

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Програмної інженерії  
(повна назва)

**АТЕСТАЦІЙНА РОБОТА**  
**Пояснювальна записка**  
рівень вищої освіти – другий (магістерський)

Дослідження нейромережевих методів для розпізнавання людського обличчя  
(тема)

Виконав: студент 2 курсу, групи ПЗМ-18-1

Магда М.А.  
(прізвище, ініціали)

спеціальності 121– Інженерія програмного забезпечення  
(код і повна назва спеціальності)

Освітньо-наукової програми  
(тип програми)

Інженерія програмного забезпечення  
(тип програми)

Керівник к.т.н., доц. Чуприна А.С.  
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри \_\_\_\_\_  
(підпис)

Дудар З.В.  
(прізвище, ініціали)

2020 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Програмної інженерії

Рівень вищої освіти – другий (магістерський)

Спеціальність 121 – Інженерія програмного забезпечення

(код і повна назва)

Тип програми освітньо-наукова програма

Освітня програма Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_

(підпис)

«\_\_\_\_\_» \_\_\_\_\_ 20\_\_ р.

## ЗАВДАННЯ НА АТЕСТАЦІЙНУ РОБОТУ

Студентові Магді Микиті Андрійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи (проекту): Дослідження нейромережових методів для розпізнавання людського обличчя

затверджена наказом по університету від " " \_\_\_\_\_ 2020 р. № \_\_\_\_\_

2. Термін подання студентом роботи (проекту)

3. Вихідні дані до роботи (проекту) скрипти для аналізу набору даних, скрипти для тренування нейромережі, пояснювальна записка. Використовувати ОС Windows.

4. Перелік питань, що потрібно опрацювати в роботі: мета роботи, аналітичний огляд методів розпізнавання зображень, аналіз нейронних мереж для класифікації зображень, експериментальне дослідження ефективності роботи нейронних мереж для знаходження обличчя на зображенні

## 5. Консультанти розділів роботи

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Спецчастина	<u>к.т.н., доц. Чуприна А.С.</u>		

**КАЛЕНДАРНИЙ ПЛАН**

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка*
1.	Аналіз предметної галузі	23 березня 2020 р.	
2.	Огляд існуючих методів	30 березня 2020 р.	
3.	Дослідження методів класифікації томографічних зображень	6 квітня 2020 р.	
4.	Підготовка пояснювальної записки	13 квітня 2020 р.	
5.	Спецчастина	20 квітня 2020 р.	
6.	Підготовка презентації та доповіді	27 квітня 2020 р.	
7.	Попередній захист	4 травня 2020 р.	
8.	Нормоконтроль, рецензування	11 травня 2020 р.	
9.	Занесення диплома в електронний архів	13 травня 2020 р.	
10.	Допуск до захисту у зав. кафедри	15 травня 2020 р.	
* заповнюється вручну після виконання чергового пункту			

Дата видачі завдання \_\_\_\_\_ 2020 р.

Студент \_\_\_\_\_  
(підпис)Керівник роботи (проекту) \_\_\_\_\_ к.т.н., доцент Чуприна А.С.  
(підпис) (посада, прізвище, ініціали)

## РЕФЕРАТ / ABSTRACT

Атестаційна робота магістра містить: 52 с., 13 рис., 26 джерел.

РОЗПІЗНАВАННЯ ОБЛИЧЧЯ, АЛГОРИТМ ВІОЛИ-ДЖОНСА, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, ЗАГОРОДЖЕННЯ, РОЗА, МАСШТАБ, ОСВІТЛЕННЯ, PYTHON.

Об'єктом дослідження є процес виявлення людського обличчя на фотографії. Предметом дослідження є методи та алгоритми виявлення обличчя.

Метою роботи є порівняння методів виявлення обличчя у різних складних ситуаціях.

Методи розробки базуються на мові програмування Python, дистрибутиві Anaconda та фреймворку TensorFlow з використанням бібліотек NumPy, SciPy, OpenCV. Під час дослідження було здійснене порівняння різних архітектур згорткових нейронних мереж між собою. Було проведено порівняння їх середньої точності при детекції обличчя на обраних категоріях зображень зі складними випадками. Було дано рекомендації щодо використання та метод розрахунку найкращої архітектури для специфічних ситуацій.

FACE DETECTION, VIOLA-JONES ALGORITHM, CONVOLUTIONAL NEURAL NETWORKS, OCCLUSION, POSE, SCALE, ILLUMINATION, PYTHON.

The object of research is the process of detecting a human face in a photograph. The subject of research is methods and algorithms for face detection.

The aim of the work is to compare methods of face detection in different complex situations.

Development methods are based on the Python programming language, Anaconda distribution and TensorFlow framework using NumPy, SciPy, OpenCV libraries. During the study, a comparison of different architectures of convolutional neural networks was performed. A comparison of their average accuracy in face detection on selected categories of images with complex cases was performed. Recommendations were given and the method of calculating the best architecture for specific situations.

## ЗМІСТ

Вступ.....	6
1 Аналіз проблемної галузі .....	8
1.1 Огляд задачі розпізнавання обличчя.....	8
1.2 Проблеми виявлення обличчя .....	12
1.3 Постановка задачі .....	15
2 Аналіз існуючих методів.....	16
2.1 Метод Віюли-Джонса.....	16
2.2 Застосування машинного навчання .....	19
2.3 Двоступеневі детектори .....	23
2.4 Одноступеневі детектори.....	22
2.5 Метрики якості.....	23
2.6 Технології та засоби .....	25
2.7 Дані для аналізу та проведення експериментів .....	27
3 Опис проведених досліджень .....	29
4 Аналіз результатів досліджень .....	32
Висновки .....	36
Перелік джерел посилання .....	37
Додаток А. Слайди презентації.....	40
Додаток Б. Тези доповіді .....	48
Додаток В. Апробація результатів роботи.....	52

## ВСТУП

Завдання виявлення людського обличчя в природній або штучній обстановці для подальшої ідентифікації завжди було одним з пріоритетних завдань для дослідників, що працюють в області систем машинного зору і штучного інтелекту. Проте, багато досліджень, що проводяться в провідних наукових центрах усього світу протягом декількох десятиліть, так і не призвели до створення реально працюючих систем комп'ютерного зору, здатних виявляти і розпізнавати людське обличчя в будь-яких умовах.

Виділення обличчя – це один з декількох кроків у процесі розпізнання обличчя та людини на фотографії або відео, а розпізнавання обличчя – одна з найпопулярніших тем в біометричних дослідженнях. Багато громадських місць зазвичай мають камери спостереження для відеозйомки, і ці камери дуже цінні для забезпечення безпеки. Широко визнано, що розпізнавання обличчя відіграло важливу роль у системах спостереження, оскільки воно не потребує співпраці об'єкта. Традиційні системи ідентифікації вимагають знання пароля, наявності ключа, ідентифікаційної картки, або іншого ідентифікує предмета, який можна забути або втратити. На відміну від них, біометричні системи ґрунтуються на унікальних біологічних характеристиках людини, які важко підробити і які однозначно визначають конкретну людину. Оскільки обличчя людини – це динамічний об'єкт, що має високу мінливість у своєму зовнішньому вигляді, що робить розпізнавання обличчя складною проблемою в комп'ютерному зорі. У цій галузі головна проблема – точність та швидкість ідентифікації. Тому покращення точності та пришвидшення є актуальними задачами сьогодення.

Кафедра програмної інженерії якраз займається розробкою математичних моделей механізмів людського зору та розробкою моделей, методів і алгоритмів розпізнавання для біометричних систем.

Також за цією темою проводяться дослідження та розробки у проблемній науковій групі «Інтелектуальний аналіз даних», відкритій у ХНУРЕ у вересні 2018 року професором Смеляковим К.С. В 2019 році вона отримала назву Data Science Lab та зараз займається аналізом даних, статистикою, класифікацією, кластеризацією, машинним навчанням, добуванням даних і передбачувальною аналітикою.

Дана робота спрямована на методи виявлення людського обличчя у різних складних ситуаціях: з різним масштабом, освітленням, перекриттям, у різних позах та виразах обличчя, з наявністю макіяжу або без нього.

В ході атестаційної роботи магістра було:

- проведено аналіз існуючих методів для розпізнавання образів;
- проведено аналіз використання згорткових нейронних мереж для задачі розпізнавання образів;
- проведено дослідження швидкості навчання згорткових мереж з різними архітектурами мереж для вилучення ознак;
- проведено дослідження якості роботи різних архітектур мереж для категорій зображень з різними позами, освітленням, загородженням та масштабом обличь;
- зроблені висновки про найкращу модель у загальному випадку та випадку використання для мобільної біометричної автентифікації.

Результати дослідження опубліковані на міжнародній науково-практичній конференції «PUBLIC COMMUNICATION IN SCIENCE: PHILOSOPHICAL, CULTURAL, POLITICAL, ECONOMIC AND IT CONTEXT» від 15-го травня 2020 року (див. додаток Б).

