КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет комп'ютерних наук та кібернетики
Кафедра теоретичної кібернетики

Кваліфікаційна робота
на здобуття ступеня бакалавра
за спеціальністю 122 Комп'ютерні науки
на тему:

Проблема та методи оцінки якості комп'ютерних зображень

Виконав студент 4-го курсу _____
Швець Олександр Вікторович (підпис)

Науковий керівник:
доцент, кандидат фіз.-мат. наук _____
Трохимчук Ростислав Миколайович (підпис)

Засвідчую, що в цій роботі немає запозичень з праць інших авторів без відповідних посилань.
Студент _____
(підпис)

Роботу розглянуто й допущено до захисту на засіданні кафедри теоретичної кібернетики «____» _____ 2021 р., протокол № ___
Завідувач кафедри
проф. Крак Ю.В. _____
(підпис)

Київ – 2021

ЗМІСТ

ВСТУП 3
РОЗДІЛ 1. АНАЛІТИЧНИЙ ОГЛЯД СУЧАСНИХ МЕТОДІВ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ 6
1.1. Постановка задачі 6
1.2. Штучні нейронні мережі. Метод опорних векторів 7
1.3 Компактні цілісні уявлення. Зниження розмірності. Метод головних компонент 13
1.4 Методи оцінки ефективності розпізнавання 16
РОЗДІЛ 2. МОДЕЛІ РЕПРЕЗЕНТАЦІЇ ОБ'ЄКТА НА ЗОБРАЖЕННІ 20
2.1 Математичний апарат моделі 20
2.2 Структура локального еквіваріантного детектора моделі 27
ГЛАВА 3. КОМПЛЕКС АЛГОРИТМІВ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ .. 34
3.1 Алгоритм виділення локальних ознак 34
3.2 Алгоритм оптичного трекінгу 38
3.3 Алгоритм розпізнавання зображень 42
ВИСНОВОК 45
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 48
ДОДАТКИ 52

ВСТУП

Актуальність роботи. Розпізнавання візуальних образів являє собою один з найважливіших компонентів систем управління і обробки інформації, автоматизованих систем і систем прийняття рішень. Завдання, пов'язані з класифікацією та ідентифікацією предметів, явищ і сигналів, що характеризуються кінцевим набором деяких властивостей і ознак, виникають в таких галузях як Робототехніка, інформаційний пошук, моніторинг і аналіз візуальних даних, дослідження штучного інтелекту. Алгоритмічна обробка і класифікація зображень застосовуються в системах безпеки, контролю і управління доступом, в системах відеоспостереження, системах віртуальної реальності та інформаційних пошукових системах. На даний момент у виробництві широко використовуються системи розпізнавання рукописного тексту, автомобільних номерів, відбитків пальців або людських осіб, що знаходять застосування в інтерфейсах програмних продуктів, системах безпеки та ідентифікації особистості, а також в інших прикладних цілях.

Інтенсивні дослідження в цій галузі мають багаторічну історію і пов'язані з роботами Д. Хьюбела і Т. Візела, [13-15], Т. Кохонена [20], М. Турка і А. Петланда [80], Д. Хінтона [12,22], Я. Лекуна [23,24] та інших. За останній час істотний прогрес в розпізнаванні візуальних образів був досягнутий з появою методів зниження розмірності [11], згорткових нейронних мереж [23,39] і констеляційних моделей [5]. Однак, незважаючи на досягнуті успіхи, сучасні дослідження підтверджують той факт, що алгоритми розпізнавання зображень досі не мають повноцінних здібностей біологічних зорових систем, таких як здатність функціонувати на широкому, не обмеженому зверху безлічі класів розпізнавання, стійкість до інваріантних перетворень і варіативності об'єктів у межах категорій.

Так, актуальною проблемою, визнаною науковим співтовариством, залишається розпізнавання зображених об'єктів під дією афінних трансформацій, здатних значним чином змінити форму зображення, не

впливаючи при цьому на приналежність об'єкта до категорії розпізнавання. Спроби вирішення цієї проблеми, що фігурує в теорії розпізнавання образів під назвою проблеми інверсії, робилися в таких методах як SIFT [27] і ORB [35], а також багатошарових згорткових мережах [24], проте зараз ці методи пропонують часткові рішення, що забезпечують стійкість до обмеженої підмножини перетворень. Актуальність даної проблеми особливо висока в галузях, де розпізнавання образів застосовується в природному середовищі (відеоспостереження, аналіз даних камер моніторингу, робототехнічні зорові системи), де зоровий сенсор може мати довільний обмежений кут огляду по відношенню до шуканого об'єкту.

Метою роботи є розгляд методу розпізнавання візуальних образів, здатного вирішувати проблему інверсії для різних галузей застосування, розпізнаючи Тривимірні об'єкти навколишнього світу з урахуванням їх інваріантних перетворень.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

1. зробити аналітичний огляд сучасних методів розпізнавання зображень;

2. розглянути моделі репрезентації об'єкта на зображенні;

3. проаналізувати комплекс алгоритмів розпізнавання зображень.

Об'єктом дослідження є системи комп'ютерного зору, що здійснюють класифікацію та ідентифікацію об'єктів на зображенні.

Предметом дослідження є проблеми та методи оцінки комп’ютерних зображень.

Методи дослідження. Для вирішення поставлених завдань використовувалися методи комп'ютерного зору, теорії оптимізації, математичної статистики, теорії штучних нейронних мереж, імовірнісних моделей, теорії планування експерименту.

Практична значимість методу полягає в здатності обробляти зображення об'єктів інваріантним чином, забезпечуючи стійке розпізнавання

в умовах різних кутів зору, а також різних видів візуального шуму (розмиття, оклюзія, часткове перекриття).

Використання методу дозволяє домогтися підвищення ефективності систем комп'ютерного зору і прийняття рішення за рахунок використання компактних ієрархічних уявлень, що вимагають значно меншої обчислювальної навантаження в порівнянні з альтернативними методами. Особливості представленої моделі дозволяють використовувати її як для вирішення вузькоспеціальних завдань, таких як розпізнавання осіб, з використанням попереднього навчання, так і для узагальненого аналізу даних – для виявлення закономірностей при відеоспостереженні і самонавчання виявленим структурам.

Обсяг і структура роботи: складається з вступу, трьох розділів, висновків та додатків. Повний обсяг роботи становить 57 сторінок. Список використаних джерел містить 37 найменувань.

РОЗДІЛ 1. АНАЛІТИЧНИЙ ОГЛЯД СУЧАСНИХ МЕТОДІВ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ

1.1. Постановка задачі

Розглядається задача знаходження єдиного, універсального способу оцінки якості зображення. Даний спосіб заснований на критеріях, які характеризують ступінь зміни яскравості, тонової насиченості і контрастності зображень. Самі по собі критерії не у всіх випадках дають можливість провести порівняльний аналіз якості зображення. У даній роботі пропонується для оцінки якості використовувати не самі критерії окремо, а функціонал, що відображає ступінь зміни перерахованих вище критеріїв в сукупності.

У першому розділі потрібно провести аналітичний огляд сучасних методів розпізнавання зображень, створити класифікацію типів існуючих підходів до розпізнавання. Визначити проблему розпізнавання тривимірних об'єктів під дією інваріантних просторових перетворень та запропонувати метод її розв'язання на основі практики відомих досліджень.

У другому розділі розглянути моделі репрезентації об'єкта на зображенні, яка дозволяє виробляти інваріантне розпізнавання об'єктів з виконанням тривимірної просторової локалізації. Сформувати принципи організації ієрархічної структури моделі і компонентів окремих рівнів ієрархії, описати математичну модель трансформуючого автоенкодера як елементарної одиниці ієрархічної структури репрезентації, здатної до формування компактних репрезентацій фрагментів зображення з урахуванням їх просторової орієнтації.

У третьому розділі проаналізувати комплекс алгоритмів розпізнавання зображень, який дозволяє витягувати і обробляти ключову для розпізнавання інформацію про тривимірну структуру об'єкта, оцінювати його положення в

просторі і локалізувати об'єкт на зображенні, використовуючи запропонований метод.

1.2. Штучні нейронні мережі. Метод опорних векторів

Один з основних підходів, який найбільш широко використовувався в області розпізнавання зображень, являє собою вживання класичних моделей-класифікаторів, що вивчаються з учителем. Для навчання таких моделей використовуються маркована вибірка даних, що складається з масиву зображень і відповідного їм масиву міток, що визначають категорію, до якої відноситься зображення. У процесі навчання масив даних розділяється на дві нерівні частини — навчальну вибірку і тестову вибірку, потім за допомогою специфічного для конкретного алгоритму правила навчання параметри моделі налаштовуються з використанням навчальної вибірки таким чином, щоб отримавши в якості вхідних даних зображення, модель на виході виробляла б мітку відповідного класу. Цей підхід представлений безліччю моделей, серед яких найбільш широко використовуваними є регресійна модель, штучна нейронна мережа (багатошаровий перцептрон), метод опорних векторів, а також дерева прийняття рішень і моделіансамблі, що представляють собою поєднання деяких перерахованих моделей [36,4].

Багатошарові перцептрони, що навчаються методом зворотного поширення помилки, широко використовуються для розпізнавання різних категорій зображень, таких як рукописні цифри [6], почерк [3], людські обличчя [8] і дані зорових сенсорів робототехнічний систем [30]. Модель багатошарового перцептрона являє собою сукупність штучних нейронів – обчислювальної одиниці моделі – об'єднаних в рівні (шари), задані в ієрархічному порядку.

Штучний нейрон являє собою модель біологічного нейрона (нервової клітини), представлену одним або декількома входами, одним виходом і функцією активації [9]. Крім цього, кожен вхід штучного нейрона має

асоційований коефіцієнт або вагу. Поведінка нейрона будується наступним чином: припустимо є т + 1 входів, значення яких рівні х0, х1, хт, а значення їх ваг рівні ш0, при цьому перший вхідний елемент, як правило, являє собою фіксоване значення зміщення х0 = 1. Тоді вихідне значення нейрона являє собою значення функції активації від зваженої суми його вхідних значень:

$$У = \varphi(\sum wixi) \qquad (1.1)$$

При об'єднанні штучних нейронів в мережу вхідні значення нейрона шару l являють собою вихідні значення нейронів попереднього шару 1-1.

При цьому нейрони першого (вхідного) шару отримують в якості вхідних значень безпосередньо дані, що підлягають розпізнаванню, які в разі розпізнавання зображення являють собою значення інтенсивності складових його пікселів (точкових елементів). Вихідний шар мережі може варіюватися в залежності від завдання, але класична архітектура має на увазі формування його числом нейронів, рівній кількості класів розпізнавання, при цьому вихідне значення кожного нейрона нормується по інтервалу {0,1}, і являє собою ймовірність приналежності вхідного зображення до відповідного класу. Як зазначають дослідники, такі багатошарові нейронні мережі здатні інкапсулювати будь-яку математичну функцію за допомогою довільного набору нейронів [23,7].

Оскільки сформулювати аналітично правило класифікації зображень за категоріями розпізнавання часто представляється скрутним, здатність навчатися на базі вибірки, що робить нейронні мережі і споріднені їм моделі придатними для розпізнавання природних зображень навколишнього світу, що відрізняються нечіткою структурою і безліччю варіацій в межах класу.

Навчання мережі методом зворотного поширення полягає в наступному: припустимо є деяка невідома функція розпізнавання д : X — Y, аргументом якої є зображення хп Е X, представлені у вигляді вектора довжини п, а значеннями функцій — безліч класів (категорій) у Е Y.

Навчальна вибірка являє собою підмножину значень цієї функції D = {Оо, Уо)> (xlt y1),... (хт> ут)}

Завдання навчання моделі розпізнавання полягає в відшуканні такої функції h: X-Y, яка б апроксимувала функцію D на всій її області визначення, в тому числі значеннях, не включених і являє собою додаток теорії оптимізації.

На кожному кроці навчання ваги нейронів інкрементуються значеннями приватних похідних відповідно до методу градієнтного спуску. Варіації навчального алгоритму включають в себе виняткові додаткових параметрів регуляризації з метою захисту від перенавчання, і використання різних оптимізаторів — методу Ньютона, методу імітації відпалу, L-BFGS та інших [2].

Багатошарові перцептрони демонструють успішні результати при використанні їх для розпізнавання зображень деяких окремих обмежених категорій, таких як символи природної мови, рукописні цифри і почерк [23]. В даний час в більшості додатків, що використовують пряме навчання з учителем для розпізнавання зображень, нейронні мережі витіснені методом опорних векторів, що пропонує більш ефективне з точки зору обсягу обчислювальних ресурсів рішення.

