

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»**

**Приладобудівний факультет**

**Кафедра автоматизації та систем неруйнівного контролю**

До захисту допущено:

Завідувач кафедри

\_\_\_\_\_ **Юрій КИРИЧУК**

«\_\_» \_\_\_\_\_ 20\_\_ р.

**Дипломна робота**

**на здобуття ступеня бакалавра**

**за освітньо-професійною програмою «Роботизовані і автоматизовані системи  
неруйнівного контролю та діагностики»**

**спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології»**

**на тему: «Підвищення ефективності роботи нейромереж методами аугментації»**

Виконав:

студент IV курсу, групи ПК-91

Овсяніков Олександр Олександрович \_\_\_\_\_

Керівник:

Проф., д.п.н., проф,

Протасов Анатолій Георгійович \_\_\_\_\_

Рецензент:

доцент, к.т.н., доцент

Добролюбова Марина Валеріївна \_\_\_\_\_

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

Студент (-ка) \_\_\_\_\_

**Київ – 2023 року**

## ВІДОМІСТЬ ДИПЛОМНОЇ РОБОТИ

№ з/п	Формат	Позначення	Найменування	Кількість листів	Примітка
1	A4		Завдання на дипломну роботу	2	
2	A4	ПК-91.09.0000.000 ПЗ	Пояснювальна записка	64	

				<b>ПК-91.09.0000.000</b>	
	ПБ	Підп.	Дата		
Розробн.	Овсяніков О.О.			Лист	Листів
Керівн.	Протасов.А.Г.			1	1
Н/контр.				КПІ ім. Ігоря Сікорського Каф. АСНК Гр. ПК-91	
Зав.каф.	Юрій КИРИЧУК				

**Пояснювальна записка  
до дипломного проекту  
на тему:  
«Підвищення ефективності роботи нейромереж методами  
аугментації»**

Київ – 2023 року

**Національний технічний університет України**  
**«Київський політехнічний інститут імені Ігоря Сікорського»**  
**Приладобудівний факультет**

**Кафедра автоматизації та систем неруйнівного контролю**

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – **151 «Автоматизація та комп'ютерно-інтегровані технології»**

Освітньо-професійна програма **«Роботизовані і автоматизовані системи неруйнівного контролю та діагностики»**

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ Юрій КИРИЧУК

«\_\_» \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ**

**на дипломну роботу студенту**

**Овсяніков Олександр Олександрович**

1. Тема роботи «Підвищення ефективності роботи нейромереж методами аугментації», керівник роботи Протасов Анатолій Георгійович, затверджені наказом по університету від «30» травня 2023р. №2057-с
2. Термін подання студентом роботи: 15.06.2023.
3. Вихідні дані до роботи. Кількість початкових зображень: 3700; Кількість отриманих зображень: 5600; Типи дефектів: алігатор, вибоїна, лінійні тріщини; Використана мережа: MobileNet; Очікувана точність: понад 90%;
4. Зміст роботи:

Вступ. Аналітичний огляд. Використання штучних нейронних мереж для автоматизації процесів. Розробка системи контролю дорожнього покриття з використанням штучного інтелекту. Експериментальна перевірка ефективності отриманих результатів. Висновки.

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо) 4 плакати: Опис об'єкту контролю та збір даних. Методи аугментації. Застосування архітектур нейронних мереж. Результати дослідження.

6. Консультанти розділів роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
-	-		

7. Дата видачі завдання: 27.01.2023.

#### Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
	Збір та огляд літературних джерел	17.04.2023	
	Проведення аналітичного огляду	24.04.2023	
	Реалізація методів аугментацій	1.05.2023	
	Імплементация нейронних мереж	8.05.2023	
	Обробка результатів дослідження	28.05.2023	

Студент

Овсяніков Олександр Олександрович

Керівник

Протасов Анатолій Георгійович

## Анотація

В даній дипломній роботі було проведено дослідження впливу аугментованих даних на ефективність згорткової нейронної мережі із додаванням у навчальну вибірку малого об'єму відсоткове відношення аугментованих даних. Було розглянуто додавання 10, 20 та 30 відсотків аугментованих даних відносно кількості даних початкової вибірки.

Застосовувалися три методи аугментації: відзеркалення; збільшення; обертання.

У якості даних використовуються три типи дефектів дорожнього одягу: сітка тріщин; вибоїна; лінійна тріщина.

Дослідження проводилися на архітектурі MobileNet із використанням підходу передавального навчання.

У результаті було розглянуто десять вибірок із аугментаціями, точність на яких практично не відмінна одна від одної, що свідчить про недійсність підходу розширення малої навчальної вибірки невеликим відсотковим відношенням аугментаційних даних.

## Abstract

In this diploma work, a study was conducted on the impact of augmented data on the efficiency of a convolutional neural network by adding a small volume of augmented data to the training set. The addition of 10, 20, and 30 percent of augmented data relative to the quantity of the initial sample data was considered.

Three augmentation methods were applied: mirroring; magnification; rotation.

The data used included three types of road surface defects: a network of cracks; potholes; linear crack.

Studies were conducted on the MobileNet architecture using the transfer learning approach.

As a result, ten samples with augmentations were considered, the accuracy of which was practically indistinguishable from each other, indicating the invalidity of the approach to extend a small training sample with a small percentage of augmentation data.

## Зміст

Скорочення та позначення.....	7
Вступ.....	8
Розділ 1. Аналітичний огляд.....	10
1.1 Загальний огляд будівництва автомобільних доріг.....	10
1.2 Загальна класифікація дефектів автомобільних доріг. Причини їх виникнення.....	13
1.3 Актуальність автоматизації контролю дорожніх покриттів.....	19
Розділ 2. Використання штучних нейронних мереж для автоматизації процесів класифікації.....	25
2.1. Огляд алгоритму штучних нейронних мереж.....	25
2.2. Базова сутність алгоритму згорткових нейронних мереж.....	28
2.3 Передавальне навчання.....	32
2.4. Контроль дорожніх покриттів з використанням алгоритмів глибинного навчання.....	33
Розділ 3. Розробка системи контролю дорожнього покриття з використанням ШІ...41	41
3.1 Аналіз та вибір методів аугментації.....	41
3.2. Застосування методів аугментації для збільшення бази даних.....	51
3.3. Вибір архітектури згорткової нейронної мережі.....	53
Розділ 4. Експериментальна перевірка ефективності отриманих результатів.....57	57
4.1. Застосування отриманих результатів для перевірки дієздатності вибраної НМ.....	57
4.2. Аналіз результатів дослідження.....	61
4.3. Пропонована подальша робота.....	62
Висновки.....	63
Список використаної літератури.....	64



## **Скорочення та позначення**

ML – Machine Learning;

DL – Deep Learning;

ANN – Artificial Neural Network;

CNN – Convolutional Neural Network;

PCI – Pavement Condition Index;

MMS – Mobile Mapping Systems;

GPS – Global Positioning System;

LiDAR – Light Detection and Ranging;

3D – Three Dimensional space;

AI – Artificial intelligence;

ІІІ — Штучний Інтелект;

CV – Computer Vision;

SVM – Support vector machine;

IRI – International Roughness Index;

ReLU – Rectified Linear Unit;

YOLO – You Only Look Once;

FCN – Fully Convolutional Network;

DA – Data Augmentation;

PCA – Principal Component Analysis;

GAN – Generative Adversarial Network;

NST – Neural Style Transfer;

CM – Confusion Matrix.

## Вступ

За всю історію людства одним із найбільш важливим питанням була логістика — пересування як самих людей так і речей різноманітніших габаритів на будь-які дистанції. Характерним явищем для доріг що знаходяться у достатньо підвищеному агресивному середовищі природи є руйнування, деформація, відхил від нормативних якостей, якостей що забезпечують працездатність транспортних ділянок. Контроль цих якостей є важливим фактором успішності практично усіх сфер діяльності людства.

Методи контролю століттями велися візуальним оглядом за участю людини спеціаліста що оцінював придатність певної ділянки дороги. Такий підхід потребує людського ресурсу який має людський фактор й може спотворити ефективність оцінки якостей. З розвиненням інформаційних технологій з'явилися більш ефективні методи та підходи з залученням цифрових пристроїв — фотоапаратів. За допомогою зображень можна отримувати інформацію про стан доріг, що фактично є оцифруванням якостей та параметрів що вже дає змогу спростити принцип оцінки контрольованих структур. Постає питання вилучення фактичної інформації з растрових зображень без участі людини спеціаліста.

Достатньої перспективності набула сфера штучного інтелекту — машинне навчання, що дає змогу виділяти з растрових зображень корисну інформацію на основі контрастності та яскравості пікселів, виділяючи ті чи інші якості зображеного. Одним із підходів машинного навчання є глибинне навчання, алгоритми згорткових нейронних мереж дають змогу аналізувати зображення, запам'ятовуючи ознаки по навченим цією мережею даними. Штучна нейронна мережа є спробою зімітувати людські нейрони що можуть запам'ятовувати подану на них інформацію та фактично аналізувати дані оперуючи попереднім досвідом, маючи змогу з великою точністю надавати вірогідні передбачення.

Використовуючи алгоритми штучного інтелекту дуже зручно автоматизувати контроль якості дорожніх покриттів, зменшивши навантаження на людський ресурс та підвищивши точність та ефективність досліджень.

Великою проблемою навчання нейронних мереж є якість та головне кількість даних. Одним із найбільш простих та базових підходів розширення бази даних є застосування методів аугментації. Аугментація даних полягає у генеруванні нових даних та додаванням їх до основної вибірки даних, збільшивши навчальну вибірку.

У цій роботі пропонується дослідити методи аугментації що можуть покращити результати у задачі класифікації дефектів дорожнього покриття штучними нейронними мережами.

## ***Розділ 1. Аналітичний огляд***

### **1.1 Загальний огляд будівництва автомобільних доріг**

Розглянемо процес будівництва асфальтованих автомобільних доріг високої якості. Процес поділяється на такі етапи [1]:

#### ***- Геологічні дослідження ґрунту та клімату:***

Важливо що саме буде знаходитися під дорожньою частиною, від цього буде залежати стабільність верхніх конструкторських шарів — дорожнього одягу. Для цього роблять проби ґрунту за допомогою свердловин, вивчають міцність порід, та що найважливіше, перевіряють на місце знаходження ґрунтових вод. Не менш важливим є кліматичні умови та специфіка природних умов, від цього залежить вологість ґрунту та подальші проєктування шарів із вибором матеріалів для них.