# 1 АНАЛІЗ ПРОБЛЕМНОЇ ГАЛУЗІ

## 1.1 Огляд задачі розпізнавання обличь

Розпізнавання обличь (РО) є визначною біометричною методикою автентифікації особи та широко застосовується в багатьох сферах, таких як військова, фінансова, громадська безпека та повсякденне життя. РО є давньою темою дослідження в спільноті CVPR (конференція з комп'ютерного зору та розпізнаванню образів). Розпізнавання обличчя – це складний багатокроковий процес, результат якого дозволяє ідентифікувати або перевірити особу на цифровому зображенні або відеокадрі [1]. Незважаючи на те, що спочатку системи розпізнавання обличь використовувались як застосунки, останнім часом все частіше використовуються на мобільних платформах та в інших технологіях, таких як робототехніка. Зазвичай воно використовується для контролю доступу в системах безпеки нарівні з іншими біометричними системами, такими як розпізнавання райдужної оболонки, відбитку пальців [2]. Хоча точність системи розпізнавання обличь як біометричної технології є нижчою, ніж розпізнавання райдужної оболонки ока та розпізнавання відбитків пальців, вона широко застосовується завдяки безконтактному та неінвазійному процесу. Останнім часом вона також стала популярною як комерційний інструмент ідентифікації та маркетинговий інструмент [3]. Інші застосунки мають такі елементи, як просунута взаємодія людини з комп'ютером, відеоспостереження, автоматичне індексування зображень та відео.

Платформи соціальних медіа застосували можливості розпізнавання обличь, щоб диверсифікувати свої функціональні можливості з метою залучення більш широкої бази користувачів серед жорсткої конкуренції з боку різних додатків.

Заснований у 2013 році, Looksey продовжував збирати гроші на додаток для модифікації обличчя на Kickstarter. Після успішного краудфандингу Looksey почав свою роботу в жовтні 2014 року. Додаток дозволяє спілкуватися у відео чаті з іншими користувачами через спеціальний фільтр для обличь, який змінює

зовнішній вигляд користувачів. Незважаючи на те, що додатки, які доповнюють зображення, такі як FaceTune і Perfect365, обмежені статичними зображеннями, Lookserу дозволив створювати відео з ефектом доповненої реальності в реальному часі. Наприкінці 2015 року SnapChat придбав Lookserу, який потім став його знаковою функцією фільтрів (рис. 1.1)[4].

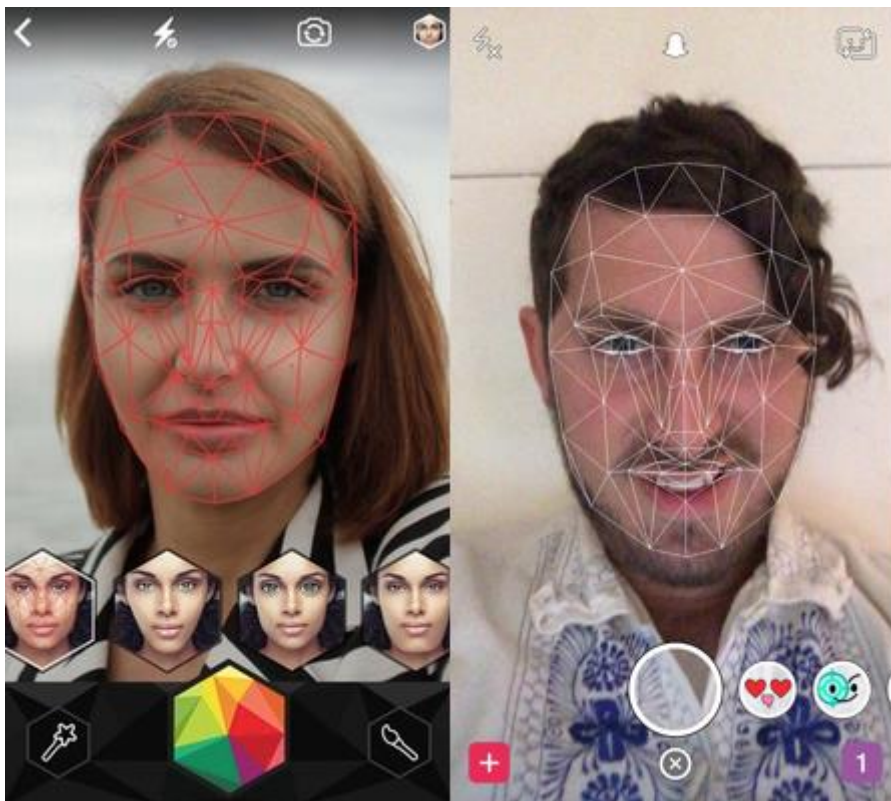


Рисунок 1.1 – Фільтри Lookserу та SnapChat

Анімовані фільтри SnapChat, які використовували технологію розпізнавання обличчя, еволюціонували й оновили селфі, дозволяючи користувачам додавати фільтри для зміни способу їх вигляду. Вибір фільтрів змінюється щодня, вони дозволяють виглядати як старий і зморщений варіант себе, перефарбовують шкіру, або навіть поміщають віртуальну квітку на вершину голови. Фільтр з собакою є найпопулярнішим фільтром, який допоміг здобути постійний успіх SnapChat. Його використовують популярні знаменитості, такі як Джіджі Хадід, Кім Кардаш'ян, що люблять регулярно розміщувати відео про себе за допомогою фільтра з собакою.

Розпізнавання обличь використовується і для забезпечення безпеки. Багато компаній зараз працюють на ринку для надання послуг ідентифікації людини для доступу до банків, власних акаунтів та корпоративних акаунтів підприємств.

Австралійські прикордонні війська і Новозеландська митна служба створили автоматизовану систему обробки кордону під назвою SmartGate, яка використовує розпізнавання особи, яке порівнює особу подорожнього з даними в електронних паспортах з мікро-чіпом [5]. Всі канадські міжнародні аеропорти використовують розпізнавання обличчя як частину програми первинної інспекції, яка порівнює обличчя мандрівника з фотографією, що зберігається в електронному паспорті. Ця програма вперше встановлена у Міжнародному аеропорті Ванкувера на початку 2017 року і була розгорнута у всіх інших міжнародних аеропортах у 2018-2019 роках. У Панамі діє система спостереження, яка використовує сотні аеропортів для розпізнавання осіб, що проходять через аеропорт.

У держдепартаменті США діє одна з найбільших систем розпізнавання осіб у світі з базою даних 117 мільйонів дорослих американців, з фотографіями, як правило, взятих з ліцензій водія. Хоча це ще далеко від завершення, його використовують у деяких містах, щоб визначити, хто був на фотографії. З 2016 року розпізнавання обличчя використовувалося для ідентифікації людей на фотографіях, зроблених поліцією в Сан-Дієго і Лос-Анжелесі (не на відео в реальному часі, а лише на фотографіях бронювання) і використання планувалося в Західній Вірджинії та місті Далласі.

У травні 2017 року чоловік був заарештований за допомогою системи автоматичного розпізнавання обличь, встановленої на фургоні, що експлуатувалася поліцією Південного Уельсу. Станом на кінець 2017 року Китай розгорнув технологію розпізнавання обличчя в Сіцзяні (рис. 1.2). Журналісти, які відвідують регіон, знайшли камери спостереження, встановлені кожні сто метрів або близько того в декількох містах, а також контрольні-пропускні пункти для розпізнавання осіб на таких місцях, як автозаправні станції, торгові центри та входи до мечеті.

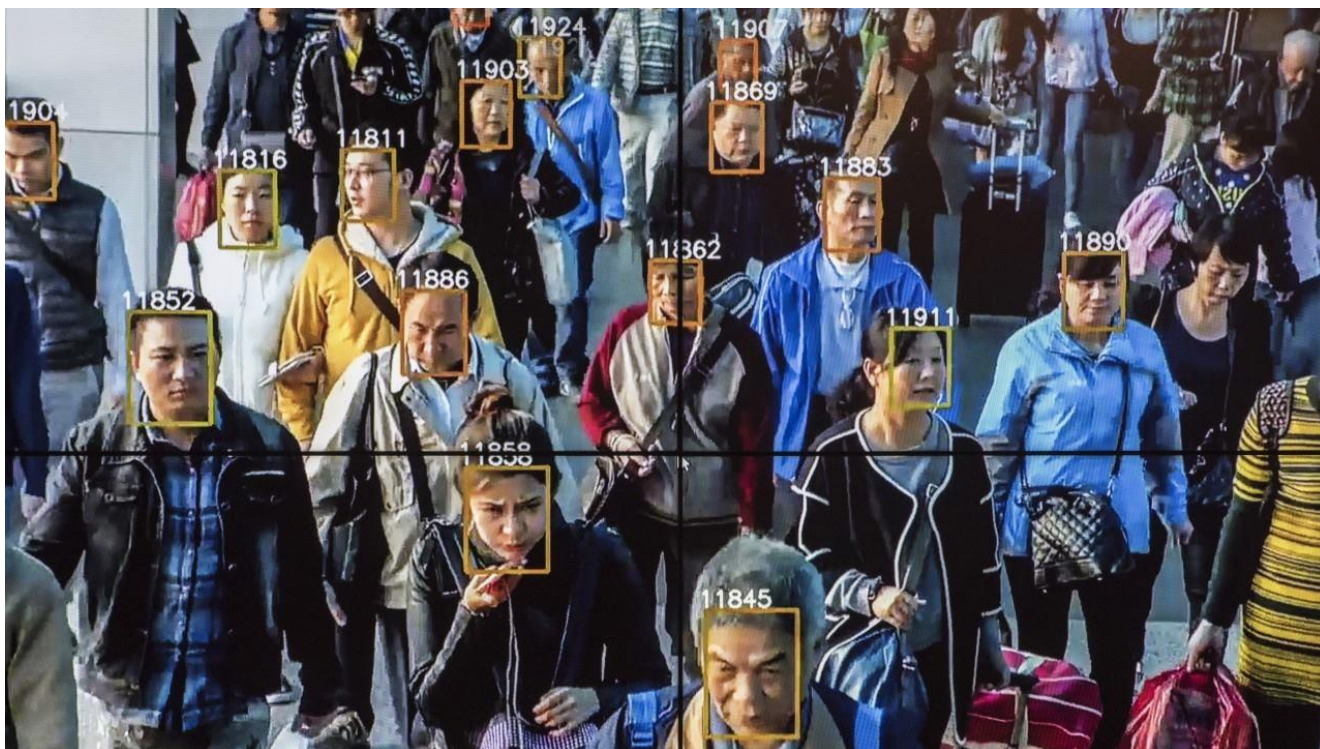


Рисунок 1.2 – Використання владою Китаю технології розпізнавання обличь у місті Сіцзян [6].

