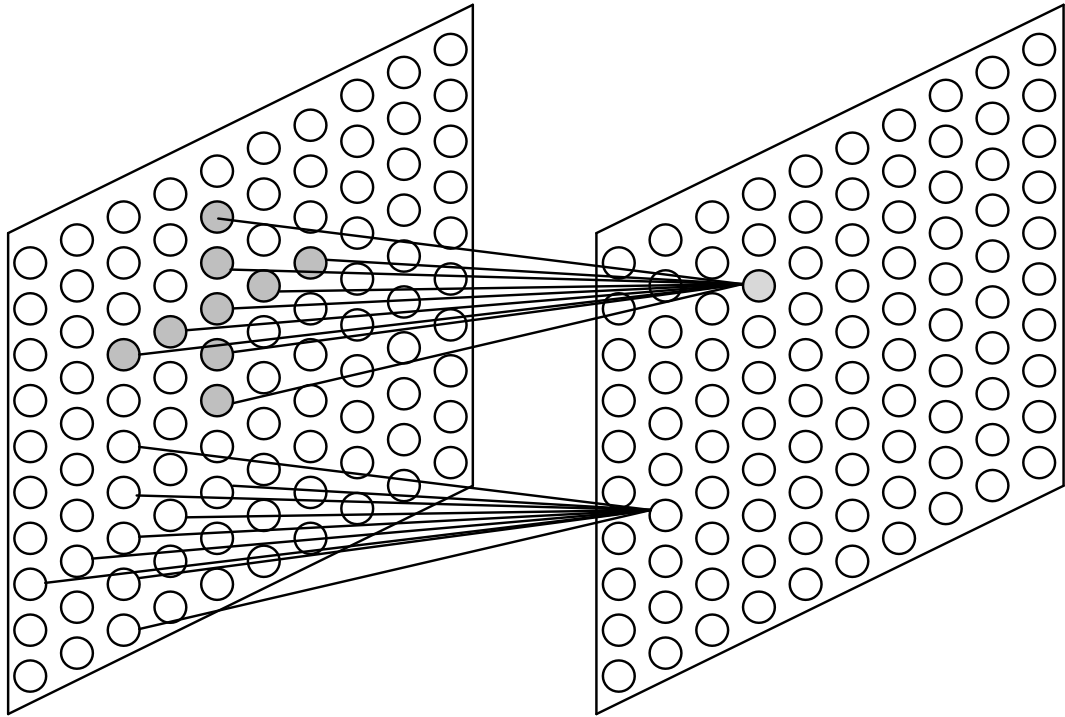


Рисунок 1.11 – Рецептивне поле складних нейронів

Просторову субдискретизацію застосовують для зменшення розмірів зображень та, як наслідок, розмірності простору ознак. Цей підхід застосовується, перш за все, для економії обчислювальних потужностей та врахування самих вагомих ознак, які поступають з попереднього шару ЗНМ.

Так, основними типами шарів, що використовуються в штучних нейронних мережах, є згортковий, субдискретизаційний та повнозв'язний шари нейронів. Шари розташовуються таким чином, щоб відповідати класичній моделі класифікації, яка складається з датчика, блока виділення ознак та класифікатора.

Датчиком є вхідний шар нейронів розміром H_1, L_1 . Він представляє цифрове зображення у вигляді тривимірної матриці значень інтенсивності в кожному пікселі зображення для кожного кольору. Блок виділення ознак

об'єктів на зображенні складається з послідовно з'єднаних шарів згортки та субдискретизації. На рисунку 1.5 наведено схематичну топологію для випадку, коли обробка здійснюється по одному кольору.

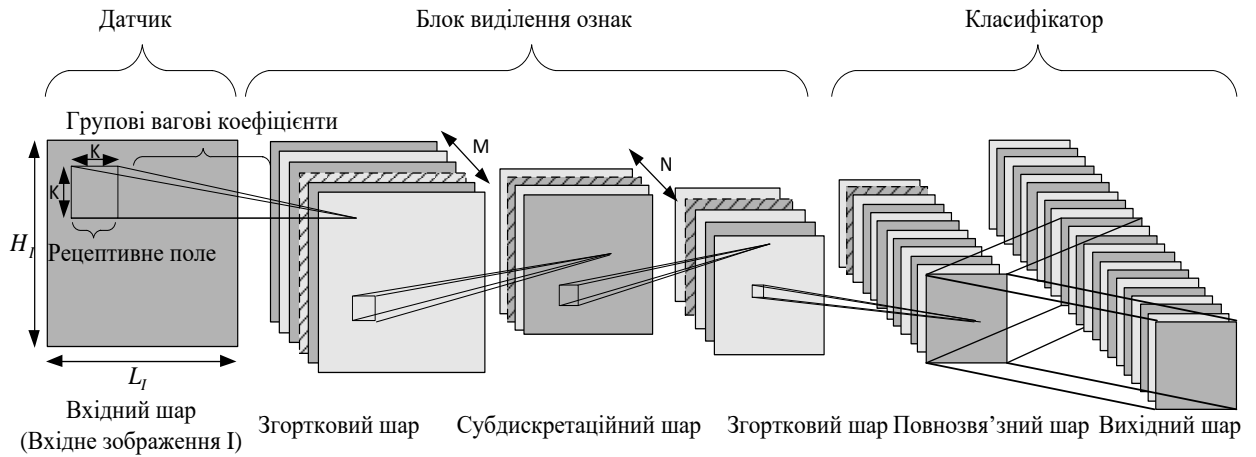


Рисунок 1.12 – Схематична топологія згорткової нейронної мережі

В мережах, що використовуються для розпізнавання складних об'єктів на зображеннях розміром від 256×256 пікселів, кількість шарів блока виділення ознак становить від 20 до 200 в залежності від кількості класів.

Якщо розглянути функціонування одного нейрона згорткового шару мережі (рисунок 1.12), то видно, що значення кожного пікселя на виході шару згортки формується із значень усіх попередніх даних M з урахуванням групових вагових коефіцієнтів W для кожної з N карт ознак.

Кожний нейрон відповідає за формування одного пікселя зображення. Виконується операція двомірної лінійної згортки зображення з ядром згортки. Воно за розмірами дорівнює рецептивному полю нейрона, а коефіцієнтами ядра згортки є вагові коефіцієнти w , які отримуються в процесі навчання.

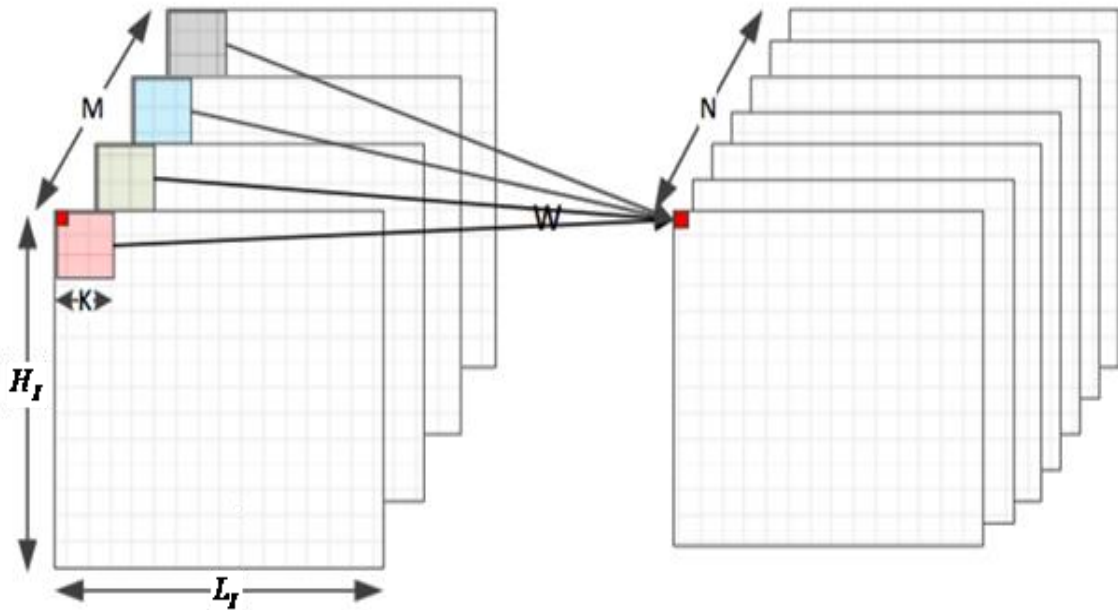


Рисунок 1.13 – Функціонування нейрона згорткового шару нейронної мережі

В аналітичному вигляді нейрон згорткового шару за моделлю нейрона Маккалоха-Піттса має вигляд:

$$y_N^{(i,j)} = \psi^{\text{node}} \left(\sum_{x_i=1}^K \sum_{y_j=1}^K w_{N,x,y} \cdot x_M^{(i+x_i, j+y_j)} \right), \quad (1.1)$$

