

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп'ютерних наук та кібернетики
Кафедра математичної інформатики

«До захисту допущено»

Завідувач кафедри

В.М.Терещенко

_____ (підпис)

« ____ » _____ 20__ р.

**Дипломна робота
на здобуття ступеня бакалавра**

за спеціальністю 122 Комп'ютерні науки

на тему:

**АНАЛІЗ МЕТОДІВ СЕГМЕНТАЦІЇ ЗОБРАЖЕНЬ ДЛЯ
ВИРІШЕННЯ ЗАДАЧ РОЗУМІННЯ ВІЗУАЛЬНИХ СЦЕН**

Виконала студентка 4 курсу

Титаренко Катерина Олексіївна

_____ (підпис)

Науковий керівник:

професор, доктор фіз-мат наук

Терещенко Василь Миколайович

_____ (підпис)

Засвідчую, що в цій дипломній
роботі немає запозичень з праць інших
авторів без відповідних посилань.

Студент

_____ (підпис)

Київ – 2021

РЕФЕРАТ

Обсяг роботи 55 сторінок, 28 ілюстрацій, 31 джерело посилань та 6 таблиць.

КОМП'ЮТЕРНИЙ ЗІР, МАШИННЕ НАВЧАННЯ, СЕМАНТИЧНА СЕГМЕНТАЦІЯ ЗОБРАЖЕНЬ, ВІЗУАЛЬНІ СЦЕНИ, НЕЙРОННІ МЕРЕЖІ, PYTHON, GOOGLE COLABORATORY

Об'єктом дослідження є задача сегментації зображень.

Предметом дослідження є алгоритми, що розв'язують дану задачу.

Метою роботи є вивчення інформації про існуючі алгоритми, виділення певного набору алгоритмів на основі зібраних знань, самостійне порівняння їх роботи на декількох різних наборах даних, аналіз отриманих результатів та надання рекомендацій щодо використання різних методів.

Основним методом та засобом дослідження є розробка програм мовою програмування Python за допомогою середовища Google Colaboratory.

У роботі було детально досліджено задачу сегментації зображень, її цілі та сфери використання, оглянуто основні методи її розв'язування, а також вже існуючі наукові роботи, що намагаються їх порівнювати. На основі отриманої інформації був зроблений обґрунтований вибір набору алгоритмів для порівняння та критеріїв, за якими їх буде порівняно. Після цього були розроблені та запущені на різних наборах даних програмні засоби з безпосередньою реалізацією обраних методів, проведений аналіз результатів та надано рекомендації щодо використання алгоритмів у різних випадках.

На відміну від більшості вже існуючих досліджень за аналогічним напрямком, ця робота охоплює відносно велику кількість різних за структурою та принципом роботи методів, а також порівнює їх роботу у кількох випадках

використання. Це надає можливість отримати значно більше корисної інформації для порівняння.

Результати цього дослідження можуть бути застосовані як певне керівництво та джерело корисної інформації при вирішенні задачі сегментації зображень. Нижче надана як детальна теоретична інформації щодо цієї задачі в загальному, так і опис та аналіз конкретного практичного експерименту.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ	6
ВСТУП	7
РОЗДІЛ 1. ОГЛЯД ІСНУЮЧИХ АЛГОРИТМІВ ТА НАУКОВИХ РОБІТ	11
1.1 Класифікація алгоритмів	11
1.2 Огляд існуючих досліджень	12
РОЗДІЛ 2. ОБРАНІ АЛГОРИТМИ	16
2.1 Загальний опис процесу сегментації зображень	16
2.2 Обґрунтування вибору	17
2.3 Опис обраних алгоритмів	18
2.3.1 Алгоритми виділення ознак	18
2.3.2 Класичні структурні алгоритми сегментації	21
2.3.3 Нейромережові методи сегментації	24
2.4 Умови та параметри порівняння алгоритмів	27
РОЗДІЛ 3. РЕАЛІЗАЦІЯ ПОСТАВЛЕННОЇ ЗАДАЧІ	30
3.1 Обрані набори даних	30
3.1.1 HAM	30
3.1.2 LaPa	31
3.1.3 Sky	33
3.1.4 CamVid	33
3.2. Реалізація мовою програмування Python	35

РОЗДІЛ 4. ОТРИМАНІ РЕЗУЛЬТАТИ	37
4.1 Набір даних НАМ	37
4.2 Набір даних LaPa	40
4.3 Набір даних Sky	42
4.4 Набір даних CamVid	44
4.5 Висновки розділу	46
4.6 Рекомендації, можливості подальших розробок	47
ВИСНОВКИ	50
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	51

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ

LEGION – Locally Excitatory Globally Inhibitory Oscillator Network

CNN - Convolutional Neural Network

RF – Random Forest

LGBM – Light Gradient Boosting Machine

SVM – Support Vector Machine

FPN - Feature Pyramid Network

VGG - Visual Geometry Group

EFB - Exclusive Feature Bundling

GOSS - Gradient-based One Side Sampling

IOU - Intersection Over Union

ВСТУП

Сегментація зображень - це завдання комп'ютерного зору, метою якого є маркування певних областей зображення відповідно до того, який саме об'єкт на них знаходиться. Мета сегментації - спростити або змінити представлення візуальної сцени на щось більш структуроване і легке для аналізу. Ця техніка часто використовується для визначення місця розташування об'єктів та їх меж на зображеннях.

Результат сегментації зображення - це набір сегментів, які разом покривають все зображення, або набір контурів, виявлених на зображенні (рис.1)

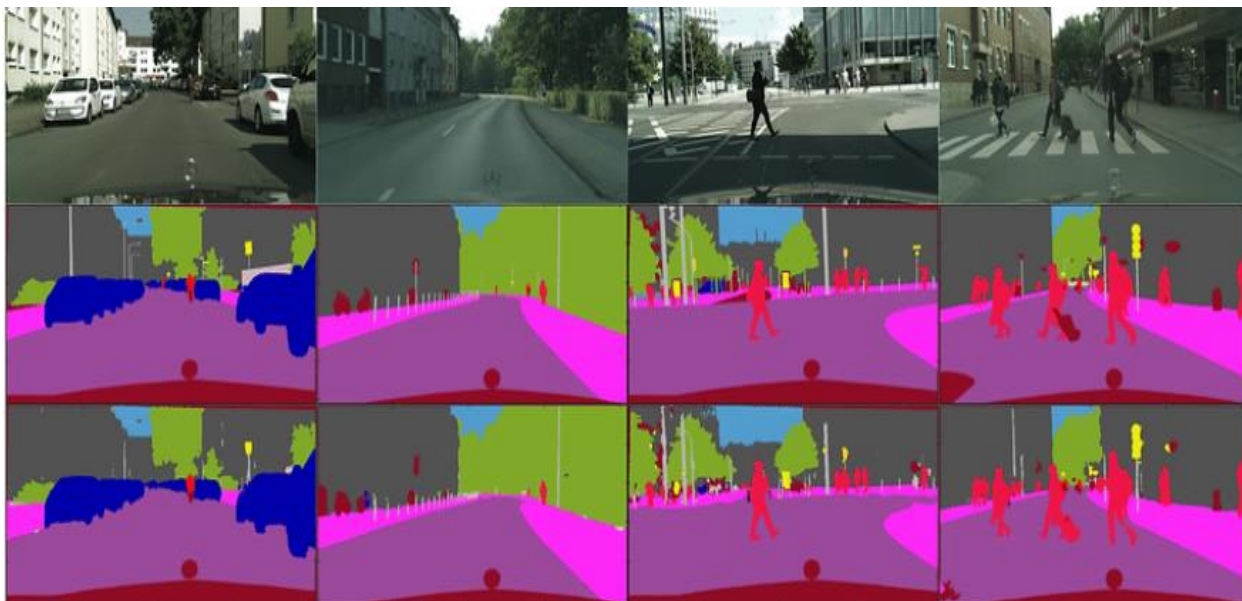


Рисунок 1 Результат сегментації на прикладі набору даних «Cityscapes»[1]

Проблематика , актуальність та практичне застосування теми роботи. Сегментація зображень наразі є дуже актуальною та широко вивчаємою задачею у сфері машинного навчання[3]. Найпопулярнішими областями її практичного застосування є:

а) Медичні знімки

1) Виявлення пухлин та інших патологій

2) Визначення обсягів тканин

3) Вивчення анатомічної структури

б) Виділення об'єктів на супутникових світлинах

в) Системи автоматизованого управління дорожнім рухом та його аналізу

г) Системи відеоспостереження

г) Дуже часто сегментація зображень є частиною систем розпізнавання, бо дозволяє відділити певний об'єкт для подальшої обробки[5].

д) Виявлення сторонніх об'єктів серед їжі[6]

е) Різноманітні додатки для обробки фотографій (накладання макіяжу, медична діагностика, фоторедакція тощо)

Отже, сегментація зображень є важливим напрямком у комп'ютерному зорі, що стала невід'ємною частиною у багатьох проектах з цієї галузі. Проте великий відсоток вже наявних практичних досліджень для цієї задачі фокусуються лише на певному вузькому класі алгоритмів[5,6,7,8]. Більш того, сегментація зображень досить активно розвивається і постійно з'являються нові, ще не широко досліджені методи або модифікації вже існуючих методів. Тому їх огляд та порівняння між собою є досить важливим.

Новизна в роботі, можливі сфери застосування. У цій роботі порівнюється відносно велика кількість різних за принципом роботи та структурою алгоритмів у кількох випадках використання, на відміну від більшості вже існуючих досліджень з аналогічною ідеєю. Це надає можливість отримати

значно більше корисних даних для порівняння, тим самим зробивши його ширшим та більш об'єктивним.

Це дослідження надає багато інформації, що може бути використана у практичних проєктах з сегментації зображень. По-перше, дана робота включає велику кількість теоретичного матеріалу, що може бути корисним для формування ґрунтовного уявлення про задачу, її актуальність, базовий процес виконання та існуючі методи розв'язування. По-друге, практична частина цього дослідження містить результати, що можуть допомогти при виборі конкретного алгоритму. Також і програмний код, що був написаний у рамках виконання даної роботи, може бути використаний для реалізації певного проєкту.

Варто зазначити, що сфера сегментації зображень є дуже широкою і важко покрити усю велику кількість нюансів які можуть виникнути при розв'язуванні цієї задачі. А тому ця робота не є джерелом беззаперечної істини та не дає універсального керівництва до того, як сегментувати зображення. Проте опис конкретного експерименту та його результатів все ще вносить свою долю у загальні дослідження для кращого розуміння вивчаємої задачі. Також наприкінці цієї роботи наводяться ідеї для можливих подальших розробок з цього напрямку.

Мета й завдання роботи. Існують наступні найпоширеніші підходи до сегментації зображень:

а) Семантична сегментація (semantic segmentation)[2]- це підхід, який для кожного пікселя визначає клас об'єкта. Наприклад, коли всі люди на фігурі сегментовані як один клас, а фон - як інший клас.

б) Сегментація примірників (instance segmentation)[3] - це підхід, який для кожного пікселя визначає примірник об'єкта. Він розрізняє кожен окремий об'єкт

на зображенні, наприклад, кожна людина на фотографії сегментується як окремий об'єкт.

в) Паноптична сегментація (panoptic segmentation) [3] - комбінація сегментації примірників і семантичної сегментації. Ми асоціюємо з кожним пікселем два значення: його мітку класу і номер примірника.

Метою цієї роботи було обрано порівняння різних алгоритмів семантичної сегментації на різних наборах даних (фотографій). В основному порівняння виконувалось для бінарної сегментації (коли потрібно виділити лише один клас певних об'єктів, наприклад, машини, а все інше сегментувати як «фон»). Але також будуть наведені результати для спроб мультикласової сегментації тими самими алгоритмами. Різницю між бінарною та мультикласовою сегментацією можна побачити на рисунку нижче:

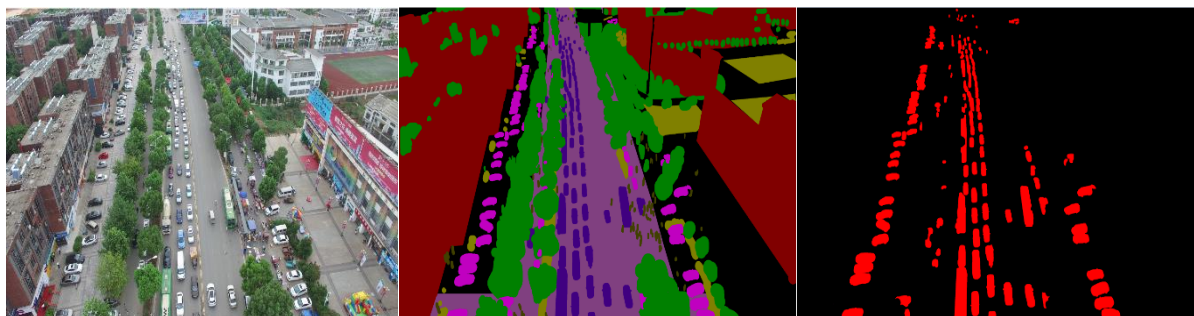


Рисунок 2 Різниця між бінарною та мультикласовою сегментацією (справа та посередині відповідно). Зліва знаходиться оригінальне зображення

РОЗДІЛ 1.