#### ***- Підготовка земляного полотна:***

Після визначення місця розташування дороги, цю територію очищують від зайвих м'яких земляних покривів та наносять шар з перемішаного ґрунту та іншими сумішами: таких як цемент, вапно, зола виносу, спінені бітуми та інші. Після нанесення шар трамбується катками, тисненням на покрив великої маси. Потім наносять шар крупного піску як перший шар для дренажу, пісок також трамбується. Часто використовують геотекстиль (Рис.1.1.а) під шаром піску для того щоб розподіляти сили тиснення з верхніх шарів, не даючи шару піску деформуватися та просідати, що спричиняє руйнування шару стисненого ґрунту. Шар земляного полотна розрахований для більшого опору на стиснення та зрушення, ці шари виконують амортизаторні властивості дороги, затухаючи напруження від коливань та інших впливів спричинених автомобілями.

#### ***- Ґрунт (основа):***

Часто використовують геогратки (Рис.1.1.б) — це сітка з металу або жорсткого полімеру із високими стінками, утворюють комірки, що підвищують міцність, жорсткість та найголовніше стійкість до бокового розповзання шарів.



а)



б)

Рис.1.1.1: а) геотекстиль, б) георатки

Підхід із використанням ґраток ще називають армуванням. Першим шаром основи, з ґратками або без, є щебінь. Щебінь використовується різних марок м'якості, але лише однієї марки при накладанні одного шару. Наприклад для високоякісних транспортних шляхів кладуть марки 1000 або 1200, зазвичай вони гранітні. Важливим є розміри гранул зернистого насипу. Першим шаром кладуть крупні гранули (40-70 мм), другий середніми (20-40 мм) та третій малими (5-20 мм) габаритами гранул. Це робиться для заповнення проміжків між гранулами більш масивних розмірів, тим самим збільшуючи міцність дороги в цілому. Форма гранул також важлива — використовують кубоподібну форму, тим самим така гранула може витримувати напруження більш якісно з усіх сторін, не руйнуючи інші гранули поряд. Такий підхід збільшує міцність та є стійким до виїмок. Щебінь не тільки є достатньо твердим шаром для подальших більш щільних покриттів, але й виконує дренажну функцію, даючи воді вихід на більш нижчі шари. Такий шар компенсує сили тиснення рівномірно. Також можуть використовувати декілька шарів щебеню із ґратками. Кількість шарів залежить від фінансових можливостей, поставлених задач та трудомісткості робіт.

Останнім шаром основи може виступати бетон що має у собі менше зв'язуючих речовин. Виконуючи роль більш м'якого шару ніж наступні, але надаючи ще більшу міцність шарам зверху за рахунок ефекту амортизації. Плоско-паралельність поверхні шару бетону спрощує накладання наступних

конструкторських покривів.

- Дорожній одяг (верхні шари):

Дорожній одяг — це шари асфальтобетону, причому кількість цих шарів різниться від призначення. Наприклад для нежорсткого дорожнього одягу можуть накладати й один шар асфальтобетону, що звісно зменшує міцність усієї конструкції та час її експлуатування. Для жорсткого одягу типово використовувати покриви асфальтобетону та цементобетону різної міцності. Найбільш розповсюдженим для жорсткого одягу є використання трьох покриттів. Перший покрив, що поміщається на основу, виконаний з крупнозернистої та пористої суміші асфальтобетонну. Другим йде така ж сама суміш, але вже більш щільна, менш пориста, з додаванням бітумів.

Бітум — це мінеральний комплекс, гірська смола, що зв'язує між собою пісок, мінеральні порошки, щебінь та інші компоненти у сумішах дорожнього покриття. Розрізняють як рідкий так і в'язкий, за рахунок чого ним легко оперувати наприклад при ремонтних роботах. Останнім шаром є найміцніший та найменш пористий покрив, що захищає усю конструкцію від тертя та різноманітніших фізичних впливів.

- Захисний шар:

Фінальним покритвом зазвичай є чіпсил (Chip Seal) — це покрив 1-2 сантиметрів виконаний з бітуму або інших смол із додаванням дрібного дисперсійного щебеню для кращого контакту машин з дорогою. Такий шар виконує захисну функцію від тертя машин та кліматичних впливів. Чіпсилу характерне стирання, але такий шар дуже легко відновлювати раз на один-два роки саме через смолу, вона надає в'язкість та відмінного контакту з поверхнями.



Рис.1.1.2: Поперечний переріз структури дорожнього покриття

## 1.2 Загальна класифікація дефектів автомобільних доріг. Причини їх виникнення

Саме по верхньому шару, асфальтобетону, у більшості випадків спостерігається дефектність дорожньої частини у багатьох варіаціях. Розглянемо найбільш розповсюдженні візуальні прояви деформацій верхнього та інших покриттів:

### - Тріщини дорожніх покриттів:

Це лінійний дефект покриття що порушує цілісність асфальтованого покриття. Тріщини можуть бути наскільки завгодно різними, виходячи з конфігурації, протяжності, ширини, форми і на сам перед геометричного розташування. Вони виникають через експлуатаційні причини з плином часу під впливом розрахованих навантажень, а також через технологічні недоліки при монтажних роботах покриттів.

Розрізняють такі типи тріщин (Рис.1.2.1):

- повздовжні по осі дороги (1);
- поперечні на усю ширину (2);
- косі (3);
- повздовжня по смузі накату (4);
- сітка тріщин (алігатори) по смузі накату (5).

Повздовжні по осі дороги тріщини (1) виникають на стику двох смуг під

час накладання покриттів. Так як усі покриття накладаються не одразу однією площею (через технологічні обмеження устаткування доводиться покривати смугами певної ширини), контакт кожної смуги під час монтажу може бути не достатньо тісним, що з часом спричиняє по всій довжині смуг тріщину. Різна температура цих окремих шарів під час експлуатації ще більше підвищує розвиток дефекту, особливо у теплу пору року.

Тріщини поперечні (2) на усю ширину покриття виникають у наслідок перепадів температур повітря та недостатньому температурному опору верхнього шару. У наслідок температурних перепадів бітум у асфальтобетоні втрачає свої гнучкі деформаційні властивості. Асфальтобетон стає крихким, матеріал змінюється у різних областях та об'ємах, змінюються й якості — гнучкість та щільність. Це призводить до надломів та тріщин на дотик цих різних по якостям об'ємів асфальтобетону. Даний прояв можна спостерігати при низьких температурах, коли саме гнучкість матеріалів порівняно мала, від чого деформації кручення та згинання не поглинаються у повному об'ємі — створюються тріщини.

Косі тріщини (3) утворюються у наслідок недостатньої міцності нижніх шарів або їх технологічної неідеальності під час монтажу. Також характерний прояв таких тріщин над трубами та іншими підземними структурами.

Повздожня тріщина по смузі накату (4) — результат напруг розтягу при неоднорідності складу матеріалу верхнього одягу. Маючи різну міцність асфальтобетон зі збільшенням напруження деформується. Такий дефект має зношуваний характер, що фактично залежить лише від якостей використаних матеріалів.

Сітка тріщин по смузі накату (5) проявляє себе у зимову пору в вигляді пагорбу. Пагорб формується у наслідок локального підтягування води до дорожнього одягу. Вода збільшується у об'ємі при замерзанні у сторону втрат тепла що спричиняє пагорб. Подібний ефект можна побачити у весінній період. У результаті надмірної вологості земляного покрову характер тріщин набуває прямокутних форм на покритті.



Головна причина більшості тріщин — зношення дорожнього одягу під впливами експлуатаційних сил, кліматичних умов та надання недостатньої міцності покривів при монтажних роботах [2].

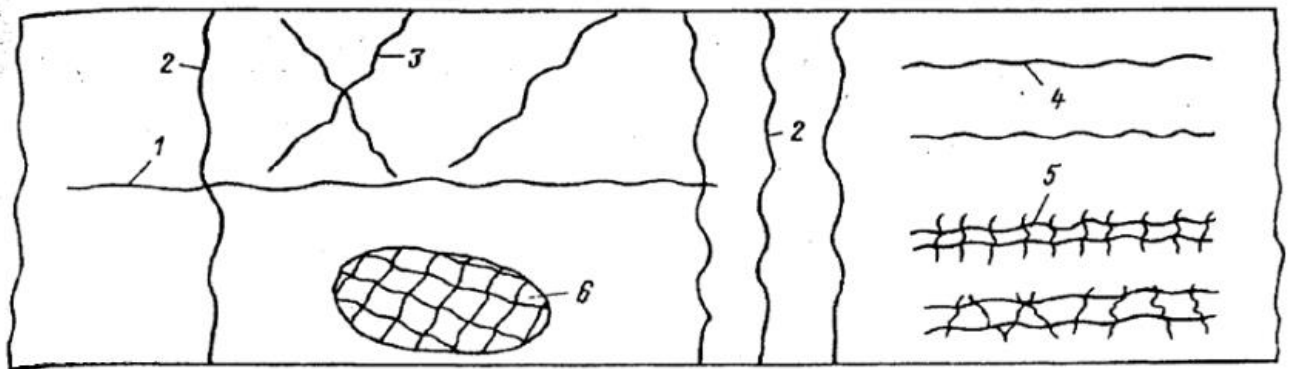


Рис.1.2.1: Тріщини дорожнього покриття

- Деформація та руйнування дорожніх покриттів:

Різноманітність деформацій теж достатньо велика [3], у загальному їх розділяють на:

- вибоїни;
- лущення;
- викришування;
- просідання;
- проломи;
- зміщення;
- руйнування країв (кромки) дорожнього одягу;
- колійність.

Вибоїна — це місцеве руйнування одягу у вигляді поглиблення з різкими вираженими крутими краями. Такі дефекти утворюються в першу чергу через недостатній контакт між мінеральними та органічними матеріалами, недостатність утрамбування покриттів, забруднення, використання неякісних матеріалів (надмірне нагрівання асфальтобетонної суміші, потрапляння необробленого щебеню чи піску у

суміш і тому подібне).

Особливо активний процес утворення вибоїн протікає у весінній період року. Кліматичні умови сприяють чергуванню від'ємних та додатних температур повітря та верхнього одягу. Вода що потрапляє у пори та мікротріщини при низьких температурах розширюється, розштовхуючи стінки та збільшуючи пустоти. Зв'язок між матеріалами при розширенні пустот послаблюються та під впливом коліс автомобілів утворюються вибоїна що швидко прогресує з часом.