де i, j – координати пікселя результату згортки;

ψ^{node} – функція активації нейрона Маккалоха-Піттса;

$w_{N,x,y}$ – вагові коефіцієнти;

$x_M^{(i+x, j+y)}$ – значення інтенсивності пікселя вхідних даних за номером.

Для згорткових нейронних мереж може застосовуватися різна функція активації нейрона Маккалоха-Піттса ψ^{node} . Головною вимогою до неї є здатність до відшукування її похідної в процесі навчання ЗНМ. У роботі [28] доведено, що використання функції активації $ReLU$ є оптимальним для використання в топології ЗНМ. Функція $ReLU$ аналітично записується:

$$\psi^{\text{node}}(x) = \max(0, x) . \quad (1.2)$$

Застосування вказаної функції за функцію активації нейрона Маккалоха-Пітса дозволяє значно підвищити збіжність стахостичного градієнтного спуску в процесі навчання ЗНМ.

Нейрон Маккалоха-Пітса для шару субдискретизації зменшує розміри X, Y вхідних даних M за рахунок усереднення значень інтенсивності пікселів, які попали в рецептивне поле нейрона K . Тобто, якщо рецептивне поле нейрона шару субдискретизації $K = 2$, то зображення вхідних даних зменшиться у 2 рази, а якщо $K = 4$, то в 4 рази. Функція активації до нейрона шару субдискретизації в згорткових нейронних мережах не застосовується. Аналітично функціонування нейрона Маккалоха-Пітса наведено у виразі:

$$y_N^{(i,j)} = \frac{1}{2K} \sum_{x_i=1}^K \sum_{y_j=1}^K x_M^{(x_i, y_j)} . \quad (1.3)$$

Таким чином, ЗНМ – це спеціальний вид нейронних мереж для обробки цифрових зображень. Їх використання дозволяє забезпечити часткову стійкість до змін масштабу, зсуву, повороту, ракурсу та інших спотворень зображень об'єктів на цифрових знімках під час виявлення їх ознак та подальшої класифікації. ЗНМ також дозволяють здійснювати паралельні обчислення за рахунок використання групових вагових коефіцієнтів, що дозволяють реалізувати обрахунки в режимі часу наближеному до реального.

2 ДОСЛІДЖЕННЯ ПІДХОДІВ ЩОДО СЕГМЕНТАЦІЇ ОБ'ЄКТІВ НА ЦИФРОВИХ ЗНІМКАХ З ВИКОРИСТАННЯМ МЕТОДОЛОГІЇ ГЛИБОКОГО НАВЧАННЯ

2.1 Дослідження основних топологій глибоких нейрон мереж для сегментації об'єктів на цифрових знімках

Розробка алгоритмів глибокого навчання, таких як згорткові нейронні мережі або глибокі автоенкодера, не тільки торкнулася типові завдання, такі як класифікація об'єктів, але також була ефективна в інших пов'язаних завданнях, в цьому випадку, сегментація зображень.

Згорткові нейронні мережі, що є одним з найбільш часто використовуваних методів в комп'ютерному зорі, взяли багато простих модифікацій, щоб також добре виконувати завдання сегментації [1-14].

Завдання класифікації зазвичай вимагають лінійного виводу в формі розподілу ймовірностей за кількістю класів. Щоб перетворити обсяги двовимірних карт активації в лінійні шари, їх часто згладжували. Згладжена форма дозволила виконати повністю пов'язані мережі, щоб отримати розподіл ймовірності. Однак цей вид зміни форми втрачає просторові відносини між пікселями на зображенні. У повністю згорткової нейронної мережі (FCN) вихід останнього згорткового блоку безпосередньо використовується для класифікації на рівні пікселів. FCN були вперше реалізовані в наборі даних сегментації PASCAL VOC 2011 і досягли точності пікселів 90,3% і середнього IOU 62,7%. Іншим способом уникнути повністю пов'язаних лінійних шарів можна застосовуючи використання повнорозмірного середнього пулу для перетворення набору двовимірних карт активації в набір скалярів.

DeepMask і SharpMask DeepMask - це ім'я, дане проекту Facebook AI Research (FAIR), що стосується сегментації зображень. Він демонстрував ту