ФБР також заснувала свою програму Next Generation Identification яка включає розпізнавання осіб, а також більш традиційну біометрію, таку як відбитки пальців і сканування райдужної оболонки ока, які можна отримати з обох кримінальних і цивільних баз даних.

У 2017 році компанія Time & Attendance ClockedIn випустила розпізнавання обличчя як форму відстеження відвідуваності для підприємств та організацій, які прагнуть мати більш автоматизовану систему відстеження відпрацьованих годин, а також для контролю безпеки та охорони здоров'я.

Згідно зі звіту про аналіз ринку 2018 року, ринок біометричного розпізнавання виросте з 13.89 мільярдів доларів у 2018 до 41.80 мільярдів у 2023 році, з яких більша частина припадає саме на розпізнавання обличчя [7]. Ринок розпізнавання обличчя матиме величезний потенціал у нових областях застосування, пов'язаних з ідентифікацією людей у розумних містах, зокрема, електронною адміністрацією, розумними будинками та розумною освітою.

## 1.2 Проблеми виявлення обличчя

Серйозною проблемою, що стоїть перед системами комп'ютерного зору, є велика мінливість візуальних образів, пов'язана зі змінами освітленості, забарвлення, масштабів, ракурсів спостереження. Крім того, люди мають звичку ходити по вулицях і в приміщенні одягненими, що призводить до суттєвої мінливості зображень однієї і тієї ж людини. Однак найбільш складним завданням комп'ютерного зору є проблема усунення неоднозначності, що виникає при проектуванні тривимірних об'єктів реального світу на плоскі зображення.

Усі фактори, які зумовлюють різний вигляд обличчя людини можна класифікувати на дві групи: внутрішні та зовнішні [8].

Внутрішні фактори зумовлені тільки фізичною природою обличчя і не залежать від спостерігача. Ці фактори можна далі розділити на два класи: внутрішньо-особистісні та міжособистісні [9]. Внутрішньо-особистісні фактори відповідають за зміну зовнішності тієї самої людини, наприклад, вік, міміка та атрибутика обличчя (волосся на обличчі, окуляри, косметика тощо). Однак міжособистісні фактори відповідають за відмінності у зовнішності обличчя різних людей, наприклад, етнічна приналежність та стать.

Зовнішні фактори – це ті, які спричиняють зміни зовнішнього вигляду обличчя завдяки взаємодії світла з обличчям та спостерігачем. До таких факторів належать параметри освітленості, пози, масштабу та зображення цілком, наприклад, роздільна здатність, фокус, шум тощо).

Розглянемо основні фактори, які є проблемою при виявленні обличчя.

Мінливість освітлення означає світлові варіації. Зміни освітлення можуть змінювати загальну величину інтенсивності світла, відбитого назад від предмета, а також схему затінення та тіні, видимі на зображенні [10]. Дійсно, зміна освітленості може призвести до більших відмінностей зображень, ніж зміни особи, або точки зору обличчя. Один і той же знімок, який знімається за допомогою однієї і тієї ж камери і бачиться з майже однаковою мімікою обличчя та позою, може виглядати

різко по-різному при зміні умов освітлення (рис. 1.3). Проблема виявлення обличчя через зміни освітленості загальновизнана як складна для людини та алгоритмів. Труднощі, що виникають при змінних умовах освітлення, залишаються важливою проблемою для систем автоматичного виявлення обличчя. Встановлено, що різниця між двома зображеннями однієї і тієї ж людини, зробленими при різному освітленні, більша, ніж різниця між зображеннями двох різних осіб, що знаходяться в одній і тій же освітленості. Зміна освітленості різко змінює зовнішній вигляд обличчя.



Рисунок 1.3 – Приклад варіації освітленості [9]

Поза змінюється з кутом огляду спостерігача та поворотом у положенні голови. Ці зміни постави завдають серйозної проблеми для ідентифікації вхідного зображення. Система розпізнавання обличчя може опрацювати випадки з невеликими кутами повороту, але це стає проблемою, коли кут повороту стає вище, а наявне зображення в базі даних може мати лише вигляд фронтальної сторони обличчя, яке може відрізнитися позою з вхідним зображенням і яке вводить в оману систему призводячи до неправильної ідентифікації або відсутності розпізнавання (рис. 1.4).



Рисунок 1.4 – Приклад варіації пози [11]

Загородження є важливим фактором для оцінки ефективності виявлення обличчя, адже є однією з найскладніших проблем. Це пов'язано з наявністю на обличчі різних перегороджень, таких як окуляри, борода, вуса тощо, або коли зображення отримується з камери спостереження, тому що деяких частин обличчя не вистачає. У реальних програмах також дуже часто зустрічаються люди, які розмовляють по телефону або в окулярах, шарфах, шапках тощо, або з якихось причин закривають обличчя руками (рис. 1.5). Така проблема може серйозно вплинути на процес знаходження обличчя на фотографії.



Рисунок 1.5 – Приклад загородження обличчя на фотографії

Проблема з низькою роздільною здатністю виникає в системі виявлення обличчя, коли роздільна здатність фотографії на якій знаходиться обличчя дуже мала, а розмір обличчя дорівнює або менше 16 пікселів у висоту або ширину. Ця проблема трапляється у багатьох програмах спостереження, таких як невеликі автономні програми для камер у супермаркетах та банках, відеоспостереження на громадських вулицях тощо, де зображення, зняті з камери спостереження, як

правило, складаються з дуже невеликої площі обличчя і не можуть забезпечити достатню роздільну здатність обличчя для визнання. Зображення обличчя низької роздільної здатності складається з дуже обмеженої інформації, оскільки більшість деталей втрачається. Це може різко знизити рівень розпізнання.

Тобто, можна сказати, що сучасні методи знаходження обличчя мають багато різних типів обставин, що погіршують якість їх роботи. Деякі пов'язані з відсутністю якоїсь кількості інформації на зображенні, інші з загальною якістю зображення або кількістю об'єктів. Це значить, що різні методи будуть по-різному працювати в залежності від типу проблемного зображення.

### 1.3 Постановка задачі

Метою магістерської атестаційної роботи є розгляд декількох архітектур згорткових нейронних мереж та проведення їх оцінки. Необхідно порівняти їх складність та час, потрібний на навчання. Також потрібно дослідити їх точність, порівняти зі значеннями розробників та визначити чи можливо використовувати у розробці; дослідити точність на різних категоріях зображень з ускладненнями та зробити висновок з рекомендаціями по використанню у різних програмних застосунках, у різних сферах. Необхідно знайти найкращу архітектуру, якщо це можливо, або дослідити у яких ситуаціях слід схилитися до певної технології. Слід розробити додаток для візуальної демонстрації роботи кожної архітектури для наочного порівняння результатів та проведення дослідів у різних ситуаціях.

## 2 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ

### 2.1 Метод Віоли-Джонса

Метод був розроблений і представлений в 2001 році Полом Віолою і Майклом Джонсом. Він до сих пор є ефективним методом для пошуку об'єктів на зображенні. Треба відмітити, що цей детектор має низьку ймовірність помилкового знаходження обличчя. Метод добре працює і визначає риси обличчя особи, навіть при спостереженні об'єкта під невеликим кутом. При куті нахилу більше 30° вірогідність знаходження обличчя зменшується. Зазначена особливість методу не дозволяє в стандартній реалізації детектувати обличчя людини, повернене під довільним кутом, що в значній мірі ускладнює або робить неможливим використання алгоритму в сучасних виробничих системах з урахуванням їх зростаючих потреб.

Для того, щоб розрахувати яскравість прямокутної ділянки зображення, використовують інтегральне уявлення [12]. Таке уявлення використовується часто і в інших методах, наприклад, в вейвлет перетвореннях, Speeded up robust feature (SURF), фільтрах Хаара і багатьох розроблених алгоритмах. Інтегральне уявлення дозволяє швидко розраховувати сумарну яскравість довільного прямокутника на даному зображенні, причому час розрахунку не залежить від площі прямокутника. Інтегральне представлення зображення являє собою матрицю, збігається за розмірами з вихідним зображенням. У кожному її елементі зберігається сума інтенсивностей всіх пікселів, що знаходяться лівіше і вище даного елемента. Елементи матриці розраховуються за формулою (1).

$$I(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (1)$$

де  $I(x, y)$  – значення точки  $(x, y)$  інтегрального зображення,  $i(x', y')$  – значення інтенсивності початкового зображення.