Наштовхуючись на невеликий дефект (Рис. 1.2.2) (1) колесо отримує поштовх, відриваючись від поверхні та створюючи недалеко від вибоїни динамічний удар, що призводить у цьому місті (4) додатних напружень. За рахунок цього на місці приземлення шини виникає ще один дефект (5), наприклад тріщина, який після певного часу може злитися із першим розглянутим дефектом, утворюючи вибоїну (6).

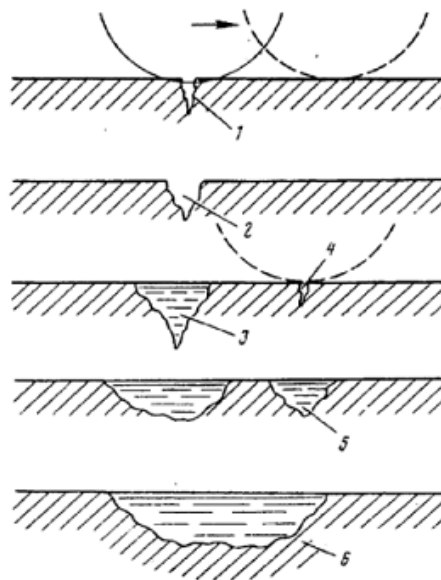


Рис.1.2.2: Динаміка розвитку вибоїни

Лущення (Рис.1.2.3.б) — це руйнування дорожнього покриття через відшарування тонких плівок та лущинок матеріалу покриття під впливом шин автомобілів, води та низької температури повітря. Характеризується утворенням мікронерівностей глибиною до 5 мм.

Викришування (Рис.1.2.3.а) — це руйнування дорожнього покриття через втрату дрібних зерен матеріалів із подальшим виникненням малих раковин глибиною до 20 мм. З плином часу площа цих дефектів розростається, сигналізуючи про початок деформації покриття.



Рис.1.2.2: а) викришування, б) лущення

Просідання (Рис.1.2.4.б) — це локальні деформації одягу у вигляді впадин з пологими схилами різного розміру, глибиною від 50 до 350 мм. Але при цій деформації не спостерігається випучування та утворення тріщин. Просідання виникають при зниженні нижніх шарів структури дороги через вологість ґрунту. Просідання можуть виникати у разі недостатнього утрамбування шарів земляного полотна або інших шарів, а також від важкої техніки, вплив якої не був розрахований на даний тип дорожньої частини.

Проломи (Рис.1.2.4.а) — руйнування дорожнього одягу у вигляді глибоких отворів і значних за площею та довжиною прорізи, зазвичай по смугам накату. Подібні деформації утворюються під дією коліс транспортних засобів, особливо в періоди зниження міцності дорожньої конструкції за рахунок температурних змін. Характеризується повною деформацією поперечного профілю. Можуть утворюватися у наслідок надмірної вологості матеріалів шарів основи та ґрунту, а також під дією надмірних вертикальних сил.



а)



б)

Рис.1.2.2: а) пролом, б) просадок

Зміщення (Рис.1.2.5.а) — це деформація асфальтобетонного покриття у формі переміщень поверхневого шару, що зсувається по нижньому шару на якому лежить перший. Відбувається внаслідок недостатності зв'язків зчеплення матеріалів а також неякісність самих покриттів. Частіше за все утворюється на місцях гальмування автомобілів. Під впливом дотичних сил відбувається зсув в матеріалах верхнього шару, або його зсув по нижній поверхні покриття з утворенням поперечних тріщин на смугах накату. Причиною цьому можна назвати підвищену пластичність матеріалу. Пластичність може бути спричинена через надмір в'язучого матеріалу у суміші а також недостатньою теплостійкістю при високих температурах, в наслідок чого й відбувається деформація розтягу.

Руйнування країв (кромки) дорожнього одягу (Рис.1.2.5.б) — це окремі тріщини та сітки тріщин уздовж країв. Спостерігаються відломлювання та дефектність поперечного профілю прикромочних смуг. Руйнування кромки відбувається в наслідок зниження міцності країв верхнього одягу, у свою чергу це відбувається через зниження товщини шару при краях через підвищення вологості земляного полотна.

Колійність (Рис.1.2.5.в) — деформація одягу у вигляді поглиблень по смугам накату автомобілів. При інтенсивному розвитку колії може утворитися пролом. Колії утворюються при накопиченні пластичних деформацій у шарах дорожнього покриття, а також при посиленому зносі верхнього шару. Зазвичай класифікують колійності глибиною до 15 мм, 15-30 мм, від 30 мм.



а)



б)



в)

Рис.1.2.2: а) зміщення, б) руйнування краю, в) колійність

### 1.3 Актуальність автоматизації контролю дорожніх покриттів

Питання якості дорожнього покриття є важливим у житті багатьох сфер діяльності людини. Дороги є логістичними шляхами що використовуються цілодобово різною технікою, підтримуючи економіку країн та запроваджуючи швидкісні та безпечні транспортні можливості для населення планети.

У загальному випадку більшість країн та компаній використовують експертно-візуальний метод для контролю покриттів. Цей метод є візуальною оцінкою та різниться від місця до місця. Усі вони засновані на американській методиці визначення індексу дорожнього стану PCI розробленою фондом USACE (United States Army Corps of Engineers), та згодом методика стала американським стандартом. PCI пропонує класифікацію дорожніх покриттів згідно рейтингу від нуля до ста, та складається з семи оцінок: failed (0-10), very poor (10-25), poor (25-40), fair (40-55), good (55-70), very good (70-85), excellent (85-100) [4]. Алгоритм розрахунку

оцінки полягає у зборі даних про кількість дефектів та ступінь дефектності кожного (малий, середній, високий) по заданим критеріям габаритних розмірів. Для кожного класифікованого дефекту вираховується щільність виснажливості згідно стандартним кореляційним таблицям та на основі щільності робиться оцінка дефектності охопленої площі. Цікавим є перший етап — усі дані збираються та оброблюються кваліфікованими спеціалістами вручну. Участь людини (або навіть бригади) у таких оцінках є не тільки неоптимальним, але й не завжди ефективним за рахунок людського фактору, що спотворює результати контролю. В Україні діє схожий алгоритм контролю якості дорожнього одягу [5]. Його сутність полягає у описаному алгоритмі, відрізняються лише розрахунки кінцевих критеріїв.

Однак із залученням цифрової техніки із часом почали використовувати автоматизовані підходи для контролю дефектності дорожнього покриття [6]. Їх типовість та функціональність різняться, але базується на фото- та відео-реєстрації — після збору зображень спеціалісти можуть оброблювати збережені дані та робити приблизні оцінки стану доріг, звісно не маючи змоги фізично їх виміряти. Наприклад такі параметри як, глибина вибоїни та колійності.

На протязі останніх десятиліть почали використовувати мобільні картографічні системи MMS [7]. Призначення таких систем полягає у розмітці тих чи інших об'єктів у просторі, позначаючи елементи простору та їх характеристики згідно сферичних координат Землі, наприклад системами глобального позиціонування GPS. Таким чином легко збирати ті чи інші оцифровані дані про об'єкти простору, за допомогою яких можливо формувати, наприклад, різноманітні інтерактивні мапи.

Першими такими системами можна назвати GPSVan<sup>TM</sup> та VISAT Van (Рис.1.3.1), 1993 та 1994 років випуску відповідно. Це фактично рухомі устаткування що фотографують простір та маючи датчики позиціонування помічають кожне зроблене зображення із відповідними координатами. GPSVan<sup>TM</sup> та VISAT Van зберігають інформацію у вигляді зображень, маючи датчики позиціонування помічають кожне зображення відповідними координатами місця фотографування. Із мітками координат можливо більш зручно розмітити ті чи інші об'єкти на

зображеннях. Використовувалися дві камери за для створення стерео ефекту за для можливості розрахунку координат об'єктів на фото. Координати місця фотографування використовували як відправну точку та визначали приблизні положення об'єктів у просторі. Позиціонування системи відтворювалось одометрами, гіроскопами та GPS, даючи точність розмітки у 0,3-3 метри [8].



Рис.1.3.1: а) GPSVan™, б) VISAT Van

Сучасні системи мають схожі принципи, використовуючи лише різні методи інтерпретації даних простору та пересування систем. Наприклад є популярною технологія Лідарів (LiDAR).

Лідарна система схожа за принципом із радарною системою, відправляючи промінь світла робиться замір часу подорожі променя. Промінь випромінюється, відбивається від першої зустрічної поверхні, та прямує назад на джерело випромінювання. За рахунок вимірюного часу та швидкості розповсюдження світлової хвилі легко знаходиться відстань від джерела до площі відбиття.

Використовуючи безпілотний літальний апарат можна ефективно збирати трьохмірну інформацію про різноманітні поверхні, поєднуючи технологію Лідарів та відеокамер [9]. Ще більш цікавими є сучасні компактні пересувні MMS що можуть бути встановлені на будь-якій поверхні. Такі системи маючи відеокамери високої якості можуть формувати зображення із полем зору у 360 градусів, а також при використанні Лідарів із високою роздільною здатністю отримуються високоякісні 3D

моделі навколишнього середовища. Подібними сучасними системами є “RIEGL VMX-2HA”, “Trimble MX9” та інші (Рис. 1.3.2).



а)



б)

Рис.1.3.2: а) RIEGL VMX-2HA, б) Trimble MX9

Усі сучасні системи MMS у загальному є системами збору та передачі даних онлайн чи їх зберігання у внутрішньому пристрої запам'ятовування. Цікавим у таких системах по відношенню до контролю дорожнього одягу є саме дані про покриття. На основі подібних MMS розробляють пересувні установи що сканують та одночасно розмічають дефекти дорожніх покриттів. Наприклад такою сучасною автоматизованою системою є ViaPPS: A Mobile Pavement Profiling System [10]. ViaPPS (Рис. 1.3.3.а) використовує комплекси сенсорів: чотири високоякісних лазерних Лідари (Z+F Profiler 9012A із вертикальним полем зору у 360 градусів, 2x Velodyne Puck Lite із горизонтальним полем зору у 360 градусів, ViaIRI Texture Laser для виміру профілю шорсткості та IRI), дві системи відеокамер (8 камер роздільністю у 5 мега пікселів, Ladybug 5+ from FLIR), комплекс з навігаційних систем. На основі синхронної роботи камер та лідарів збираються дані про стан дороги, утворюються карти розміток проїжджої частини за координатними мітками GPS. Система рухаючись зі швидкістю 60 км/год детектує дефекти та загальний стан проїжджої ділянки за міжнародними стандартами, отримуючи показники PCI та IRI (International roughness index). ViaPPS є ефективним засобом контролю доріг, потребуючи мінімальної



кількості персоналу для функціонування.