Метод опорних векторів розглядає кожен екземпляр даних (зображення) як точку в n-мірному просторі, де п відповідає розмірності даних або загальному числу пікселів зображення. Кожна з точок належить до деякого класу (категорії). При цьому завдання розпізнавання представляється у вигляді завдання по знаходженню такої гіперплощини в і-мірному просторі, яка б відокремлювала всі точки, відповідні зображенням даного класу, від інших, що не належать йому. Припускаючи, що таких гіперплощин може існувати багато, метод опорних векторів ставить за мету відшукання площини, відстань до якої від найближчої точки максимально в межах безлічі можливих варіантів — так звану оптимальну гіперплоскість, яка розділяє і відповідний їй оптимальний Класифікатор.

Вхідні дані, таким чином, мають вигляд:

$$\{(x0, Уо), (xi, Уі),... (xm, ym)\} \quad (1.2)$$

де xi — і-те зображення, а yi — і-тий клас, представлений цілим числом. Значення xL, що представляють собою п-мірний вектор, нормалізуються в межах інтервалу {0,1}. Розділяє площину задається параметром w — перпендикуляром (нормальним вектором) від точки до площини, і описується рівнянням wx — b = 0. Таким чином, завдання зводиться до мінімізації ||w||. За теоремою Куна-Таккера [10] гіперплощинність може бути представлена в якості лінійної комбінації векторів навчальної вибірки:

$$w = \sum aiyixi \quad (1.3)$$

де ai – деякі множники Лагранжа. Знаходження значення w таким чином дозволяє отримати лінійні гіперплоскості, тому такий метод відноситься до розділу т.зв. лінійних опорних векторів. Класифікуюча функція при цьому дорівнює F (x) = sign ((w,x) + b), де b — допоміжний параметр зміщення. На практиці випадки, де дані в задачі розпізнавання можуть бути розділені лінійно, досить рідкісні. У таких випадках застосовується метод використання ядер, запропонований Б. Босером, І. Гійон і В. Вапником, і полягає в тому, що елементи вибірки вкладаються в простір x' більш високої розмірності за допомогою спеціального відображення φ: Rn — x'. При цьому відображення вибирається так, щоб в просторі X' вибірка була розділена лінійно. Ядром класифікатора називається вираз x (x,x') = (φ (x), φ (x')), що задає відображення вибірки в новий простір, і його роль, як правило, може виконувати будь-яка позитивно визначена симетрична функція двох змінних. На практиці зустрічаються такі ядра:

поліноміальне, радіальна базисна функція, гаусова базисна функція, сигмоїда [10].

Навчання моделі, що використовує метод опорних векторів проводиться методами квадратичного програмування, такими як послідовна мінімальна оптимізація [7].

Метод опорних векторів має деякі переваги і недоліки по відношенню до використання багатошарових перцептронів:

1. Багатошаровий перцептрон являє собою модель з безліччю прихованих параметрів, що залежать від числа нейронів мережі. Параметризована модель потенційно здатна до інкапсуляції більш складних, високорівневих функцій, але при цьому вимагає більше часу і обчислювальних ресурсів для навчання та налаштування параметрів. Метод опорних векторів використовує вектори, відібрані з навчальної вибірки, при цьому кількість параметрів обмежена зверху розміром вибірки, а на практиці може бути проріджена за рахунок використання інженерії ознак [10].

2. На відміну від навчання нейронної мережі, яке здійснюється за допомогою методу градієнтного спуску (і його варіацій) і оцінки помилки мережі, навчання моделі опорних векторів включає в себе не тільки оцінку помилки, але і метрику складності отриманої гіперплощини. Пошук оптимального значення нейронної мережі вразливий до наявності локального мінімуму, здатного зупинити процес градієнтного спуску, при цьому метод опорних векторів при коректному виборі метапараметрів гарантує знаходження глобального рішення [25].

3. Навчена нейронна мережа вимагає мінімальних обчислювальних ресурсів для роботи в режимі розпізнавання (передбачення категорій). Метод опорних векторів у деяких випадках, коли число векторів велике порівняно з розміром вибірки, будує передбачення істотно повільніше [10].

4. У порівнянні з нелінійним (використовують ядра) методом опорних векторів, нейронна мережа демонструє розширені здібності до онлайн-

навчання, коли розмір вибірки не фіксований і поповнюється за рахунок надходження нових даних.

У більшості сучасних додатків алгоритмів розпізнавання і машинного навчання зараз віддано перевагу методу опорних векторів [6, 34] за рахунок скорочення часу навчання і стійкості до локального мінімуму. Метод опорних векторів також широко використовується для розпізнавання зображень, таких як людські обличчя, демонструючи високу точність розпізнавання (80-85% успішно розпізнаних зображень) [32] для вирівняної вибірки.

Особливість завдання розпізнавання зображень полягає в тому, що дані, що представляють собою візуальні сигнали, демонструють вкрай низьку інформаційну ємність – тобто, велика частина точок растрового зображення (наприклад, відповідні ділянкам однотонного або рівномірно розподіленого фону) не містить інформації, що впливає на розпізнавання [38]. При цьому розмірність зображень, що використовуються в системах обробки інформації, як правило, досить велика — сучасні засоби мультимедіа, графічні дисплеї і сенсори забезпечують масове поширення зображень (фотографій, кадрів відео, комп'ютерної графіки високого дозволу, розмірність яких вимірюється мільйонами точок. Для класичних методів розпізнавання образів характерна пряма залежність між розмірністю (числом параметрів) даних навчальної вибірки і часом навчання, а також показниками збіжності при оптимізації моделі. Наявність великого числа параметрів, основна частина яких не містить істотної для розпізнавання інформації, негативно впливає на продуктивність моделі, і крім вимоги значно більш високих обчислювальних ресурсів веде до появи проблеми перенавчання [3], коли функція розпізнавання, апроксимована моделлю, задовільно класифікує навчальну вибірку, але є при цьому не генералізованою і демонструє низьку точність у тестовій вибірці. Для вирішення цієї проблеми використовується підхід пошуку компактного представлення зображення – виділення обмеженого

числа генералізованих ознак, що містять основну інформацію, необхідну для розпізнавання.

1.3 Компактні цілісні уявлення. Зниження розмірності. Метод головних компонент

Одна з особливостей розпізнавання зображень в порівнянні з іншими додатками теорії розпізнавання образів полягає в тому, що зображення в растровому вигляді (у вигляді двовимірної матриці пікселів, кожен з яких має деяке значення яскравості або кольору), мають високу розмірність — середньостатистична фотографія може бути представлена вектором довжини ~ 106. Дані, представлені розмірністю таких порядків, вимагають виняткових обчислювальних ресурсів, і практично не піддаються обробці на сучасних персональних комп'ютерах (ситуація, відома як «прокляття розмірності» [1]). При цьому, однак, лише невелика частина цих параметрів критична для завдання розпізнавання, що дозволяє зображенням демонструвати низьку чутливість до випадкового шуму і глобальних спотворень. Ця особливість успішно використовується в алгоритмах стиснення з втратами – за допомогою алгоритму JPEG зображення може бути стиснуто аж до 10%, при цьому зміни залишаються непомітні для людського ока. Враховуючи цю особливість, стає можливим застосування до природних зображень статистичних методів зниження розмірності, таких як метод головних компонент [17]. Суть методу полягає в тому, щоб представити вхідні дані у вигляді лінійної суми компонент з деякими коефіцієнтами.

Класичний метод головних компонент, однак, непридатний для більшості зображень через обчислювальної складності побудови коваріаційної матриці. Турк і Пентланд [41] у 1991 р. запропонували алгоритм розпізнавання Eigenfaces, де використовували альтернативний, прийнятний для сучасних комп'ютерів метод розрахунку власних векторів. У їхньому прикладі метод використовувався на фронтальних фотографіях

людських облич. Підтверджуючи припущення про те, що розмірність зображення може бути значно знижена, зберігаючи при цьому достатньо інформації для успішного розпізнавання людиною, вони показали, що кожне з осіб вибірки можна представити за допомогою обмеженого (<10) набору головних компонент.

Рисунок 1.1 Приклади головних компонент алгоритму Eigenfaces [41]

Для розпізнавання тестові зображення проектувалися на базис обраних головних компонент, тобто представлялися у вигляді лінійної суми p доданків. Потім на представлених таким чином даних тренували модель, що використовує навчання з учителем (багатошаровий перцептрон або SVM), і таким чином, завдання зводилося до класичного. Використання Eigenfaces дозволяло ефективно розпізнавати обличчя при різному освітленні і давало деяку стійкість до орієнтації; однак, алгоритм погано працював на обличчях різного розміру (варіації масштабу). Крім того, алгоритм був розрахований на те, що вхідні дані будуть являти собою особи, зорієнтовані відповідним чином, не пропонуючи методу відшукання фрагмента особи, що цікавить серед зображення композитної сцени.

Крім перерахованих, метод головних компонент мав і інші обмеження, які сприяли появі нових методів представлення зображень. Б. Ольшозен у своїй роботі [33] показав, що алгоритм, названим ним розрідженим кодуванням здатний ефективніше представляти внутрішню структуру зображення і об'єктів в ньому, при цьому демонструючи деякі властивості, вражаюче схожі з властивостями клітин зорової кори головного мозку (так званих «простих клітин» зони V1). Цей алгоритм, однак, на противагу PCA,

представляв дані у вигляді надповного базису векторів, кожен з яких, таким чином, не був лінійно незалежним від інших.

Для розпізнавання зображень успішно застосовувалися глибокі моделі, що складаються з обмежених машин Больцмана — так звані глибокі мережі довіри [12]. Використання ієрархічних уявлень дозволяє такій моделі навчатися складним, масштабним об'єктам, забезпечуючи додаткові рівні стійкості до інваріантних перетворень на кожному шарі представлення. Так, глибока модель, навчена на базі людських облич, здатна розпізнавати значно більш істотні спотворення, ніж модель Eigenfaces, що включають в себе обертання об'єкта в межах обмежених кутів. Глибокі моделі також можуть будуватися і на базі методів розрідженого кодування — одним з найбільш відомих є глибокий автоенкодер [11], що навчаються пошарово, жадібним чином. В цілому глибокі моделі забезпечують більш гнучкі і багаті уявлення, які підходять для об'єктів зі складною структурою. Зворотною стороною цієї переваги є ускладнений процес навчання, в окремих випадках (для глибоких мереж довіри) вимагає розробки окремих алгоритмів, і в загальному випадку — споживає більше обчислювальних ресурсів.

Компактні цілісні уявлення дозволяють позбутися від «прокляття розмірності», перетворюючи складні в обробці, об'ємні зображення в компактний вигляд, забезпечуючи при цьому деяку стійкість до варіативності. Методи, що здійснюють нелінійні перетворення, такі як розріджене кодування, можуть використовуватися для отримання багаторівневих уявлень, використовуючи глибоке навчання і властивість стаціонарності природних зображень (той факт, що статистичні характеристики локальних ділянок зображень, як правило, розподілені рівномірно). При цьому підхід відшукання компактних цілісних уявлень демонструє високі результати для об'єктів, що мають в цілому схожу форму (як людські обличчя), але не здатний справлятися з об'єктами, що мають значні візуальні відмінності (наприклад, відносити до одного класу Автомобілі різних моделей) [3].

Більш того, оскільки розпізнавані об'єкти зазвичай мають тривимірну природу, вони здатні істотно змінювати форму під впливом геометричних трансформацій (Так, зображення особи в профіль не може бути представлено сумою компонентів, отриманих декомпозицією зображення обличчя анфас). В силу умови цілісності отримані уявлення вразливі до проблеми неповних даних — ситуацій, коли частина об'єкта загороджена або невиразна через шум. Для отримання компактних цілісних уявлень, таким чином, необхідна строго підібрана вибірка об'єктів, вирівняних по загальній орієнтації. Складання подібної вибірки має на увазі участь експериментатора і обробки вихідних зображень людиною.

Ці особливості та обмеження методу зниження розмірності призвели до розвитку альтернативного підходу до розпізнавання, специфічного для сфери розпізнавання зображень і використовує виявлення локальних ознак, що представляють собою стійкі компоненти (частини) зображеного об'єкта.

1.4 Методи оцінки ефективності розпізнавання

Актуальним питанням при розпізнаванні зображень є оцінка ефективності роботи методу розпізнавання. Для отримання чисельного значення оцінки широко використовуються як загальні методи математичної статистики, так і специфічні показники, що застосовуються для оцінки алгоритмів машинного навчання.

Однією з найбільш простих метрик оцінки ефективності є процентна частка коректно розпізнаних зображень [36,4,25]. Коректним розпізнаванням вважається отримання на виході алгоритму класу, відповідного попередньо заданому класу. Для оцінки використовується вибірка, спроектована аналогічно навчальній вибірці, але містить зображення, до яких алгоритм не міг мати доступ в процесі навчання. Для цього, як правило, вихідна загальна вибірка розділяється на дві нерівні частини (розмір тестової вибірки при цьому може відрізнятися, і складати 20-30% розміру загальної вибірки [4]).