ОГЛЯД ІСНУЮЧИХ АЛГОРИТМІВ ТА НАУКОВИХ РОБІТ

1.1 Класифікація алгоритмів

У науковій літературі зустрічається багато різних способів сегментації зображень. Нижче наведені декілька з них.

За природою алгоритму[6]:

а) класичні підходи (в основному базуються на математичних та статистичних методах). Сюди входять методи з використанням гістограм, розростання областей, встановлення рівня тощо. Термін «класичні алгоритми» використовується у подальших розділах цієї роботи саме у цьому значенні.

б) методи на основі штучного інтелекту (наприклад, самоорганізаційна карта Кохонена, нейронна мережа Гопфілда, різноманітні варіації LEGION та CNN) [11].

в) змішані техніки, а також такі, що не потрапляють до жодної з вищезазначених категорій [6].

За даними, що використовуються для сегментації[7]:

а) структурні методи (базуються на інформації про певні ділянки (регіони) зображення, що сегментується)

б) стохастичні методи (сегментують зображення лише на основі даних про кожен окремий піксель)

в) змішані методи (використовують інформацію як про окремі пікселі, так і про регіони)

За визначальними критеріями сегментації[7]:

а) такі, що сегментують на основі інформації про різницю або «розриви» між певними регіонами зображень(наприклад, сюди потрапляють всі алгоритми, що базуються на знаходженні кутів).

б) такі, що сегментують на основі інформації про схожість між пікселями, що належать до одного сегменту. Сюди входять методи встановлення рівнів, розростання областей та усі методи кластеризації.

Деякі джерела також виокремлюють наступні класи алгоритмів в залежності від їх основної концепції[7][8]:

а) базуються на виділенні регіонів

б) на виділенні кутів

в) на кластеризації

г) на встановленні рівнів

г) на часткових похідних

д) на штучному інтелекті

Також, як і багато інших алгоритмів зі сфери машинного навчання, методи сегментації зображень можна розділити на ті, що «навчаються з учителем» (supervised learning) та «без учителя» (unsupervised learning).

1.2 Огляд існуючих досліджень

Задача сегментації зображень є актуальною досить давно, тому було зроблено чимало досліджень з метою описання та порівняння різних методів її вирішення [6, 7, 8, 11]. Але, водночас, ця сфера розвивається дуже швидко, велика

кількість широко вживаних сьогодні алгоритмів були розроблені зовсім нещодавно (наприклад, LinkNet[13] - у 2017 році). Тому деякі джерела, що були актуальні у свій час, наразі не описують усіх ефективних алгоритмів, що використовуються у сучасних задачах комп'ютерного зору. Також деякі роботи були написані тоді, коли нейронні мережі та глибинне навчання (deep learning) застосовувалися значно менше[7][8], тому дослідження не приділили їм належної уваги.

Узагальнюючи всі джерела, вивчені для написання цієї роботи, можна зробити наступні висновки:

а) Всі роботи свідчать, що наразі немає універсальних правил щодо вибору методів сегментації. Ця сфера є досить широкою, кожен алгоритм має багато специфічних нюансів. Більше того, зображення, що використовуються для сегментації, також мають велику кількість різноманітних параметрів (кольори, кількість та розміри зображених об'єктів, наявність шумів, насиченість, контрастність тощо). Деякі алгоритми були розроблені для певних типів зображень (наприклад Unet[14] - для біомедичних знімків), тому результати їх застосування на даних різного характеру можуть бути дуже неоднорідними. Проте, не дивлячись на всю неоднозначність, вивчення та порівняння різних методів сегментації та нюансів їх використання, що і є метою даної роботи, має великий зміст для розвитку сфери машинного навчання.

б) Багато класичних методів є досить специфічними та накладають обмеження на зображення, для яких можуть бути успішно використані (наприклад, деякі погано працюють на картинках з низьким контрастом кольорів чи великою кількістю кутів). Також вони можуть мати велику кількість складних обчислень, потребуючи багато обчислювальних ресурсів. Водночас, на певних

наборах даних, алгоритми, засновані на штучному інтелекті, можуть працювати повільніше за рахунок того, що потребують час на тренування.[7]

в) Виділяються наступні тенденції у розвитку сфери сегментації зображень[8]:

1) Використання декількох технік разом досягнення кращих результатів.

2) Все більше використання штучного інтелекту та нейронних мереж, навіть у комбінації з класичними підходами (наприклад, вибір параметру k за допомогою нейронних мереж для подальшого використання методу k середніх). Останні розробки свідчать, що нейронні мережі показують дійсно вражаючі результати у напрямку сегментації.

3) Швидка розробка нових алгоритмів, постійне збільшення їх кількості.

г) Не дивлячись на те, що було зроблено чимало досліджень в області сегментації зображень, лише невелика їх частина є практичними, що порівнюють роботу різних алгоритмів та дають підстави для висновків та рекомендацій щодо специфіки їх використання у різних випадках. Звичайно, проводилися деякі порівняння, але більшість з тих, що були знайдені та вивчені у рамках підготовки до написання цієї роботи є або обмеженими до певного класу алгоритмів (наприклад, порівнюють лише методи глибинного навчання чи лише різні варіації SVM), або не мають достатньо даних для повноцінних висновків. Приклад зі статті [11] можна побачити на рисунку 1.1 нижче.

Це дає ще більше підстав для проведення подібного дослідження і більш глибокого вивчення різних алгоритмів у рамках цієї роботи.

Model	2012 PASCAL VOC (mIoU)	PASCAL-Context (mIoU)	2016 COCO (AP)	2016 COCO (AR)	2017 COCO (AP)	Cityscapes (mIoU)
FCN	62.2	X	X	X	X	X
ParseNet	69.8	40.4	X	X	X	X
Conv & Deconv	72.5	X	X	X	X	X
FPN	X	X	X	48.1	X	X
PSPNet	85.4	X	X	X	X	80.2
Mask R-CNN	X	X	37.1	X	41.8	X
DeepLab	79.7	45.7	X	X	X	70.4
DeepLabv3	86.9	X	X	X	X	81.3
DeepLabv3+	89.0	X	X	X	X	82.1
PANet	X	X	42.0	X	46.7	X
EncNet	85.9	52.6	X	X	X	X

Overview of the scores of the models over the 2012 PASCAL VOC dataset (mIoU), the PASCAL-Context dataset (mIoU), the 2016 / 2017 COCO datasets (AP and AR) and the Cityscapes dataset (mIoU)

Рисунок 1.1 Приклад результатів дослідження, що порівнює лише методи, засновані на нейронних мережах (зі статті [11])

РОЗДІЛ 2.

ОБРАНІ АЛГОРИТМИ

2.1 Загальних опис процесу сегментації зображень

Задачу семантичної сегментації зображень можна звести до задачі класифікації: алгоритм повинен класифікувати кожен піксель картинки як такий, що належить до певного класу. Як відомо, методи класифікації отримують числові дані на вхід, а отже першим етапом процесу сегментації зображень є виділення з картинок числових даних, що будуть використовуватися для подальшої роботи. Цей процес називається виділенням ознак (feature extraction) і існує велика кількість різних методів (як класичних, так і нейромережевих) для розв'язування такої задачі. Після цього починається робота самого алгоритму, яка, в залежності від його особливостей, може складатися з різних етапів. Загальну схему процесу сегментації зображень можна побачити на рисунку нижче:

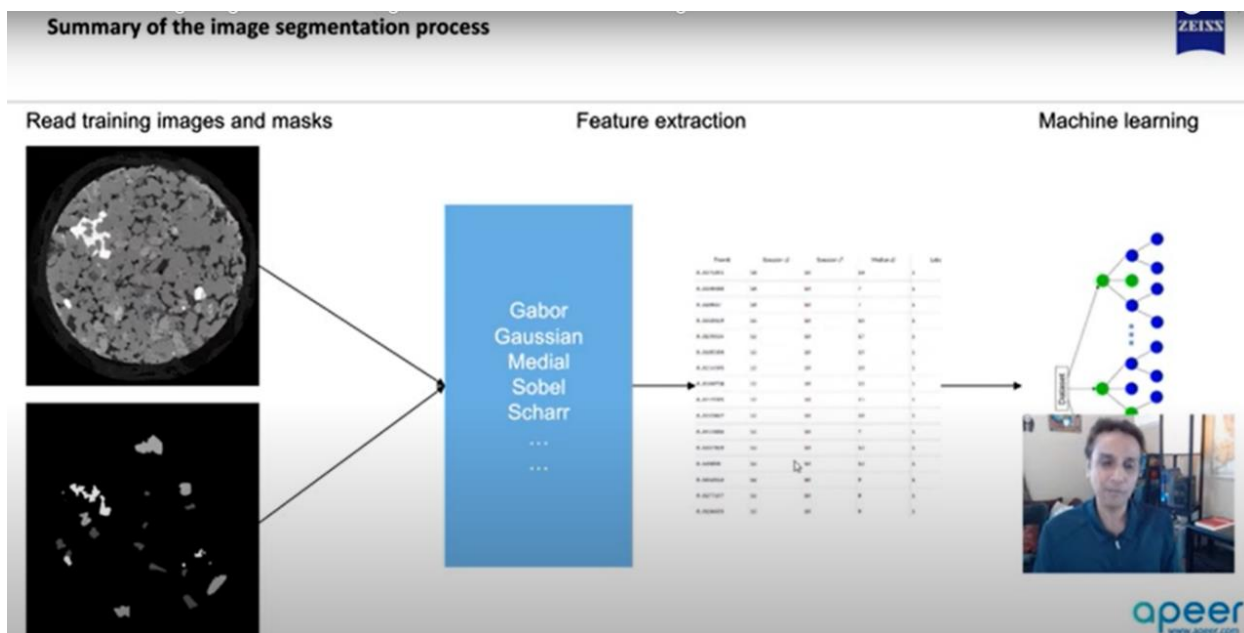


Рисунок 2.1 Процес сегментації зображень (з джерела [15])

2.2 Обґрунтування вибору

Як зазначалося раніше, великий відсоток вже наявних практичних досліджень фокусуються на певному вузькому класі алгоритмів (наприклад, порівнюють лише методи глибинного навчання). А тому при виборі методів, що були застосовані у рамках даної роботи, хотілося врахувати більш широкий спектр підходів та порівняти між собою дуже різні (як принципово, так і структурно) алгоритми. Для цього було обрано 3 основних підходи: виділення ознак класичними методами та застосування їх для класичних алгоритмів сегментації, виділення ознак нейромережевими методами та застосування їх для класичних алгоритмів сегментації, виділення ознак нейромережевими методами та застосування їх для нейромережових методів сегментації.

В якості конкретних класичних алгоритмів сегментації було обрано Random forest (RF), Support vector machine (SVM) та Light gradient boosting machine (LGBM). Серед нейромережових методів сегментації було використано 3 види CNN: Unet, LinkNet та FPN.

Для виділення ознак класичними методами було обрано фільтри Габора, Гауса, медіанні фільтри а також декілька інших фільтрів для виділення кутів. В якості нейромережових алгоритмів виділення ознак використовувалися VGG та efficientnet.

Усі обрані алгоритми є досить актуальними та використовуваними, а тому було доцільно та корисно порівняти їх у рамках даної роботи. Більш детальний опис усіх названих методів наведено у наступному розділі.

На платформі Youtube є декілька експериментальних відео з порівнянням деяких з наведених вище алгоритмів[9][12]. Вони дають корисні та цікаві результати, але, як і в багатьох інших дослідженнях, ці експерименти

проводилися лише на одному наборі зображень, а також класичні і нейромережеві методи не порівнювалися між собою.