Рис.1.3.3: а) Загальний вигляд системи ViaPPS, б) Фрагмент результатів зібраних та оброблених даних

На рівні з PCI використовують більш сучасний інтернаціональний індекс нерівності (шорсткості) IRI (International Roughness Index). Вимірювання проводиться більш складними методами, із використанням профілометрів. Після аналізу вертикальних коливань та горизонтальних зсувів дорожнього покриття результати оброблюються алгоритмами виходячи з середніх вертикальних коливань рухомої системи по контрольованому покриву. Ці виміри здійснюються спеціальними датчиками та системами що детектують коливання транспортного засобу. Після збору даних одного кілометра дорожньої частини робиться оцінка та визначення рейтингу, від 10 до 0, де 10 — найгірший стан дорожнього покриття, 0 — найкращий. Цей показник є намір зібрати загальну характеристику про стан дороги, абстрагуючись від класифікації дефектів. І це має сенс — саме через дефектність дорожнього покриву зазвичай автомобільний транспорт має більші коливання, ніж на бездефектному. Така оцінка є зручною для загального маркування тих чи інших частин доріг не зважаючи на характер деформацій. Високий показник IRI лише свідчить про те що на певній ділянці потрібно провести технологічний огляд стану покриву.

Але проблема усіх подібних систем у коштовності, нехай вони і є ефективними, той самий IRI потребує системи з акселерометрів, гіроскопів та

лазерних датчиків, або інших систем.

Деяким вирішенням може слугувати зменшення кількості та якості коштовного устаткування. Таким є підхід використання алгоритмів штучного інтелекту AI, а тобто його підгалузь машинного навчання ML. Сенс підходу заключений у використанні при скануванні дорожньої частини лише фото- та відео-камер, залучаючи до системи розпізнавання лише растрові зображення та алгоритми ML. У наш час такі алгоритми здатні аналізувати зображення із порівняно низькою роздільною здатністю, що наприклад можна зустріти практично у будь-якому смартфоні.

Дуже зручно мати змогу замінити комплексні системи датчиків на більш дешеві, у найоптимальнішому випадку на камеру смартфона. Останні розробки із пропонованою ідеєю детектування дефектів доріг зі зображень знятих на камеру смартфона мають достатньо перспективні результати, маючи точності класифікування та позиціонування близькі до високих рівнів [11, 12, 13].

Всі вище названі методи так чи інакше дають змогу збирати растрові зображення дорожніх частин, що може бути використано для будь-якої майбутньої системи заснованої на методах ML або DL.

## ***Розділ 2. Використання штучних нейронних мереж для автоматизації процесів класифікації***

### **2.1 Огляд алгоритму штучних нейронних мереж**

Коротко розглянемо принцип алгоритму штучної нейронної мережі (ANN) [28].

ANN це спроба зімітувати систему біологічної нейронної мережі що притаманні багатьом тваринам. Така мережа складається зі зв'язків, синапсисів між нейронами (Рис.2.1.1). Зв'язки зв'язують нейрони, даючи змогу передавати нейронам один одному сигнали. Ці сигнали є дійсними числами  $x_{ij}^{(L)}$ . Вихід нейрону розраховується за певною нелінійною активаційною функцією (Activation Function), наприклад ReLU або сигмоїдальна функція. Кожному зв'язку відповідає певне число — вага  $w_{ji}^{(L)}$ , що характеризує силу зв'язку або кількість передаточного сигналу. Зазвичай нейрони групують у шари, розділяючи: вхідний шар, що характеризує вхідні дані  $x_{1j}^{(1)}$ ; проміжні шари, що характеризують навчальні здібності; та вихідний шар. Вихідний шар характеризує результат виходу алгоритму ANN  $x_{ij}^{(L+N)}$ , зазвичай роблячи ті чи інші передбачення. Також кожному зв'язку відповідає значення офсету, або ще називають поправкою активаційної функції  $b_i^{(L)}$ . Найпростіша ANN представлена на Рис. 2.1.1.

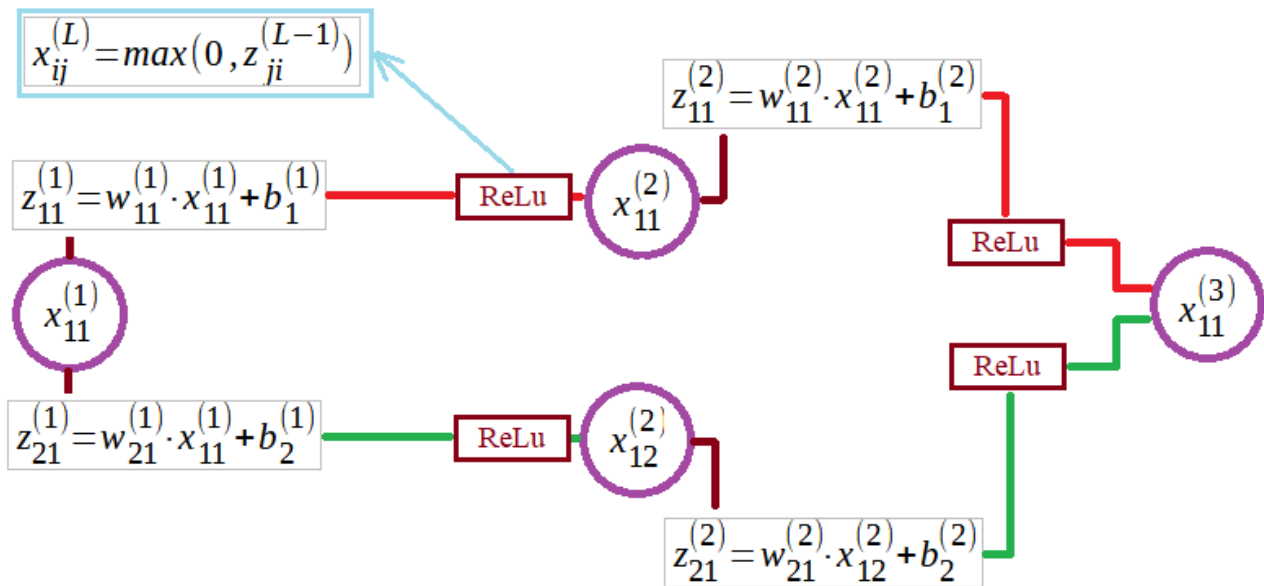


Рис.2.1.1: Найпростіша ANN

Задача навчання полягає у корегуванні всіх ваг та значень офсету ANN. Це можна пояснити тим що параметри вихідних прошарків залежать від усіх ваг та офсетів, тобто ваги та офсети є аргументами кожного окремого вихідного нейрону.

Частіше за все при навчанні із вчителем (Supervised Learning) вводять функцію втрат (Loss Function). Функція втрат показує, наскільки вихідні дані співпадають із очікуваними. Однією із простих таких функцій є середньоквадратична похибка:

$$MSE_i = (y_i - \hat{y}_i)^2,$$

де  $y_i$ - і-те вихідне значення нейрону;

$\hat{y}_i$ - і-те відповідне очікуване значення виходу нейрону.

Сума усіх значень функцій втрат кожного нейрону є значення функції загальної втрати (Cost Function):

$$C = MSE = \frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2,$$

де  $y_i$ - і-те вихідне значення нейрону;

$\hat{y}_i$ - і-те відповідне очікуване значення виходу нейрону.

$n$ - кількість значень виходів нейронів.

Дуже часто використовують іншу Cost Function, а точніше функцію перехресної ентропії (Cross-entropy).

ANN буде робити правильні передбачення у тому випадку, коли значення функції Cost Function буде мінімальним. Тобто задача навчання ANN полягає у мінімізації Cost Function, причому мінімізація відносно кожного аргументу цієї функції, а тобто усіх ваг та офсетів.

Змінення навчальних параметрів ANN здійснюється методом зворотного поширення помилки Backpropagation. Сутність полягає у послідовному диференціюванні значень ANN за для знаходження оптимальних значень, що будуть задовольняти перелічені умови. Така задача має назву градієнтного спуску (Gradient Descent) для кожної з функції, тобто значення ваг та офсетів, для яких знаходяться мінімуми залежно від вхідних аргументів. Наприклад для того що б знайти функцію зміни Cost Function від параметру першої ваги з прикладу на Рис.(2.1.1) потрібно спростити наступний вираз:

$$\frac{dC}{dw_{11}^{(1)}} = \frac{dC}{dx_{11}^{(3)}} \cdot \frac{dx_{11}^{(3)}}{dz_{11}^{(2)}} \cdot \frac{dz_{11}^{(2)}}{dx_{11}^{(2)}} \cdot \frac{dx_{11}^{(2)}}{dz_{11}^{(1)}} \cdot \frac{dz_{11}^{(1)}}{dw_{11}^{(1)}}$$

Цей вираз можна отримати застосувавши правило диференціювання складної функції (Chain Rule) [29]. Як можна побачити алгоритм розрахунку є досить комплексний та потребує великих обчислювальних потужностей для більш громіздких архітектур ANN.

Сутність ANN полягає у знаходженні подібних асоціативних параметрах між даними. Ці асоціації проявляються у вигляді потрібних параметрів штучної нейронної мережі, за рахунок яких мережа і запам'ятовує інформацію та маючи змогу робити передбачення з вхідних даних.

## 2.2 Базова сутність алгоритму згорткових нейронних мереж

Розглянемо базову сутність та принципи згорткових нейронних мереж (CNN) [30].

CNN дуже схожа на ANN, відмінність у тому що на вході такої мережі є не масив чисел, а растрове зображення. Кожен піксель зображення характеризується яскравістю. Якщо зображення чорно-біле, значення будь-якого пікселю буде ранжуватися від 0, що відповідає чорному кольору, до 255, що відповідає білому кольору. Значення між цим діапазоном дають відтінки сірого. Наприклад зображення розмірністю  $100 \times 100 \times 1$  пікселів можливо інтерпретувати як матрицю зі ста рядків та ста стовпців, значення комірок яких є яскравість пікселів. Якщо застосувати підхід із ANN, та створити на вході такої мережі  $100 * 100 = 10000$  нейронів, буде достатньо складно вирахувати усі параметри ANN. Такий підхід може давати певні результати, але час на їх отримання робить метод абсолютно не оптимальним, вже не кажучи про перенавчання, перенавантаження вагів та інші проблеми.