На основі застосування інтегрального представлення зображення обчислення ознак однакового виду, але з різними геометричними параметрами, відбувається за однаковий час. Кожен елемент матриці  $I(x, y)$  являє собою суму пікселів в прямокутнику від  $i(0,0)$  до  $i(x, y)$ , тобто значення кожного елемента  $I(x, y)$  дорівнює сумі значень усіх пікселів вище та лівіше від нього. Розрахунок матриці займає лінійний час, пропорційний числу пікселів в зображенні і його можна виразити формулою (2).

$$I(x, y) = i(x, y) - I(x - 1, y - 1) + I(x, y - 1) + I(x - 1, y). \quad (2)$$

Інтегральне уявлення має цікаву особливість: за допомогою інтегральної матриці можна дуже швидко вирахувати суму пікселів довільного прямокутника.

Використовуючи інтегральне зображення, можна дуже легко і швидко обчислювати ознаки Хаара. В алгоритмі Віюлі-Джонса використовуються прямокутні ознаки, або примітиви Хаара, зображені на рисунку 2.1:

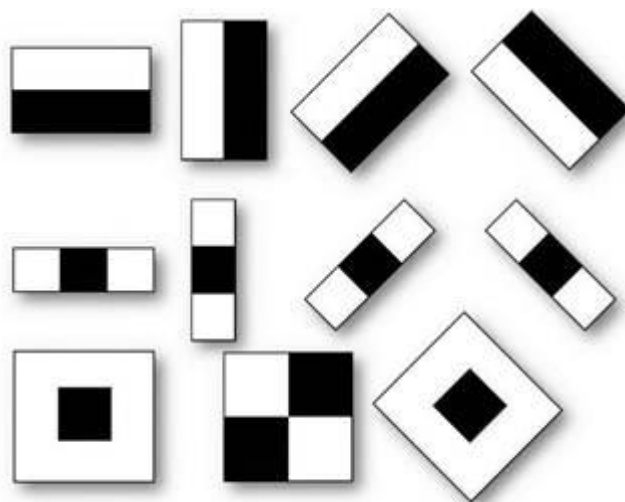


Рисунок 2.1 – Примітиви Хаара [13]

Значення обчислюваної ознаки буде різниця суми яскравостей пікселів білої і чорної області ознаки.

Надалі потрібно зробити обчислення ознак на зображенні. Для цього використовується ковзне вікно. Вибирається невелике вікно сканування, всередині нього обчислюються різні ознаки. Потім вікно пересувається по зображенню з кроком в 1 комірку. Після знаходження всіх ознак змінюється масштаб скануючого вікна, і потім знову відбувається розрахунок ознак. Далі знайдені ознаки подаються на вхід класифікатору, який в подальшому виносить рішення чи є виявлені ознаки людським обличчям. Одним з найважливіших і самих довгих етапів є навчання класифікатора [14].

Для навчання класифікатора в методі Віоли-Джонса використовується технологія бустінга. Бустінг – це комплекс методів, що дозволяють з композиції простих класифікаторів побудувати більш складний. Безпосередньо для методу Віоли-Джонса алгоритм класифікації виглядає таким чином:

- а) Маючи вибірку даних, на яких вказано необхідний об'єкт класифікації, обчислюємо різні примітиви Хаара.
- б) Для кожної ознаки вибирається відповідний поріг для класифікації.
- в) Відбираються найкращі примітиви і відповідний поріг для кожного примітиву.

Приклад каскадної моделі сильних класифікаторів показаний на рисунку 2.2:

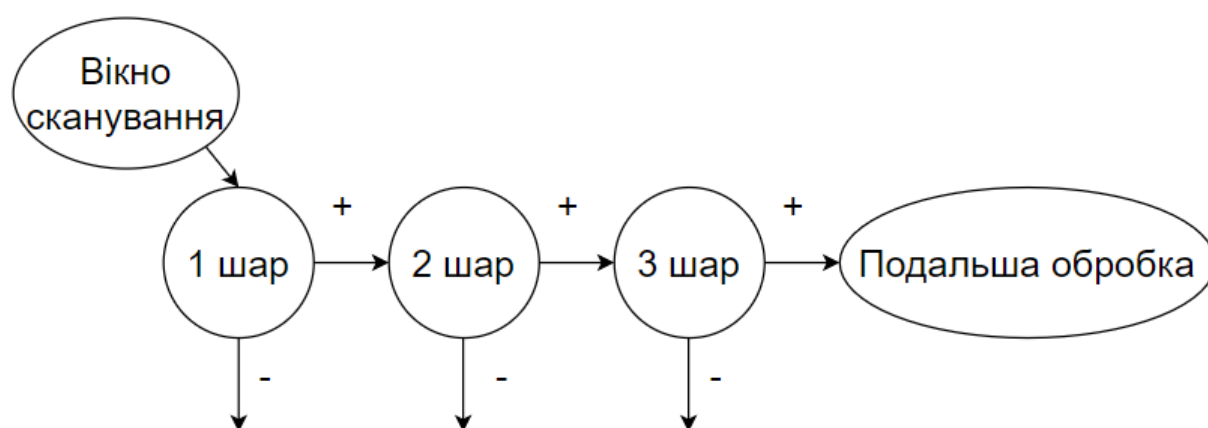


Рисунок 2.2 – Схема каскадного детектору

При застосуванні до зображення на першому етапі використовуються прості класифікатори для відкидання частини скануючих вікон. До позитивних значень

першого класифікатора застосовується другий і так далі. Таким чином застосовується ланцюжок з класифікаторів, і помилка стає все менше і менше.

Головним недоліком методу є тривалість його навчання, а також використання великої кількості даних для навчання класифікатора. Також алгоритм не дає спрацьовування при повороті особи на кут більше 30 градусів. Крім того, при досить сильному освітленні або затемненні, алгоритм також може не спрацювати.

## 2.2 Застосування машинного навчання

Ключовим завданням виявлення обличчя є виявлення на зображенні декількох присутніх обличчя незважаючи на масштаб, освітлення, перекриття, ті інші проблеми, які зустрічаються у реальному житті. Віола та Джонс вказували на те, що алгоритми навчання, такі як бустинг або метод опорних векторів, та функції зображення, такі як вейвлети та примітиви Хаара, недостатньо сильні, а отримані класифікатори безнадійно неточні. Однак, з нещодавнім розвитком глибокого навчання та обчислень за допомогою графічних процесорів, стало можливим використати високу потужність згорткових нейронних мереж для вилучення ознак та тренування набагато точніших класифікаторів.

Згорткові нейронні мережи нещодавно показали неабияку ефективність у різних завданнях комп'ютерного зору, таких як виявлення та розпізнавання обличчя, класифікація об'єктів та виявлення об'єктів [15]. Зокрема, тренувана 8-шарова нейронна мережа під назвою AlexNet показала, що глибокі згорткові нейронні мережі можуть значно перевершити інші методи [16]. Мережа AlexNet 30 вересня 2012 року перемогла у змаганні по розпізнаванню образів ImageNet, досягнувши похибки top-5 у 15.3%, що майже на 11% нижче ніж попередники [16]. Тому для вирішення завдання виявлення обличчя були обрані згорткові нейронні

мережі, тому що вони забезпечують часткову стійкість до змін масштабу, зсувів, поворотам, перекриттям та інше.

Архітектура згорткової нейронної мережі складається з багатьох шарів [14]. Шари бувають двох типів: згорткові та агрегувальні, вони чергуються один з одним як показано на рисунку 2.3.

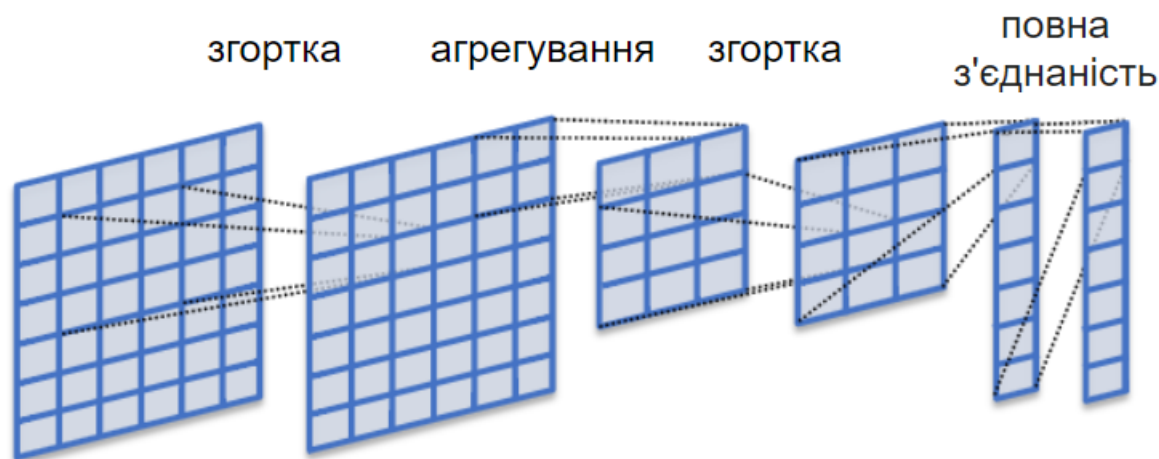


Рисунок 2.3 – Приклад архітектури згорткової нейронної мережі

Нейрони в межах шару організовані в площинні. У кожному шарі є набір з декількох площин, причому нейрони однієї площини мають однакові вагові коефіцієнти синапсів, проведені до всіх локальних частин попереднього шару. Кожен нейрон шару отримує входи від деякої області попереднього шару (локальне рецептивне поле), тобто вхідне зображення попереднього шару як би сканується невеликим вікном і пропускається крізь набір синоптичних коефіцієнтів, а результат відображається на відповідний нейрон поточного шару. Набір площин являє собою карти характеристик, і кожна площина знаходить «свої» ділянки зображення в будь-якому місці попереднього шару. Розмір локального рецептивного поля вибирається самостійно в процесі розробки нейронної мережі. Агрегуючий шар зменшує масштаб площин шляхом локального усереднення значень виходів нейронів. Таким чином досягається ієрархічна організація. Наступні шари витягують найбільш загальні властивості, менше залежать від спотворень зображення.