Також варто зауважити, що обраний набір конкретних алгоритмів не є вичерпним або єдиним вірним. Як вже було сказано, сфера сегментації зображень є дуже широкою та швидко розвиваючоюся, а тому цілком можливо провести аналогічні експерименти з іншими методами. Також і до результатів описаного у цій роботі дослідження варто ставитися як до таких, що мали місце на конкретних техніках та наборах даних за конкретних налаштувань та початкових умов. Вони не дають універсального керівництва до того, як сегментувати зображення та не покривають усю велику кількість нюансів які можуть виникнути при розв'язуванні цієї задачі.

2.3 Опис обраних алгоритмів

2.3.1 Алгоритми виділення ознак

2.3.1.1 Класичні структурні підходи

Для виділення ознак класичними методами було обрано фільтри Габора, Гауса, медіанні фільтри а також декілька інших фільтрів для виділення кутів (оператор Собеля, оператор Робертса тощо).

Фільтр Гаусса[17] - фільтр розмиття зображення, який використовує нормальний розподіл (також відомий як гаусівський) для обчислення перетворення, що застосовується до кожного пікселя зображення. Приклад його роботи зображено на рисунку 2.2



Рисунок 2.2 Приклад роботи гаусівського фільтра (з джерела[17])

Фільтр Габора[16] є лінійним фільтром, що використовується для аналізу текстури зображень. Деякі джерела стверджують, що аналіз зображень за допомогою фільтрів Габора схожий на їх сприйняття в зоровій системі людини. Було виявлено, що вони особливо підходять для відображення текстури і розрізнення.

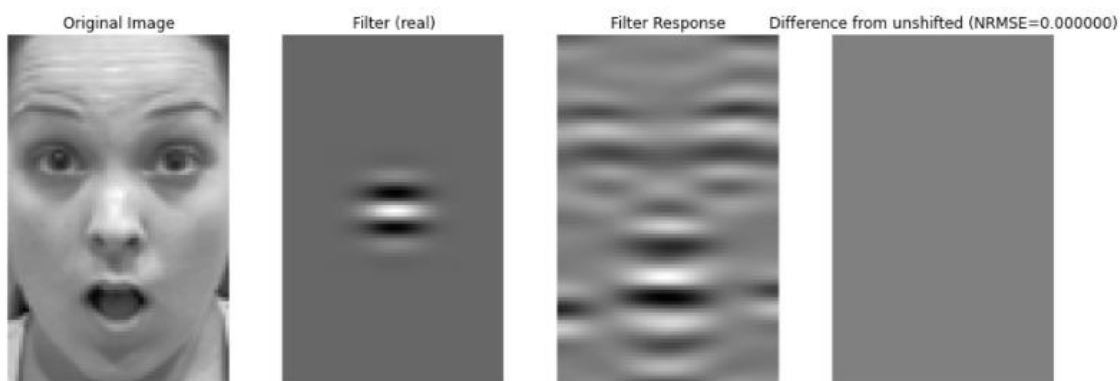


Рисунок 2.3 Приклад роботи фільтра Габора (з джерела [16])

Медіанний фільтр[17] - один з видів нелінійного усереднення. Значення пікселів записуються в ряд, ряд сортується і визначається медіана ряду. Застосування цього фільтра дає хороші результати для збереження перепадів відтінків. У порівнянні з лінійними фільтрами медіанний краще зберігає контури зображення.

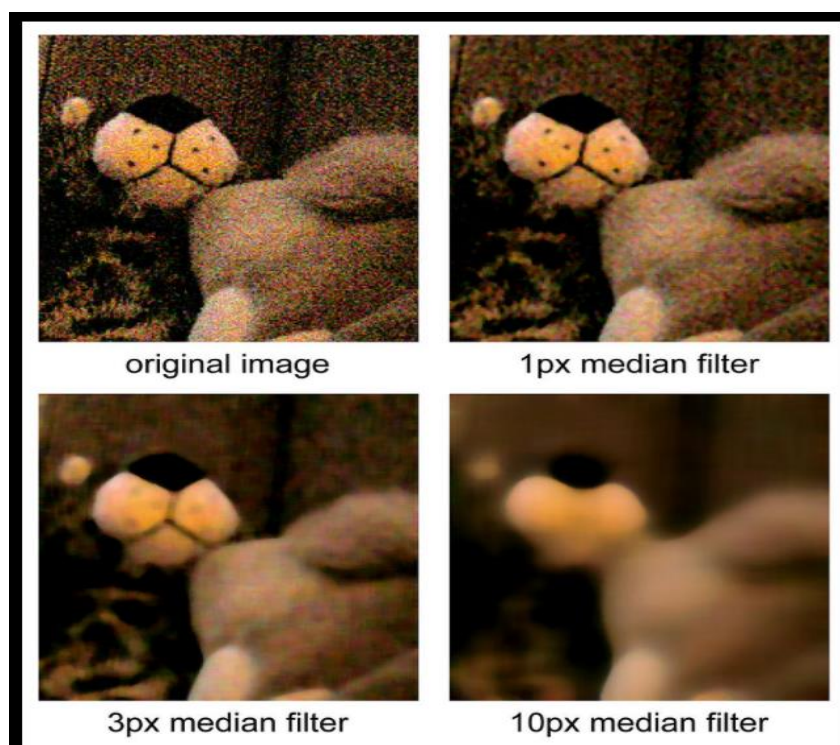


Рисунок 2.4 Приклад роботи медіанного фільтра (з джерела [18])

2.3.1.2 Підходи, засновані на застосуванні нейронних мереж

В якості нейромережових алгоритмів виділення ознак використовувалися згорткові(convolutional) шари попередньо натренованих VGG16 та efficientnetb3. Обидва алгоритми відносяться до CNN. Вони були натреновані на базі даних imagenet[19], що містить понад 14 мільйонів картинок. Виділенні згорткові шари по суті своїй є комплексними фільтрами. Приклад роботи такого фільтра можна побачити на рисунку 2.5 нижче.

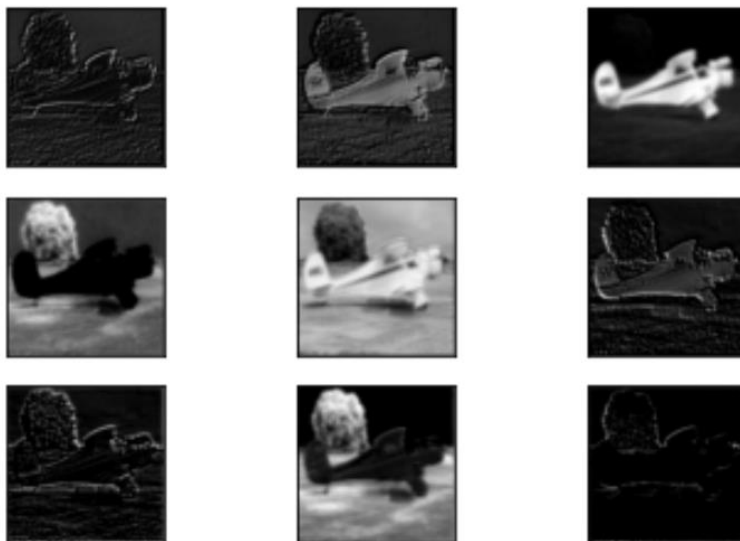


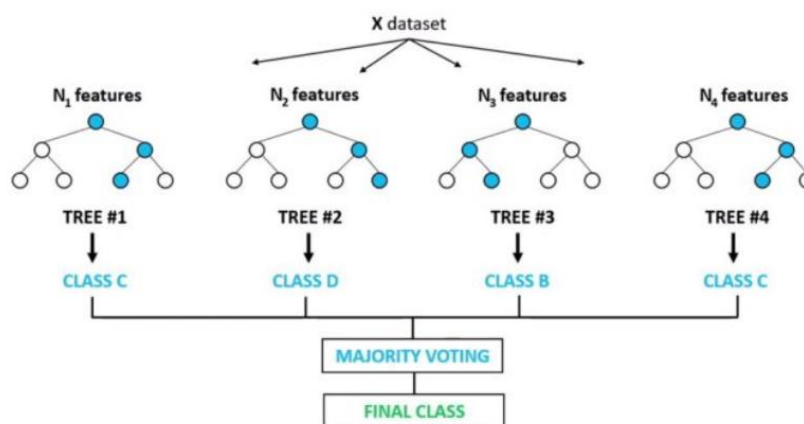
Рисунок 2.5 Використання згорткових шарів з VGG16 для виділення ознак з зображення(з власних розробок)

2.3.2 Класичні структурні алгоритми сегментації

2.3.2.1 Random forest

Random Forest є досить відомим та широко застосовним методом машинного навчання. В тому числі його багато використовують і для сегментації зображень[20]. Він являє собою групу дерев рішень, що були натреновані незалежно один від одного. Дерева рішень є відомими вже давно та показують непогані результати, але вони часто мають проблеми з перенавчанням (overfitting). Основна ідея RF – зменшити цей ефект за рахунок комбінування результатів з декількох дерев рішень, кожне з яких було натреновано на певній унікальній підмножині з початкового набору даних. RF демонструє кращі результати ніж звичайне дерево рішень, при цьому залишаючись достатньо обчислювально ефективним [20].

Random Forest Classifier



Рисункок 2.6 Схематичне зображення алгоритму RF

2.3.2.2. Support vector machine

SVM є не менш відомим та популярним класичним алгоритмом, в тому числі і для задач сегментації. Однією з найчастіших сфер використання є застосування до медичних зображень. Для цих цілей використовується велика кількість модифікованих алгоритмів, заснованих на SVM[21].

Основною ідеєю класифікації за допомогою SVM є знаходження такої границі між примірниками кількох класів, яка максимізує сумарну відстань до найближчих до неї елементів. Приклад такого поділу можна побачити на рисунку 2.7 нижче.

Варто додати, що SVM є достатньо обчислювально складним, тому очікується, що він буде працювати досить повільно, порівняно з іншими алгоритмами та вимагатиме більшої кількості обчислювальних ресурсів.

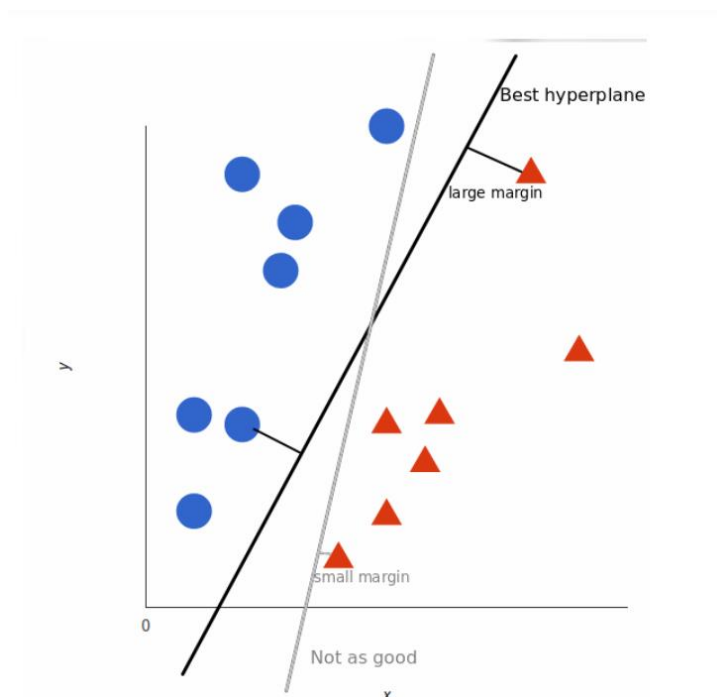


Рисунок 2.7 Поділ двох класів за допомогою алгоритму SVM

2.3.2.3 Lightweight gradient boosting machine

LGBM[22] – алгоритм, заснований на деревах рішень, який був розроблений компанією Microsoft. Основна ідея LGBM – оптимізувати обчислювальну складність та використовувану пам'ять дерев рішень, при цьому не зазнавши значних втрат в ефективності. Для цього було створено та використано 2 нові техніки: Gradient-based One Side Sampling(GOSS) та Exclusive Feature Bundling (EFB)

Техніка GOSS заснована на тому, що не всі примірники у наборі даних для тренування(training set) мають однакову «вагу». Примірники з вищим градієнтом мають вищу роль у зменшенні ентропії. GOSS передбачає залишати у наборі даних ці примірники з високим градієнтом та випадковим чином «викинути» певний відсоток інших.

EFB дозволяє зменшити кількість ознак (features). Його ідея полягає в тому, що часто ознаки бувають взаємовиключними (або майже взаємовиключними) у тому сенсі, що майже не приймають ненульове значення одночасно. ESB «збирає» такі ознаки в одну задля зменшення розмірності.

В цілому LGBM є досить новим та багатообіцяючим алгоритмом, що показує вражаючі результати, в тому числі і у сфері сегментації зображень[23].