З часом знайшовся нових підхід, він полягає у зменшенні кількості вхідних нейронів для ANN. Зменшення відбувається на основі обробки зображення та вилучення з зображення великої кількості первинних ознак. Цими ознаками є низькорівнева інформація про об'єкти, наприклад частини контурів що характеризують об'єкт. Отримання такої інформації хоч і у деякому сенсі спотворює дані, але для алгоритму ці втрати практично несуттєві. Опісля операції згортки та деякої кількості ітерацій вилучень отримується набагато менший обсяг чисел що репрезентують все ті ж самі вхідні дані, але з якими простіше працювати надалі звичайній ANN. Розглянемо більш детально описані процеси. Для цього необхідно розібратися у складових структури CNN.

CNN складається з наступних частин (шарів):

- Згорткові шари (Convolutional layers);
- Шари активації (Activation layers);
- Пакетна нормалізація (Batch Normalization);
- Шари пулінгу (Pooling layers);
- Повністю з'єднані шари (Fully connected layers).

Згорткові шари представляють з себе той самий процес зменшення та виділення корисних ознак для CNN. Для цього використовуються маски, або фільтри,

частіше їх називають ядром (Kernel). Ядро, або ж фільтр, представляє з себе невелику матрицю з певними значеннями у комірках (Рис.2.2.1.б). Значення цих комірок й формують профіль та характер фільтра.

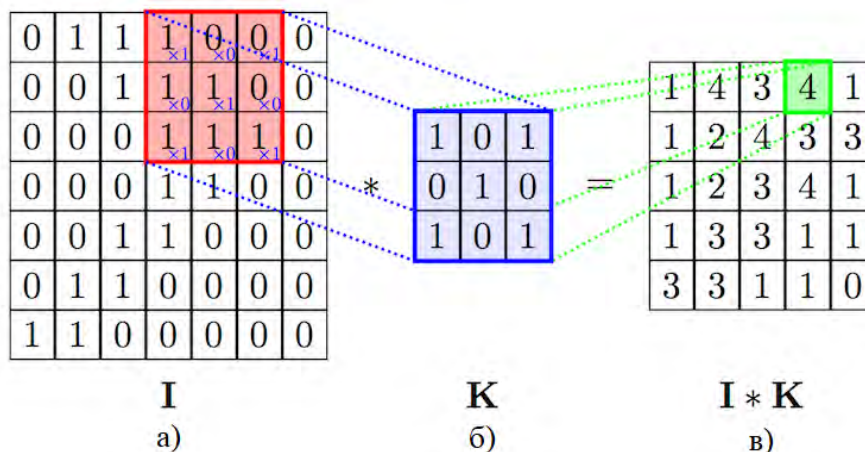


Рис.2.2.1: Ілюстрація згортки фільтром

а) вхідний масив даних, б) фільтр, в) матриця ознак

Розмір фільтрів може варіюватися від 3x3 до більших розмірів. Переміщуючись по зображенню фільтр обирає певну ділянку на вхідній матриці (Рис.2.2.1.а). Ця виділена ділянка має таку ж розмірність як і фільтр. Виділена ділянка стає тимчасовою матрицею. Тимчасова матриця множиться на матрицю фільтра та отримується скалярне число. Це скалярне число характеризує піксель нового зображення, задаючи його яскравість. Після проходження фільтром усього зображення отримується нова матриця із виділеними первинними ознаками, що має назву Feature Maps – матриця розмітки ознак, або просто матриця ознак (Рис.2.2.1.в).

Маючи зображення у 100x100 пікселів після згортки фільтром 3x3 буде отримана матриця 98x98. У випадку коли результат згортки повинен дати більш менші зображення використовується збільшення кроку переміщення фільтра. Це називається страйдом (Stride). Якщо взяти крок переміщення 2, у випадку із матрицею 100x100 після обробки тим самим фільтром вихідне зображення буде мати 49x49 пікселів.

Звісно після зменшення розмірності певна частина корисної інформації втрачається, але усі пікселі були використані у формуванні нової матриці ознак, тому втрати інформативності переважно малі. Важливим є характер фільтру, а тобто значення його комірок. Значення матриці фільтру задає параметри формування первинних ознак. Кожен піксель фільтру є ваговим коефіцієнтом що визначає вклад відповідного пікселя тимчасової матриці у загальне число первинної ознаки. Наприклад якщо стоїть задача детектування конкретних об'єктів, слід звернути увагу які саме контури притаманні об'єкту, та вибрати відповідний фільтр.

Чим більше число первинної ознаки після множення матриці фільтром із частиною зображення, тим більш подібні у цій області зображення та фільтр. Такий ефект акцентує важливість цієї області для запам'ятовування.

Основна ідея усіх згорткових мереж будується саме на концепті фільтрів що збирають корисні первинні ознаки у матриці ознак. Таких послідовних згорток може бути багато. Якщо мережа має більше ніж п'ять згорткових шарів — така мережа відноситься до DL.

Шарами активації є проміжні проходження значень комірок матриць (ознак чи інші) через функцію активації, наприклад ReLU. Такі нелінійні функції потрібні за для того щоб нейронна мережа знаходила нелінійні кореляції між даними. Без такої функції що застосовується до значень нейронів мережа не зможе виробити складні кореляційні відношення між даними. У алгоритмі CNN функція ReLU застосовується опісля кожного шару пакетної нормалізації.

Пакетна нормалізація це техніка оптимізування та нормалізації виходів нейронних шарів. Сутність полягає у обчисленні середнього значення та варіаційності вихідних значень нейронів, відніманні отриманого середнього значення від кожного значення виходу нейрону та масштабуванні кінцевих отриманих значень на основі попередньо отриманої варіації. Це допомагає зменшити внутрішню коваріаційну зміну системи. Також цей прошарок додає два навчальних параметри: зсув та масштабування. За допомогою них здійснюється нормалізація. Ці параметри є такі ж як і ваги та офсети ANN, тобто вони тренуються разом із усією системою.



Шари пулінгу дуже схожі за своєю функціональністю до згорткових. Їх використовують у випадку коли потрібно за одну ітерацію згортки суттєво зменшити габарити матриці ознак. Відмінність лише у фільтрі та страйді. Існують три типи фільтрів для пулінгу: Max pooling, Average pooling, Global pooling. Вони є ситуативними але у більшості випадків використовуються за для збільшення варіативності та стійкості системи до перенавчання.

Максимальний пулінг обирає максимальні значення з області сканування фільтром. Це допомагає зберегти найважливіші деталі, відкинувши всі інші.

Середній пулінг обчислює середнє арифметичне значення комірок сканування. Це згладжує та перемішує важливі ознаки.

Глобальний пулінг вираховує середнє арифметичне для кожного каналу зображення, утворюючи одномірний масив з такою ж розмірністю як і кількість каналів зображення.

Важливим є розуміння що у випадку із різнокольоровими зображеннями (Рис.2.2.2) алгоритми та принципи описані вище не змінні. Збільшується лише розмірність первинного зображення та масивів ознак. Наприклад різнокольорове зображення розмірністю 100x100 буде мати вигляд трьохмірної матриці — 100x100x3, де 3 це кольорові канали RGB. Кожна підматриця, або канал, має у собі яскравість конкретного кольору — червоного, зеленого та синього. Зміниться також розмірність фільтрів у третьому вимірі, наприклад 3x3x3.

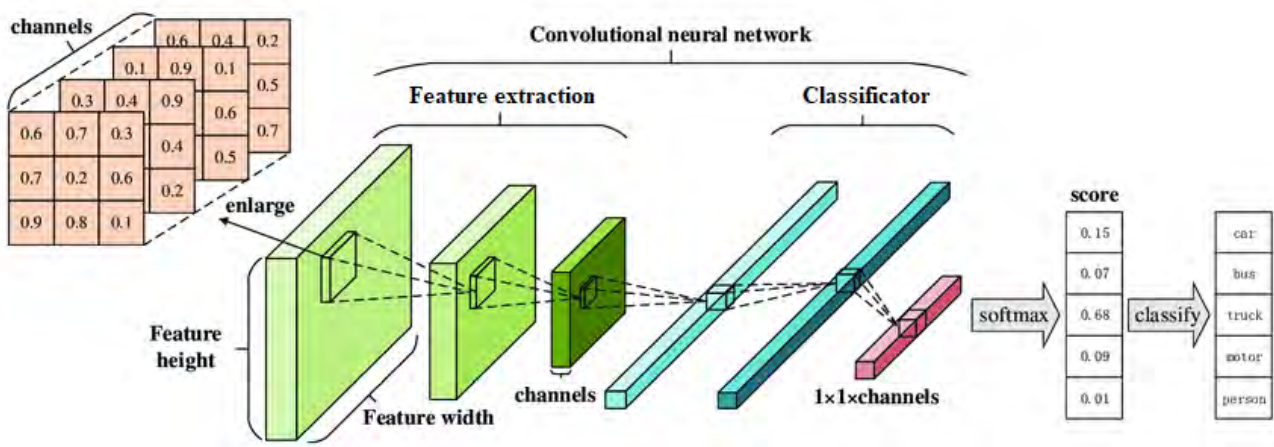


Рис.2.2.2: Ілюстрація структури CNN

Повністю з'єднані шари (Fully connected layers) це звичайна ANN що буде класифікувати отримані дані з попередніх шарів. Але перед цим важливо перетворити останній двох- або трьох-вимірний фінальний шар ознак у одновимірний масив. Це досягається звичайним переставлянням елементів матриці у рядок. Вже цей масив вилучених ознак оброблюються за вже розглянутим алгоритмам ANN: де вихідний шар видає результати, наприклад, класифікації даних — растрового зображення.

Як можна побачити на Рис.2.2.2 CNN ділиться на дві частини: Feature extraction та Classifier. Feature extraction виконує виділення особливостей даних Feature maps – матриці нейронів із первинними ознаками. Classifier у свою чергу класифікує спрощені дані та робить передбачення.

Ще частіше у багатьох системах та алгоритмах CNN різних архітектур використовують як основні алгоритми для вилучення первинних ознак з даних, тобто Feature extraction. Такі алгоритми в загальному називають Backbone networks.