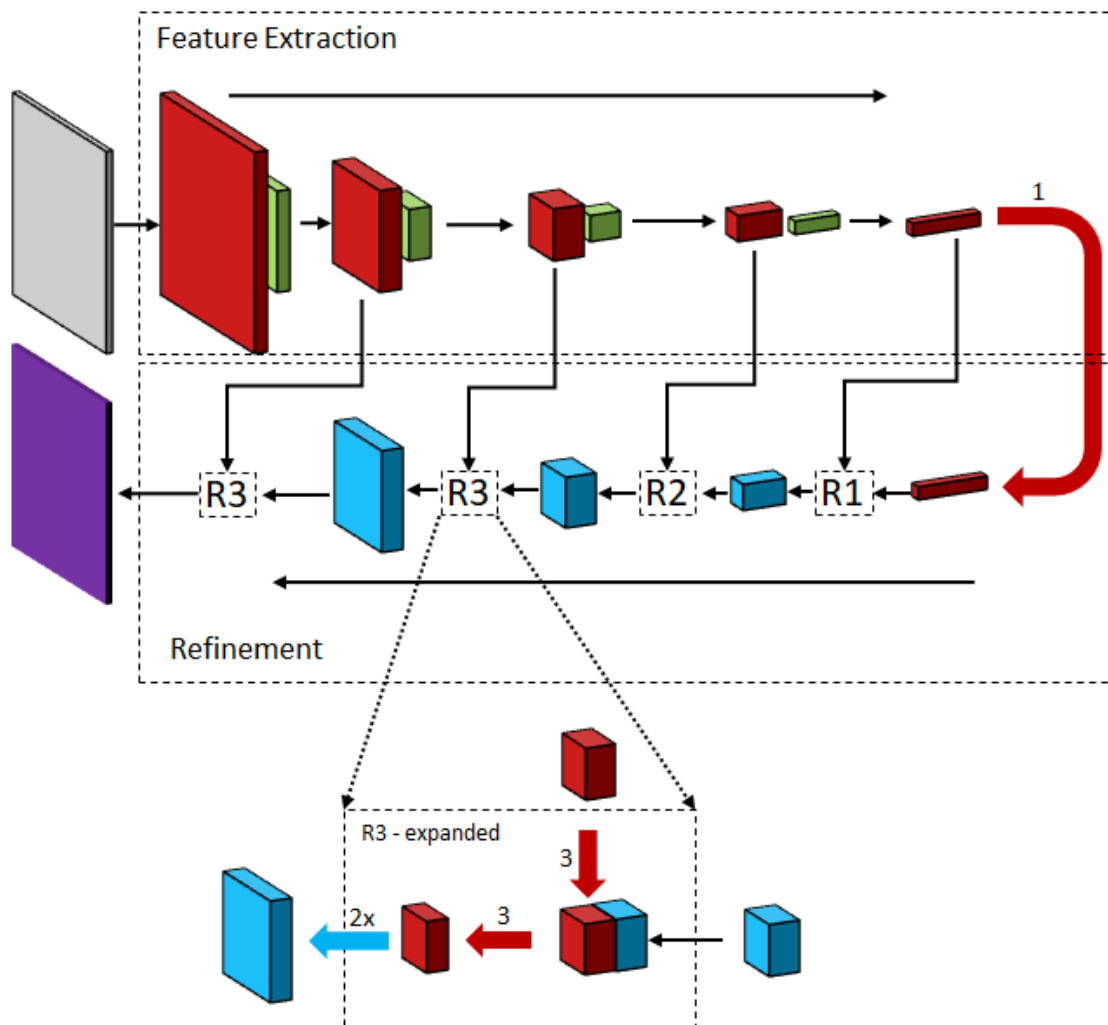


Рисунок 2.2 – Мережа Sharpmask

RCNN (регіональні згорткові нейронні мережі) Впровадження CNN підняло багато нових питань в області комп'ютерного зору. Одним з них є, перш за все, можливість розширення мережі, такий як AlexNet, для виявлення присутності більш ніж одного об'єкта. CNN на основі областей або більш широко відомий як R-CNN використовував метод вибіркового пошуку, щоб запропонувати ймовірні області об'єкта, і виконав класифікацію на обрізаному вікні, щоб перевірити розумну локалізацію на основі розподілу ймовірності на виході. Техніка селективного пошуку аналізує різні аспекти, такі як текстура, колір або інтенсивності кластеризації пікселів в об'єкти. Обмежують блоки, які відповідають цим сегментам, передаються через класифікують мережі, щоб скласти короткий список деяких з найбільш

розумних блоків. Нарешті, за допомогою простої лінійної регресійної мережі можна отримати більш жорсткі координати. Основним недоліком методу є його обчислювальна вартість. Мережа повинна обчислювати прямий прохід для кожної пропозиції обмежувальної рамки. Проблема зі спільним використанням обчислень для всіх блоків полягала в тому, що блоки були різних розмірів і, отже, об'єкти однакового розміру були недосяжні. В оновленому Fast R-CNN було запропоновано об'єднання областей інтересу (ROI), в якому області інтересів були динамічно об'єднані для отримання вихідних даних фіксованого розміру. Відтепер, мережа була в основному вузьким місцем за допомогою методу вибіркового пошуку для пропозиції регіону кандидата. У Faster-RCNN замість залежних від зовнішніх ознак проміжних карт активації використовувалися для пропозиції обмежувальних рамок, що прискорювало процес вилучення ознак. Обмежувальні рамки представляють розташування об'єкта, однак вони не надають сегменти на рівні пікселів. Мережа Faster R-CNN була розширена як маска R-CNN з паралельної гілкою, яка виконувала двійкову класифікацію об'єктів рівня пікселів для забезпечення точних сегментів. З Mask-RCNN середня точність 35,7 була досягнута на тестових зображеннях COCO. Сімейство алгоритмів RCNN зображено на рисунку 2.3. Регіональні мережі пропозицій часто об'єднувалися з іншими мережами, щоб дати сегментації рівня екземпляра. RCNN був вдосконалений під назвою HyperNet за рахунок використання функцій з декількох рівнів екстрактора функцій. Регіональні мережі пропозицій також були реалізовані для конкретної сегментації. Як згадувалося раніше, можливості виявлення об'єктів в таких підходах, як RCNN, часто поєднуються з моделями сегментації, щоб генерувати різні маски для різних екземплярів одного і того ж об'єкта.

Хоча сегментація на рівні пікселів була ефективною, на продуктивність як і раніше впливали дві доповнюють проблеми. По-перше, ядра меншого розміру не змогли зібрати контекстну інформацію.

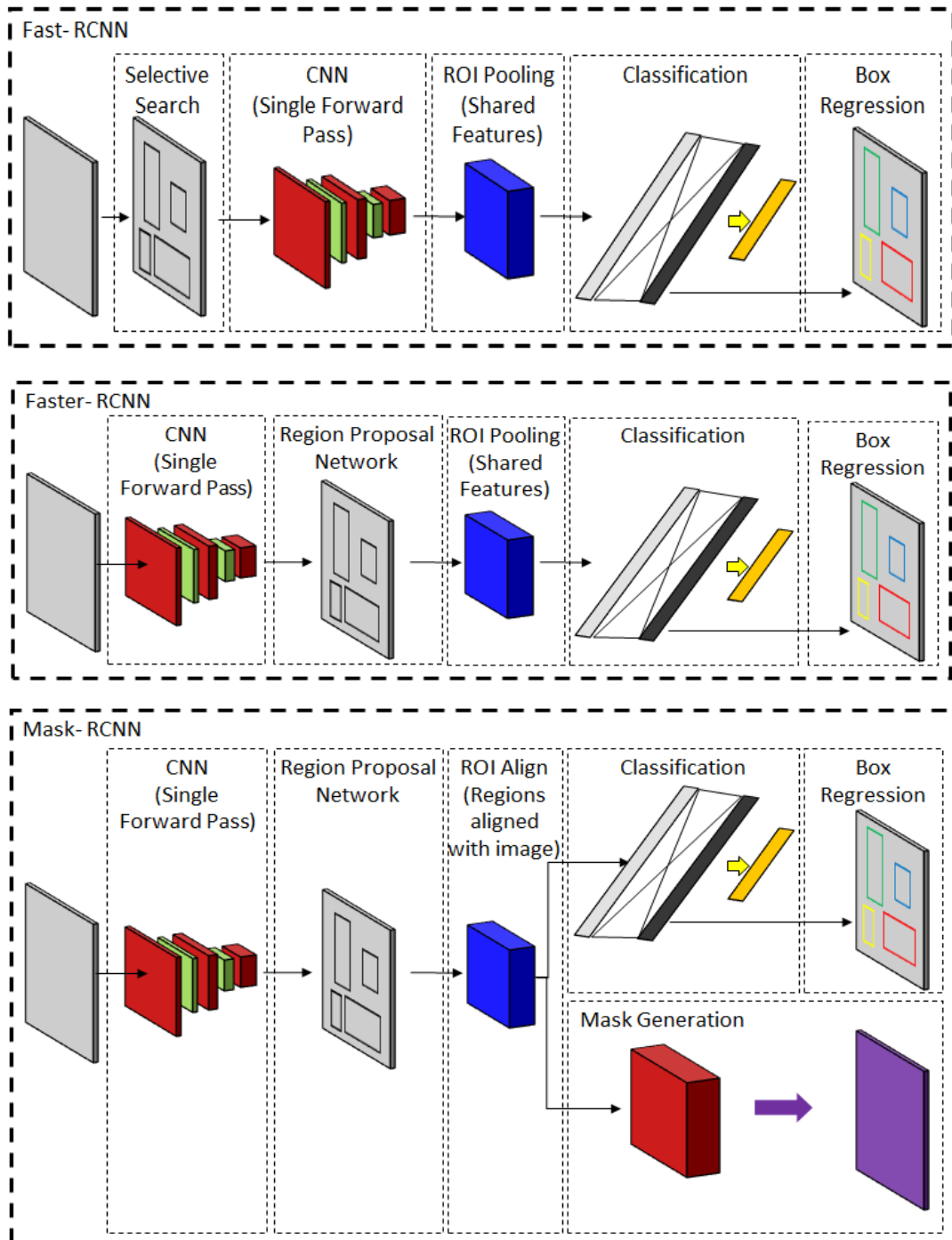


Рисунок 2.3 – Сімейство мереж локалізації та сегментації RCNN

У задачах класифікації це вирішується з використанням об'єднаних шарів, які збільшують сенсорну область ядер в порівнянні з вихідним зображенням. Але в сегментації це знижує різкість сегментованого виведення. Альтернативне використання великих ядер, як правило, повільніше через значно більшої кількості учнів параметрів.