2.3.3 Нейромережеві методи сегментації

В якості нейромережевих методів сегментації було використано 3 види CNN: U-net, LinkNet та FPN. Їх структуру можна побачити на рисунку нижче.

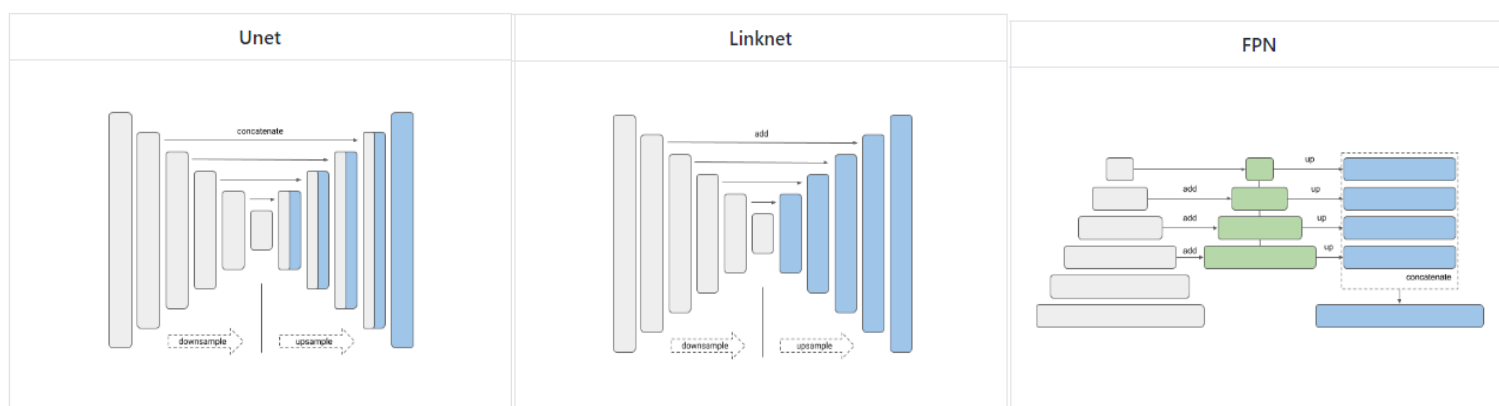


Рисунок 2.8 Структура обраних нейромережевих методів сегментації (з джерела[24])

2.3.3.1 U-Net

U-Net [14] – одна з варіацій згорткових(convolutional) нейронних мереж, яка була розроблена в першу чергу для обробки біомедичних зображень та показує найкращі результати саме на них. Вона була побудована таким чином, що працює досить швидко та дає кращу точність навіть за малої кількості даних, ніж

більшість інших відомих на той час CNN. Архітектуру U-Net можна побачити на рисунку 2.9 (складається з шарів для звуження (ліва частина) та розширення (права частина)):

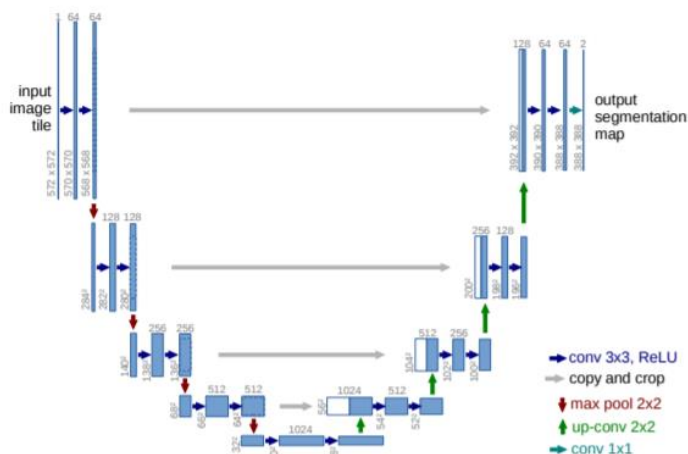


Рисунок 2.9 Архітектура U-Net

Однією з інновацій у U-Net є те, що в частину розширення внесено велику кількість каналів ознак, які дозволяють мережі поширювати контекстну інформацію на шари більшої роздільної здатності. Тобто звуження є типовою згортковою мережею, під час застосування якої просторова інформація зменшується, а інформація про ознаки збільшується. В частині, де відбувається розширення, поєднується просторова інформація та ознаки за допомогою згортки, які збільшують розмірність і конкатенації з ознаками, які відповідають високій роздільній здатності на шляху звуження.

2.3.3.2 LinkNet

Алгоритм LinkNet [13] був створений спеціально для сегментації зображень, в тому числі з метою його використання на мобільних пристроях, а тому дуже великий фокус при розробці було направлено на швидкість та обчислювальну складність методу, особливо на мінімізацію кількості параметрів мережі. Зазвичай в CNN просторова інформація втрачається в кодуючому блоці (encoder block), яка

відновлюється пізніше з використанням індексів об'єднання або повної згортки. Основна ідея LinkNet полягає у тому, що замість вищезазначеної техніки застосовується передача просторової інформації безпосередньо від encoder блоку до відповідного decoder блоку, що підвищує точність, а також значно скорочує час обробки.

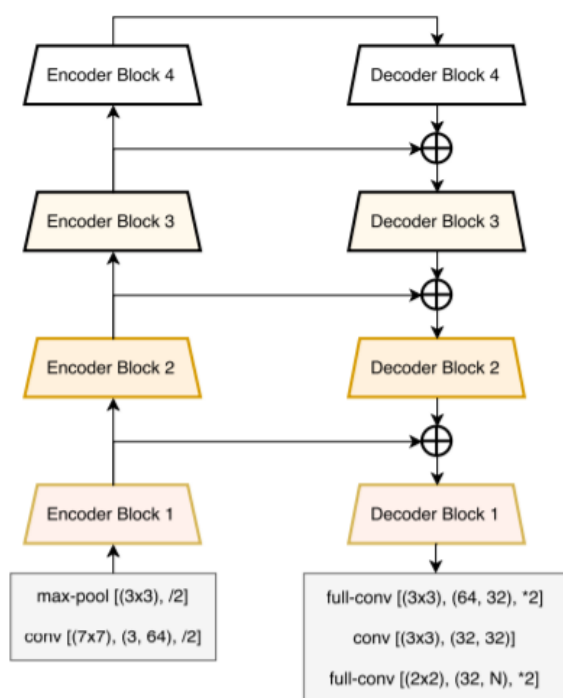


Fig. 1: LinkNet Architecture

Рисунок 2.10 Схематичне зображення архітектури LinkNet

2.3.3.3 FPN

FPN [25] використовує пірамідальну ієрархію глибоких згорткових мереж (deep CNN) для побудови так званих «мап ознак» (feature maps). Він складався із шляхів «знизу вгору» та «зверху вниз» (див. рис 2.11)

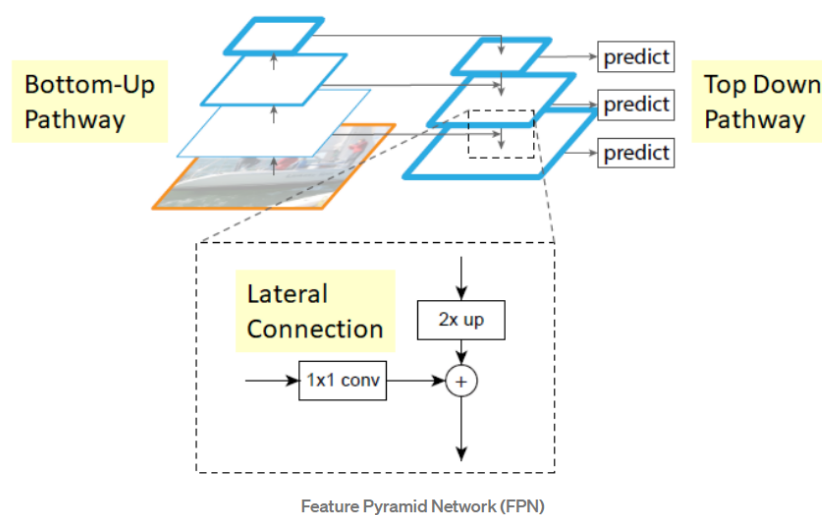


Рисунок 2.11 Схематичне зображення архітектури FPN

Така архітектура демонструє досить гарні результати, особливо в таких задачах, як виявлення об'єктів і сегментація зображень. Але відносно інших алгоритмів цей метод вважається досить повільним. Очікується, що він працюватиме значно довше, ніж інші описані вище нейромережеві техніки.

2.4 Умови та параметри порівняння алгоритмів

Як вже було зазначено, алгоритми у цій роботі переважно використовувалися для бінарної сегментації (тобто на зображеннях треба було виокремити лише об'єкти одного конкретного класу, а все інше позначити як «фон»). Мультикласова сегментація наведена лише для одного набору даних. Таке рішення було прийняте задля забезпечення більшої наглядності та простоти порівняння.

Ключовим етапом є визначення основної метрики якості роботи алгоритмів, на основі якої вони можуть бути об'єктивно порівняні. Точність у відсотках (ассурасу) є популярною величиною для великої кількості задач машинного навчання. Проте вона не підходить для сегментації зображень. Іноді, особливо у

випадку бінарної сегментації, «цільовий об'єкт» може займати дуже маленьку частину зображення (близько 5-10%) і тому, навіть визначивши усе як «фон» та взагалі не сегментувавши «цільового об'єкту» на картинці, алгоритм може отримати точність близько 90%, при цьому, фактично, зробивши досить погану роботу. Тому у даній роботі використовується інша метрика, що є значно більш популярною у наукових роботах, пов'язаних з сегментацією зображень – Intersection over union (IOU) (також відомий як коефіцієнт Жаккара). Зокрема, для оцінки якості роботи алгоритмів враховувався як індивідуальний IOU для окремих класів, так і середній IOU. У випадку з семантичною сегментацією зображень IOU являє собою відношення ділянки перетину (йдеться про перетин ground true сегменту для певного класу та того сегменту, що був визначений для нього алгоритмом) до ділянки об'єднання. Таким чином, для наведеного вище прикладу з маленьким «цільовим об'єктом», IOU для його класу буде дорівнювати нулю у випадку, якщо алгоритм зовсім не сегментує його.

Окрім цього до уваги брався час роботи алгоритму, обчислювальна складність, а також деякі нюанси та особисті враження від використання. Варто зауважити, що у даному контексті час не є абсолютно об'єктивною величиною. Алгоритми запускалися у середовищі Google Colaboratory [26] (першочергово тому, що власних ресурсів персонального комп'ютера автора було недостатньо). Ця програма не надає користувачам сталих обчислювальних ресурсів, а тому навіть один і той самий алгоритм може мати різний час роботи в залежності від часу запуску, а також використання додаткових «прискорювачів». Проте, якщо у великій кількості випадків певний метод працює швидше або повільніше за інші, це можна вважати досить значущим показником.

Також необхідно зазначити, що для більшої «чистоти експерименту» алгоритми було використано у максимально можливому «чистому» вигляді.

Тобто без використання розширення даних (data augmentation) або інших допоміжних технік, що покращують результати роботи. Використовувалися лише певні обмеження для зменшення ефекту перетренування (overfitting). Кількість ітерацій для кожної задачі була підібрана методом «проб і помилок» з метою отримати найкращий результат.

РОЗДІЛ 3.

РЕАЛІЗАЦІЯ ПОСТАВЛЕННОЇ ЗАДАЧІ

3.1 Обрані набори даних

Всього для практичної частини у рамках даної роботи було використано 4 набори даних для сегментації зображень (деякі з них застосовувалися у кількох варіаціях). Усі вони містять відповідні фотографії та ground true маски з сегментованими ділянками. Всі набори даних є у вільному доступі в мережі Інтернет. Нижче наведено більш детальний опис кожного з них.

3.1.1 НАМ

Набір даних «НАМ10000»[27] є великою колекцією (близько 10000 зображень лише у тренувальному наборі) дерматоскопічних зображень із поширеними пігментними ураженнями шкіри. Приклад зображення та ground true маски для нього можна побачити на рисунку нижче.