Тоді якщо функція C(Xi) являє собою функцію-індикатор коректності розпізнавання і-того зображення, тобто:

$$C(Xi)\begin{cases} 1, h(xi) = yi \\ 0, \text{в іншому випадку} \end{cases} \quad (1.4)$$

Розглянутий показник є найбільш узагальненим і підходить для безлічі завдань розпізнавання з обмеженою кількістю класів. У задачах, де кількість класів не фіксована, і завдання розпізнавання являє собою завдання ідентифікації об'єкта певної категорії серед безлічі інших, потенційно необмежених категорій, замість неї застосовуються такі показники як точність і повнота оцінки. Їх використання дозволяє розрізняти помилково-позитивні (Класифікатор прийняв позитивне рішення по зображенню, що не містить шуканого об'єкта) і помилково-негативні (Класифікатор не розпізнав об'єкт на зображенні, де він був присутній) помилки розпізнавання, або помилки першого і другого роду. Таким чином, точність оцінки в межах класу являє собою частку зображень, дійсно належать даному класу щодо всіх зображень які система віднесла до цього класу. Повнота системи – це частка знайдених класифікатором зображень, що належать класу щодо всіх зображень цього класу в тестовій вибірці [36].

Показники точності і повноти широко використовуються в області обробки інформації, і як правило, розраховуються спільно. При цьому існує кілька методик зіставлення двох показників:

- кожен показник враховується індивідуально,

- для випадків, коли між показниками може спостерігатися залежність, проводиться оцінка одного з показників при фіксації іншого (наприклад, оцінка точності при рівні повноти в 0,75) ,

- обидва показники можуть бути скомбіновані в один.

Серед прикладів агрегованих метрик для третьої методики зустрічаються такі показники як P-міра, коефіцієнт кореляції Меттьюза,

регресійні коефіцієнти ΔP та ΔP′. [36]. Також, для деяких цілей розглядається зважене арифметичне середнє показника точності і значення, зворотного показником повноти, і навпаки [37]. З точки зору теорії ймовірності ці показники можуть інтерпретуватися таким чином: точність відповідає ймовірності того, що випадково вибране з множини позитивно упізнаних зображень дійсно розпізнано коректно, при цьому повнота являє собою ймовірність того, що випадково вибрана із загальної зображення буде коректно класифіковано алгоритмом (позитивно або негативно). Так, в залежності від додатка завдання, до продуктивності методу розпізнавання можуть бути пред'явлені вимоги, що стосуються максимізації повноти (для випадків, коли деяка кількість помилково-позитивних рішень допускається), так і збалансованого значення двох показників.

Крім перерахованих показників, важливу роль в оцінці ефективності методів розпізнавання відіграють показники, що дозволяють оцінити метрики роботи алгоритмів, не пов'язаних безпосередньо з прийняттями класифікуючих рішень. Серед цих показників можна виділити розмір довірчого інтервалу для вибірки зображень, оцінка числа ітерацій і часу роботи алгоритму при навчанні, необхідних для досягнення збіжності, темп збіжності і т.д. [25].

Для оцінки розміру вибірки, що забезпечує статистично достовірні результати, використовується метод розрахунків довірчих інтервалів. Для параметра Θ-числа зображень вибірки, що представляє собою випадкову величину X з рівнем довіри p, таку, що X = (x1, x2,.., xп) довірчий інтервал представляється за допомогою таких меж l (xl, x2,.. xп) і u(x1,x2,.., xп), які є реалізаціями випадкових величин L (Xl, X2,.., Xп), U (Xl, X2,.., Xп), таких, що:

$$P (L < \Theta < U) = p \qquad (1.5)$$

при цьому граничні точки довірчого інтервалу I і U являє собою довірчі межі. Рівень довіри p вибирається виходячи з специфіки поставленого

завдання і вимог, що пред'являються до системи обробки інформації (а також пов'язаних з нею ризиків).

Існує набір методик оптимізації, що дозволяють варіювати розміри цього порогового значення виходячи з темпів збіжності на конкретній ділянці області визначення функції, заданої моделлю [4]. У деяких випадках, коли досягнення збіжності ускладнене (наприклад, при оцінці метапараметрів моделі на етапі крос-валідації), число ітерацій навчання задається фіксованим значенням. Для сучасних алгоритмів оптимізації, таких як L-BFGS, minFunc і т.д., це значення, як правило, не перевищує двох тисяч ітерацій [30].

РОЗДІЛ 2. МОДЕЛІ РЕПРЕЗЕНТАЦІЇ ОБ'ЄКТА НА ЗОБРАЖЕННІ

2.1 Математичний апарат моделі

В рамках дослідження була розглянута ієрархічна модель репрезентації об'єкта на зображенні, що використовує оцінку тривимірних просторових характеристик об'єкта. Дана модель може використовуватися для формування репрезентацій об'єктів в загальному випадку будь-яких категорій природних зображень (при цьому в даному дослідженні детально розглядається Категорія Людських осіб). У цьому розділі представлено аналітичне формулювання моделі та її компонентів.

Основу запропонованої моделі складають два рівні абстракцій: локальний детектор і рівень (шар) репрезентації.

Модель являє собою впорядковану ієрархічну послідовність рівнів репрезентації Lo, L1... Lc. Кожен рівень складається з деякої кількості локальних еквіваріантних детекторів ознак DL. Кожен рівень здійснює обробку вхідних даних, отриманих на попередньому рівні, і обчислює результат обробки, відправляючи його на вхід наступного рівня. Результат являє собою значення функції ідентифікатора (активацію детектора) і параметри локалізації об'єкта, оцінені локальним детектором.

Вхідні дані для рівня Lo являють собою набір покадрових послідовностей локальних ділянок зображень, і відповідних їм трансформацій, що витягуються з відеофрагменту зорового сенсора. На базі кожної покадрової послідовності навчається один еквіваріантний детектор. Модель навчається порівнево, жадібним способом. Після того, як перший рівень моделі навчений, вихідні композиції відеофрагменту представляються у вигляді констеляційних графів, де вузлами графів є детектори, навчені на першому рівні. Кожен і-тий рівень для $i \neq 0$, крім цього, здійснює кластеризацію (локальне угруповання) даних, об'єднуючи окремі вузли вхідних графів за деяким евристичним правилом.

Наступні рівні навчаються аналогічним чином на отриманих репрезентаціях даних, після чого процес повторюється для наступних рівнів. Повністю навчена модель здатна представити повнорозмірне зображення у вигляді обмеженої (1-3) кількості детекторів вищого рівня, кожен з яких містить ієрархічне представлення комплексного об'єкта. При цьому модель є специфічною для конкретного класу зображень, що використовуються при її навчанні.

Таким чином, модель характеризується наступними параметрами:

- кількість рівнів репрезентації L0, L1 ... Lc;

- кількість і локальне розташування (координати (xi,yi;) на площині зображення) детекторів ознак Di;

- кількість кластерів сцени Кс для кожного рівня Lc;

- внутрішні параметри локальних детекторів ознак.

Використання моделі для розпізнавання зображень здійснюється наступним чином. Припустимо є шукана функція розпізнавання д: X Y, аргументом якої є зображення Хп ∈ X, представлені у вигляді вектора довжини п, а значеннями функцій — безліч класів (категорій) у ∈ Y, варійоване в залежності від поставленого завдання. Є підмножина пар аргументів і значень функції D = {(x0, y0),.. , (xm, ym)}. Представлена модель, таким чином, реалізує функцію h: X→ Y, яка б апроксимувала функцію g на всій її області визначення, в тому числі в значеннях, не включених в D. Для обчислення значення h (x) зображення надходить на вхід першого шару навченої моделі, потім виконується послідовна активація локальних детекторів на кожному їх рівнях. Вихідна активація моделі являє собою бінарне значення, що визначає приналежність зображення до класу, при цьому вихідний рівень моделі також виробляє оцінку параметрів локалізації зображеного об'єкта, якщо значення активації дорівнює 1 (зображення успішно розпізнано як належить класу).

У початковому вигляді модель використовується для вирішення задачі унарної класифікації, коли безліч класів у ∈ Y представлено двома

елементами і дорівнює {0,1}. Функція g, таким чином, дорівнює 1 у випадках, коли зображення, що служить її аргументом, містить об'єкт, що належить цікавлячому класу, і 0 в іншому випадку. Для випадків, коли потрібно розпізнати зображення серед декількох можливих класів (завдання мультиклассовой класифікації), проводиться навчання моделі для кожного конкретного класу, і потім проводиться почергова перевірка зображення на позитивну відповідність цим екземплярам моделі. У цьому випадку розглядається функція g', визначена на множині x', значеннями якої є множина класів Y ', таке що для обраного i-того класу Y ' = {yj U j ≠ i Yi}.

Схема моделі для двох рівнів репрезентацій наведена на рисунку 2.1.

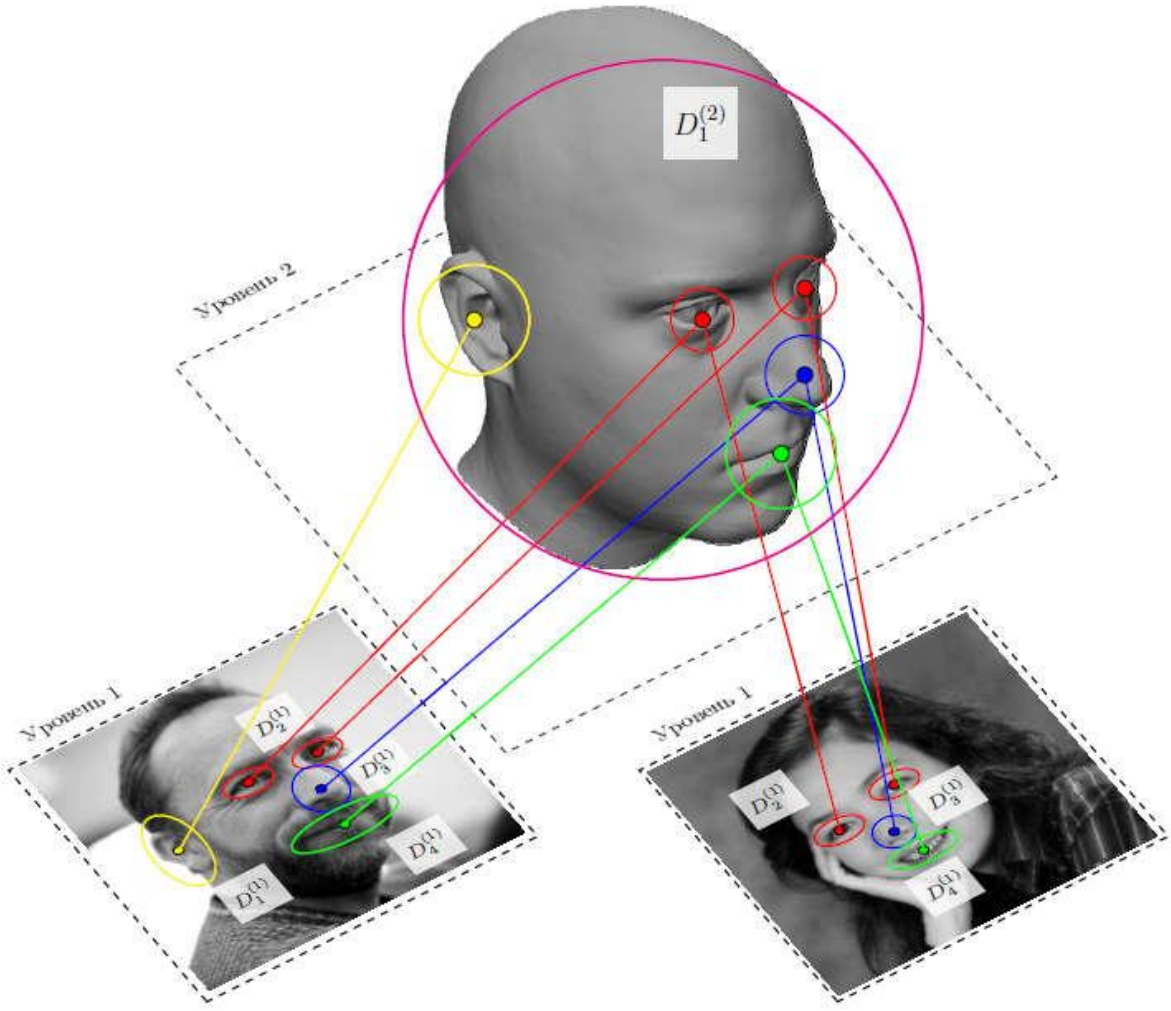


Рисунок 2.1 – Модель представлення об'єктів на прикладі людських осіб, що містить два рівні репрезентації

Основним компонентом моделі є локальний еквіваріантний детектор ознак, що представляє собою відокремлену математичну конструкцію в межах одного рівня ієрархії моделі. Роль локальних детекторів в даній моделі близька до ролі нейронів штучної нейронної мережі – окремий детектор є частиною ієрархічного ланцюжка, отримуючи в якості вхідних даних результати детекторів нижнього рівня, і відправляючи вихідні дані на наступний рівень моделі.