Поступово нейронна мережа навчається виділяти ключові характеристики обличчя в поданих на вхід зображеннях.

В епоху глибокого навчання виявлення об'єктів можна поділити на 2 категорії: двоступеневі детектори та одноступеневі детектори, де перші роблять «грубий», «неточний» процес визначення, а другі – завершують в один крок.

### 2.3 Двоступеневі детектори

Спочатку набули популярності двоступеневі детектори. Першим та значним був R-CNN [17]. Ідея, що стоїть за R-CNN, проста: вона починається з вилучення набору об'єктних пропозицій (об'єктів-кандидатів) шляхом вибіркового пошуку [18]. Потім кожен пропозицію змінюють у масштабі до фіксованого розміру і подають у модель CNN, підготовлену на ImageNet (скажімо, AlexNet) для отримання функцій. Нарешті, лінійні класифікатори SVM використовуються для прогнозування наявності об'єкта в кожній області та для розпізнавання категорій об'єктів. Хоча R-CNN зробила великий прогрес, вона була дуже повільною через те, що потрібно було запустити CNN на великій кількості об'єктних пропозицій (близько 2000). Це приводить до дуже повільної роботи навіть на графічних процесорах (близько 14 секунд на одне зображення). Пізніше SPPNet [19] вирішив цю проблему обчислюючи карти характеристик тільки один раз для всього зображення і тоді представлення довільних областей фіксованої довжини можуть бути сформовані для навчання детекторів, що дозволяє уникнути багаторазового обчислення згорткових ознак.

Наступним кроком у розвитку була мережа Fast R-CNN [20], яка об'єднала найкраще від R-CNN та SPPNet [19]. Fast R-CNN дозволяє одночасно тренувати детектор та регресор обмежувальних контурів з тією ж конфігурацією мережі. Це дозволяє не втрачаючи точність (близько 65% на VOC07 датасеті) мати приріст швидкості у 200 разів порівняно з R-CNN.

У 2015 році була запропонована Faster R-CNN [21], яка використовувала Region Proposal Network – нейронну мережу для об'єктних пропозицій без втрати швидкості. Це стало причиною того, що Faster R-CNN стала першою нейронною мережею яка працює майже в реальному часі (17 кадрів у секунду).

## 2.4 Одноступеневі детектори

Наступним кроком розвитку була поява одноступеневих детекторів. Першим був запропонований YOLO (You only look once) в 2015 році [22]. Він надзвичайно швидкий: швидка версія спроможна обробляти близько 150 зображень на секунду, що в десятки разів перевищує двоступеневі алгоритми зберігаючи точність у 53% на VOC07 датасеті. Замість використання парадигми об'єктних пропозицій та їх верифікації, автори вирішили однією нейронною мережею розділяти зображення на регіони та прогнозувати граничні поля та ймовірності для кожного регіону одночасно. Хоча YOLO здається вирішенням проблеми детекції об'єктів, нажаль, вона дуже сильно страждає від падіння точності локалізації для невеликих об'єктів. Вирішити цю проблему намагалися наступні версії YOLO [23, 24] та SSD [25].

SSD (Single Shot Detector) [25] був запропонований у 2015 році та був другим одноступеневим детектором. Його головною перевагою була більша точність для маленьких об'єктів порівняно з іншими одноступеневими детекторами. Автори досягли точності у 77% на тому ж датасеті за допомогою детекції об'єктів на різних рівнях масштабу.

У 2017 році представлено RetinaNet [26]. Автори виявили причини, через які одноступеневі детектори були позаду двоступеневих за точністю. За їх словами, це через те, що при навчанні не приділялося уваги на великий контраст між об'єктами та фоном. Також вони представили нову функцію втрат «focal loss», що при навчанні приділяла більше уваги складним, не розпізнаним прикладам. Це все призвело до значного росту точності одноступеневих детекторів.

## 2.5 Метрики якості

Для порівняння різних методів виявлення обличчя треба спочатку визначити якість кожного методу. Для цього потрібно обрати метрику, за якою будемо оцінювати ефективність. Згідно з дослідженням [17] мета метрик для виявлення об'єктів – отримати значущу оцінку здатності системи виконувати завдання з виявлення об'єктів. Метрики включають кількість правильно виявлених об'єктів, помилково виявлених об'єктів або взагалі не виявлених об'єктів. Широко використовуваними метриками виявлення є точність виявлення та повнота.

Точність показує частку правильно знайдених об'єктів, виразимо формулою (3).

$$precision = \frac{|relevant \cap retrieved|}{retrieved} \quad (3)$$

де *precision* – точність,

*relevant* – кількість релевантних об'єктів,

*retrieved* – кількість знайдених об'єктів.

Повнота – це кількість правильно розпізнаних об'єктів відносно загальної кількості об'єктів. Виразимо формулою (4).

$$recall = \frac{|relevant \cap retrieved|}{relevant} \quad (4)$$

де *recall* – повнота,

*relevant* – кількість релевантних об'єктів,

*retrieved* – кількість знайдених об'єктів.

Для задач класифікації використовують терміни True positives, True negatives, False positives та False negatives. True positive означає, що класифікатор прогнозує

позитивне значення, та воно є вірним, true negative – теж саме, але для негативного значення. False positive значить, що прогнозоване негативне значення невірне, False negative те ж саме, але для негативного значення.

Можемо виразити точність та повноту через введені вище терміни:

$$\begin{aligned} \textit{precision} &= \frac{\textit{True positives}}{\textit{True positives} + \textit{False positives}} \\ \textit{recall} &= \frac{\textit{True positives}}{\textit{True positives} + \textit{False negatives}} \end{aligned} \quad (5)$$

Варіюючи значення довіри, можна також створити криву PR (рисунок 2.4).

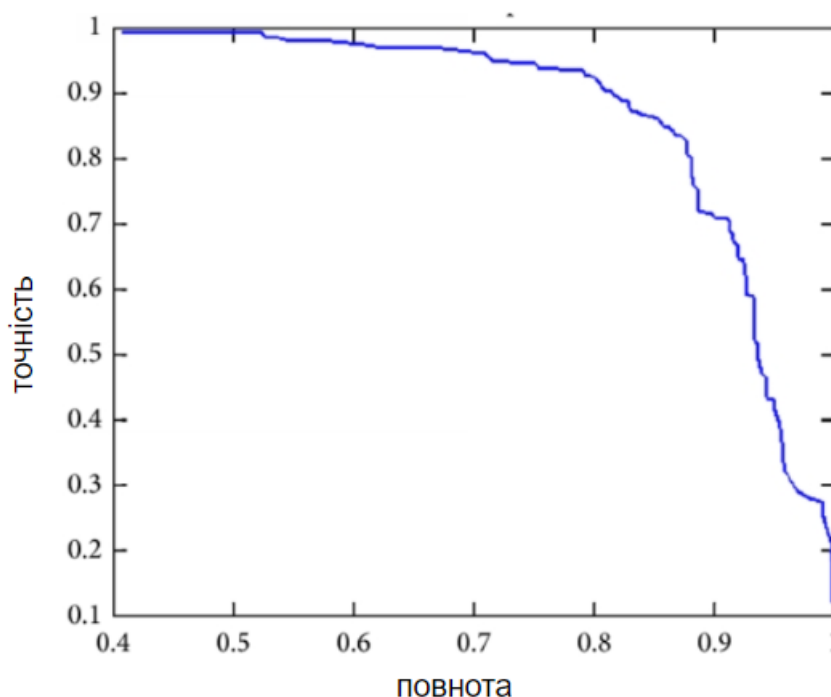


Рисунок 2.4 – Приклад PR кривої

У розпізнаванні образів та пошуку інформації точність (також називається позитивною прогножною величиною) – це частка отриманих екземплярів, які є релевантними, тоді як повнота (також відома як чутливість) – це частка відповідних екземплярів, які отримані класифікатором. Тому точність і повнота є метриками релевантності.

Ще один спосіб порівняння продуктивності детекторів об'єктів – це обчислення площі під PR кривою. Оскільки PR криві часто є зигзагоподібними кривими, що йдуть вгору і вниз, порівнювати різні криві (різні детектори) на одному й тому ж графіку зазвичай не є легким завданням, оскільки криві, як правило, перетинаються одна з одною дуже часто. Ось чому середня точність, числовий показник, також може допомогти нам у порівнянні детекторів. На практиці середня точність – це середнє значення усіх значень точності між значеннями повноти від 0 до 1.

11 точкова інтерполяція намагається узагальнити форму PR кривої шляхом усереднення точності на наборі одинадцяти рівних відстаней значень повноти [0, 0.1, 0.2, ..., 1]:

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} p_{interp}(r) \quad (6)$$

$$p_{interp}(r) = \max p(r'),$$

де AP – середня точність,

$p(r')$  – значення точності при повноті  $r$ .