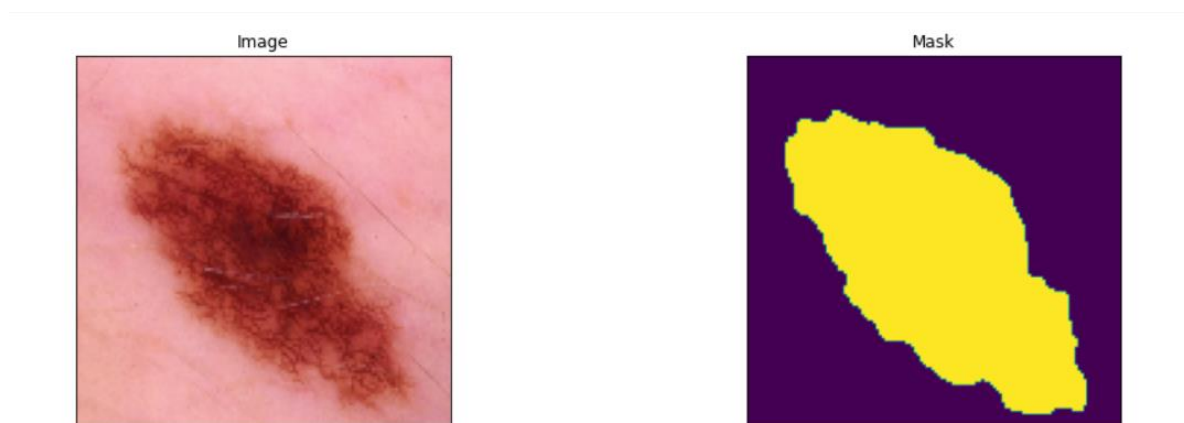


Рисунок 3.1 Зображення з набору даних «НАМ»

Набір даних одразу містить зображення та маски для бінарної сегментації, тому попередня обробка не є необхідною. Для того, щоб зменшити час на кількість необхідних ресурсів набір даних було зменшено до 20 зображень у тренувальному

наборі, 10 у тестувальному та 11 у перевірочному (validation set). Знімки у цьому наборі даних є досить схожими між собою, тому очікується, що цієї кількості має бути достатньо для отримання непоганих результатів.

3.1.2 LaPa

Набір даних «LaPa»[28] містить більше 22000 фотографій облич різних людей та ground true маски для 11 класів (див. рис. 3.2).

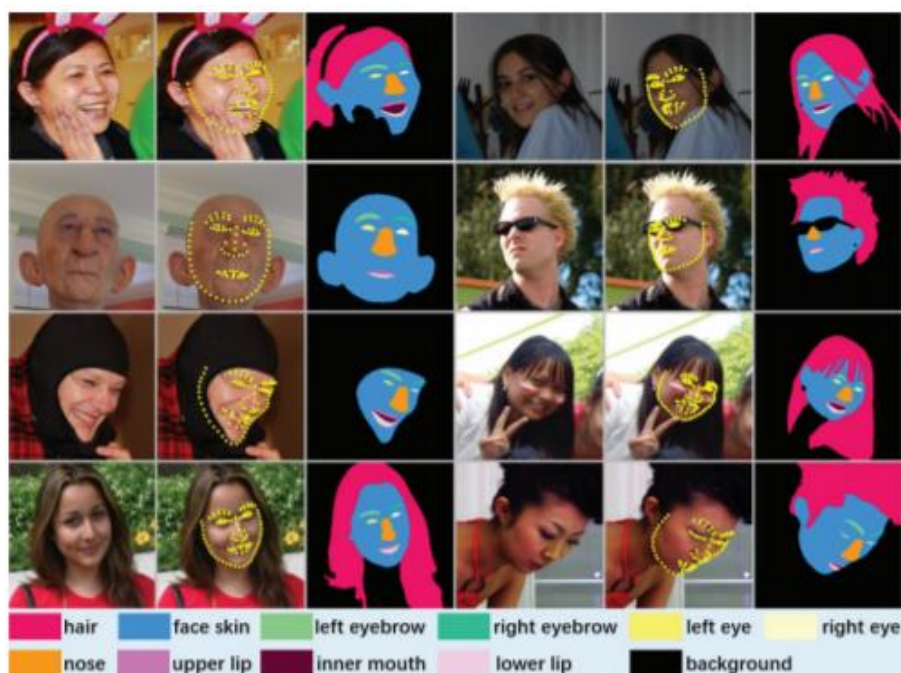


Рисунок 3.2 Зображення з набору даних «LaPa»

Експерименти застосування різних алгоритмів сегментації на такій великій кількості фотографій зайняли б величезну кількість часу та обчислювальних ресурсів, тому довелося зменшити набір даних до 100 зображень у тренувальному наборі, 24 у тестувальному і 11 у перевірочному. Оригінальний набір даних містить фото людей різного віку, статі та раси. Але важко забезпечити хороші результати сегментації на маленькій підмножині картинок дуже різноманітного

характеру. Тому у зменшеному наборі було залишено лише фотографій молодих світлошкірих жінок.

Як зазначено вище, набір даних «LaPa» має ground true маски для 11 класів, але для забезпечення бінарної сегментації було використано лише кілька з них. Зокрема, цей набір даних застосовується у двох варіаціях: для бінарної сегментації шкіри обличчя та для волосся. Приклади фотографій та ground true масок для цього можна побачити на рисунках 3.3 та 3.4.

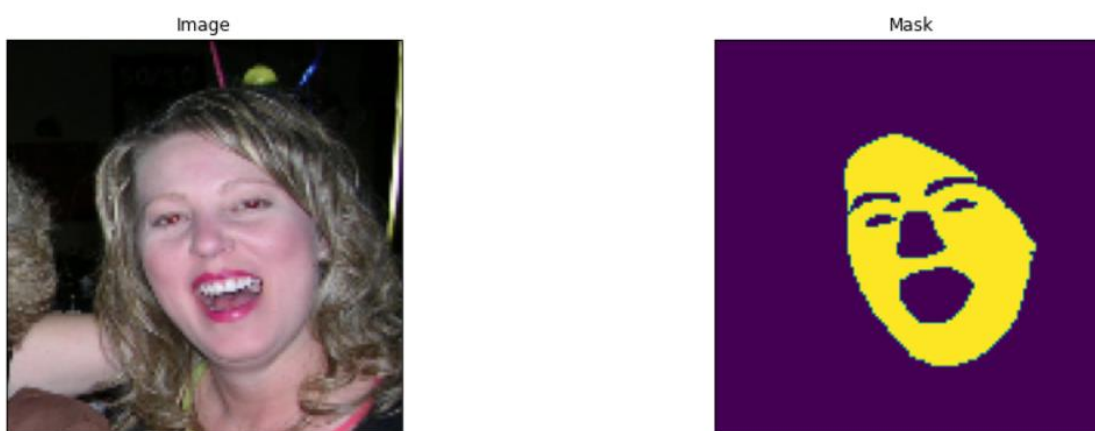


Рисунок 3.3 Зображення з ground true маскою для шкіри обличчя

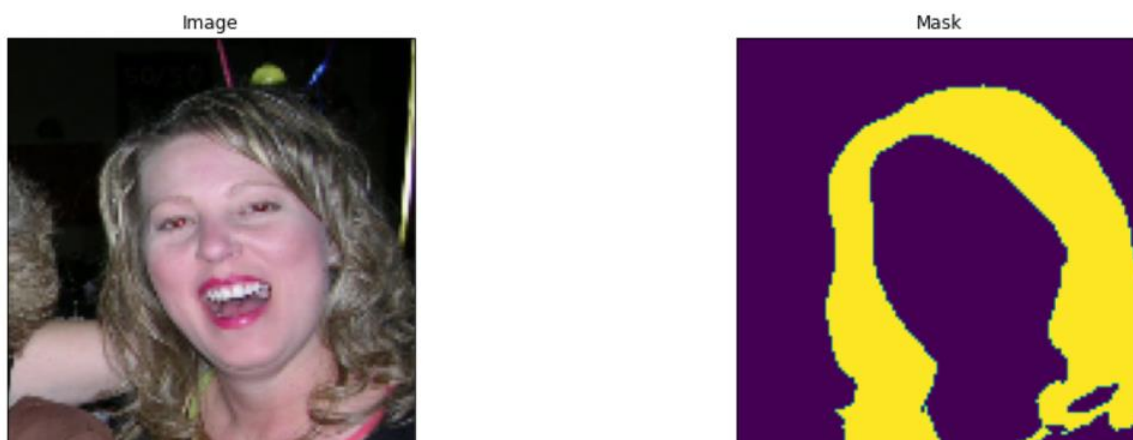


Рисунок 3.4 Зображення з ground true маскою для волосся

3.1.3 Sky

Набір даних «Sky»[29] містить 60 зображень з ground true масками для сегментації неба



Figure 1: Sky dataset images and their respective ground truths

Рисунок 3.5 Фотографії з набору даних «Sky»

Для практичної частини в рамках даної роботи ці 60 зображень були розбиті на 48 фотографій у тренувальному наборі даних, 10 у тестувальному та 2 у перевірконому. Цей набір даних одразу підходить до бінарної сегментації. Також очікується, що алгоритми покажуть одні з найкращих результатів на цих зображеннях за рахунок великого розміру «цільової ділянки», а також відносної легкості сегментації неба.

3.1.4 CamVid

CamVid[30] - це колекція знімків вуличних сцен із семантичними мітками для 12 класів(див. рис. 3.6).

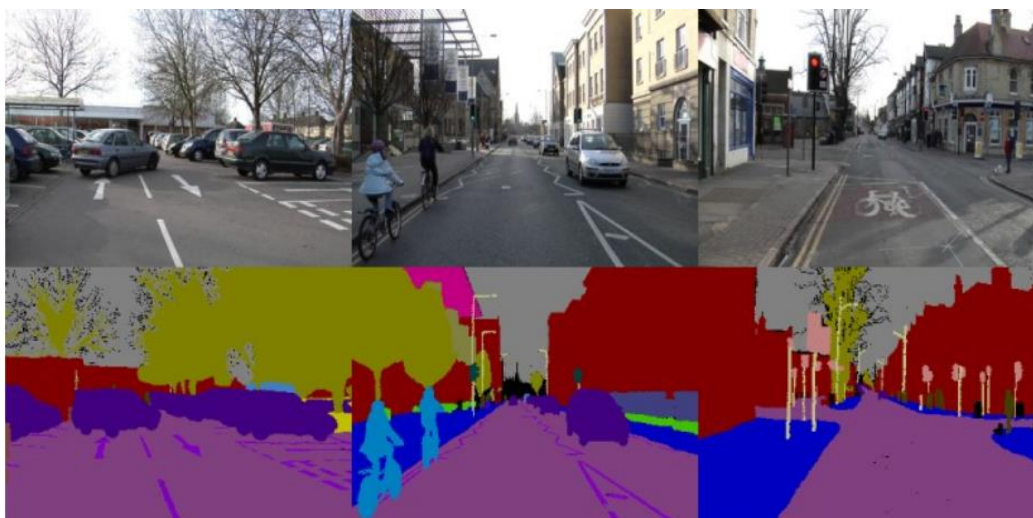


Рисунок 3.6 Зображення з набору даних «CamVid»

CamVid містить 367 фотографій у тренувальному наборі, 233 у тестувальному та 101 у перевірконому. Саме на цьому наборі проводився експеримент з «мультикласовою» сегментацією, а також з бінарною сегментацією для класу автомобілів. Приклад зображення та ground true маски для такої сегментації можна побачити на рисунку 3.7 нижче.

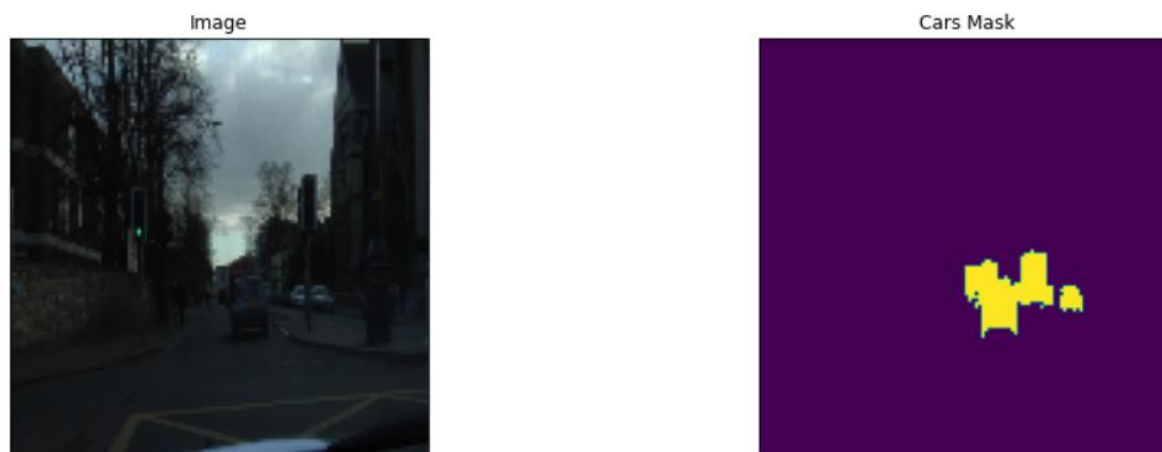


Рисунок 3.7 Зображення з ground true маскою для автомобілів

Очікується, що результати на цьому наборі даних будуть найгіршими через найбільшу кількість зображень у ньому, маленький розмір «цільових об'єктів» а також через те, що, як можна побачити на рисунку вище, певна кількість знімків є

досить темними, «цільові об'єкти» на них важко виокремити навіть людському оку.