Дамо визначення локального еквіваріантного детектора:

1. Припустимо p – множина n тривимірних точок, заданих координатами (X, Y,Z), і представлених у формі матриці розміру 3 x n.

2. Проекція точок P на площину камери описується за допомогою операції перспективи як Pr(P) = CP, де C — матриця камери. Результатом проекції є безліч двовимірних точок p, заданих координатами (x,y).

3. Розглянемо операцію афінної трансформації в тривимірному просторі Tr (T, P) = TP, де T – матриця трансформації.

4. Припустимо p0 = Pr(P) – деяка проекція об'єкта. Розглянемо безліч довільних матриць трансформацій T1, T2,.. Tm, і безліч проекцій P1, P2, …Pm, таких що Pj = Pr(Tr(Tj, P)).

5. Тоді модель, інкапсулюючі функції Dj (pj) = p0 і Dt (pj) = Tj є еквіваріантним детектором.

Таким чином, еквіваріантний детектор оперує на безлічі форм — проекцій тривимірного об'єкта, і здатний для даної проекції PJ визначити як вихідну форму об'єкта за замовчуванням, так і трансформацію об'єкта на даній проекції. Враховуючи, що позиції об'єктів в тривимірному просторі симетричні, вибір p0 в загальному випадку є довільним. Узагальнення формулювання еквіваріантного детектора дозволяє позбутися концепції форми за замовчуванням для тих випадків, коли її розгляд не є доцільним. Так, якщо Pi = TiP, Pj = TjP, а pi і pj – проекції для відповідних точок, то вірні такі рівності:

$$\begin{cases} \text{Di (Pi)} = \text{Di (Pj)} \\ \frac{DT(Pi)}{DT(Pj)} = \frac{Ti}{Tj} \end{cases} \quad (2.1)$$

Розширимо поняття еквіваріантного детектора для випадків, коли об'єкти належать різним класам. Припустимо Po, P1,…..Pn – безліч тривимірних об'єктів, а Yo, Y1,..Yn – безліч класів. Функція класифікації У(P) = Y визначає приналежність об'єкта до класу. Для деякого класу Yk розглянемо функцію унарної класифікації yk′, таку що

$$\text{yk}'(\text{Pi}) = \begin{cases} 1, \text{y(Pi)} = \text{Yk} \\ 0, \textit{в іншому випадку} \end{cases} \quad (2.2)$$

Тоді (D_I^k,D_T^k) — дискримінативний еквіваріантний детектор для класу Yk, якщо він заданий функціями (D_I^k,D_T^k, такими, що

$$\begin{cases} D_I^k(Pi) = \text{yk}'(\text{Pi}) \\ D_T^k(Pi) = \begin{cases} Ti, D_I^k(Pi) = 1 \\ \emptyset, \text{в іншому випадку} \end{cases} \end{cases} \quad (2.3)$$

Таким чином, дискримінативний детектор не тільки виконує розпізнавання об'єкта, але і визначає конкретну трансформацію, застосовану до зображеного об'єкту. Функція трансформації Тг в загальному випадку може представляти різні операції, не обмежені умовою афінності або виражені у формі, відмінній від матричної. Так, надалі в роботі в демонстраційних цілях будуть розглянуті функції трансляції на площині Тг ((Δx, Δy),P) і функція просторового обертання, виражена кутами обертання Ейлера Tr(((φx,(φy,(φz),P). На малюнку 2.2 наведені приклади деяких вхідних і вихідних значень дискримінативного детектора.

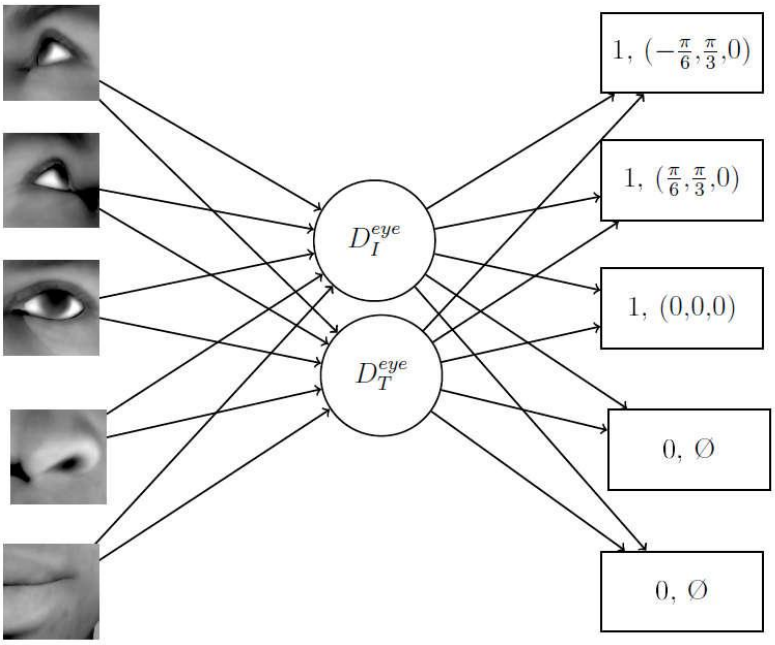


Рисунок 2.2 – Схема дискримінативного еквіваріантного детектора для зображення ока. Вихідні дані детектора являють собою кортеж із значень двох його функцій. Трансформація T в даному випадку представлена вектором трьох кутів Ейлера

Для оцінки ефективності роботи еквіваріантного детектора визначимо поняття функції помилки відновлення трансформації як середньоквадратичне відхилення передбаченої детектором трансформації від вихідної трансформації T:

$$J(Dt,p,T)=\frac{1}{2}\sum_{i=1}^{m}|Dt\left(p^{(i)}\right)-T^{(i)}|^2 \qquad (2.4)$$

Слід зазначити, що ключовим елементом визначення эквиваріантного детектора є той факт, що в якості вхідних даних він приймає тільки двовимірні (спроектовані) зображення, не володіючи доступом до координат об'єкта в тривимірному просторі — в іншому випадку, визначення трансформації T являло б собою тривіальну задачу знаходження різниці між

двома станами об'єкта. З цим обмеженням пов'язані деякі наступні властивості:

- Розглянемо два зображення трансформованих об'єктів pi і pj, яким відповідають трансформації Ti і Tj }. Для випадку, коли pi ≈ pj, але Ti ≠ Tj, детектор не в змозі коректно відновити трансформації по зображенню. Таким чином, DT (pi) ≈ DT (pj).

- Для випадку, коли тривимірні об'єкти P I Q, що належать різним класам дають схожі проекції, так, що p≈ q, аналогічна поведінка спостерігається для функції ідентифікації детектора. Тобто Dj (p) ≈ Dj (q).

З урахуванням перерахованих вище властивостей, очевидно, що існують такі поєднання зображень і трансформацій, для яких теоретично неможливо побудувати точний еквіваріантний детектор (такий, що J→ 0). Фактор варіативності об'єктів збільшує середнє значення помилки в межах класу за рахунок того, що відхилення тривимірних форм об'єктів можуть негативним чином позначатися на розрізненості їх проекцій — зображень. Можна припустити існування негативної кореляції між розміром і різноманітністю об'єкта / зображення і точністю відновлення трансформації. Таким чином, репрезентація цілого об'єкта за допомогою єдиного еквіваріантного детектора представляється завданням, що перевершує за складністю завдання розпізнавання, і надлишкової з точки зору поставленої мети.

Локальні дискримінативні еквіваріантні детектори являють собою основний елемент пропонованої моделі представлення об'єктів на зображенні. Перший рівень моделі представлений детекторами, які можуть реагувати на ділянки зображень у різній просторовій орієнтації. Ансамблі детекторів і їх значення на першому рівні для кожного окремого зображення різні, але при цьому служать матеріалом для навчання високорівневого детектора, що відшукує уявлення для композицій детекторів першого рівня. Таким поданням є тривимірна карта ознак-детекторів першого рівня, розміщених на ній відповідно до даних еквіваріантних трансформацій.

Функція ідентифікації детектора другого рівня, таким чином, визначає остаточну приналежність об'єкта до категорії.

Більш докладний опис моделі включає в себе розгляд її основного елемента – еквіваріантного детектора.

2.2 Структура локального эквіваріантного детектора моделі

Для реалізації елементарної одиниці моделі представлення об'єкта – локального еквіваріантного детектора, розглянемо модель специфічної нейронної мережі – трансформуючого автоенкодера, що є підвидом більш загального класу нейронних мереж зниження розмірності (автоенкодерів). Особливістю трансформуючого автоенкодера є здатність формувати не тільки стійкі до варіативних змін компактні уявлення ділянок зображення, але і проводити оцінку параметрів афінної трансформації, якій піддається зображений об'єкт. Далі розглянемо архітектуру і принципи роботи трансформуючого автоенкодера.

Трансформуючий автоенкодер являє собою нейронну мережу, що навчається методом зворотного поширення помилки. В основі його лежить наступний принцип: вихідний рівень мережі структурно дорівнює вхідному, а в якості еталонних значень для навчання використовуються значення на вході автоенкодера — таким чином, нейронна мережа навчається передбаченню тих же самих даних, що і отримує на вході. Функція, інкапсуліруєма такою мережею, в загальному випадку є тривіальною і являє собою C=f(x)=x, проте в разі автоенкодера на мережу накладається додаткове обмеження — наявність «пляшкового горла» в одному з проміжних (прихованих) шарів, тобто шару з числом нейронів меншим, ніж у вхідному шарі. Нейрони такого шару (для найпростішого випадку автоенкодера може існувати один прихований компактифікований шар), таким чином, являють собою репрезентацію вхідних даних. Враховуючи використання нелінійних функцій активації нейронів і безлічі шарів автоенкодера, таке уявлення може

бути компактним і точним, на відміну від лінійного методу головних компонент. Так, для випадку, коли вхідні дані являють собою порівняно невеликі зображення $x \in R^{28x28=784}$, їх репрезентація може бути представлена прихованим шаром розміру порядку 30, тобто $c = f(x'') \in R^{30}$ [11].

На відміну від класичних багатошарових перцептронів, що мають однорідну структуру в межах шару, трансформуючий автоенкодер являє собою гетерогенну мережу, що складається з декількох мереж меншого розміру. Кожна така мережа носить назву капсули і інкапсулює уявлення конкретної візуальної сутності, представляючи, таким чином, одну параметризовану ознаку об'єкта. Всі капсули автоенкодера мають однакову структуру, і характеризуються наступними особливостями:

1. Кожна капсула включає в себе один вирішальний нейрон p, що приймає значення в діапазоні [0,1], відповідні ймовірності того, що об'єкт присутній на зображенні, і деяка кількість нейронів інстанціювання (залежить від трансформації, яким навчається мережа), що кодують позицію, або параметри інстанціювання об'єкта в просторі.

2. Навчання трансформуючого автоенкодера відбувається наступним чином: вхідні дані являють собою вихідне зображення, вихідні — зображення, піддане трансформації T. В ході навчання автоенкодер пророкує трансформоване зображення, порівнюючи його з вихідними даними і обчислюючи помилку як середньоквадратичне відхилення або кроссентропію.

3. Під час навчання автоенкодеру не повідомляються параметри інстанціювання вихідного об'єкта. Замість він отримує на вході значення трансформації T (наприклад, для випадку зсуву в площині — значення Δx і Ду), і додає їх до вихідних значень нейронів інстанціювання.

4. Автоенкодер навчається компактному коду, відповідному розміром числу нейронів інстанціювання (як правило, для афінних трансформацій — 2-9 вимірювань) кожної капсули. Навчений автоенкодер може витягувати подання зображень прямим поширенням; при результат мережі складається з

суми результатів окремих капсул. Оскільки кожна капсула містить вирішальний нейрон, що реагує на присутність відповідної візуальної сутності, то «зайві» капсули деактивуються значенням p, близьким до нуля, і таким чином, не вносять свій внесок у загальне уявлення автоенкодера.