### **2.3. Передавальне навчання**

Передавальне навчання, або ж Transfer Learning, це дослідницька задача або підхід, що дає змогу використати навчені параметри якогось алгоритму AI для іншої задачі []. Типовим у такому підході є використання вже навчених параметрів CNN у задачі класифікації інших об'єктів. Наприклад ваги навченої нейронної мережі на розпізнаванні котів та собак можуть бути використані для розпізнавання щось схожого до названих об'єктів, наприклад інших тварин — ведмедів та білок.

Ефективність такого підходу може різнитися від навчених параметрах, вони можуть як негативно так і позитивно вплинути на точність алгоритму [27]. Але якщо існує кореляційна схожість із даними двох задач, результативність буде відчутно більша, ніж навпаки. Наприклад використовувати параметри з задачі розпізнавання тварин не доцільно при вирішенні задачі розпізнавання типів будівель чи машин.

Дуже часто метод використовується коли є брак даних для тренування.

Transfer Learning компенсує це, даючи можливість перенести знання з однієї задачі на іншу, при цьому використовуючи малий обсяг даних з нової задачі.

Метод реалізації для CNN достатньо простий. Береться лише перша частина CNN — Feature extraction, ваги якої роблять незмінними, їх “заморожують”. Друга — Classifier, відкидається. Основна задача передавального навчання це навчити новий класифікатор для нових даних. Вже параметри нового класифікатора встановлюють кореляційні зв’язки із вхідними даними та мітками класів.

## **2.4 Контроль дорожніх покриттів з використанням алгоритмів глибокого навчання**

Задача контролю дорожніх покриттів архітектурами CNN може бути реалізована чотирма підходами комп’ютерного зору (CV) [14]:

- Класифікація зображень;
- Детектування об’єктів на зображенні;
- Сегментація зображення;
- Сегментація об’єктів на зображенні.

Класифікація зображень є найпопулярнішою задачею. Існує велика кількість статей присвячених класифікації дефектів дорожніх покриттів. Усі вони так чи інакше використовують різноманітні, але подібні по своїй сутності дані, тренуючи різноманітні архітектури CNN та використовуючи різні обробки даних, їх інтерпретації та способи класифікацій. Наприклад є статті у яких ставиться задача класифікації по двом класам: зображення із дефектом та без [15]. Звісно результативність такого алгоритму не є дуже ефективною, саме через брак інформативності про саму деформацію. Щось більш цікавим є класифікація присутності та відсутності одного типу дефекту, наприклад вибоїни [12]. Результативність є вже більш цікавою: можна зрозуміти який дефект був ідентифікований та на якому зображенні. Але це все ще не дає нам розуміння ступеню дефектності, вже не кажучи про такі метрики як PCI.

Більш поширеними є класифікація з номенклатури, як мінімум декількох дефектів. Задача полягає у тому ж самому — віднесення вхідного фото до певного класу. Це вже більш інформативно, але все ще не ідеально для метрик та повних висновків про стан одягу. Алгоритм такого підходу є достатньо простий — звичайна CNN оброблює вхідні дані із маркуванням зображень. Фактично якщо не брати до уваги підготовку даних, увесь алгоритм залежить лише від навчання параметрів нейронної мережі.

Детектування об'єктів вже дає більше інформації. Такий алгоритм здатен розбирати у зображенні певну кількість дефектів, класифікувати та маркувати координати кожного з певною точністю (Рис.2.4.1) [16]. Частіше за всього у цьому випадку вхідне зображення розбивають на сітку підзображень. Кожне таке зображення конкретна CNN класифікує на приналежність до бажаного класифікованого об'єкту або відсутність будь-якого. Складність алгоритму полягає у тому, що кожне підзображення потребує аналізу та обробкою згортковим шаром CNN. Задача зводиться до тренування CNN на ознаки об'єктів в підзображеннях, що є більш довготривалою задачею та призводить до більш нестабільних результатів [17]. Використовують різні фреймворки для вирішення цього завдання — найпопулярнішим на зараз є YOLO (You Only Look Once).

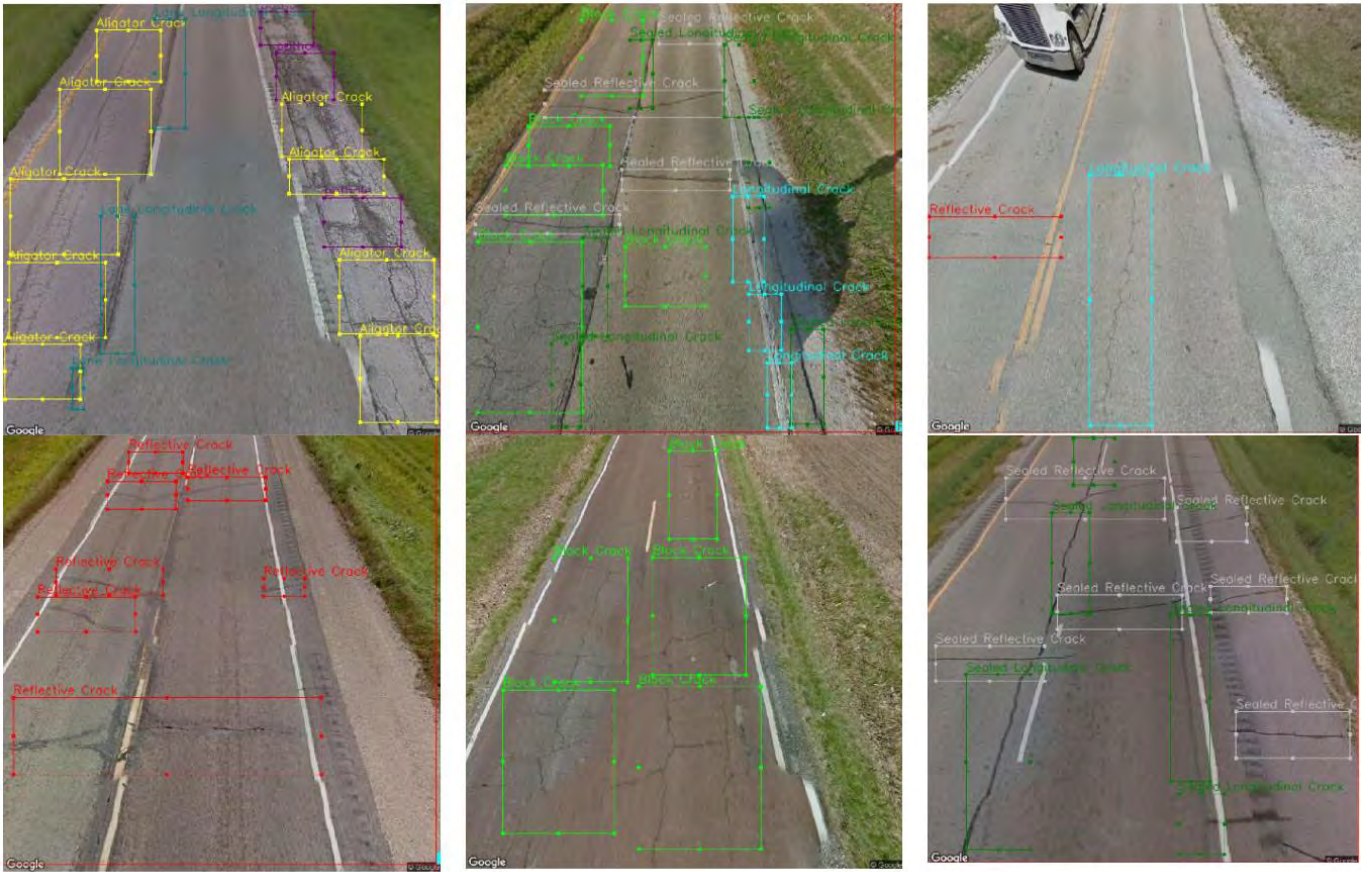


Рис.2.4.1: Ілюстрація результату маркування R-CNN детектованих дефектів дорожнього покриття

Також існує підзадача детектування об'єктів — локалізація одного об'єкта (Localisation). Є тим самим ж, але лише з одним об'єктом на фото.

Сегментація зображення є достатньо традиційною задачею CV. Вона полягає у виділенні на зображенні довільну кількість сегментів що виділяють ті чи інші об'єкти та образи. Сегментація виконується багатьма методами, не тільки DL та ML. Використовуючи DL цю задачу вирішують різноманітнішими шляхами [14].

Задача реалізацій полягає у семантичній сегментації зображення — тобто кожний піксель, що ідентифікують той чи інший об'єкт, маркують, або класифікують відповідно до приналежності до утворення певного класу об'єкта на зображенні. Наприклад у бінарній семантичній класифікації існує всього дві групи пікселів — бажаного об'єкту та всі інші що не є об'єктом (Рис. 2.4.2). Зазвичай білим кольором маркують пікселі об'єкту, а чорним всі інші.

Стандартним підходом сегментації у DL є використання повністю згорткових нейронних мереж (Fully Convolutional Network). Такі мережі сегментують об'єкти на зображеннях. Причому особливість такої мережі є те що вона може приймати на вхід зображення будь-якого розміру. Це пов'язано саме із структурою FCN, вона спочатку згортає зображення, навчаючи параметри, а потім розгортає із тим самим розміром зображення, та використовуючи параметри при згортці видає сегментоване зображення на виході [18].

Використовують і інші алгоритми у сфері DL, наприклад архітектури U-net є більш перспективними у задачах сегментації [14]. Для сегментації можливо використовувати будь-які архітектури CNN, вони всі виконують задачу виділення асоціативних ознак з зображень.

У загальному задача детектування країв тріщин є більш цікавою ніж попередньо розглянуті у контролі доріг. Так як задача полягає у виділенні пікселів об'єктів, ми можемо приблизно оцінювати площу тріщин [19]. Сучасні методи дають змогу з високою точністю передбачити контури дефекту, причому на комплексному зображенні із іншими об'єктами та заднім фоном. Наприклад розроблена система на основі архітектури STRNet (Semantic Transformer Representation Network) [20] має змогу у реальному часі на відеозйомці репрезентувати на виході сегментоване зображення (або потік зображень) (Рис.2.4.2).