Для вирішення цієї проблеми сімейство алгоритмів DeepLab продемонструвало використання різних методологій, таких як згорткові згортки, піраміди просторового об'єднання і повністю пов'язані умовні випадкові поля, для виконання сегментації зображення з великою ефективністю. Алгоритм DeepLab зміг досягти середнього значення 79,7 для набору даних PASCAL VOC 2012.

Розмір ядер згортки в будь-якому шарі визначає область сенсорної реакції мережі. У той час як більш дрібні ядра витягають локальну інформацію, більші ядра намагаються зосередитися на більш контекстній інформації. Однак більші ядра зазвичай мають більшу кількість параметрів.

Одна з основних проблем, пов'язаних з сегментацією зображення для природних зображень сцени, полягає в тому, що розмір об'єкта, що цікавить дуже непередбачуваний, оскільки в реальному світі об'єкти можуть мати різні розміри, а об'єкти можуть виглядати більше або менше в залежності від положення об'єкта. об'єкт і камера. Природа CNN диктує, що тонкі дрібномасштабні об'єкти фіксуються на ранніх рівнях, тоді як при переміщенні по глибині мережі функції стають більш специфічними для більших об'єктів. Наприклад, крихітний автомобіль в сцені має набагато меншу вірогідність бути захопленим в верхніх шарах через таких операцій, як об'єднання в пул або знижує вибірка. Часто корисно витягати інформацію з карт характеристик різного масштабу для створення сегментації, що не залежать від розміру об'єкта на зображенні. Моделі багатомасштабного автокодуювальника розглядають активації різних дозволів, щоб забезпечити виведення сегментації зображення.

PSPNet Мережа парсинга сцени піраміди була побудована на основі мережі класифікації пікселів на основі FCN. Карти функцій з мережі ResNet перетворюються в активації різних дозволів за допомогою ретельного багатомасштабного об'єднання пулів, які згодом піддаються додатковій дискретизації і об'єднуються з оригінальною картою функцій для виконання сегментації (рисунок 2.4). Процес навчання в глибоких мережах, таких як

ResNet, був додатково оптимізований за допомогою допоміжних класифікаторів. Різні типи модулів пулу орієнтовані на різні області карти активації. Які об'єднують ядра різних розмірів, такі як 1×1 , 2×2 , 3×3 , 6×6 , вивчають різні області карти активації, щоб створити піраміду просторового об'єднання. Одна з проблем, пов'язаних з аналізом сцени ImageNet, дозволила PSPNet отримати середнє значення IoU 57,21 по відношенню до 44,80 в FCN і 40,79 в SegNet.

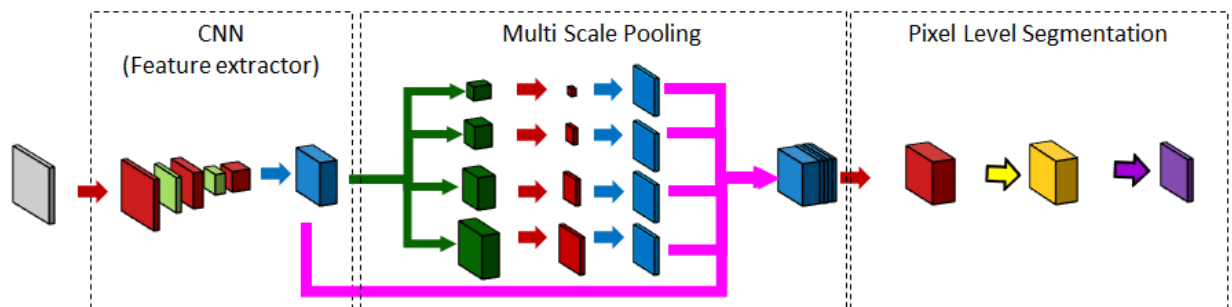


Рисунок 2.4 – Схематичне представлення PSPNet

Архітектура U-Net, запропонована в 2015 році, виявилася досить ефективною для вирішення різних завдань, таких як сегментація нейронних структур, рентгенографія і завдання відстеження клітин. Мережа характеризується кодером з послідовністю рівнів згортки і максимального пулу. Шар декодування містить дзеркальну послідовність транспонованих згорток. Він веде себе як традиційний автокодувальник. Щоб розглянути різні рівні абстракції, U-Net реалізує пропущені з'єднання для копіювання незжатих активацій з кодуєть блоків в їх дзеркальні копії серед блоків декодування, як показано на рисунку 2.5. Екстрактор функцій U-Net також може бути модернізований для поліпшення карт сегментації. Мережа під назвою «Тірамісу ...» застосувала концепцію U-Net з використанням екстрактора функцій на основі щільної мережі. Інші сучасні варіації включають використання капсульних мереж поряд з локальної маршрутизацією.

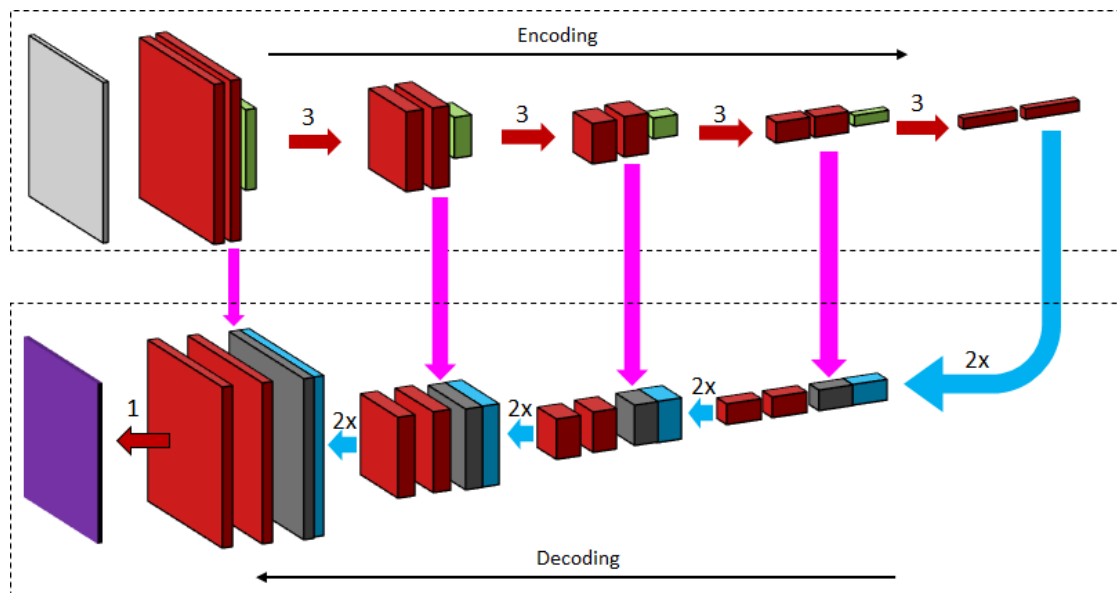


Рисунок 2.5 – Схематичне представлення U-Net

U-Net був обраний в якості переможця конкурсу ISBI. У наборі даних PhC-U373 середнє значення IoU склало 0,9203, тоді як друге місце було на рівні 0,83. У наборі даних DIC-HeLa він набрав середнє значення IoU 0,7756, що було значно краще, ніж другий кращий підхід, який набрав лише 0,46.

До ознакових класифікаторів також відносять, перш за все, глибокі (багатошарові) нейронні мережі і зокрема ЗНМ. Вони безпосередньо можуть здійснювати як виділення ознак, так і розпізнавання об'єктів, у тому числі і в рамках нейромережових ансамблів. Задачу формалізації знань про розпізнавання об'єктів на цифрових аерофотознімках за допомогою глибоких нейронних мереж умовно розділимо на наступні підзадачі:

- формалізація процесу вилучення ознак об'єктів на цифровому аерофотознімку;
- формалізація процесу виявлення і виділення об'єктів на цифровому аерофотознімку;
- формалізація процесу класифікації за набутими ознаками виявлених і виділених об'єктів;
- формалізація процесу підвищення якості цифрового аерофотознімка низької здатності у випадку, якщо об'єкт виявлений, але не класифікований.