Основна функціональна частина архітектури, що дозволяє трансформуючому автоенкодеру формувати осмислений код в умовах трансформацій – це додавання значення трансформацій до значень нейронів інстанціювання. Таким чином вчитель повідомляє мережі інформацію про те, як змінюються параметри позиціонування об'єкта на зображення. Розглянемо ситуацію, коли параметри позиціонування являють собою координати об'єкта на площині або матрицю афінної трансформації. При навчанні трансформуючого автоенкодера на таких параметрах відбувається наступне: в разі, коли об'єкт на зображенні зрушений, автоенкодер активує той же набір капсул, що і для оригінального зображення, але вихідне значення нейронів інстанціювання для них інкрементується значенням, в точності рівним величині зсуву. Таким чином, капсула кодує просторове положення об'єкта в компактній формі, що відповідає обраному поданню координат (для тривимірних трансформації це матриця розміру 3X3). Таким чином, архітектура мережі дозволяє отримати не тільки компактне уявлення об'єкта, але і явно задати для автоенкодера смислове значення кожного елемента коду.

Оскільки код, сформований трансформуючим автоенкодером, являє собою параметри позиціонування об'єкта, то в тих випадках, коли інформація про позицію доступна, з'являється можливість проводити явне навчання з учителем, порівнюючи апроксимоване значення автоенкодера з даними значеннями позиціонування. Ця інформація не використовується в ході навчання автоенкодера – як правило, точна позиція об'єкта є для зорової системи невідомою величиною (в разі, якщо вона функціонує методом проекції тривимірних об'єктів на плоску поверхню сенсора), і не може

використовуватися для зіставлення. Однак, вона виявляється корисною при оцінці точності роботи автоенкодера.

Механізм роботи трансформуючого автоенкодера може бути продемонстрований за допомогою прикладу трансляцій двовимірних зображень – рукописних цифр. Для цього випадку вхідні дані $w^i \in R^{784}$ являють собою зображення розміру 28X28.

На рисунку 2.3 зображені схема трансформуючого автоенкодера і відповідних шарів.

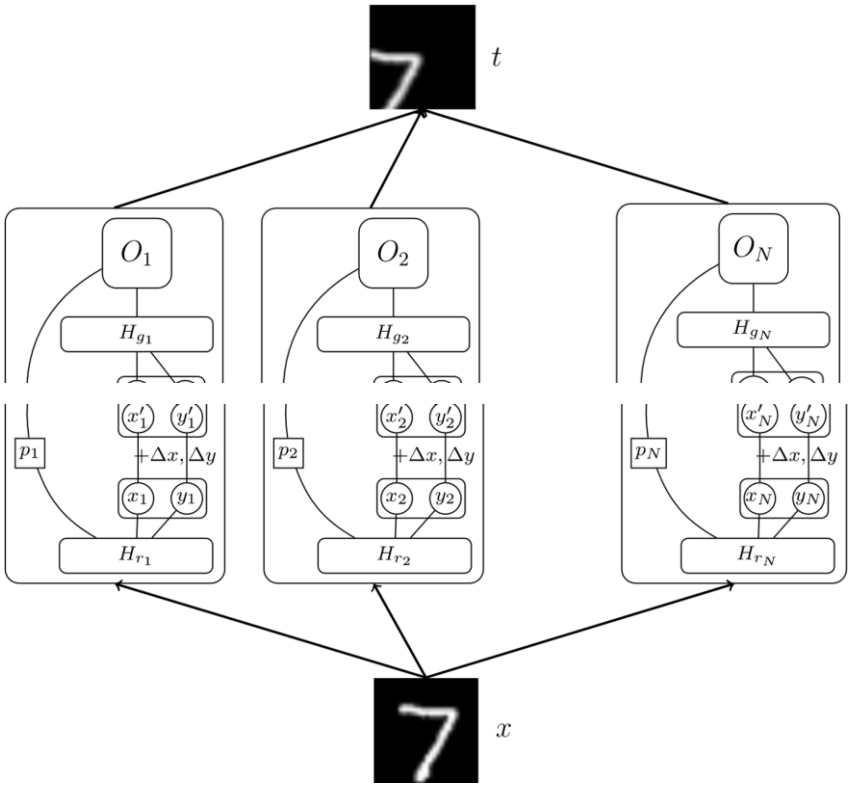


Рисунок 2.3 – Структурна схема трансформуючого автоенкодера для навчання трансляціям

Навчання автоенкодера відбувається шляхом мінімізації помилки (наприклад, середньоквадратичного відхилення) між O і T — таким чином, метою навчання є апроксимація які трансформують автоенкодером такої функції F, де O = F(s,x), що F являє собою (невідому) функцію трансформації T, за допомогою якої отримано трансформоване зображення t = T(s,x). Супутнім обмеженням, що накладається на функцію F, є використання

обмеженої кількості капсул і витяг компактного набору параметрів інстанціювання $c \in R^2$ для кожної капсули.

Слід звернути увагу на те, що C' = C + s. цю особливість архітектури можна інтерпретувати наступним чином: код (параметри інстанціювання), що витягується кожною капсулою, відчуває на собі вплив шуму, який виражається вектором трансляції s. c і p являють собою значення, отримані тільки за допомогою вхідних даних, при цьому вихідні дані передбачаються капсулою за допомогою значення c' (зашумленного) і P. Таким чином, c' в процесі навчання повинна являти собою деяку характеристику вхідних даних, і при цьому містити достатньо інформації для реконструкції трансформованого зображення. Характеристикою, яка обмежена цими вимогами, природним чином є координати об'єкта на зображенні — більш того, малоймовірно існування будь-якої іншої характеристики, що змінюється відповідно трансформації об'єкта і здатної представляти як вхідне, так і вихідне зображення. При цьому трансформуючому автоенкодеру в процесі навчання невідома точна величина c – замість цього він користується тільки інформацією про зміну цієї величини s.

Трансформуючий автоенкодер навчається методом зворотного поширення помилки. Для вибірки {x, t, s} при навчанні мінімізується функція ціни.

В якості додаткової умови, що накладається на функцію ціни автоенкодера, пропонується використовувати критерій розрідженості. Використання розрідженого коду (тобто, такого, який для конкретного набору вхідних даних формується за рахунок активності відносно невеликої підмножини нейронів, в той час як інша частина нейронів неактивна і дорівнює нулю) особливо ефективно для трансформуючого автоенкодера, оскільки в загальному випадку ознаки, яким навчаються його нейрони, є незалежними. Умова розрідженості сприяє тому, що кожен нейрон автоенкодера інкапсулює окрему ознаку вхідного зображення, не змішуючи їх з іншими нейронами відповідного шару.

Критерій розрідженості являє собою додатковий член, включений у формулювання функції ціни автоенкодера J (x), і являє собою обмеження, що накладається на нейрони породжує і розпізнає шару. В якості математичного вираження умови розрідженості для моделі використана дивергенція Кульбака-Лейблера або інформаційна дивергенція:

$$S = \sum_{j=1}^{Ln} KL(s\backslash\backslash sj) = \sum_{j=1}^{Ln} S\log\frac{s}{sj} + (1-s)\log\frac{1-s}{1-sj} \qquad (2.7)$$

де L_N число нейронів відповідного шару, Sj-середнє значення активації нейрона j, s — параметр розрідженості.

Використання цього параметра дозволяє обмежити середнє значення активації нейронів, встановивши його пропорційно не більшим s. Таким чином можна варіювати вплив критерію розрідженості на роботу трансформуючого автоенкодера, отримуючи, в залежності від величини s, більш незалежні ознаки (активації нейронів).

Розглянемо функцію ціни автоенкодера з точки зору оптимізації параметрів, тобто як функцію від значень ваг мережі W, b, то з урахуванням додаткового члена розрідженості вона являє собою наступний вираз:

$$Js(W, b-) = J(W, b-) + \beta\sum_{j=1}^{L}\sum_{j=1}^{Ln} KL(s\backslash\backslash sj) \qquad (2.8)$$

де W і b — відповідно, загальні матриці ваг і зсувів мережі, β – метапараметр, контролюючий вплив критерію розрідженості. При цьому значення Sj неявним чином залежить від W, Ь, як середня активація нейрона j.

Оскільки значення функції ціни для розрідженого коду включає в себе змінні оптимізації, то алгоритм зворотного поширення помилки також вимагає внесення додаткового члена, що представляє собою похідну критерію розрідженості.

Сформульована Архітектура трансформуючого автоенкодера являє собою відокремлену модель, здатну функціонувати у відриві від основної моделі для інкапсуляції і розпізнавання окремих (невеликих) зображень, таких як рукописні цифри. Для використання трансформуючого автоенкодера в складі моделі необхідно розширити формулювання трансформації T за межі двовимірних трансляцій.

ГЛАВА 3. КОМПЛЕКС АЛГОРИТМІВ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ

3.1 Алгоритм виділення локальних ознак

Алгоритм виділення локальних ознак виконує завдання вилучення вибірки з потоку даних для навчання детекторів моделі. Ключовим вимогам до такої вибірки є інформаційна ємність витягуваних фрагментів зображення – отримана вибірка повинна максимізувати точність розпізнавання при мінімізації обчислювальних ресурсів, витрачених на формування і навчання локальних детекторів.

Для оцінки фрагментів зображення з використанням даного критерію інформаційної ємності введемо поняття помітності – дискримінативної характеристики сприйняття, що відображає ступінь відмінності фрагмента зображення від сусідніх ділянок (фону) та інших об'єктів в полі зору. Розглянутий алгоритм виробляє оцінку показника помітності за допомогою ансамблю двох математичних методів, що взаємно доповнюють один одного:

1. Метод обчислення локальної інформаційної ентропії по Шеннону.

2. Метод знаходження екстремумів масштабованого уявлення.

Обидва методи застосовуються до зображення паралельно, що дозволяє розподілити обчислювальні ресурси з використанням паралельних або асинхронних обчислень. Результат алгоритму являє собою перетин фрагментів, знайдених кожним з методів. Далі розглянемо по черзі обидва компоненти ансамбля і функцію їх об'єднання.

Для пошуку локальних ознак, що представляють собою обмежені замкнуті фрагменти зображення, пропонується виконати кластеризацію простору помітності, що може бути зроблено за допомогою наступних кроків:

1. Застосування функції глобального порогу до зображення (вибір пікселів зображення, інтенсивністю не нижче It).

2. Вибір глобальної точки максимуму в просторі помітності Y.

3. Для знайденої точки максимуму проводиться пошук до найближчих сусідів в просторі помітності, де к – задана константа.

4. Для кожної з К точок визначається коефіцієнт варіації Vk і відстань до центральної точки Dk.

5. Кожна точка, для якої виконуються умови Dk > s і Vk < Vt, додається в масив виявлених фрагментів. Тут s – середнє значення коефіцієнта масштабування фрагмента, Vk – задане порогове значення коефіцієнта варіації.

6. З К точок вибирається наступна за значенням помітності, і процес повторюється заново з кроку 2.

Отримана безліч фрагментів високої помітності буде являти собою «слабкий» результат алгоритму, що включає в себе безліч фрагментів, що характеризуються високою помітністю в порівнянні з навколишнім фоном, але не є структурно унікальними в межах зображення. Для знаходження структурно унікальних ділянок зображення використовується друга частина алгоритму, що полягає в знаходженні екстремумів масштабованого уявлення.

Пошук екстремумів масштабованого уявлення проводиться за допомогою оцінки двох показників — різниці гауссіан зображення і визначника Гессе.

Для знаходження ділянок структурної помітності (що представляють собою локальну концентрацію високодеталізованих елементів зображення) використовуємо обчислення різниці гауссіан. Уявімо зображення I у вигляді функції від двох змінних f (x,y). Розглянемо масштабоване представлення зображення L(x, y, t) — результат згортки f (x, y) функцією Гаусса:

$$g(x,y,t) = \frac{1}{2\pi t2} e^{\frac{x2+y2}{2t^2}} \qquad (3.1)$$

т.е. *L(x,y,t) = g(x,y,t) * f(x,y).*

Застосування оператора Лапласа до даного масштабованого поданням дозволяє отримати детектор, що приймає позитивні значення в областях яскравих ділянок неоднорідності радіусу 2t. Для визначення ділянок різного радіусу використовується мульти-масштабний підхід і різні значення t. Шуканий показник, таким чином, являє собою оператор Лапласа від L(x,y,t), нормалізований по масштабу:

$$\nabla^2_{norm} \;\; L(x, y, t) = t(Lxx + Lyy) \qquad (3.2)$$

Оскільки отримане масштабоване представлення задовольняє рівнянню дифузії, стає можливою апроксимація нормалізованого лаплассіана за допомогою різниці двох уявлень, згладжених гаусовими функціями з параметрами t — Δt і t + Δt.:

$$\nabla^2_{norm} \;\; L(x, y, t) \approx \frac{\mathrm{t}}{\Delta \mathrm{t}} (\mathrm{L}(\mathrm{x}, \mathrm{y}, \mathrm{t} + \Delta \mathrm{t}) — \mathrm{L}(\mathrm{x}, \mathrm{y}, \mathrm{t} — \Delta \mathrm{t})) \qquad (3.3)$$

Отримана апроксимація називається являє собою різницю гауссіан функції *f(x,y).* Використовуємо цю різницю для пошуку ділянок високої деталізації на зображенні. Для цього застосуємо до зображення фільтр Канні [16], що використовується для виділення кордонів, конвертуємо отримане зображення в бінарне, де *f(x,y')* = 0 в разі негативного значення детектора Канні, і 1 в іншому випадку. Тоді ділянки зображення, що відповідають високій концентрації кордонів Канні, будуть оброблятися методом різниці гауссіан як ділянки неоднорідності. Таким чином різниця гауссіан можна використовувати для знаходження морфологічних деталей зображеного об'єкта, що не відрізняються забарвленням (інтенсивністю) і не дають помітних ефектів освітлення, але характеризуються відокремленою концентрацією деталізованих елементів текстури.