3.2.Реалізація мовою програмування Python

Для реалізації практичної частини даної роботи, а саме випробування різних методів сегментації зображень на різних наборах даних, було обрано мову програмування Python. Python наразі є однією з найпопулярніших мов, що використовується для розв'язування задач зі сфери машинного навчання. На ній розроблено велику кількість бібліотек, що імплементують різноманітні алгоритми. Таким чином, це надає можливість використати вже реалізовані бібліотечні методи сегментації зображень, а також виділення ознак, без необхідності самостійно писати код для них (що зазвичай є досить комплексним та складним).

Детально з кодом, що був написаним у рамках виконання практичної частини даної роботи, а також з результатами, отриманими за його допомогою можна ознайомитися на платформі GitHub за посиланням <https://github.com/EkaterinaTytarenko/image-segmentation>. Наведений репозиторій поділений на папки для кожного набору даних (див. рис. 3.8), де кожен алгоритм представлений у вигляді окремої програми у форматі jupyter notebook.

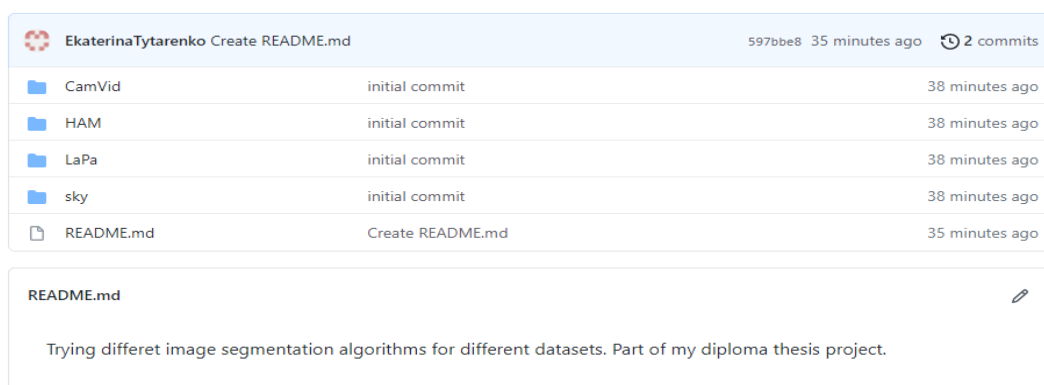


Рисунок 3.8 Структура github-репозиторію з розробленими програмами

Окрім цього, було розроблено декілька невеликих допоміжних програм для виконання простих функцій попередньої обробки даних (наприклад, конвертація картинок формату jpeg у формат png, або перейменування деяких файлів). Проте код цих програм не є суттєво важливим для даної роботи, тому його не було наведено у репозиторії.

Як вже згадувалося, для виконання розроблених програм використовувалося середовище Google Colaboratory[26]. Google Colaboratory дозволяє запускати код, написаний на Python, у браузері без жодних додаткових налаштувань. Також це середовище надає безкоштовний доступ до певної кількості обчислювальних ресурсів та можливість легко ділитися своїм кодом з іншими користувачами, надаючи їм доступ до нього. Основною мотивацією використання цього середовища для даної роботи слугувала наявність безкоштовних обчислювальних ресурсів. Алгоритми машинного навчання зазвичай є достатньо обчислювально складними, тому потужності персонального комп'ютера автора було недостатньо для того, щоб запускати деякі з них. Також це середовище надає можливість виконувати декілька програм одночасно та з певною частотою використовувати так звані «пришвидшувачі» (додаткові обчислювальні ресурси, що дозволяють виконувати код ще швидше).

Варто також зазначити, що, задля додаткового пришвидшення роботи алгоритмів та зменшення кількості обчислювальних ресурсів, усі зображення, що використовувалися, були зменшені до розмірності 128 на 128 пікселів. Звичайно, наслідками такого зменшення є значна втрата якості зображень, а отже і гірша робота методів сегментації. Проте це не має суттєво впливати на результати порівняння алгоритмів, адже абсолютно у всіх випадках картинки зменшуються перед початком роботи з ними.

РОЗДІЛ 4.

ОТРИМАНІ РЕЗУЛЬТАТИ

У цьому розділі представлені результати використання усіх описаних вище методів сегментації зображень для кожного набору даних. Вони будуть наведені у вигляді таблиць з числовими показниками, отриманих на відповідних тестувальних наборах даних, доповнених деякими коментарями.

4.1 Набір даних НАМ

Таблиця 4.1.1 – Результати роботи алгоритмів для набору даних «НАМ»

Методи	Основні параметри	IOU(1-й клас)	IOU(2-й клас)	Середній IOU	Час тренування
efficientnet+Unet	30 епох	≈0.927	≈0.744	≈0.835	≈0:08:30.16
efficientnet+Linknet	40 епох	≈0.926	≈0.74	≈0.833	≈0:09:39.69
efficientnet+FPN	20 епох	≈0.931	≈0.766	≈0.849	≈0:09:49.89
VGG+RF	55 дерев	≈0.899	≈0.68	≈0.79	≈0:02:47.94
VGG+SVM	До 13 ітерацій	≈0.879	≈0.633	≈0.756	≈0:00:02.83
VGG+LGBM	30 ітерацій	≈0.905	≈0.69	≈0.798	≈0:00:05.26
Gabour+RF	100 дерев	≈0.887	≈0.626	≈0.756	≈0:01:17.9
Gabour+SVM	До 750 ітерацій	≈0.896	≈0.626	≈0.76	≈0:01:26.29

Gabour+LGBM	100 ітерацій	≈0.89	≈0.634	≈0.76	≈0:00:05.51
-------------	--------------	-------	--------	-------	-------------

Можна побачити, що усі алгоритми досить гарно виконали сегментацію(з «візуальними» результатами можна ознайомити за посиланням наведеном у попередньому розділі), в тому числі за рахунок невеликої кількості зображень у наборі даних, їх схожості між собою, великому розміру «цільових об'єктів», а також високій якості та контрастності знімків.

Методи, що використовують нейронні мережі показали трохи кращі результати. Причому повністю нейромережеві - кращі за ті, що використовують нейронні мережі лише для виділення ознак, які у свою чергу трохи кращі за ті, що використовують лише класичні алгоритми. Проте швидкість тренування нейромережевих алгоритмів виявилася трохи гіршою. Але тут варто зауважити, що кількість епох, обрана для тренування деяких моделей була трохи надлишковою та аналогічних результатів можна було досягти і за менший час. Це демонструють і графіки зміни IOU та втрат для відповідних моделей.

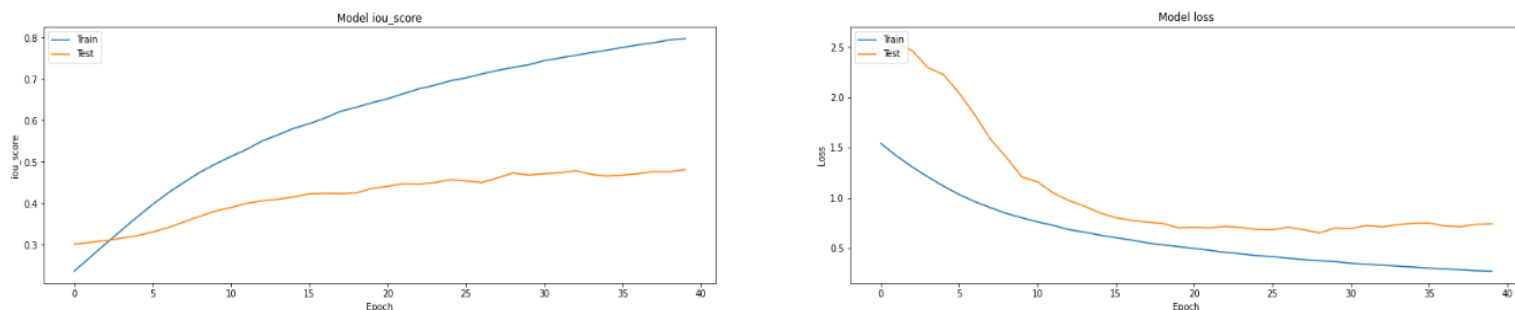


Рисунок 4.1 Графіки зміни IOU та втрат для Linknet

Варто зауважити, що, не дивлячись на те, що усі моделі на даному наборі даних зробили хорошу роботу, вони показали різну стійкість до шумів(наприклад, волосся), на знімках. Перші три алгоритми (суто нейронні мережі) виявилися найбільш стійкими. Друга трійка(нейромережі для виділення ознак + тренування

класичними методами) показали найменшу стійкість. І останні три алгоритми(суто класичні методи) показали середній результат. Приклади можна побачити на рис. 4.2.

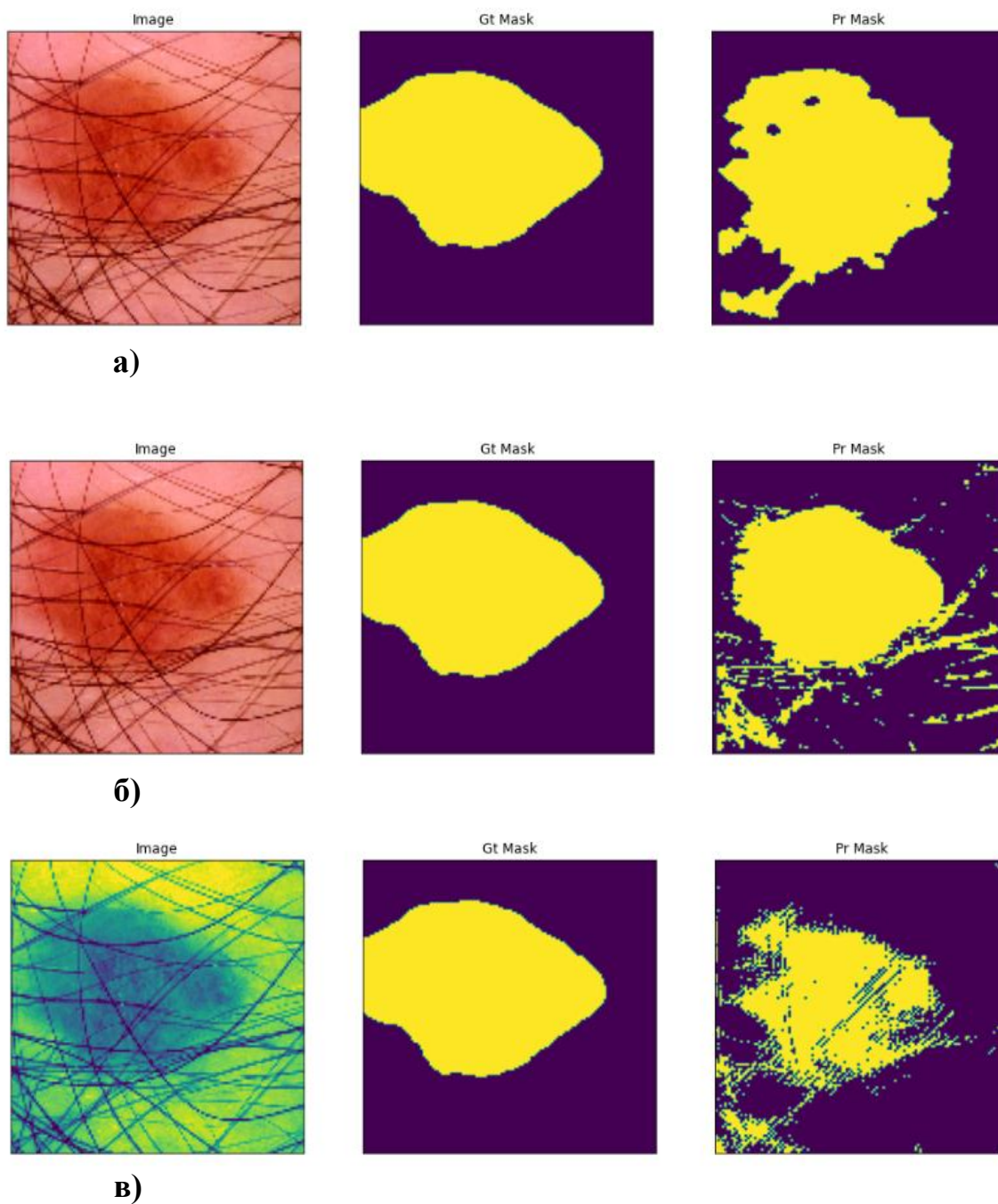


Рисунок 4.2 Приклади сегментації : а) для FPN; б) для VGG+LGBM; в) для Gabour+SVM

4.2 Набір даних LaPa

Як вже зазначалося, цей набір даних застосовувався у двох варіаціях: для бінарної сегментації шкіри обличчя та для волосся. Наведемо результати окремо для кожного з цих експериментів.