Перехід від приналежності до неприналежності розпізнаного об'єкту до класів відповідних об'єктів є поступовим, а не стрибкоподібним. Вимога щодо знаходження однозначної класифікації досліджуваних елементів є достатньо грубою. Тому для класифікації за набутими ознаками виявлених і виділених об'єктів можливо використовувати нечіткі класифікатори (системи класифікації) на основі нейронних нечітких продукційних мереж. Вони ослаблюють вимогу однозначної класифікації за допомогою визначення нечітких ступенів приналежності об'єктів до відповідних заздалегідь заданих класів.

Безпосередньо ЗНМ можна розглядати за модель представлення знань для формалізації задачі розпізнавання об'єктів, а саме у вигляді її навчальної вибірки та синаптичних карт (фільтрів), сформованих за результатами визначення оптимальної топології та навчання ЗНМ. На даний час відома значна кількість ЗНМ різної архітектури. Вони розрізняються додатковими шарами, різною глибиною, наприклад, LeNet, AlexNet, VGG-19, різновиди Inception і ResNet та ін., а також функціональністю. Зокрема, розглянуті ЗНМ використовуються для класифікації об'єктів. А, наприклад, мережі R-CNN (її розширення Fast R-CNN, Faster R-CNN), YOLO, SSD, MultiBox GoogleNet та ін. знайшли застосування для одночасного виділення і класифікації об'єктів в процесі їх розпізнавання на цифрових зображеннях. Такі мережі фактично є ансамблем нейромереж і називаються нейромережевими детекторами.

Топологію SSD (Single Shot MultiBox Detector) забезпечує швидкий і точний пошук гіпотез про місцеположення об'єкту, що виділяється на цифровому знімку, наприклад, порівняно з архітектурою YOLO, R-CNN, Fast R-CNN і MultiBox GoogleNet. При цьому SSD за точністю практично відповідає Faster R-CNN, але є набагато більш швидкодіючою. Це досягається використанням вбудованих в топологію детектора нейромережеских генераторів прогнозу місцерозташування і класифікаторів об'єктів, які:

- є такими, що навчаються, на відміну, наприклад, від топологій Fast R-CNN і R-CNN, де використовується евристичний метод Selective Search;
- використовують карти ознак різного масштабу для забезпечення високої точності розпізнавання з урахуванням різної роздільної здатності карт ознак.

Топологія ЗНМ Inception-v2. Дана архітектура має більшу глибину в порівнянні з вбудованим виділителем ознак SSD на основі VGG-19 і такої архітектури, як LeNet, AlexNet. Крім того, архітектура ЗНМ Inception-v2 має меншу кількість параметрів, що навчаються, у порівнянні з архітектурою AlexNet, VGG-19, ResNet, Inception-v1 і меншу обчислювальну складність порівняно з новими архітектурами Inception-v4, Inception-ResNet. Дані переваги забезпечують, з одного боку, підвищення точності класифікації, а з іншого боку, зменшення часу навчання ЗНМ і ресурсних витрат.

Якщо цифровий аерофотознімок не задовольняє вимогам за детальністю, він додатково обробляється з використанням ЗНМ з ефективною обробкою долей пікселів (Efficient Sub-Pixel CNN (ESPCN)). Використання даної мережі дозволяє скоротити обчислювальну складність рішення задачі підвищення якості цифрового аерофотознімка в автоматичному режимі, сформувати карти ознак в просторі низької роздільної здатності і підвищити роздільну здатність за рахунок використання субпіксельного згорткового шару.

2.2 Оцінка необхідної кількості навчальних прикладів вибірки зображень для навчання ЗНМ

Кількість навчальних прикладів повинна забезпечувати репрезентативність навчальної вибірки. Розподіл кількості навчальних прикладів зображень по класах повинен бути рівномірний. Це досягається за рахунок формування однієї кількості навчальних прикладів для кожного з класів. Дотримання умови рівномірності дає змогу застосувати міру Вапніка-Червоненкіса для оцінки

достатньої кількості навчальних прикладів [42, 43]. Так, достатня кількість навчальних прикладів для навчання ЗНМ обчислюється за формулою:

$$l_{\text{дост}} \leq \frac{\ln N_{\text{ЗНМ}} - \ln \eta}{2\varepsilon^2}, \quad (2.1)$$

де $l_{\text{дост}}$ – достатня кількість навчальних прикладів одного класу для навчання ЗНМ;

$N_{\text{ЗНМ}}$ – кількість вагових коефіцієнтів ЗНМ, що навчаються;

η – надійність правильного навчання ЗНМ;

ε – точність функціонування ЗНМ [59].

Наведена в додатку А топологія ЗНМ передбачає навчання $N_{\text{ЗНМ}} = 5978677$ вагових коефіцієнтів. Візьмемо $\eta = 1$ та $\varepsilon = 0.05$, тоді $l_{\text{дост}} \leq 3120$ навчальних прикладів для кожного класу.

2.3 Визначення загальних положень процедури навчання ЗНМ

Для навчанням ЗНМ виявленню та класифікації об'єктів на цифрових знімках використовується навчання з учителем (supervised learning). Для навчання використовується навчальна вибірка з $l_{\text{дост}}$ навчальних прикладів. У процесі навчання ЗНМ пред'являється навчальний приклад, оцінюється результат її функціонування, який порівнюється з істинним результатом, оцінюється помилка E та корегуються вагові коефіцієнти w_j^{node} . Навчальна вибірка зображень для навчання ЗНМ поділяється на 2 частини у співвідношенні: навчальна частина 75%, контрольна 25%. Поділ навчальної вибірки виконується для навчання та контролю правильності функціонування ЗНМ.

Для оцінки якості навчання ЗНМ використовується помилка E правильної сегментації об'єкта ОЕПР на цифровому знімку земної

поверхні. Вона утворюється сумою двох незалежних помилок та знаходиться за формулою:

$$E = E_{box} + E_{cov}, \quad (2.2)$$

де E_{box} – помилка між реальними та визначеними в процесі функціонування ЗНМ пікселями;

E_{cov} – помилка між реальними та визначеними в процесі функціонування ЗНМ пікселями об'єкта на зображенні.

Помилка E_{cov} визначається за формулою:

$$E_{cov} = \frac{c_{px}}{a_{px} + b_{px} - c_{px}}, \quad (2.3)$$

де c_{px} – кількість пікселів цифрового знімку, які одночасно виявлені ЗНМ та реально належать зображенню об'єкта;

a_{px} – кількість пікселів цифрового знімку, яка виявлена ЗНМ, як ті, що належать об'єкту

b_{px} – кількість пікселів цифрового знімку, яка належить зображенню об'єкта.

Помилка E_{box} визначається за формулою:

$$E_{box} = \sqrt{\left(X_{px}^{class} - \bar{X}_{px}^{class}\right)^2 + \left(Y_{px}^{class} - \bar{Y}_{px}^{class}\right)^2} + \sqrt{\left(l_{px}^{class} - \bar{l}_{px}^{class}\right)^2 + \left(h_{px}^{class} - \bar{h}_{px}^{class}\right)^2}, \quad (2.4)$$

де $X_{px}^{class}, Y_{px}^{class}$ – піксельні координати центра виявленого ЗНМ об'єкта ОЕПР на зображенні;

$\bar{X}_{px}^{class}, \bar{Y}_{px}^{class}$ – реальні піксельні координати центра об'єкта ОЕПР на зображенні;

$l_{px}^{class}, h_{px}^{class}$ – висота та ширина виявленого ЗНМ об'єкта в пікселях на цифровому знімку;

$\bar{l}_{px}^{class}, \bar{h}_{px}^{class}$ – висота та ширина реального об'єкта в пікселях на цифровому знімку.