Для виявлення точок неоднорідної інтенсивності використовуємо обчислення оператора Гессе. Для точки зображення з координатами (x, y)і

обраного масштабу t визначник Гессе задається наступним чином:

$$\det HL\ (x,y,t) = t^2(L_{xx}Lyy - L^2xy) \qquad (3.4)$$

де HL являє собою матрицю Гессе від масштабованого представлення L. Точки максимуму для такого визначника являють собою координати і масштабний розмір ділянок неоднорідності, дозволяючи визначнику Гессе виступати в якості диференціального детектора:

$$(x,y,t) = \underset{x,y,t}{\operatorname{argmax}}(\det HL\ (x, y, t)) \qquad (3.5)$$

Координати отриманих ділянок неоднорідності (x, y) і їх радіус t, отримані за допомогою визначника Гессе, коваріантні по відношенню до трансляції, обертання і масштабування зображення. Для ділянок зображення, отриманих таким чином, крім того, характерно дещо більша стійкість до афінних перетворень у порівнянні з різницею гауссіан. Крім того, метод визначника Гессе дозволяє виявити як яскраві, так і темні ділянки неоднорідності в зображенні, і має значно нижчу обчислювальну складність, що робить його підходящим кандидатом для запропонованого комбінованого методу.

Об'єднання методів ансамблю для алгоритму виділення локальних ознак проводиться шляхом знаходження перетинаючих фрагментів, виявлених обома методами. Використання ансамблю методів дозволяє відшукувати фрагменти зображення, що характеризуються властивостями помітності (максимальною локальною ентропією) та унікальності як за структурним змістом, так і за порівняльною інтенсивністю. На рисунку 3.1 наведено приклади фрагментів, виявлених за допомогою ансамблю методів алгоритму виділення ознак. Комбінація цих ознак являє собою ділянки, відповідні елементам людського обличчя, і таким чином, дозволяє домогтися ефективного представлення зображення для завдання розпізнавання.

а) б)

Рисунок 3.1 - Фрагменти зображення, виявлені алгоритмом виділення локальних ознак. а) фрагменти, що характеризуються максимальною помітністю (локальної ентропією). б) фрагменти, що характеризуються відмінними показниками інтенсивності і структурної складності

3.2 Алгоритм оптичного трекінгу

Після того як локальні ознаки для окремого кадру відеофрагменту виявлені алгоритмом виділення ознак, наступна стадія навчання моделі являє собою оптичний трекінг або відстеження переміщень відповідних фрагментів в потоці наступних кадрів. Для цього представлений алгоритм оптичного трекінгу, що виконує оцінку значення трансформації для кадрів відеопотоку, і здатний виявляти зміщення фрагментів в потоці.

Розглянемо кадри I_t, $I_{t + \Delta t}$, I точку першого кадру I (x, y, t), де t — момент часу. При наявності відносного руху спостерігача, проекція оригіналу вихідної точки в другому кадрі буде зміщена на деякі величини Δx, Δy.

$$I(x,y,t)\_I(x + \Delta x, y + \Delta y, t + \Delta t) \qquad (3.6)$$

Процес трекінгу для потоку даних цілком складається з наступних етапів:

1. Ініціалізувати буфер активних (відстежуваних) ділянок помітності. Буфер являє собою безліч послідовностей Cl, C2,..Cb, де кожна послідовність CL містить ділянки помітності, виявлені у відповідних кадрах.

2. Якщо буфер не порожній, то для кожної CL витягти останній елемент послідовності bj і знайти для нього bj+1 по вищеописаному алгоритму і додати в послідовність.

3. Для поточного кадру Ij провести пошук нових ділянок помітності, які не є частиною існуючих послідовностей (ділянка не належить послідовності, якщо вона має частину площі, не більшу p, загальну з однією з відстежених ділянок bj+1). Виявлені b1, b2, .. bj формують нові послідовності в загальному буфері.

Розглянутий алгоритм трекінгу в чистому вигляді має ряд обмежень, що допускають помилки у визначенні зсувів точок. Основним з них є так звана проблема діафрагми. Під діафрагмою або вікном сприйняття розуміється область зображення, в рамках якої алгоритм розрахунку оптичного потоку здійснює пошук ймовірних кандидатів зміщення точки. Завдяки цьому обмеженню, оцінка руху точки стає неоднозначною – в залежності від розмірів вікна, може існувати безліч варіантів рухомих контурів, що демонструють однакову поведінку в межах вікна. У граничному випадку, коли шукана точка 1 (x, y, t) належить до однорідної поверхні і не несе статистично помітних ознак відмінності від сусідів, обмежених розміром діафрагми, питання визначення положення такої точки в кадрі It + Δt являє собою повну визначеність і не може бути вирішене аналітично без збільшення меж вікна до прийнятних розмірів (що тягне за собою багаторазове зростання обчислювального навантаження).

Крім цього, ключовою функцією трекінгу є підтримка ідентичності відстежуваних фрагментів – реєструючи той факт, що дві ділянки помітності в сусідніх кадрах пов'язані оптичним потоком, модель поміщає їх в одну

вибірку для навчання відповідного еквіваріантного детектора, обходячись, таким чином, без даних, наданих вчителем. Ідентичність фрагментів, що відслідковуються, однак, може порушуватися в тих випадках, коли оптично простежити переміщення шуканої ділянки зображення представляє неможливим — наприклад, в ситуаціях, коли об'єкт повернути до камери протилежною стороною. В такому випадку при появі шуканої ділянки перед камерою в наступному разі, його ідентичність буде втрачена, а сама ділянка і його наступні трансформації будуть оброблятися моделлю як нова послідовність, на основі якої буде формуватися вибірка і навчатися додатковий надлишковий детектор.

Для обробки таких випадків пропонується кілька методів, кожні з яких може застосовуватися як окремо, так і разом з іншими:

- При появі нових відстежуваних локальних фрагментів попередньо перевіряти їх за участю вже існуючих детекторів. При наявності позитивної відповіді від одного з них, ділянка вважається приналежним відповідному детектору.

- Метод працює тільки для тих випадків, коли Детектори моделі знаходяться в стійкому, навченому стані.

- Хешувати в пам'яті кадри, в яких трекінгова послідовність обривається, забезпечуючи можливість при поверненні в ту ж точку спостереження відновити приналежність фрагментів до послідовностей. Даний метод не підходить для ситуацій, коли повернення у вихідну точку не відбувається (наприклад, камера здійснює повний обхід навколо об'єкта), але в експериментах показав ефективність при обробці перешкод, пов'язаних з випадковими мікро-рухами камери.

- Використовуючи дані трекінгу, здійснювати побудову тривимірної карти ознак, відстежуючи, таким чином, приналежність кожної ділянки помітності до відповідної йому (єдиної) локації на карті. Найнадійніший метод з точки зору можливих зміни кута огляду; однак, для відновлення тривимірної структури об'єкта може знадобитися додаткова інформація.

Розглянемо докладніше розширення алгоритму оптичного трекінгу за допомогою побудови тривимірної карти ознак. Для кожної тривимірної точки об'єкта (Xi, Yi, Zi) проекція такої точки на зображення визначається як Xi = Xi f/Z, Yi = Yi f/Z, де f — фокусна відстань камери (рис. 3.2).

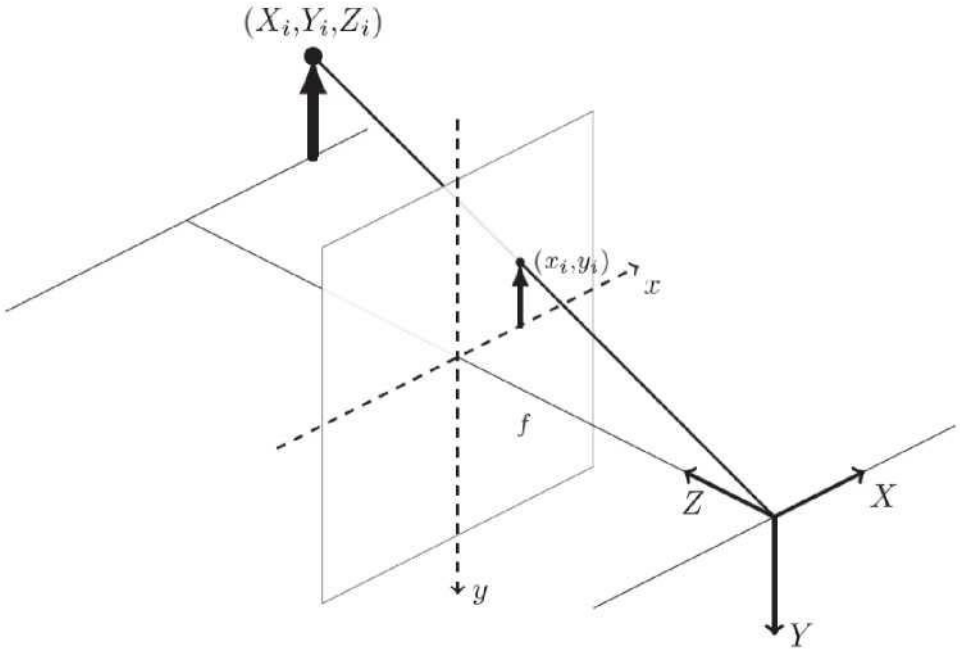


Рисунок 3.2 – Схема проекції тривимірних точок на площину сенсора. Математична операція проекції використовується при побудові тривимірної карти ознак

Операція проекції являє собою незворотне відображення (існує безліч точок (Xj,Yj,Zj), відповідних точці (Xi/Yi) на площині. З цього випливають дві властивості:

- Об'єкти з різною тривимірною формою можуть мати однакові відображення на площині;

- Двовимірна форма одного і того ж об'єкта може відрізнятися в залежності від афінної трансформації об'єкта (або камери).

Оскільки операція проекції тривимірних точок є незворотною, не існує однозначного способу відновити тривимірну форму об'єкта за обчисленими зміщеннями точок Vx, Vy. Запропонований алгоритм оптичного трекінгу використовує комбінацію різних методів оцінки положення точки в просторі,

в залежності від граничних умов завдання і доступу до інформації про положення камери.

3.3 Алгоритм розпізнавання зображень

Алгоритм розпізнавання зображень є ключовим елементом методу, що здійснює безпосередню процедуру класифікації зображень за категоріями з використанням попередньо навченої моделі. Алгоритм розпізнавання складається з послідовної активації детекторів моделі, починаючи з першого рівня, на заданому зображенні.

Припустимо дана навчена модель M_i для деякого класу зображень c_i (наприклад, таким класом можуть виступати людські обличчя), або кілька моделей для завдання мультикласового розпізнавання, і зображення I, яке необхідно розпізнати. Алгоритм розпізнавання складається з наступних кроків:

1. Задамо загальну ієрархічну пам'ять HM, представлену безліччю рівнів, розміром рівному максимальному числу рівнів навченої моделі.

2. Для l-того рівня моделі L_l, починаючи з першого, і для кожного детектора відповідного рівня застосуємо ідентифікуючу функцію детектора до всіх локальних ділянок зображення I(x.. .x + w,y. . .y + h).

3. Якщо для деякого рівня L_l активації всіх детекторів цього рівня негативні (дорівнюють нулю), то зображення не належить до класу c_i;

4. Якщо завдання розпізнавання є мультикласовим, то повторюємо алгоритм, починаючи з кроку 1, для моделі наступної категорії m_{i+1}. В іншому випадку алгоритм вважається зупиненим.

5. Якщо Детектори шару L_l активовані, то продовжимо послідовну активацію детекторів наступних шарів, повторюючи алгоритм з кроку 2.

6. Якщо активація останнього рівня позитивна, то зображення належить до класу c_i;

7. Для мультикласового розпізнавання перераховані вище кроки

повторюються для безлічі навчених екземплярів моделі M_o, M_1,.. $M_п$, при цьому приналежність зображення до класу визначається по максимальному активному рівню.