Таблиця 4.2.1 – Результати роботи алгоритмів для набору даних «LaPa» для бінарної сегментації волосся

Методи	Основні параметри	IOU(1-й клас)	IOU(2-й клас)	Середній IOU	Час тренування
efficientnet+Unet	50 епох	≈0.927	≈0.759	≈0.843	≈0:43:48.19
efficientnet+Linknet	50 епох	≈0.912	≈0.724	≈0.818	≈0:37:24.34
efficientnet+FPN	35 епох	≈0.922	≈0.735	≈0.828	≈1:02:18.57
VGG+RF	60 дерев	≈0.798	≈0.37	≈0.588	≈ 0:15:26.94
VGG+SVM	До 2500 ітерацій	≈ 0.675	≈ 0.25	≈ 0.464	≈ 0:53:10.17
VGG+LGBM	500 ітерацій	≈0.804	≈0.387	≈ 0.595	≈ 0:03:21.06
Gabour+RF	70 дерев	≈ 0.777	≈ 0.293	≈ 0.535	≈0:07:13.75
Gabour+SVM	До 3500 ітерацій	≈0.751	≈0.254	≈0.5	≈0:55:34.27
Gabour+LGBM	1000 ітерацій	≈0.782	≈0.311	≈0.546	≈0:03:42.83

Таблиця 4.2.2 – Результати роботи алгоритмів для набору даних «LaPa» для бінарної сегментації шкіри обличчя

Методи	Основні параметри	IOU(1-й клас)	IOU(2-й клас)	Середній IOU	Час тренування
efficientnet+Unet	50 епох	≈0.973	≈0.851	≈0.912	≈0:44:22.71
efficientnet+Linknet	50 епох	≈0.956	≈0.778	≈0.867	≈0:36:20.07
efficientnet+FPN	35 епох	≈0.971	≈0.838	≈0.904	≈1:01:46.13
VGG+RF	50 дерев	≈0.865	≈0.391	≈0.628	≈0:11:48.29
VGG+SVM	До 800 ітерацій	≈0.825	≈0.36	≈0.592	≈0:13:44.29
VGG+LGBM	50 ітерацій	≈0.865	≈0.393	≈0.629	≈0:00:33.38
Gabour+RF	75 дерев	≈0.843	≈0.186	≈0.515	≈0:06:27.58
Gabour+SVM	До 100 ітерацій	≈0.507	≈0.219	≈0.363	≈0:01:41.00
Gabour+LGBM	1000 ітерацій	≈0.842	≈0.189	≈0.516	≈0:03:33.22

Для цього набору даних спостерігається значна різниця в ефективності методів: перша трійка алгоритмів виявилася найкращою для обох випадків, решта спрацювали приблизно однаково у випадку з волоссям. У випадку ж зі шкірою обличчя друга трійка алгоритмів дала трохи кращі результати, ніж останні три.

Для часу тренування спостерігаються такі самі тенденції, як і для попереднього набору даних: суто нейромережеві методи тренувалися повільніше за ті, що використовували нейронні мережі лише для виділення ознак, які, у свою чергу, є повільнішими за суто класичні.

Для отримання детальнішої інформації на деяких алгоритмах були виміряні IOU для тренувального набору даних (конкретні числові результати можна побачити у коді роботи за посиланням з попереднього пункту). На основі отриманих вимірів можна зробити висновки, що деякі з класичних методів показують значне перетренування(overfitting). Зокрема, найбільше перетренування показав RF для обох випадків використання набору даних та обох способів виділення ознак. Схожий ефект спостерігався у комбінації VGG+LGBM для сегментації волосся. Решта експериментів не показують суттєвого перетренування.

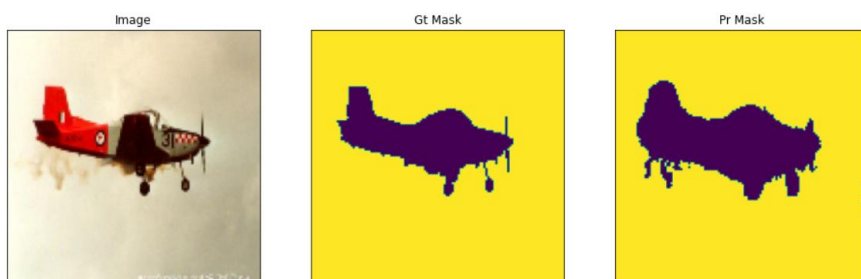
4.3 Набір даних Sky

Таблиця 4.3.1 – Результати роботи алгоритмів для набору даних «Sky»

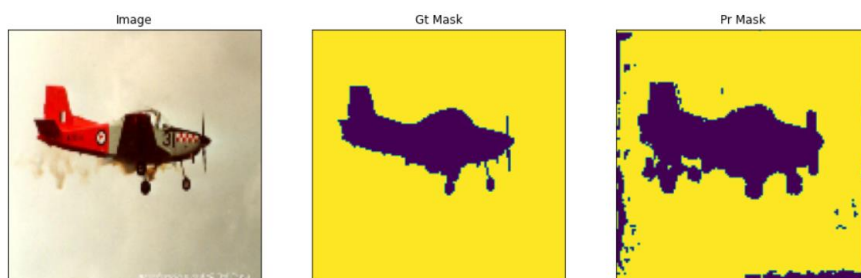
Методи	Основні параметри	IOU(1-й клас)	IOU(2-й клас)	Середній IOU	Час тренування
efficientnet+Unet	40 епох	≈ 0.9	≈ 0.889	≈ 0.895	≈ 0:14:01.95
efficientnet+Linknet	50 епох	≈ 0.9	≈ 0.887	≈ 0.894	≈ 0:14:46.63
efficientnet+FPN	20 епох	≈ 0.853	≈ 0.816	≈ 0.834	≈ 0:14:09.67
VGG+RF	45 дерев	≈ 0.851	≈ 0.826	≈ 0.838	≈ 0:04:48.19
VGG+SVM	До 120 ітерацій	≈ 0.864	≈ 0.843	≈ 0.853	≈ 0:00:46.86

VGG+LGBM	50 ітерацій	≈ 0.873	≈ 0.848	≈ 0.86	$\approx 0:00:18.75$
Gabour+RF	50 дерев	≈ 0.898	≈ 0.885	≈ 0.891	$\approx 0:01:47.89$
Gabour+SVM	До 200 ітерацій	≈ 0.872	≈ 0.865	≈ 0.869	$\approx 0:01:06.8$
Gabour+LGBM	50 ітерацій	≈ 0.894	≈ 0.881	≈ 0.888	$\approx 0:00:07.72$

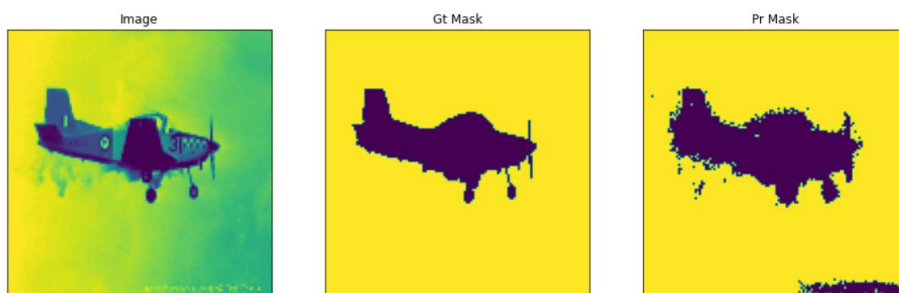
На даному наборі даних усі алгоритми мають приблизно однакові показники IOU. Можна сказати, що у такому випадку, класичні методи мають певну перевагу, тому що працюють значно швидше. Проте вони також продемонстрували один недолік на наборі даних Sky: класичні алгоритми сегментують зі значно більшою кількістю шумів. Приклад цього можна побачити на рис. 4.3.



a)



б)



в)

Рисунок 4.3 Приклади сегментації : а) для Unet; б) для VGG+RF; в) для Gabour+RF

4.4 Набір даних CamVid

Як і очікувалось, цей набір даних виявився найважчим для сегментації. За рахунок відносно великої кількості картинок у тренувальному наборі більшість з класичних алгоритмів не змогли натренуватися. Вони або потребували забагато обчислювальних ресурсів і середовище Google Colaboratory зупинило виконання програми з повідомленням, що був перевищений ліміт доступної оперативної пам'яті. Або ж алгоритми не показували жодних результатів за адекватний час (IOU був майже нульовим). Це одразу показує певну перевагу нейромережових методів. Нижче будуть наведені результати лише для тих алгоритмів, що хоча б мінімально змогли натренуватися. Окремо розглянуті 2 випадки: «мультикласова» сегментація та бінарна сегментація автомобілів. Зауважимо, що «мультикласова» сегментація виконувалася для 12-ти різних класів, тому для простоти порівняння буде наведено лише середній IOU, а не показник для кожного окремого класу.

Таблиця 4.4.1 – Результати роботи алгоритмів для набору даних «CamVid» для бінарної сегментації автомобілів

Методи	Основні параметри	IOU(1-й клас)	IOU(2-й клас)	Середній IOU	Час тренування
efficientnet+Unet	70 епох	≈ 0.987	≈ 0.695	≈ 0.841	$\approx 3:36:35.82$
efficientnet+Linknet	70 епох	≈ 0.987	≈ 0.687	≈ 0.837	$\approx 3:00:17.86$
efficientnet+FPN	40 епох	≈ 0.987	≈ 0.69	≈ 0.839	$\approx 4:13:00.27$
VGG+RF	40 дерев	≈ 0.966	≈ 0.255	≈ 0.61	$\approx 0:59:45.39$

Таблиця 4.4.2 – Результати роботи алгоритмів для набору даних «CamVid» для мультикласової сегментації

Методи	Основні параметри	Середній IOU	Час тренування
efficientnet+Unet	70 епох	≈ 0.397	$\approx 3:27:36.88$
efficientnet+Linknet	85 епох	≈ 0.378	$\approx 3:53:53.11$
efficientnet+FPN	50 епох	≈ 0.403	$\approx 4:57:03.24$
VGG+RF	50 дерев	≈ 0.252	$\approx 1:50:26.74$

Можна побачити, що в даному випадку чисто нейромережеві алгоритми спрацювали значно краще, але суттєвої різниці між ними самими помічено не

було. RF показав незадовільний результат, для нього знову було виміряно IOU на тренувальному наборі даних та знову помічений ефект перетренування.

4.5 Висновки розділу

Підсумовуючи описані у цьому розділі експерименти можна виокремити наступні результати:

а) однозначними лідерами за якістю сегментації та метриками IOU виявилися методи, де і виділення ознак і безпосередня сегментація виконувалися за допомогою нейронних мереж. Причому принципової різниці між якістю роботи Unet, Linknet та FPN помічено не було, крім того нюансу, що FPN у більшості випадків потребує трохи більше часу на тренування.

б) у більшості випадків SVM показував гірші результати, ніж інші класичні методи. До того ж, за параметрів «за замовчуванням», поведінка класифікатора з бібліотеки sklearn є досить непередбачуваною. При різних запусках програми на однаковому наборі даних та з однаковою кількістю ітерацій алгоритм показує достатньо різні результати, що ускладнює використання SVM. Тому можна сказати, що RF та LGBM у рамках даного експерименту показали себе краще. Якщо ж порівнювати їх між собою, то LGBM в середньому давав кращі результати, був менш вразливим до перетренування, а також тренувався значно швидше. Проте він не зміг показати більш-менш прийнятних результат на наборі даних «CamVid», тому можна прийти до висновку, що він спрацював трохи краще, ніж RF, але різниця не є колосальною.