Процедура навчання зводиться до одночасного зменшення E_{cov} та E_{box} за рахунок корегування w_j^{node} . Зменшення помилок E_{cov} та E_{box} здійснюється з використанням градієнтного методу зворотного поширення похибки. Градієнтний метод зворотного поширення похибки ітеративний. Тому корегування w_j^{node} здійснюється при кожному пред'явленні ЗНМ навчального прикладу. Величина, на яку корегуються вагові коефіцієнти одного нейрона (одного пікселя ядра згортки) w_j^{node} , знаходиться за формулою:

$$\Delta w_j^{node} = -\lambda \cdot (g_t + \rho \cdot w_{j-1}^{node}) + \mu \cdot \Delta w_{j-1}^{node}, \quad (2.5)$$

де Δw_j^{node} – величина, на яку корегуються вагові коефіцієнти ЗНМ;

λ – коефіцієнт швидкості навчання;

g_t – коефіцієнт врахування попередніх значень градієнта;

ρ – коефіцієнт регуляризації;

w_{j-1}^{node} – значення вагових коефіцієнтів попереднього шару нейронів;

μ – коефіцієнт моменту;

Δw_{j-1}^{node} – величина, на яку корегувалися вагові коефіцієнти попереднього шару нейронів.

Коефіцієнт врахування попередніх значень градієнта залежить від помилок E_{cov} та E_{box} на попередніх ітераціях та знаходиться за формулою:

$$g_t = \frac{D_t}{1-\beta} \cdot \sqrt{\frac{1-\alpha}{S_t}}, \quad (2.6)$$

де α – константа яка дорівнює 0,999;

β – константа рівна 0,9;

S_t – коефіцієнт врахування попередніх значень градієнта помилки E_{cov} ;

D_t – коефіцієнт врахування попередніх значень градієнта помилки E_{box} .

Коефіцієнти S_t та D_t знаходяться за наступними формулами:

$$S_t = \alpha \cdot S_{t-1} + (1-\alpha) \cdot E_{cov}^2; \quad (2.7)$$

$$D_t = \beta \cdot D_{t-1} + (1-\beta) \cdot E_{box}. \quad (2.8)$$

Таким чином, визначена процедура навчання ЗНМ з використанням методу градієнтного зворотного поширення помилки. Вона забезпечує навчання ЗНМ по навчальний вибірці цифрових зображень.

3 МОДЕЛЮВАННЯ СЕГМЕНТАЦІЇ ОБ'ЄКТІВ МІСЬКОГО СЕРЕДОВИЩА НА ЦИФРОВИХ АЕРОФОТЗНІМКАХ

3.1 Розробка інформаційної технології автоматизованого виявлення та класифікації простих об'єктів оптико-електронної повітряної розвідки

Детальний аналіз глибоких нейронів моделей для семантичної сегментації в різних предметних областях наведено в другому розділі. Найбільш ефективними для вирішення завдань семантичної сегментації в рамках теорії глибокого навчання є різні реалізації автокодувальника (autoencoder). Для реалізації семантичної сегментації в рамках глибокого навчання був проведений аналіз основних архітектур глибоких нейронних мереж, кожна з яких в загальному випадку розглядається як мережа кодувальника, за якою слідує мережу декодера: FCN, SegNet, U-Net, FC-Densenet, E-Net & Link-Net, RefineNet, PSPNet, Mask-RCNN, G-FRNet, DecoupledNet, GAN Based Approaches. В якості критеріїв вибору архітектури глибоких нейронних мереж для сегментації об'єктів міського середовища на цифрових аерофотзнімках використовувалися [33, 42, 56, 56]:

- прикладна спрямованість архітектури;
- забезпечення передбачуваної точності розпізнавання;
- мінімальна навчальна вибірка для забезпечення передбачуваної точності розпізнавання;
- мінімальна кількість епох для навчання;
- швидкість розпізнавання в режимі реального часу.

Визначено, що найбільш повно критеріям відповідає архітектура згортосної нейронної мережі, призначена для сегментації біомедичних зображень U-Net. Зокрема аналіз результатів реалізації U-Net використовуючи Keras показав, що при використанні на тридцяти зображеннях розміром 512 * 512 при виконанні додаткової аугментации

навчальної вибірки після 10 епох розрахункова точність U-Net в реальному масштабі часу становить близько 0,97. 2.

Аналіз апаратно-програмної платформи програмного макета для сегментації об'єктів міського середовища на цифрових аерофотзнімках проводився за наступними напрямками:

- досліджено основні існуючі в даний час фреймворкі з відкритим кодом для глибокого навчання: фреймворкі для символічних обчислень (MXNET (<https://github.com/apache/incubator-mxnet>), TensorFlow (<https://www.tensorflow.org/>), Theano (<https://github.com/Theano/Theano>)); несимвольні фреймворкі (Torch (<https://github.com/torch/>), Caffe (<http://caffe.berkeleyvision.org/>));

- в якості критеріїв вибору фреймворка для реалізації U-Net для сегментації об'єктів міського середовища на цифрових аерофотзнімках використовувалися:

- сумісність з високорівневою бібліотекою Keras для швидкої реалізації Deep Learning алгоритмів; підтримка безлічі GPU;

- швидкість функціонування; наявність зручного для користувача інтерфейсу;

- можливості по моделюванню глибоких згортальних мереж і їх модифікацій;

- супроводження розробником і наявність достатньої кількості якісних посібників для вивчення та тренувальних матеріалів;

- визначено, що найбільш критеріям відповідає фреймворк машинного навчання Google TensorFlow, який в рамках реалізації другої версії включає Keras як стандартний високорівневий API;

- в якості базового інтерактивного середовища моделювання для швидкої підготовки високоточної глибокої нейронної мережі для задач сегментації пропонується використовувати хмарний сервіс Google Colab (<https://colab.research.google.com/notebooks/intro.ipynb>).

3.2 Обґрунтування і формування репрезентативної навчальної вибірки при малому обсязі даних

Формування навчальної вибірки при глибокому навчанні вимагає попередню підготовку даних, їх розмітку (доразметку) і вивантаження в необхідному форматі. При розробці даного класу систем рішення відповідної задачі є трудомістким процесом і може займати велику частину часу, особливо в разі виконання «ручний» розмітки, наприклад з використанням Adobe Photoshop.

У зв'язку з цим проведено аналіз і обґрунтування вибору програмної утиліти для підготовки навчальних даних, розробка класифікації об'єктів на знімку для вирішення завдання автоматизованого формування навчальної вибірки і безпосередньо для сегментації об'єктів міського середовища на цифрових аерофотзнімках. За результатами аналізу:

- досліджено основні існуючі в даний час інструменти анотації, які можна використовувати для розмітки медичних зображень:
- спеціалізовані програмні утиліти для розмітки зображень при вирішенні завдань глибокого навчання: онлайн-інструмент анотації для створення баз даних зображень для дослідження комп'ютерного зору [LabelMe];
- онлайн сервіс управління, анотування і підготовки датасета для глибокого навчання Supervise (Supervise.ly).
- основний недолік LabelMe – відсутність інструментів для виконання попиксельной сегментації. Вільна версія Supervise представляє можливості по онлайн сервіс по роботі з датасета і необхідні інструменти анотирования цифрових знімків, в тому числі попиксельно сегментацію;
- розроблена методика створення репрезентативної навчальної вибірки при малому обсязі даних для виявлення посудин з використанням Supervise:
- з використанням інструментарію Supervise анотує перша частина знімків, які є найбільш чіткими і включають об'єкти розпізнавання (приклад

класифікації об'єктів для розпізнавання наведено нижче). При цьому інструментарій Supervise дозволяє створювати шаблон кожного класів, які в подальшому використовуються для прискорення розмітки з подальшим доопрацюванням їх в ручну, що підвищує якість і швидкість розмітки.