Замість використання точності, що спостерігається в кожній точці, AP отримують, інтерполюючи значення точності лише на 11 рівнях, приймаючи максимальну точність, значення повноти якої перевищує  $r$ .

## 2.6 Технології та засоби

Найпопулярнішою мовою програмування серед розробників систем розпізнавання образів є C++ та Python. Ці дві мови відрізняються швидкістю, C++ через свою низькорівневість, а Python, тому що написаний на C++ та більшість модулів або пакетів, які він використовує, теж написані на C++.

Але все ж Python є лідером через те, що має дуже багато фреймворків, написаних спеціально для математичних розрахунків за допомогою графічних процесорів. Також, безперечною перевагою Python є швидкість прототипування, адже ця мова інтерпретована. Це означає, що можна виконувати програму частинами, що полегшує розробку.

Найвідомішим дистрибутивом Python є Anaconda. Це вільно та відкрито розповсюджуваний дистрибутив мов Python та R, який спеціалізується на наукових обчисленнях, науці про дані, методах машинного навчання та інше. Дистрибутив має інтегрованим менеджер пакетів та модулів (більше 1500), тому використовується близько 15 мільйонами користувачів. Anaconda має вбудований редактор коду Jupiter Notebook, який працює в браузері (рисунок 2.5).

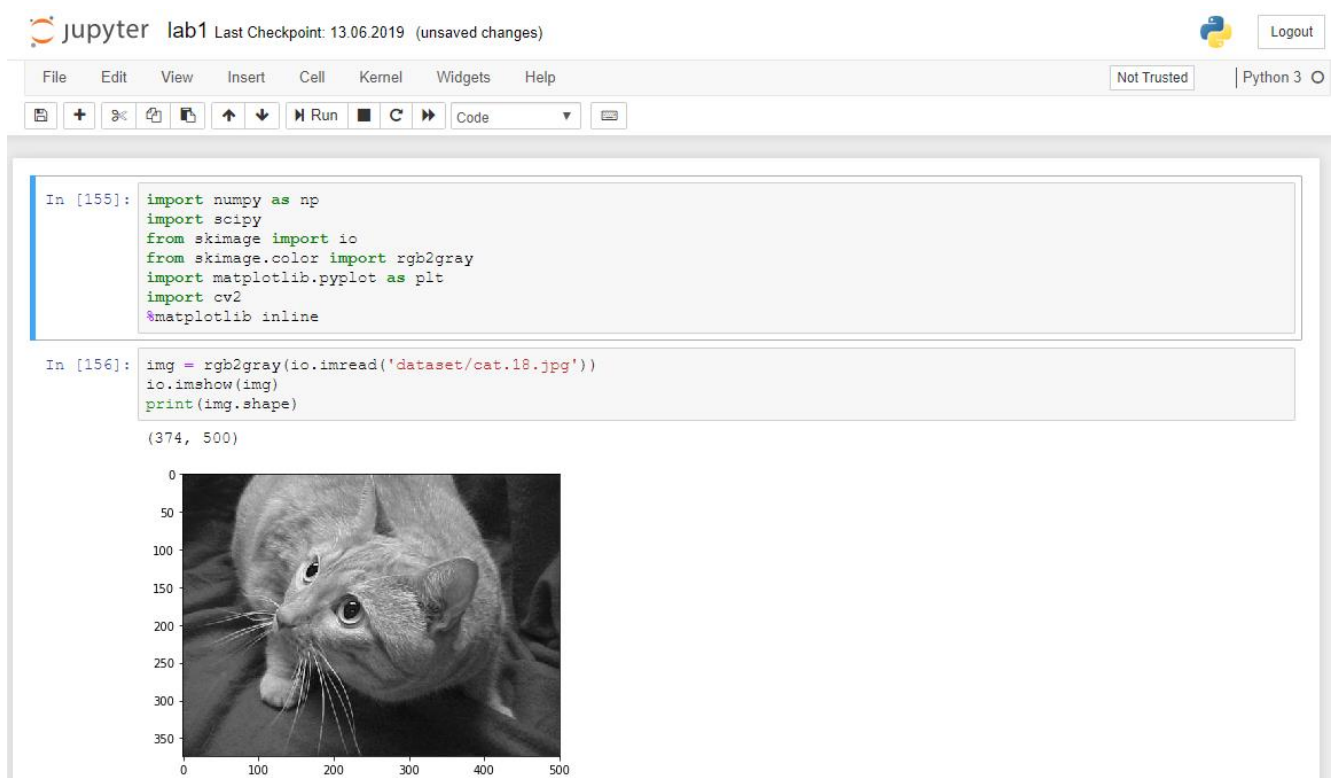


Рисунок 2.5 – Вікно редактору Jupiter Notebook

Окрім редактору та обраної мови програмування потрібен фреймворк. Найпопулярнішими фреймворками є TensorFlow, PyTorch, Keras, MXNet, Caffe, Caffe2 та CNTK. Кожен з фреймворків має свої плюси на мінуси, але всі вони мають

можливість для створення глибоких нейронних мереж, навчання їх та тестування. Кожен з них має вже реалізовані архітектури нейронних мереж, набори даних та інструменти для розрахунку точності та продуктивності отриманих моделей.

## 2.7 Дані для аналізу та проведення експериментів

Для того, щоб навчити класифікатор або згорткову нейронну мережу виявляти обличчя, треба дуже багато інформації. У нашому випадку ця інформація – це фотографії людей, на яких видно обличчя та дані про те, скільки на кожній фотографії людей та де ці обличчя знаходяться. У цій роботі пропонується використання набору даних WIDER FACE [11]. Це найбільший набір зображень з розміченими обличчями. Він містить 32203 зображень на яких знаходяться 393703 обличчя (рис. 2.6).

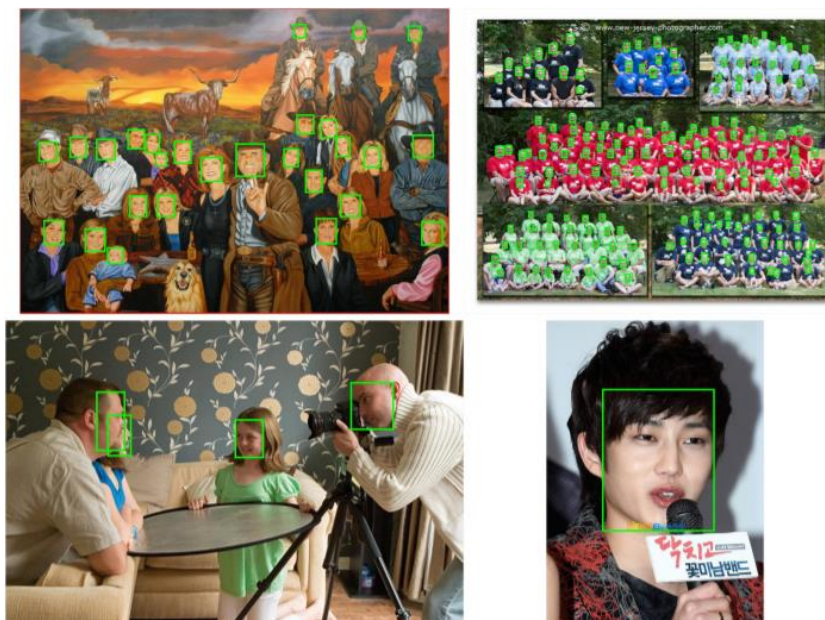


Рисунок 2.6 – Приклад зображень з датасету WIDER FACE

Для того, щоб визначити як кожен з алгоритмів працює зі складними ситуаціями (перекриттям, освітленням та ін.) потрібно поділити дані на категорії складності. Розробники датасету у своїй праці [11] привели графіки складності

тематики зображень за категоріями ускладнення детекції. Наприклад, для проблеми масштабу обличч найскладнішими тематиками є Traffic (фотографії авто-трафіку) Festival та Parade (фотографії фестивалей та парадів) (рис 2.7).