Використовуючи той факт, що низькорівневі локальні ділянки зображення (взяті з достатнім масштабом), як правило, являють собою краї, межі і кути, слід зазначити має місце тенденцію до зростання різноманітності серед еквіваріантних детекторів на більш високих рівнях моделі. Експерименти демонструють, що кількість детекторів першого рівня при навчанні не перевищує ~ 102, при цьому моделі, навчені на об'єктах різних категорій, здатні розділяти між собою частину детекторів першого рівня, демонструючи ефект, що нагадує трансферне навчання або переднавчання без вчителя.

Алгоритм розпізнавання з використанням дворівневої моделі може ефективно застосовуватися для вирішення проблем розпізнавання таких об'єктів як людські обличчя, в деяких випадках (певний масштаб) — силуети, автомобілі. У загальному випадку евристичне правило вибору кількості рівнів для моделі залежить від кількості ступенів свободи цікавить об'єкта, і не може бути однозначно встановлено по зображенню в силу впливу шуму і обмеженого простору спостереження.

Серед недоліків алгоритму в ході експериментів було виявлено незначне (до 5%) зниження точності розпізнавання в ситуації, коли об'єкти на зображенні піддавалися оклюзії. Помилки такого роду узгоджуються з принципом роботи моделі, що вимагає для позитивної активації детектора вищого рівня наявності активованих детекторів нижнього рівня. У ситуації, коли частина об'єкта не відповідає структурі навченого детектора, нижній рівень моделі активується лише частково, і результуюче відхилення може позначитися на точності розпізнавання. Для вирішення проблеми оклюзії пропонується використовувати методику «відсіву» для нейронних мереж, періодично в ході навчання відключаючи випадковим чином вибрані детектори.

Алгоритм розпізнавання відповідає вимогам поставленого завдання і дозволяє здійснювати розпізнавання зображених об'єктів з обчисленням параметрів їх локалізації в тривимірному просторі. Застосування ієрархічного підходу в реалізації алгоритму дозволяє здійснювати розпізнавання за короткий час, не вимагаючи значних обчислювальних ресурсів, шляхом активацій обмеженого числа локальних ділянок зображення (для деяких категорій зображень, як, наприклад, людські обличчя, це число не перевищує «10». Використання алгоритмом розпізнавання еквіваріантних детекторів дозволяє алгоритму об'єднувати в одній категорії зображення, різні з інформатико-теоретичний точки зору (такі, як зображення людського обличчя в профіль і в фас), але відповідні при цьому когнітивним класам розпізнавання.

Використання локальних детекторів дозволяє досягти стійкості до оклюзії (перекритті частини зображення стороннім фоном). Оскільки Детектори моделі являють собою специфічні дискримінативні нейронні мережі, що навчаються розрідженим репрезентаціям даних, кожен локальний детектор забезпечує стійкість до ефектів випадкового шуму і розмиття в своїй локальній області, дозволяючи алгоритму розпізнавання успішно справлятися з класичними проблемами розпізнавання образів – нечіткості вхідних даних і внутрішньокласової варіативності.

ВИСНОВОК

Таким чином можна зробити висновок.

Алгоритмічна обробка і класифікація зображень застосовуються в системах безпеки, контролю і управління доступом, в системах відео-спостереження, системах віртуальної реальності та інформаційних пошукових системах. На даний момент у виробництві широко використовуються системи розпізнавання рукописного тексту, автомобільних номерів, відбитків пальців або людських осіб, що знаходять застосування в інтерфейсах програмних продуктів, системах безпеки та ідентифікації особистості, а також в інших прикладних цілях.

Порівняльний аналіз методів розпізнавання зображень показав, що незважаючи на різноманітність підходів до навчання і виділення класифікуючих ознак, велика частина методів не розглядає проблему розпізнавання зображень об'єктів під дією інваріантних тривимірних перетворень. Існуючий клас методів, орієнтованих на розпізнавання тривимірних об'єктів, таких як глибокі згорткові мережі і констеляційні моделі, здатний вирішувати завдання інваріантного розпізнавання, проте демонструє обмежену здатність до локалізації об'єктів на зображенні і визначення параметрів їх просторового розташування, що є важливим для цілого ряду прикладних завдань обробки інформації.

Один з основних підходів, який найбільш широко використовувався в області розпізнавання зображень, являє собою вживання класичних моделей-класифікаторів, що вивчаються з учителем. Для навчання таких моделей використовуються маркована вибірка даних, що складається з масиву зображень і відповідного їм масиву міток, що визначають категорію, до якої відноситься зображення. У процесі навчання масив даних розділяється на дві нерівні частини — навчальну вибірку і тестову вибірку, потім за допомогою специфічного для конкретного алгоритму правила навчання параметри моделі налаштовуються з використанням навчальної вибірки таким чином,

щоб отримавши в якості вхідних даних зображення, модель на виході виробляла б мітку відповідного класу. Цей підхід представлений безліччю моделей, серед яких найбільш широко використовуваними є регресійна модель, штучна нейронна мережа (багатошаровий перцептрон), метод опорних векторів, а також дерева прийняття рішень і моделі-ансамблі, що представляють собою поєднання деяких перерахованих моделей.

Незважаючи на різноманітність підходів до навчання і виділення класифікуючих ознак, велика частина методів не розглядає проблему розпізнавання зображень об'єктів під дією інваріантних тривимірних перетворень. Існуючий клас методів, орієнтованих на розпізнавання тривимірних об'єктів, таких як глибокі згорткові мережі і констеляційні моделі, здатний вирішувати завдання інваріантного розпізнавання, проте демонструє обмежену здатність до локалізації об'єктів на зображенні і визначення параметрів їх просторового розташування, що є важливим для цілого ряду прикладних завдань обробки інформації.

Розглянуто метод навчання моделі на основі потоку даних, що дозволяє здійснювати навчання без наявності маркованої вибірки. Даний метод підвищує практичну цінність моделі, роблячи можливим її використання в автоматичних системах управління і обробки інформації, в умовах відсутності контролю з боку людини. Використання потоку даних являє собою доступне рішення для різних видів інформаційних систем, що мають доступ до змінюються в часі даними, що дозволяє використовувати представлену модель в додатках, пов'язаних з обробкою відеофрагментів, моніторингом і відео-спостереженням.

Розглянуто архітектуру елементної бази моделі, що представляє собою локальні еквіваріантні Детектори на основі трансформуючих автоенкодерів. Представлені елементи моделі здатні вирішувати завдання локальних репрезентацій, забезпечувати стійкість моделі до часткових перешкод і оклюзії на зображенні, ефективно реагувати на тривимірні трансформації зображених об'єктів і проводити оцінку параметрів їх локалізації.

Представлений комплекс алгоритмів, що виконують завдання навчання моделі і розпізнавання зображень за допомогою розробленої моделі. Запропоновані алгоритми дозволяють повною мірою реалізувати поставлене завдання розпізнавання зображень з виконанням тривимірної просторової локалізації об'єктів. Розглянуто алгоритм виділення локальних ознак, заснований на використання інформатико-теоретичних характеристик зображення і виробляє вибір найбільш інформаційно-ємних фрагментів зображення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bellman, R. E. Perturbation techniques in mathematics, engineering and physics / R. E. Bellman. — Courier Corporation, 2003. — 214 pp.

2. Bengio, Y. Learning deep architectures for AI / Y. Bengio // Foundations and trends in Machine Learning. — 2009. — no. 1. — Pp. 1-127.

3. Bengio, Y. Representation learning: A review and new perspectives / Y. Bengio, A. Courville, P. Vincent // Pattern Analysis and Machine Intelligence. — 2013.— no. 35(8). — Pp. 1798-1828.

4. Bishop, C. M. Pattern recognition and machine learning / C. M. Bishop. — New York: Springer, 2006. — 12 pp.

5. Cao, L. Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes / L. Cao, L. Fei-Fei // Computer Vision (ICCV), 2007 IEEE International Conference. — 2007. — Pp. 1-8.

6. Deng, L. The MNIST database of handwritten digit images for machine learning research / L. Deng // IEEE Signal Processing Magazine. — 2012. — no. 29.6. — Pp. 141-142.

7. Duin, R. P. W. Open issues in pattern recognition / R. P. W. Duin, E. Pekalska // Computer Recognition Systems. — 2005. — Pp. 27-42.

8. Fergus, R. A sparse object category model for efficient learning and exhaustive recognition / R. Fergus, P. Perona, A. Zisserman // Computer Vision and Pattern Recognition. — 2005. — no. 1. — Pp. 380-387.

9. Grauman, K. Visual object recognition / K. Grauman, B. Leibe. — Morgan & Claypool Publishers, 2010. — 165-186 pp.

10. Hearst, M.A. Support vector machines / M.A. Hearst, S.T. Dumais, E. Osman et al. // Intelligent Systems and their Applications. — 1998. — no. 4. — Pp. 18-28.

11. Hinton, G.E. Reducing the dimensionality of data with neural networks / G.E. Hinton, R.R. Salakhutdinov // Science. — 2006. — no. 313(5786). — Pp. 504-507.

12. Hinton, G.E. Transforming auto-encoders / G.E. Hinton, A. Krizhevsky, S. D. Wang // Artificial Neural Networks and Machine Learning-ICANN 2011. — 2014. — Pp. 44-51.

13. Hubel, D. H. Brain and visual perception / D. H. Hubel, T. N. Wiesel. — ISBN13, 2005. — 36-46 pp.

14. Hubel, D. H. Eye, brain, and vision / D. H. Hubel. — New York: Scientific American Library, 1988. — 85-87 pp.

15. Hubel, D. H. Receptive fields and functional architecture of monkey striate cortex / D. H. Hubel, T. N. Wiesel // The Journal of physiology. — 1968. — no. 195(1). — Pp. 215-243.

16. Ji, Q. 3D face pose estimation and tracking from a monocular camera / Q. Ji // Image and vision computing. — 2002. — no. 20(7). — Pp. 499-511.

17. Jolliffe, I. Principal component analysis. / I. Jolliffe. // John Wiley and Sons, 2002. — 13-16 pp.

18. Kadir, T. An affine invariant salient region detector / T. Kadir, A. Zisserman, M. Brady // Computer Vision-ECCV. — 2004. — Pp. 228-241.

19. Keysers, D. Comparison and combination of state-of-the-art techniques for handwritten character recognition: topping the mnist benchmark / D. Keysers // arXiv. — 2007. — no. 0710.2231. — Pp. 21-27.

20. Kohonen, T. Self-organization and associative memory / T. Kohonen // Springer-Verlag Berlin Heidelberg New York. — 1988. — no. 8(1). — Pp. 13-27.

21. Kreutz-Delgado, K. Dictionary learning algorithms for sparse representation / K. Kreutz-Delgado // Neural computation. — 2003. — no. 15.2. — Pp. 349-396.

22. Krizhevsky, A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // Advances in neural information processing systems. — 2012. — Pp. 1097-1105.

23. LeCun, Y. Backpropagation applied to handwritten zip code recognition / Y. LeCun // Neural computation. — 1989. — no. 4. — Pp. 541-551.

24. LeCun, Y. Comparison of learning algorithms for handwritten digit recognition / Y. LeCun // International conference on artificial neural networks. — 1995. — no. 60. — Pp. 111-115.

25. Lee, H. Efficient sparse coding algorithms / H. Lee // Advances in neural information processing systems. — 2006. — Pp. 801-808.

26. Lindeberg, T. Scale-space theory: A basic tool for analyzing structures at different scales / T. Lindeberg // Journal of applied statistics. — 1994. — no. 21.1-2. — Pp. 225-270.

27. Lowe, D.G. Object recognition from local scale-invariant features / D.G. Lowe // Computer Vision (ICCV). The proceedings of the seventh IEEE international conference. — 1999. — no. 2. — Pp. 1150-1157.

28. Lucas, B.D. An iterative image registration technique with an application to stereo vision / B.D. Lucas, T. Kanade // IJCAI. — 1981. — no. 81. — Pp. 25-34.

29. Matsugu, M. Subject independent facial expression recognition with robust face detection using a convolutional neural network / M. Matsugu // Neural Networks. — 2003. — no. 16(5). — Pp. 555-559.

30. Murphy-Chutorian, E. Head pose estimation in computer vision: A survey. / E. Murphy Chutorian, M.M. Trivedi // Pattern Analysis and Machine Intelligence, IEEE Transactions on 31.4. — 2009. — Pp. 607-626.

31. Ng, A. The Importance of Encoding Versus Training with Sparse Coding and Vector Quantization / A. Ng, A. Coates // Workshop on Learning Architectures, Representations, and Optimization for Speech and Visual Information Processing / International Conference on Machine Learning. — 2011. — 06.

32. Ng, A.Y. An analysis of single-layer networks in unsupervised feature learning / A.Y. Ng, H. Lee, A. Coates // International Conference on Artificial Intelligence and Statistics. — 2011. — Pp. 215-223.