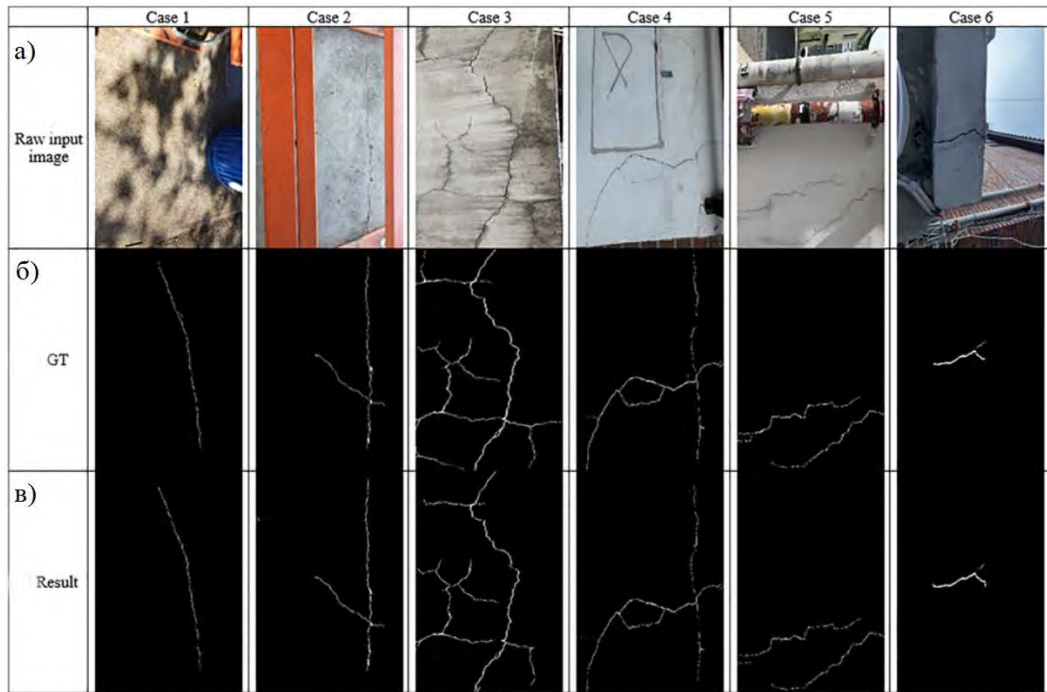


Рис.2.4.2: а) Оригінальне зображення, б) правдива сегментація (Ground Trust), в) передбачення імплементації з STRNet

Існує багато подібних систем алгоритмів заснованих на інших імплементаціях мереж. Для їх реалізацій використовують як базові CNN (Xception [21], ResNet [22]), так і більш комплексні, що розроблені під вузьку спеціалізацію, як STRNet, SegNet [23].

Як можна побачити такі системи є більш ефективними через свою інформативність, можливо розробити або розрахувати критерій дефектності на основі профілю та геометрії детектованого контуру, або навіть за допомогою виділеної та розрахованої площі.

Зазвичай сегментацію використовують для детектування тріщин, але саме тріщини це один із найбільш складних типів дефектів для алгоритмів CV для розпізнавання. Якщо алгоритми будуть здатні виокремлювати дрібніші мікротріщини у поверхнях, інші деформації будуть не проблемою — саме через їх геометричні прояви на растрових зображеннях.

Сегментація об'єктів на зображенні є найбільш перспективною задачею у CV, вже не кажучи про контроль якості, також її ще називають багатокласовою

сегментацією. Задача полягає у сегментації багатьох об'єктів/класів на зображенні. Наприклад у поставленій задачі було б зручно виділять одразу всі дефекти різними кольорами — вибоїни, тріщини та інші. У задачах із сегментацією тріщини можуть бути виділені із різними товщинам, наділяючи їх різними класами дефектності, наприклад товщини малих, середніх та великих розмірів маркуються трьома кольорами (Рис.2.4.3) [24].

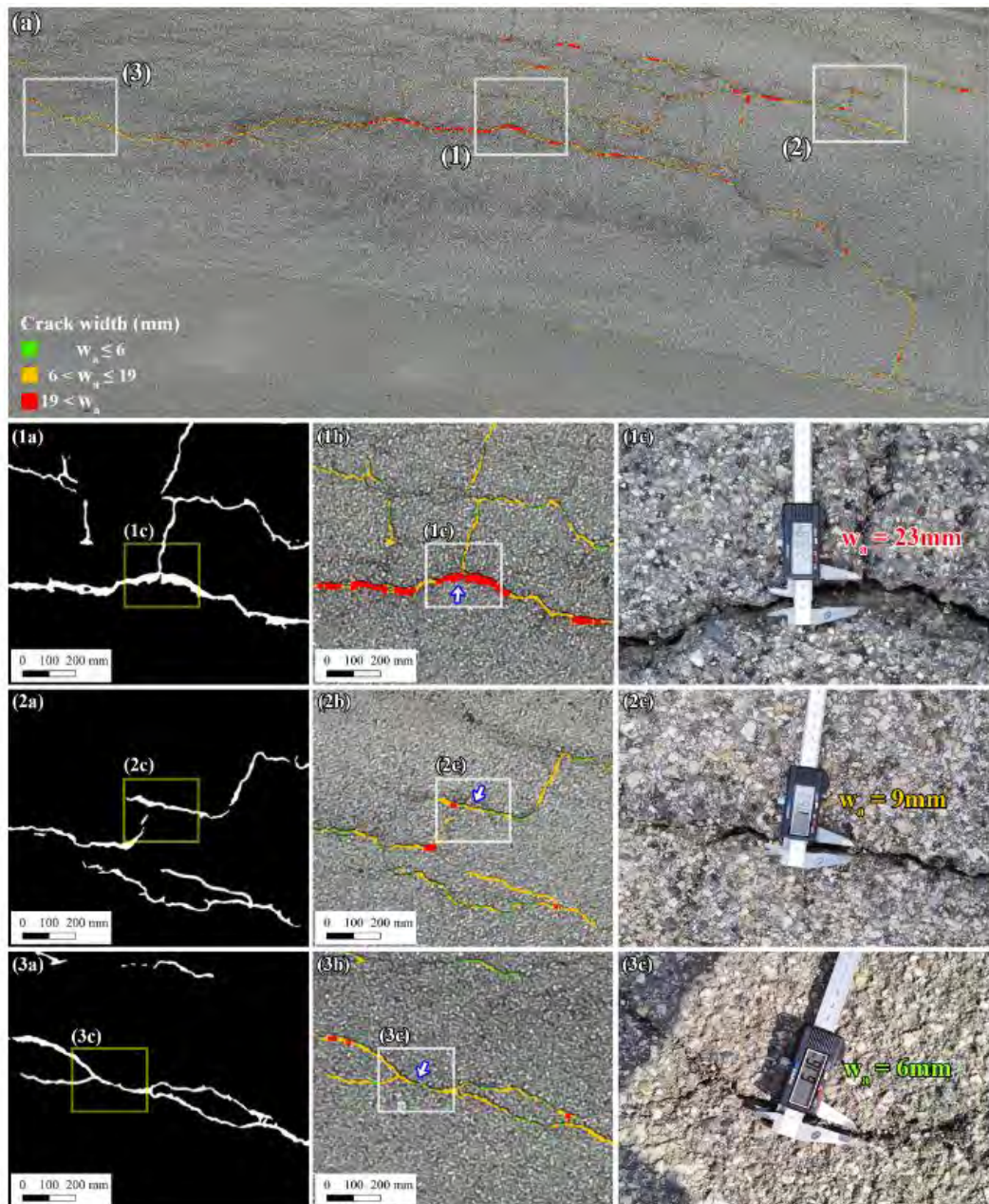


Рис.2.4.3: а) Оригінальне зображення, б) правдива сегментація (Ground Truth), в) передбачення імплементації з STRNet



Класова сегментація потребує більш комплексних підходів саме через складність задачі. Потрібно виділяти на фоні пікселі одразу декількох класів, це створює певні труднощі у класифікації підзображень що аналізуються CNN для класифікування [25].

Будь-який підхід із використанням штучного інтелекту потребує навчальних даних. Це одна з найбільших проблем всієї дослідницької діяльності, не тільки в області задачі контролю дефектності доріг. Наприклад у більшості статтях та роботах присвячені контролю доріг є дисбаланс кількості тренуваних зображень у кожних класах. Проблемою є зображення вибоїн — не завжди вибоїна може бути використана для тренування — наприклад після осадів вибоїни бувають заповнені дощовою водою. Вибоїна із водою частіше за всього не підходить для тренування (якщо це не передбачено поставленою задачею), саме через гладкість поверхні води. Це стосується й інших дефектів, наприклад колійностей та проломів, вони можуть бути засмічені якоюсь матерією не тільки під час осадів. Також важливий час зняття дефектів за для уникання тіней, типу суміші покриву, за рахунок якого колір відмінний, та інші проблеми. Недостатність даних може приводити не тільки до малої точності, але й до перенавчання, не даючи можливості розширювати асоціативний ряд таких алгоритмів.

Існує натуральний метод “роздугтя даних” — аугментація даних (Data Augmentation). Його використання почалося з початком появи ідеї класифікації як такої [26]. Суть методу полягає у натуральних трансформаціях оригінальних даних, тобто таких трансформацій що можуть зустрічатися у реальних даних, але вони є штучними [31]. Трансформоване з оригінального зображення для алгоритмів задач CV є фактично новим, відмінним від вихідного. Тобто застосувавши геометричні чи фотометричні алгоритми можливо доповнювати базу даних для навчання з вихідних даних. Наприклад взявши фото вибоїни, її можна повернути на якийсь кут, або відобразити паралельно, чи все разом — це зображення може стати потенційним зразком для навчання.

Більшість робіт що присвячені автоматичному розпізнаванню дефектності

доріг методами DL використовують DA. DA є невибагливим до обчислювальної потужності, простим та ефективним підходом покращення результатів нейронних мереж. У цій роботі пропонується дослідити методи DA.

### **III**

#### **3.1 Аналіз та вибір методів аугментації**

Мета усіх методів аугментації [31] досягти якнайбільш різноманітності даних, тобто отримання алгоритмом AI повної генералізації даних навчання, з метою уникнення перенавчання — тобто уникнення різниці між тренувальними даними, та даними що алгоритм буде бачити вперше. Саме методи аугментації і надають можливості збільшити та урізноманітнити навчальні дані, це досягається “роздуттям” даних з оригінальної тренувальної вибірки. Різні методи можуть корегувати та надавати ті чи інші поправки у асоціації алгоритмів. Коли потрібно надати алгоритму інформацію що до різноманітності положень об’єктів на зображеннях. Наприклад у задачі розпізнавання облич зазвичай оригінальні зображення центровані, навчання на таких даних дасть асоціації алгоритму що обличчям є лише зображення із усім обличчям в кадрі. За для надання мережі розуміння що обличчям може бути і його частина, на оригінальних даних застосовується наприклад трансформація зміщення або обертання. Такі трансформовані зображення зазвичай додаються до основної вибірки навчання.