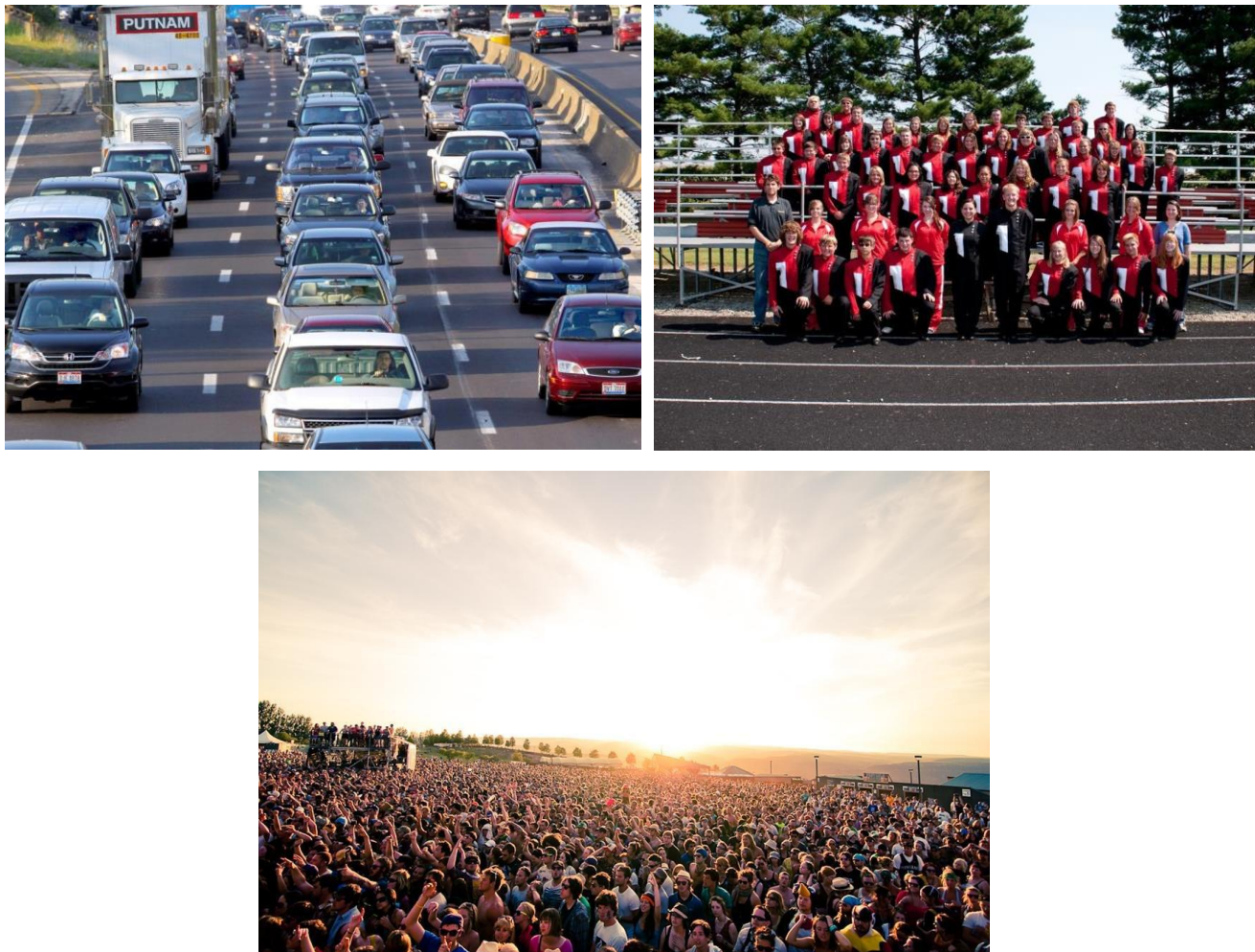


Рисунок 2.7 – Приклад зображень с маленьким масштабом обличч

Те ж саме стосується категорій перекриття та пози, усе що пов'язано з освітленням та якістю потрібно вибирати власноруч. Для проведення експериментів потрібно вибрати близько 100 зображень для кожної категорії.

### 3 ОПИС ПРОВЕДЕНИХ ДОСЛІДЖЕНЬ

Дослідження будуть проводитися для наступних архітектур мереж для детекції об'єктів:

- Faster R-CNN (InceptionV2)
- R-FCN (ResNet101)
- RetinaNet (ResNet50)
- SSD (MobileNetV1)

Faster R-CNN та R-FCN обрано як найкращих представників двоступеневих детекторів, а RetinaNet та SSD – як одноступеневих. Хоча R-FCN є покращеним варіантом R-CNN, а RetinaNet за запевненням авторів краще за швидше SSD, це все може бути хибним через специфіку задачі виявлення обличчя, адже ми маємо тільки 1 клас, який мережа повинна навчитися визначати. Також архітектури були обрані з різними мережами для вилучення ознак.

Побудова нейронної мережі для використання у продуктах або сервісах починається з приготування даних та навчання самої мережі. Дуже важлива швидкість навчання, адже для великих повнозв'язних мереж може знадобитися кілька тижнів роботи багатьох графічних процесорів для виходу на прийнятний рівень похибки. Тому перше дослідження стосуватиметься потрібного часу на навчання кожної архітектури мережі. Цей час залежить від складності (кількості параметрів) та архітектури.

Навчання буде проводитися на графічному процесорі NVIDIA GTX 1070 з 8 гігабайтами пам'яті. Через те, що час навчання на одній відеокарті становить декілька днів, звизимо кількість зображень у тренувальному датасеті до прийнятної кількості – 460. Розмір батчу (кількість зображень опрацьованих одночасно) буде дорівнювати 1, кількість епох – 100, а скорочення швидкості навчання 0.001.

Зображення для навчання не відіграють значної ролі, тому було обрано перші 460 зображень у категорії парад.

Результати дослідження швидкості навчання наведено у таблиці 3.1

Таблиця 3.1 – Результати дослідження швидкості навчання.

Модель	Розмір (MB)	Час навчання (хвилин)
Faster-RCNN	641 MB	40
R-FCN	623 MB	74
RetinaNet	203 MB	42
SSD	22 MB	53

Наступне дослідження передбачає оцінку якості роботи різних моделей на різних категоріях ускладнень. Для цього було обрано по 100 валідаційних зображень з датасету WIDER FACE та розподілено у категоріях «Поза» (різний поворот голови), «Перекриття» (коли частина обличчя перекрита та не видна), «Освітлення» (коли обличчя знаходиться в тіні або на яскравому світлі), «Малі обличчя» (коли обличчя на зображенні дуже мале, розміром близько 15 пікселів у висоту та ширину).

Кожна модель була запущена на цих зображеннях з параметром IoU = 0.5. Отримані результати mAP було занесено в таблицю 3.2.

Таблиця 3.2. Результати оцінки якості роботи моделей.

Модель	Поза	Перекриття	Освітлення	Малі обличчя	Швидкість опрацювання одного зображення
Faster-RCNN	0.56	0.70	0.58	0.60	416 ms
R-FCN	0.67	0.69	0.55	0.62	170 ms
RetinaNet	0.73	0.91	0.65	0.89	91 ms
SSD	0.72	0.68	0.6	0.67	22 ms

Висновок моделі для передбачуваного обмежувального поля навряд чи є точним первинним обмежувальним полем. Тому для вимірювання того, наскільки

точним є об'єкт, визначений у зображенні або кадрі, ми можемо використовувати метрику IOU (intersection over union).

Це дає нам можливість вважати об'єкт виявленим чи ні. IoU – це простий спосіб оцінювання нашої навчальної моделі та обмежувальна скринька з її роботою на тестовому наборі.

Параметр IoU зі значенням 0.5 означає, що знайдені прямокутники положення обличчя будуть вважатися вірними, якщо перетин знайденої площі та розміченої буде дорівнювати хоча б половині від об'єднання знайденої та розміченої області. Візуальний приклад цієї метрики зображено на рисунку 3.1:

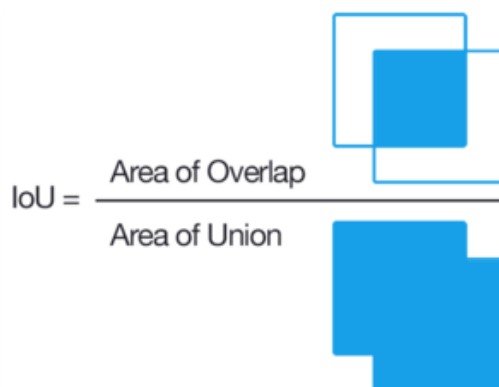


Рисунок 3.1 – Приклад візуальної репрезентації метрики IoU

Також було виміряно середню швидкість опрацювання моделлю одного зображення у мілісекундах.

Цих даних достатньо для аналізу, оцінки якості та швидкості роботи моделей, надання рекомендацій щодо застосування у прикладних задачах.

#### 4 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕНЬ

Маючи всі дані проведених досліджень, потрібно визначити кращу модель з запропонованих до розгляду. Це багатокритеріальна задача прийняття рішення. Перенесемо усі значення до однієї таблиці, нормуємо, а швидкість навчання та опрацювання одного кадру, розмір моделі – обернемо (таблиця 4.1).

Таблиця 4.1 – Таблиця нормалізованих результатів.

Модель	Поза	Перекр.	Освіт.	Малі обличчя	Швидкість роботи	Час навчання	Розмір
Faster-RCNN	0.56	0.70	0.58	0.60	0	1	0
R-FCN	0.67	0.69	0.55	0.62	0.08	0	0.03
RetinaNet	0.73	0.91	0.65	0.89	0.19	0.94	0.71
SSD	0.72	0.68	0.6	0.67	1	0.62	1

Оскільки всі критерії в даному випадку є порядковими, мають різні одиниці виміру і є рівнозначними за важливістю, то скористаємося методом лінійної адитивної згортки з нормувальними множниками.

Формула лінійної адаптивної згортки з ваговими коефіцієнтами представлена нижче:

$$Z^* = \max \sum_{j=1}^n \alpha_j \beta_j a_{ij}, \quad (7)$$

де  $\alpha_j$  – нормуючі множники,

$\beta_j$  – вагові коефіцієнти, що відображають відносний внесок окремих критеріїв до загального критерію.

Визначимо вагові коефіцієнти простим ранжуванням. Якість роботи в пріоритеті, за нею йде швидкість роботи, розмір та останній – час навчання:  $(1, 2, 3, 4) > 5 > 7 > 6$ . Тоді очки розподіляються наступним чином:

- якість роботи – 5.5;
- швидкість – 3;
- розмір моделі – 2;
- час навчання – 1.

Сума очків дорівнює 28, тоді вагові коефіцієнти будуть дорівнювати:

Таблиця 4.2 – Вагові коефіцієнти.

Параметр	Значення
Поза	0.20
Перекриття	0.20
Освітлення	0.20
Малі обличчя	0.20
Швидкість роботи	0.10
Час навчання	0.07
Розмір	0.03

Отримаємо результат згортки:

Таблиця 4.3 – Результат згортки.