- з використанням інструментарію Supervise для кожного анотованого знімка виконуються n операцій аугментации (автоматизованого перетворення знімка за рахунок зміни масштабу, поворотів і зміщення зображень), мінімум три операції.

- формується безліч з N початкових знімків і на основі моделі U-Net і початкової навчальної вибірки виконується грубе сегментування. Результати сегментування додатково допрацьовуються з використанням інструментарію Supervise і над ними виконуються операції аугментірованія. Таким чином отримуємо репрезентативну навчальну вибірку в розмірі $N * n$ знімків, що теоретично є прийнятним для архітектури U-Net.

- класифікацію об'єктів на знімках пропонується розглядати для двох випадків:

- для випадку розробки навчальної вибірки;
- безпосередньо для вирішення завдання сегментації об'єктів міського середовища на цифрових аерофотзнімках.

При цьому класи розглядаються, як класи-елементи що належать і не належать об'єктам міського середовища. Приклад сегментованих зображень міського середовища (будівля та не будівля) наведено на рисунку 3.1.

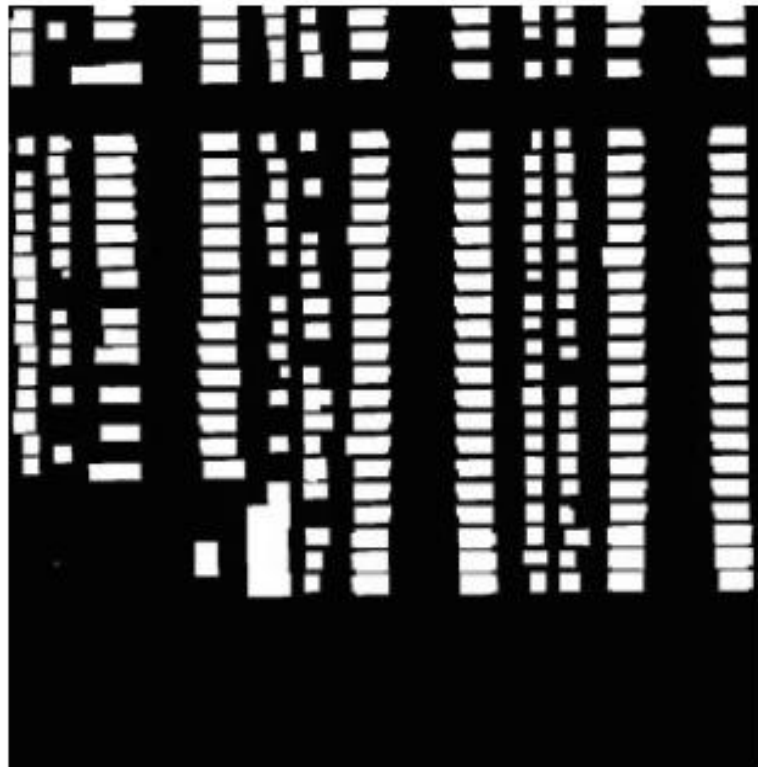


Рисунок 3.1 – Приклад зображення та його маски с набору даних Inria

3.3 Опис експерименту

Експерименту полягав в сегментації на основі глибокої нейронної мережі з постобробкою результатів передбачення з використанням операцій бинаризації на передбачених знімках. Короткий опис експерименту:

- в якості вхідних даних для навчання нейронної використовуються знімки розміром 960x1280 з аугментацією «fleeper» з використанням інструментарію *supervisely* (фрагменти зображень).

- в якості нейронної мережі використовується топологія мережі U-net зі входом 480x640 (кількість карт ознак однієї гілки: 32 → 64 → 128 → 256 → 256 + 256 (*keras* і *tensorflow* з використанням мови програмування *python*)).

Середній час навчання 9 – 11 годин (рис. 3.2).

- постобробка результатів прогнозування включає обробку з використанням суб'єктивного порогу бинаризації на основі відповідних операцій бібліотеки *opencv* для *python*.

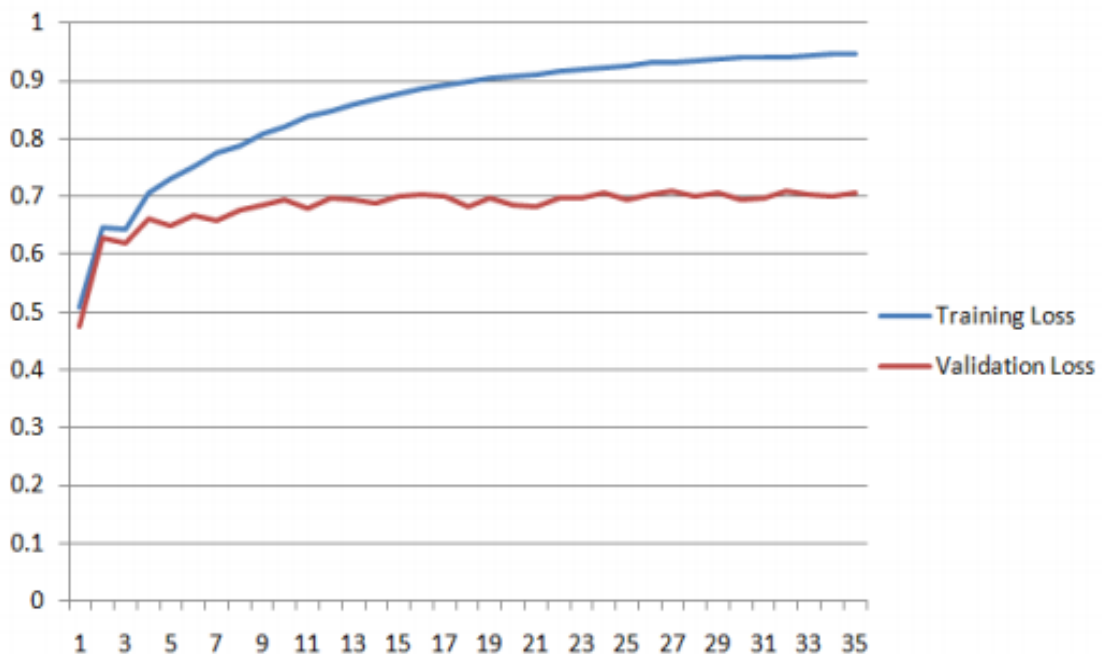


Рисунок 3.2 – Процес навчання модифікованої мережі

Поріг бинаризації: «Global Thresholding», значення порога 127.

Точність навчання: основі валідаційної вибірки 8% по метриці «dice» з функцією навчання «binary crossentropy».

У загальному вигляді основні етапи розвитку експерименту полягали в наступному:

- з метою подальшого підвищення точності сегментації на основі нейронних мереж застосовується технологія «transfer learning». Для цього в якості енкодера базової моделі U-net використовувалася з бібліотеки моделей Keras предобученная модель, наприклад VGG-16, на бібліотеці зображень ImageNet.

- навчання отриманої гібридної нейронної мережі в рамках технології «transfer learning» виконувалося в два етапи:

- перший етап – для зменшення часу навчання використовується технологія «bottleneck features» згідно з якою на виході предобученої моделі знімаються ознаки (в даному випадку у енкодера в форматі VGG-16) і подаються на вхід навченою частини мережі (в даному випадку це декодер в форматі U- net) і в подальшому навчається лише друга частина (т. е. в даному випадку тільки декодер). Особливістю моделі U-net є наявність проміжних зв'язків між енкодером і декодером. Для декодера має місце кілька входів (в даному випадку 5), що значно ускладнює використання технології «bottleneck features» при семантичній сегментації (на відміну від нейронного класифікатора у якого тільки один вхід). В ході дослідження розроблено унікальний підхід до вирішення даної проблеми (візуалізація топології отриманої гібридної нейронної мережі представлена в додаток А). Даний підхід дозволив скоротити одну епоху навчання на однакових даних з 1 години до 5 хвилин (при цьому використовувати всі переваги попередньо навченої моделі на базі даних ImageNet). В якості попереднього навчання декодера використовувалася база зображень, яка в даному експерименті містить 3600 знімків;

- другий етап – власне тонка настройка. На етапі розморожуються 1-2 шари енкодера, з'єднуються ваги енкодера (попередньо навченої моделі VGG-16 на бібліотеці зображень ImageNet) і отримані на першому етапі ваги декодера і виконується тонка настройка розморожених шарів енкодера і шарів декодера.