в) порівнюючи використання класичних фільтрів або згорткових шарів VGG16 для виділення ознак, бачимо наступну поведінку: на маленьких наборах даних результуючий IOU є приблизно однаковим, методи з класичними

фільтрами працюють швидше, але дають трохи більше «шуму» в результаті сегментації. На великих наборах даних методи з VGG16 спрацювали краще.

г) чим більше певний метод застосовує нейронні мережі тип більше часу в середньому витрачається на тренування. Проте, тут варто врахувати, що кількість епох для тренування деяких моделей була обрана трохи «з запасом» та аналогічних результатів можна було досягти і за менший час. Тобто, вони все ще залишаються більш повільними, але різниця не така колосальна.

г) для маленьких за розміром та неважких для сегментації наборів даних (таких як Sky або HAM) класичні методи показують досить прийнятні результати, витрачаючи при цьому значно менше часу на тренування.

д) Як вже зауважувалося, дані результати варто інтерпретувати як такі, що були отримані конкретних техніках та наборах даних за конкретних налаштувань та початкових умов. Вони не дають універсального керівництва до того, як сегментувати зображення та не покривають усю велику кількість нюансів які можуть виникнути при розв'язуванні такої задачі.

4.6 Рекомендації, можливості подальших розробок

Як свідчать наукові джерела і як вже згадувалося у цій роботі, сфера сегментації зображень є дуже широкою. Існує велика кількість різних алгоритмів, результати яких варіюються в залежності від конкретного набору даних та способу використання. Тому найкраща стратегія для вирішенні цієї задачі - експериментувати з кількома різними методами та обирати найкращі. Це тим паче важливо тому, що сегментація зображень часто використовується у дуже відповідальних сферах(наприклад, системи автопілоту або виокремлення ракових пухлин на знімках). Проте на основі результатів, отриманих у ході написання

даної роботи, можна надати наступні рекомендації, що можуть бути корисними у процесі вибору алгоритмів для подальших експериментів:

а) Техніки, що використовують CNN як для виділення ознак, так і для безпосередньої сегментації, показали себе найкраще, а тому, за наявності достатньої кількості обчислювальних ресурсів та відсутності дуже жорстких часових обмежень, рекомендується застосовувати саме їх. До того ж, цей напрямок зараз активно розвивається, тому очікується, що у найближчому майбутньому буде з'являтися ще більше методів, заснованих на CNN, з удосконаленою якістю сегментації та оптимізованим часом. Серед усіх оглянутих алгоритмів (Unet, FPN, Linknet) явного «лідера» виділити не вдалося, проте FPN, в середньому, потребує більше часу на тренування, тому, якщо це є принциповим фактором, рекомендується звернути більшу увагу на інші два алгоритми.

б) Для невеликих та неважких для сегментації наборів даних класичні методи демонструють гарні результати, витрачаючи при цьому значно менше часу на тренування, ніж ті, що хоча б частково застосовують нейронні мережі(зауважимо, що у такому випадку зображення мають бути достатньо схожими між собою, щоб навіть їх маленької кількості було достатньо щоб добре «натренувати» модель). Тому у деяких випадках може бути доцільним вибір класичних алгоритмів замість CNN. Наприклад, якщо досить критичним є час тренування та заради нього можна незначно пожертвувати якістю сегментації. Опираючись на результати описаного вище експерименту можна сказати, що в першу чергу варто звернути увагу на LGBM, а також, за можливості, спробувати RF.

в) Для вибору між класичними та нейромережевими методами виділення ознак загальні рекомендації досить схожі: нейронні мережі, в середньому, роблять кращу роботу, проте працюють повільніше, а також використання класичних

алгоритмів зазвичай достатньо для хороших результатів на невеликій кількості зображень.

Проведення досліджень, аналогічних до практичної частини цієї роботи, є дуже важливим, адже сфера сегментації зображень є надзвичайно цінною, корисною, швидко розвиваючоюся та широко застосовною. А тому подальші експерименти можуть розглянути більше випадків, що, у свою чергу, дозволить помітити ще більше закономірностей та розширити або уточнити надані рекомендації.

В першу чергу, було б корисно провести дослідження на більш потужних машинах зі значно більшими наборами даних (що налічують тисячі, а іноді і десятки тисяч картинок). Це б значно краще відповідало більшості способів використання алгоритмів сегментації у реальних практичних задачах. Також варто було б перевірити їх роботу з ще більшою кількістю зображень іншої природи (наприклад, фотографії їжі або знімки з мікроскопів). Також і список методів, розглянутих у рамках виконання цієї роботи, не був вичерпним та корисно було б провести практичні дослідження з ще більшою кількістю актуальних і широко застосованих алгоритмів (наприклад, DeepLab[31]). За наявності достатньої кількості обчислювальних та часових ресурсів, можна також спробувати порівняти різні методи сегментації зображень, використовуючи певні техніки їх «покращення» (наприклад, data augmentation). Можливо, це також дозволило б виявити певні нові закономірності.

ВИСНОВКИ

В процесі написання даної дипломної роботи було проведено детальний аналіз сфери сегментації зображень: вивчено основні піднапрямки та підзадачі, що існують у цій області, розглянуто загальний процес сегментації, досліджені існуючі алгоритми та їх класифікація, а також наукові роботи, що вже займалися їх вивченням.

На основі зібраної інформації був вибраний певний перелік алгоритмів та наборів даних для їх практичного випробування. За допомогою мови програмування Python було створено програми, що реалізували сегментацію різних зображень різними методами та вимірювали певні показники якості роботи. Детально з кодом та результатами виконання цих програм можна ознайомитися за посиланням <https://github.com/EkaterinaTytarenko/image-segmentation>.

Далі був проведений детальний аналіз отриманих вимірювань, разом з деякими іншими особливостями роботи алгоритмів (візуальна оцінка якості сегментації, нюанси з досвіду використання). На основі проведеного аналізу були надані рекомендації щодо вибору методів для практичного використання, а також щодо можливостей подальших розробок у області сегментації зображень.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The Cityscapes Dataset for Semantic Urban Scene Understanding,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, Pages 1-11
2. An overview of semantic image segmentation [Електронний ресурс] –2018 - Режим доступу до ресурсу: <https://www.jeremyjordan.me/semantic-segmentation/#representing>
3. Object Detection and Instance Segmentation: A detailed overview [Електронний ресурс]/Shaunak Halbe –2020 - Режим доступу до ресурсу: <https://medium.com/swlh/object-detection-and-instance-segmentation-a-detailed-overview-94ca109274f2>
4. Image segmentation [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Image_segmentation
5. Why Image Segmentation is Needed: Image Segmentation Techniques [Електронний ресурс] – 2021 - Режим доступу до ресурсу: <https://becominghuman.ai/why-image-segmentation-is-needed-image-segmentation-techniques-ee52b92e651a>
6. A REVIEW ON NEURAL NETWORK-BASED IMAGE SEGMENTATION TECHNIQUES [Електронний ресурс]/ Amza, Catalin Gheorghe –2021 - Режим доступу до ресурсу: https://www.researchgate.net/publication/228873725_A_REVIEW_ON_NEURAL_NETWORK-BASED_IMAGE_SEGMENTATION_TECHNIQUES

7. Various Image Segmentation Techniques: A Review, Dilpreet Kaur, Yadwinder Kaur. International Journal of Computer Science and Mobile Computing, Vol. 3, Issue. 5, May 2014, pg.809 – 814
8. Image Segmentation Algorithms Overview [Электронный ресурс]/Song Yuheng, Yan Hao – Режим доступа до ресурсу:.
9. Comparing Random Forest, XGBoost and LGBM for semantic image segmentation [Электронный ресурс] – 2021 - Режим доступа до ресурсу:
https://www.youtube.com/watch?v=pbdseWOKG1U&ab_channel=Apeer_micro
10. A Review of Computer Vision Segmentation Algorithms [Электронный ресурс]/ Dingding Liu, Bilge Soran, Gregg Petrie, and Linda Shapiro - Режим доступа до ресурсу: <https://courses.cs.washington.edu/courses/cse576/12sp/notes/remote.pdf>
11. Review of Deep Learning Algorithms for Image Semantic Segmentation [Электронный ресурс] – 2018 - Режим доступа до ресурсу:
https://medium.com/@arthur_ouaknine/review-of-deep-learning-algorithms-for-image-semantic-segmentation-509a600f7b57
12. U-Net vs LinkNet for multiclass semantic segmentation [Электронный ресурс] – 2021 - Режим доступа до ресурсу:
https://www.youtube.com/watch?v=NUvmHYTQxrs&ab_channel=DigitalSreeni
13. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation [Электронный ресурс]/Abhishek Chaurasia, Eugenio Culurciello, Pages 1-5 – 2017- Режим доступа до ресурсу: <https://arxiv.org/pdf/1707.03718.pdf>
14. U-Net: Convolutional Networks for Biomedical Image Segmentation [Электронный ресурс]/Olaf Ronneberger, Philipp Fischer, and Thomas Brox. , Pages 1-8 – 2015 - Режим доступа до ресурсу: <https://arxiv.org/pdf/1505.04597.pdf>

15. Image segmentation using traditional Machine Learning [Электронный ресурс] – 2020 - Режим доступа до ресурсу:
https://www.youtube.com/watch?v=VCrLaEHShDo&ab_channel=Apeer_micro
16. Gabor Filters [Электронный ресурс]/Tomoki Tsuchida – 2015 - Режим доступа до ресурсу: <http://ttsuchi.github.io/2015/08/26/gaborfilters.html>
17. Особенности применения фильтров обработки изображений перед поиском объектов на изображениях / В. В. Бутенко. Технические науки: теория и практика: материалы III Междунар. науч. конф. (г. Чита, апрель 2016 г.). — С. 1-3.
18. Median filter [Электронный ресурс] – Режим доступа до ресурсу:
https://www.wikiwand.com/en/Median_filter
19. About ImageNet [Электронный ресурс]/ Stanford Vision Lab, Stanford University, Princeton University –2020 - Режим доступа до ресурсу:
<http://www.image-net.org/about>
20. Object Class Segmentation using Random Forests [Электронный ресурс]/F.Schroff, A. Criminisi, A. Zisserman. - Режим доступа до ресурсу:
http://www.wisdom.weizmann.ac.il/~vision/courses/2010_2/papers/schroff_BMVC08.pdf
21. SVM Methods in Image Segmentation, Tian Chi Zhang, Jian Pei Zhang, Jing Yang, Jing Zhang. COLLA 2016 : The Sixth International Conference on Advanced Collaborative Networks, Systems and Applications. Pages 62-65
22. LightGBM (Light Gradient Boosting Machine) [Электронный ресурс] –2020 - Режим доступа до ресурсу: <https://www.geeksforgeeks.org/lightgbm-light-gradient-boosting-machine/>

23. Gradient Boosting Decision Trees for Echocardiogram Images ,Vinícius Veloso de Melo, Regina Célia Coelho, Daniela Mayumi Ushizima.. Conference: 2018 International Joint Conference on Neural Networks (IJCNN)
24. Segmentation models [Электронный ресурс] –2020 - Режим доступа до ресурсу: https://github.com/qubvel/segmentation_models
25. Feature Pyramid Network for Multi-Class Land Segmentation [Электронный ресурс]/Selim Seferbekov, Vladimir Iglovikov, Vladimir Iglovikov, Alexey Shvets. – Режим доступа до ресурсу: https://openaccess.thecvf.com/content_cvpr_2018_workshops/papers/w4/Seferbekov_Feature_Pyramid_Network_CVPR_2018_paper.pdf
26. Google Colaboratory [Электронный ресурс] – Режим доступа до ресурсу: <https://colab.research.google.com/notebooks/intro.ipynb>
27. Набір даних «HAM10000» [Электронный ресурс]/ Tschandl Philipp, Rosendahl Cliff, Kittler Harald –2018 - Режим доступа до ресурсу: https://www.researchgate.net/publication/324078202_The_HAM10000_Dataset_A_Large_Collection_of_Multi-Source_Dermatoscopic_Images_of_Common_Pigmented_Skin_Lesions
28. A New Dataset and Boundary-Attention Semantic Segmentation for Face Parsing. Yinglu Liu, Hailin Shi, Hao Shen, Yue Si, Xiaobo Wang, Tao Mei. In AAAI, 2020
29. Набір даних «Sky» [Электронный ресурс] – Режим доступа до ресурсу: <https://www.ime.usp.br/~eduardob/datasets/sky/>
30. Набір даних «CamVid» [Электронный ресурс] – Режим доступа до ресурсу: <http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/>

31. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L, TPAMI 2017