	$\alpha_1\beta_1a_{i1}$	$\alpha_2\beta_2a_{i2}$	$\alpha_3\beta_3a_{i3}$	$\alpha_4\beta_4a_{i4}$	$\alpha_5\beta_5a_{i5}$	$\alpha_6\beta_6a_{i6}$	$\alpha_7\beta_7a_{i7}$	$Z^*$
1	0.04179	0.05223	0.04328	0.04477	0	0.02611	0	0.20820
2	0.05	0.05149	0.04104	0.04626	0.00298	0	0.00033	0.19212
3	0.05447	0.06791	0.04850	0.06641	0.00709	0.02455	0.00794	0.27690
4	0.05373	0.05074	0.04477	0.05	0.03731	0.01619	0.01119	0.26395

Отже, за результатом згортки, найкращою альтернативою є альтернатива за найбільшим значенням результату  $Z^*$ , а саме – RetinaNet.

Ці результати можна використовувати для підбору найкращої моделі у специфічних ситуаціях, наприклад, скористаємося нею для визначення найкращого варіанту для якогось випадку.

Припустимо ми розробляємо мобільним додаток, який передбачає біометричну автентифікацію за обличчям. Одним з кроків розпізнавання обличчя є його детекція. При цьому, найважливішим параметром є те, наскільки точно модель знаходить обличчя при різних параметрах освітлення. Другим за важливістю параметром йде розмір моделі, адже мобільні пристрої мають обмежений запас пам'яті. Третім параметром йде швидкість роботи, адже ми не хочемо змушувати користувача чекати. Оскільки ми при використанні дивимося на телефон майже рівно, наше обличчя досить близько розташовано та рідко чимось загороджене. Ми можемо враховувати інші результати в останню чергу. Тоді очки розподіляються наступним чином:

- освітлення – 7;
- розмір – 6;
- швидкість роботи – 5;
- поза, перекриття, масштаб, час навчання – 2.5.

Перерахуємо згортку, використовуючи нові вагові коефіцієнти:

Таблиця 4.3 – Нові вагові коефіцієнти.

Параметр	Значення
Поза	0.09
Перекриття	0.09
Освітлення	0.25
Малі обличчя	0.09
Швидкість роботи	0.18
Час навчання	0.09
Розмір	0.21

Таблиця 4.4 – Новий результат згортки.

	$\alpha_1\beta_1a_{i1}$	$\alpha_2\beta_2a_{i2}$	$\alpha_3\beta_3a_{i3}$	$\alpha_4\beta_4a_{i4}$	$\alpha_5\beta_5a_{i5}$	$\alpha_6\beta_6a_{i6}$	$\alpha_7\beta_7a_{i7}$	$Z^*$
1	0.01880	0.02350	0.05410	0.02014	0	0.03358	0	0.15014
2	0.0225	0.02317	0.05130	0.02082	0.00537	0	0.00235	0.12552
3	0.02451	0.03056	0.06063	0.02988	0.01276	0.03156	0.05563	0.24556
4	0.02417	0.02283	0.05597	0.0225	0.06716	0.02082	0.07835	0.29182

За результатами згортки, для описаного вище випадку найкраща для використання модель – SSD.

При використанні цього метода слід не забувати відкидати Парето неоптимальні варіанти. Також використання нормування з урахуванням  $\min$  і  $\max$  не завжди дає оптимальний результат. Недолік полягає в тому, що ми допускаємо компенсування нестачі одного показника іншим, наприклад, швидкість роботи моделі її розміром. Це не завжди має сенс та є правильним, тому залежно від ситуації може знадобитися використання нелінійної нормалізації.

Також є сенс визначати вагові коефіцієнти пропорційним методом, замість простого ранжування, якщо цього потребує ситуація. Це може підвищити точність, коли після згортки виходять декілька оцінок  $Z^*$ , які дуже слабо відрізняються або навіть однакові.

## ВИСНОВКИ

В ході атестаційної магістерської роботи було досліджено ефективність використання згорткових нейронних мереж для задачі розпізнання обличчя на зображенні.

У звіті було зроблено огляд задачі розпізнання образів та підзадачі виявлення обличчя на фотографії або відеокадрі. Проаналізовано існуючі проблеми, їх види та сучасні методи вирішення. Було виділено основні алгоритми детекції обличчя: метод Віоли-Джонса та згорткові нейронні мережі; описано алгоритм роботи обох методів.

Були описані основні та найбільш важливі метрики, за допомогою яких будуть порівняні алгоритми, представлені усі необхідні для розрахунків формули. Описано інструменти реалізації, мову програмування, фреймворки та бібліотеки, які будуть використані при розробці програми. Сформульовано задачу атестаційної роботи. Визначено джерело даних для опрацювання, обрано датасет та описано його структуру та параметри.

Виконана робота дозволяє проводити всі необхідні експерименти для виконання сформованої задачі та реалізації її у програмному забезпеченні.

Результати експериментів дають змогу зробити висновок про те, чи вирішувана задача, які з запропонованих методів ефективніше вирішують проблему виявлення обличчя в обраній категорії проблем (перекриття, освітлення, макіяж та інше). Також можна зробити висновок о доцільності використання алгоритмів на слабких IoT або мобільних пристроях.

За результатом роботи було визначено, що в загальному випадку мережа RetinaNet показує кращий загальний результат, а для використання на мобільних пристроях краще використовувати мережу SSD.

У майбутньому результат роботи можна розширювати додаючи нові алгоритми та архітектури нейронних мереж та тестуючи інші категорії даних.

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ**

1. W.Zhao, R.Chellappa, P.J.Phillips. Face recognition: A literature survey, ACM Computing Surveys (CSUR), 2003.
2. M. M. Abdelwahab, S. A. Aly, I. Yousry. Efficient Web-Based Facial Recognition System Employing 2DHOG, Computer Vision and Pattern Recognition, 2012.
3. A. Suman. Automated face recognition: Applications within law enforcement, Market and technology review – «NPIA», 2006.
4. Snapchat Acquires Looksery To Power Its Animated Lenses [Електронний ресурс] / TechCrunch – Startup and Technology News. – Режим доступу: <https://techcrunch.com/2015/09/15/snapchat-looksery/> – 23.03.2020 р. – Назва з екрана.
5. Australia smartgate [Електронний ресурс] / Wikipedia. – Режим доступу: <https://en.wikipedia.org/wiki/SmartGate> – 23.03.2020 р. – Назва з екрана.
6. CHINA USING FACIAL RECOGNITION TECHNOLOGY TO TRACK MUSLIM MINORITY GROUP [Електронний ресурс] / Big Data, Analytics and Insight. – Режим доступу: <https://www.analyticsinsight.net/china-using-facial-recognition-technology-to-track-muslim-minority-group/> – 23.03.2020 р. – Назва з екрана.
7. Biometrics technology market analysis report by end-use. Grand View Research – San Francisco, CA, USA, 2018.
8. S. Gong, S. J. McKenna, and A. Psarrou. Dynamic Vision: From Images to Face Recognition. – Imperial College Press (World Scientific Publishing Company), 2000.
9. T. Jebara. 3D Pose Estimation and Normalization for Face Recognition. – Center for Intelligent Machines, 1996.
10. K. Smelyakov, A. Chupryna, M. Hvozdiev and D. Sandrkin. Gradational Correction Models Efficiency Analysis of Low-Light Digital Image. – 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 2019.
11. S. Yang, P. Luo, C. C. Loy, X. Tang. WIDER FACE. – IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

12. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
13. Дмитрієв Е.А., Танаєв І.В., Швейкіна В.В. Метод Віоли-Джонса. Наукове товариство студентів: зб. ст. по мат. VI міжнар. студ. наук.-практ. конф. № 3 (6), 2006.
14. М.А. Новотарський, Б.Б. Нестеренко. Штучні нейронні мережі: обчислення // Праці Інституту математики НАН України. – Т50. – Київ: Ін-т математики НАН України, 2004. – 408 с.
15. A. Arsenov, I. Ruban, K. Smelyakov, and A. Chupryna. Evolution of convolutional neural network architecture in image classification problems. in Proc. Sel. Papers 18th Int. Sci. Practical Conf. IT Secur. (ITS) CEUR Workshop Process., 2018.
16. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012.
17. R. Girshick, J. Donahue, T. Darrell and J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014.
18. K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W. Smeulders. Segmentation as selective search for object recognition. 2011 IEEE International Conference on. IEEE, 2011.
19. He K., Zhang X., Ren S., Sun J. (2014) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8691. Springer, Cham.
20. R. Girshick. Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015.
21. S. Ren, K. He, R. Girshick and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, 2017.

22. J. Redmon, S. Divvala, R. Girshick and A. Farhadi. You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016.
23. J. Redmon and A. Farhadi. YOLO9000: Better, Faster, Stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017.
24. [1804.02767] YOLOv3: An Incremental Improvement [Электронный ресурс] / arXiv.org e-Print archive. – Режим доступа: <https://arxiv.org/abs/1804.02767> – 23.03.2020 г. – Назва з екрана.
25. Liu W. et al. SSD: Single Shot MultiBox Detector. Computer Vision. ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905, 2016.
26. [1708.02002] Focal Loss for Dense Object Detection [Электронный ресурс] / arXiv.org e-Print archive. – Режим доступа: <https://arxiv.org/abs/1708.02002> – 23.03.2020 г. – Назва з екрана.