33. Olshausen, B. A. Emergence of simple-cell receptive field properties by learning a sparse code for natural images / B. A. Olshausen // Nature. — 1996.

— no. 6583. — Pp. 607-609.

34. Pedregosa, F. Scikit-learn: Machine learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort et al. // The Journal of Machine Learning Research. — 2011. — no. 12. — Pp. 2825-2830.

35. Rublee, E. ORB: an efficient alternative to SIFT or SURF / E. Rublee // Computer Vision (ICCV), 2011 IEEE International Conference. — 2011. — Pp. 2564¬2571.

36. Sebe, N. Machine learning in computer vision / N. Sebe. — New York: Springer Science & Business Media, 2005. — 29 pp.

37. Serre, T. Object recognition with features inspired by visual cortex / T. Serre, L. Wolf, T. Poggio // Computer Vision and Pattern Recognition. — 2005. — no. 2. — Pp. 994-1000.

38. Simoncelli, E. P. Natural image statistics and neural representation / E. P. Simoncelli, B. A. Olshausen // Annual review of neuroscience. — 2001. — no. 24(1). — Pp. 1193-1216.

39. Szegedy, C. Going deeper with convolutions / C. Szegedy // arXiv. — 2014. — no. 1409.4842.

40. Thaler, L. Neural correlates of natural human echolocation in early and late blind echolocation experts / L. Thaler, S. R. Arnott, M. A. Goodale // PLoS One. — 2011. — no. 6(5). — P. e20162.

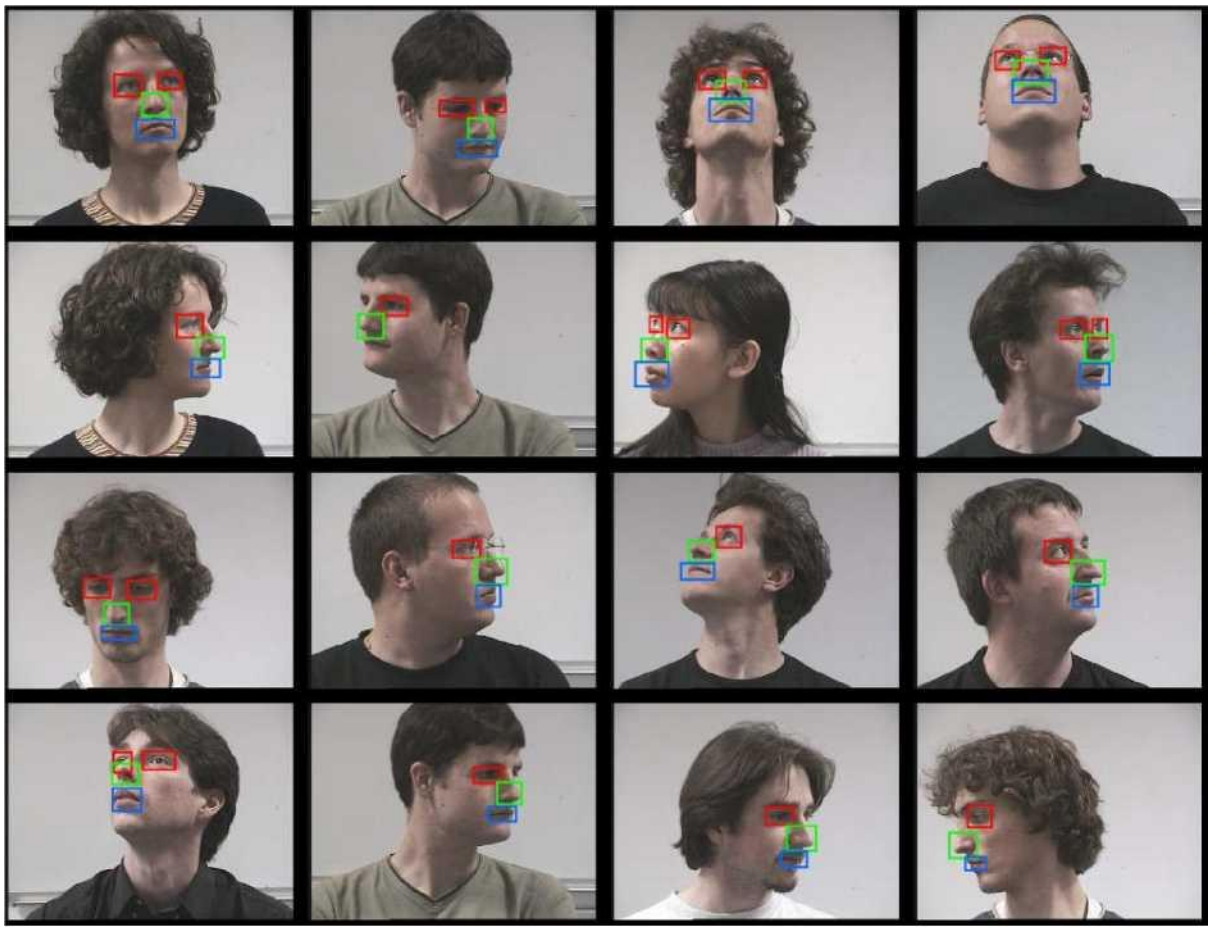
41. Turk, M. A. Face recognition using eigenfaces / M. A. Turk, A. P. Pentland // Computer Vision and Pattern Recognition. — 1991. — no. Proceedings CVPR'91., IEEE Computer Society Conference. — Pp. 586-591.

42. Yahia, S. Human detection based on integral Histograms of Oriented Gradients and SVM / S. Yahia, M. Atri, R. Tourki // Communications, Computing and Control Applications. — 2011. — Pp. 1 - 5.

43. Zuo, F. Fast facial feature extraction using a deformable shape model with haar- wavelet based local texture attributes / F. Zuo, P. H. N. de With // Image Processing. — no. 3. — Pp. 1425-1428.

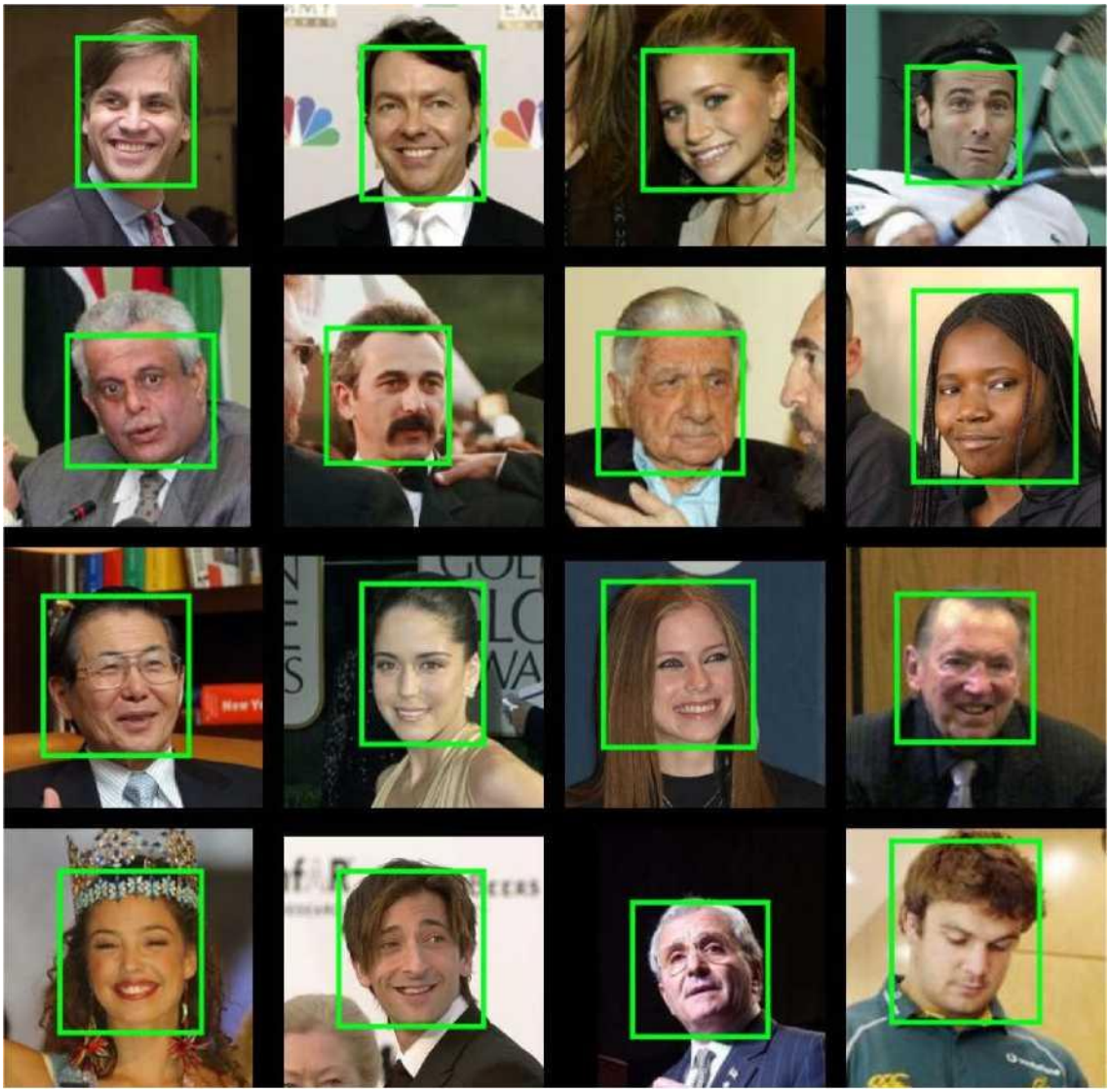
ДОДАТКИ

Додаток А



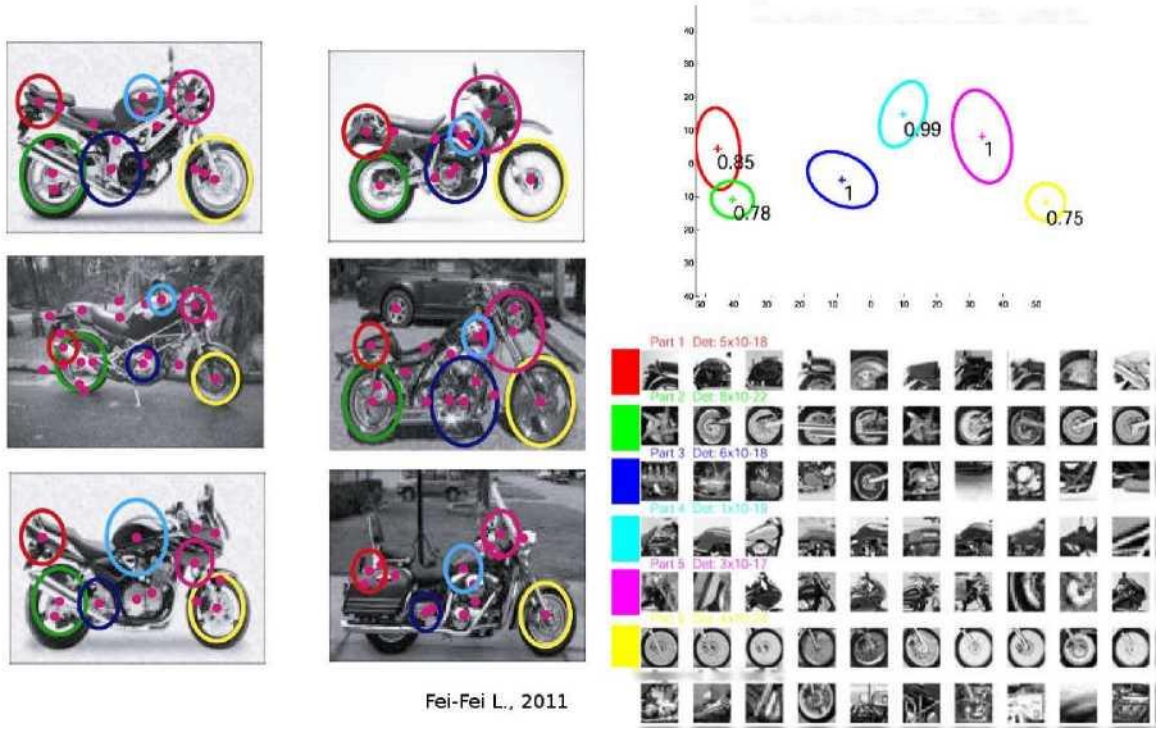
Результати активацій локальних детекторів моделі, відповідних структурним елементам людських облич

Додаток Б



Приклади успішно розпізнаних облич

Додаток В



Справа внизу: приклади зображень для навчання складових частин моделі. Справа вгорі: візуалізація параметрів в. зліва: приклади успішно розпізнаних зображень [5]