Таблиця 3.1 – Результати оцінки сегментації об'єктів

Тип нейромережевої моделі	Точність	Повнота
SegNet	82,7	81,6
FCN	72,7	75,0
U-Net	82,3	82,2
Розроблений в роботі підхід на основі модифікованої U-Net	84,8	84,1

Завдяки отриманим результатам очікується, що може бути підвищена точність сегментації на основі нейронних мереж в порівнянні з базовими методами (за рахунок використання використання ваг предобученної моделі), при цьому зменшується час навчання за рахунок використання технології «bottleneck features» (таблиця 3.1).

ВИСНОВКИ

1. В результаті проведеного аналізу відомих методів для автоматизації виявлення так класифікації об'єктів на цифрових знімках. Встановлено, що методи, які реалізують піксельно-орієнтований підхід доцільно застосовувати для виявлення та класифікації об'єктів на цифрових знімках через їх більшу точність, а з об'єктно-орієнтованих методів support vector machine не може бути реалізований в режимі часу, наближеному до реального, а детектори HOG, Haar, LBP, SIFT не дозволяють здійснювати одночасне виявлення та класифікації об'єктів на зображеннях.

2. Аналіз основних напрямків підходів з автоматизації процесів обробки даних виявлення так класифікації об'єктів на цифрових знімках показав необхідність застосування знання-орієнтованих методів для семантичної сегментації об'єктів міського середовища.

3. Досліджено можливості використання згорткової нейронної мережі для автоматизації процесів виявлення та класифікації об'єктів на цифрових аерофотознімках. Їх використання дозволяє забезпечити часткову стійкість до змін масштабу, зсуву, повороту, ракурсу та інших спотворень зображень об'єктів на цифрових знімках під час виявлення їх ознак та подальшої сегментації.

4. У модифікованій моделі U-Net була випробувана технологія "battlenack" в поєднанні з технологією «transfer learning». Валідаційні помилка при цьому зменшилася до 7 – 8%. З огляду на дуже обмежений обсяг вибірки було применіна кілька сучасних підходів (навчання з самого початку, з використанням отриманих ваг, та з використанням попередньо навчених моделей Keras різних топологій у якості енкодера, технології "batellnack", поетапного розморожування шарів і т.д.). Для того, щоб підвищити точність розпізнавання необхідно додатково збільшувати навчальну вибірку. Зокрема в даний час існують підходи, які дозволяють

синтезувати "штучні" зображення по заданій тематиці для навчання нейронних мереж (не аугментація старих, а отримання саме нових зображень). На даний момент в індустрії "deep learning" це ціле окремий напрямок, який вважається досить перспективним.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Pearl Mary Samuel and Thanikaiselvan Veeramalai Multilevel and Multiscale Deep Neural Network for Retinal Blood Vessel Segmentation. // Symmetry 2019, 11(7), 946; <https://doi.org/10.3390/sym11070946>.
2. Understanding Deep Learning Techniques for Image Segmentation, arXiv:1907.06119v1.
3. Fully Convolutional Networks for Semantic Segmentation, CVPR 2015, <https://arxiv.org/abs/1411.4038>.
4. A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, 2015, <https://arxiv.org/abs/1511.00561>.
5. Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015, <https://arxiv.org/abs/1505.04597>.
6. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation, 2016, <https://arxiv.org/abs/1611.09326>.
7. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation, 2016, LinkNet: Feature Forwarding: Exploiting Encoder Representations for Efficient Semantic Segmentation, 2017, <https://arxiv.org/abs/1606.02147>.
8. Multi-Path Refinement Networks for High-Resolution Semantic Segmentation, CVPR 2017, <https://arxiv.org/abs/1611.06612>.
9. Pyramid Scene Parsing Network, CVPR 2017, <https://arxiv.org/abs/1612.01105>.
10. Mask-RCNN: extends Faster R-CNN, 2017, <https://arxiv.org/abs/1703.06870>.
11. Gated Feedback Refinement Network for Dense Image Labeling, CVPR 2017, <http://www.cs.umanitoba.ca/~ywang/papers/cvpr17.pdf>.
12. Decoupled Deep Neural Network for Semi-supervised Semantic Segmentation, NIPS 2015, <https://arxiv.org/abs/1506.04924>.

13. Semi and Weakly Supervised Semantic Segmentation Using Generative Adversarial Network, 2017, <https://arxiv.org/abs/1703.09695>.
14. <https://github.com/zhixuhao/unet#implementation-of-deep-learning-framework-unet-using-keras/>
15. http://brainiac2.mit.edu/isbi_challenge/.
16. <https://itk.org/>.
17. <http://www.fil.ion.ucl.ac.uk/spm/>.
18. <https://sourceforge.net/projects/niftyseg/>.
19. <http://labelme.csail.mit.edu/Release3.0/>.
20. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems. pp. 1097-1105, 2012.
21. Y. LeCun, and Y. Bengio, "Convolutional networks for images, speech, and time series", The handbook of brain theory and neural networks. vol.3361, no 10, 1995.
22. I. Goodfellow, Y. Bengio, and A. Courville, Deep learning The MIT Press, 2016. ISBN 978-0-262-03561-3.
23. ДСТУ 4758:2007, Дистанційне зондування Землі з космосу. Оброблення даних. Україна, Київ: Держспоживстандарт України, 2007.
24. И. И. Юсипов, А. И. Калякулина, и С.Ю. Филочева, Классификация объектов городского ландшафта по аэрофотоснимкам высокого разрешения. [Електронний ресурс]. – Режим доступа: <http://www.uic.unn.ru/~zny/ml/Projects/YusipovKalyakulinaFilicheva.pdf>
25. R. L. Davis, "The Joint Service Imagery Processing System (JSIPS)," in IEEE Aerospace and Electronic Systems Magazine, [Електронний ресурс]. Режим доступа: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=177446&isnumber=4481>
26. Joint Service Imagery Processing System [JSIPS]. [Електронний ресурс]. Режим доступа: <http://www.globalsecurity.org/intell/systems/jsips.htm>.

27. I. Goodfellow, Y. Bengio, and A. Courville, Deep learning The MIT Press, 2016. ISBN 978-0-262-03561-3.

28. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems. pp. 1097-1105, 2012.

29. P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, pp. I-511-I-518 vol.1. doi: 10.1109/CVPR.2001.990517

30. D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors", European conference on computer vision. Springer Berlin Heidelberg, pp. 340-35, 2012.

31. К. В. Воронцов, Машинное обучение [Электронный ресурс]. – Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение

32. Е. В. Бодянский, и О. Г. Руденко. Искусственные нейронные сети: архитектуры, обучение, применение. Харьков, Украина : ТЕЛЕТЕХ, 2004. ISBN 966-95416-2-2.

33. W. S. McCulloch, and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", Bulletin of mathematical biology vol.52, no.1-2, pp. 99-115, 1990.

34. R. Hecht-Niselen, "Kolmogorov's mapping neural network existence theorem", in First IEEE International Joint Conference on Neural Networks, Inst. Of Electr. and Electron. Eng., San Diego, Calif., June 1987, pp. 21-24

35. D. H. Hubel, and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", The Journal of physiology. vol. 160 no. 1, pp. 106-154, 1962.

36. K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", Biological cybernetics. vol. 36, no 4, pp. 193-202, 1980.