Сутність усіх методів полягає у імітуванні деяких натуральних явищ що можуть зустрітися у реальних даних, особливо коли реальних даних із подібними явищами бракує. Поділяють на прості (базові) та комплексні методи.

Виділяють наступні класи та підкласи базових методів [31]:

- геометричні трансформації (Geometric transformations):
  - перевертання, або відзеркалення (Flipping);
  - обрізання (Cropping);
  - обертання (Rotation);
  - зміщення (Translation);
  - зашумлення (Noise injection).
- колірні трансформації, або фотометричні (Color space transformations);
- фільтрові трансформації (Kernel filters);

- міксування зображень (Mixing images);
- випадкове стирання, або вирізання (Random erasing).

Геометричні трансформації є найбільш базовими методами. Їх сутність полягає у маніпуляціях із положенням пікселів на растровому зображенні. Вони не потребують таких кількостей обчислювальних потужностей як у випадку навчання, легко імплементуються у системи та алгоритми.

Перевертання (Рис.3.1.1.б) переставляє відповідні пікселі по одній з осей, отримуючи наприклад дзеркальне відображення від оригінального у випадку вертикального перевертання. Фактично дає розуміння асоціації різного положення об'єкту в просторі. Наприклад якщо об'єкт до гори ногами — це все ще конкретний клас об'єкту, і цей об'єкт потрібно коректно ідентифікувати.

Обрізання (Рис.3.1.1.в) є методом вилучення з основного зображення його основної частини, або будь-якої іншої, зменшуючи при цьому розмір зображення у пікселях. Метод може допомогти виділити якусь конкретну частину на зображенні, наприклад збільшене зображення основного об'єкту, або його частини. Є одним із найбільш ефективним методом поміж більшості інших [33]. Перетворення не зберігає просторові розміри зображення, але це можливо корегувати розтягом обрізаної частини під розміри оригінального.

Обертання (Рис.3.1.1.г) зміщує пікселі по колу, у напрямі за чи проти часовою стрілкою, на певний градус, наприклад 10 та -10 градусів. Схоже на перевертання, але більш параметризоване за рахунок двох довільних напрямів обертання. Перетворення не змінює розміри оригінального зображення, але опісля повертання можуть з'явитися пусті області що заповнюються контрастним кольором чи шумом.

Зміщення (Рис.3.1.1.д) зміщує усі пікселі зображення на певний крок у певний бік, створюване пусте місце після зміщення можуть заповнювати контрастними кольорами чи шумом.

Зашумлення (Рис.3.1.1.е) є методом додавання до матриці пікселів матриці

шуму. Це спричиняє зображення до деякого спотворення — зашумлення. Цей метод достатньо корисний, так як додає стійкості при зашумленні даних, спотворені наприклад апаратурою зйомки.

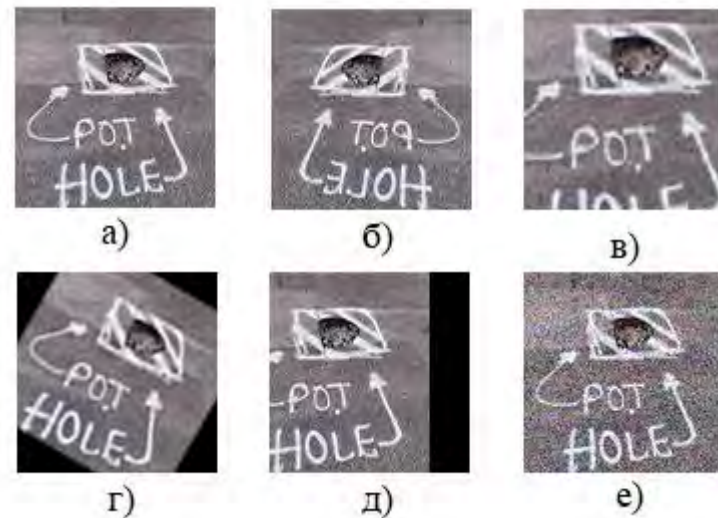


Рис.3.1.1: а) оригінал, б) відзеркалення, в) обрізання, г) обертання, д) зміщення, е) зашумлення

Існує безліч методів геометричних перетворень, способів переставляти пікселів достатня кількість, але у загальному частіше за всього використовують саме названі.

Колірні трансформації полягають у маніпуляції зі значеннями матриць кольорових каналів зображень. Різноманітність маніпуляцій дуже складно класифікувати, їх безліч саме через варіативність обчислювальних методів, але можливо виділити основні що застосовуються у DL. Найчастішим є застосування різноманітних фільтрів що усереднюють, підсилюють, зменшують ті чи інші кольорові канали (Рис.3.1.2). Такі фільтри накладаються виходячи з утворень гістограм кольорових каналів зображень, на основі них можливо зрозуміти який колірний розподіл має фото. У аугментаціях зазвичай використовують випадкові підсилення кольорових каналів (Color Jittering), затемнення та зміщення країв (Edge Enhancement, Vignette), що не сильно спотворює зображення, але надає йому нових ознак та шумів.

Популярним є метод головних компонент (Principal Component Analysis). Він полягає у зменшенні розмірності даних (у даному випадку матриць кольорів) витрачаючи найменші кількості інформації. У задачі аугментації змінює інтенсивність каналів RGB відповідно до природних варіацій зображень, усереднюючи відстань різниць між кольоровими каналами. Таким чином зображення для CNN може виглядати кардинально по іншому.

Такі методи можуть збільшити варіативність вибірки та генералізацію параметрів саме через різноманіття кольорових варіацій зображень. Можливо симулювати як нормальне освітлення, так і з засвітлення — все залежить від типу та якості даних і алгоритму модифікації матриць кольорових каналів. Наприклад зображення дефектності доріг дуже часто мають тіні від різноманітних інших об'єктів фону. Саме такими методами можливе деяке усунення, або ж навпаки погіршення впливу тіней на зображеннях.

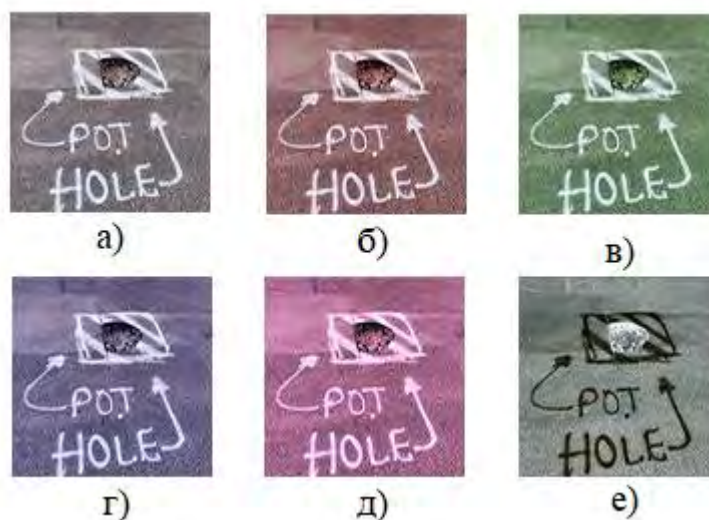


Рис.3.1.2: а) оригінал, б) підсилення R каналу, в) підсилення G каналу, г) підсилення B каналу, д) довільні RGB канали, е) інверсія кольорів

Фільтрові трансформації представляють з себе аналогічний механізм трансформації зображення як і у згорткових шарах CNN. Застосовується фільтр, наприклад для розмиття зображення (Рис.3.1.3.б), схожий на Average pooling або Гауссового шуму, що усереднює значення прилеглих пікселів. Підвищення різкості

(Рис.3.1.3.а) також реалізується на основі фільтрів, роблячи вертикальні та/чи горизонтальні краї більш виразними.

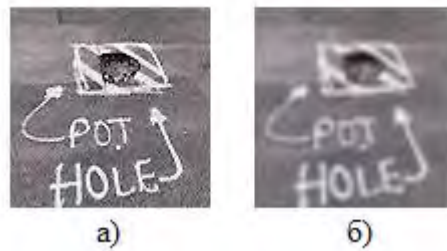


Рис.3.1.3: а) різкість, б) розмиття

Популярним є метод випадкового перемішування значень тимчасових матриць обраними фільтром (PatchShuffle Regularization), та утворення з перемішуваних нової частини зображення (Рис.3.1.4). Результативність цього методу набагато покращила результати точності класифікації на деяких дата сетах [32].

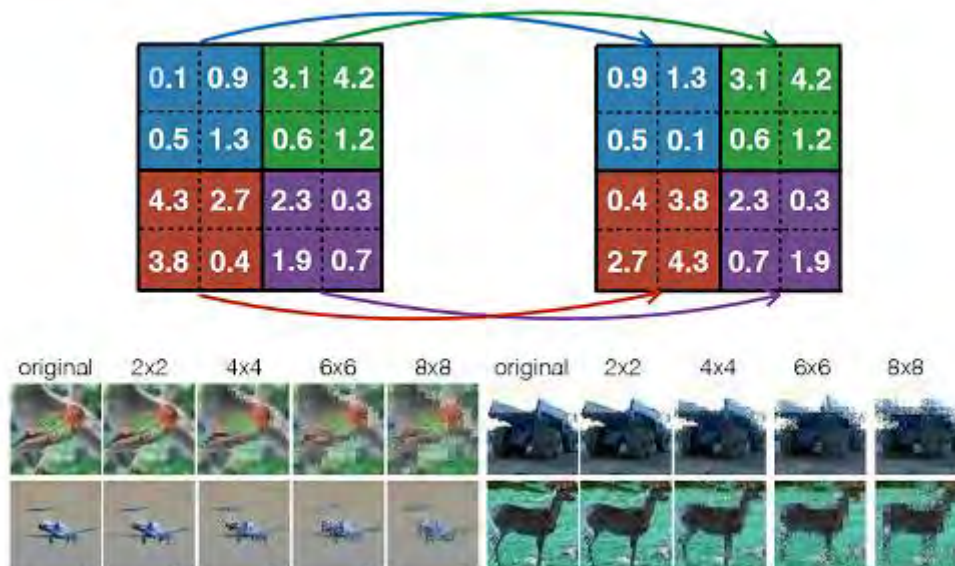


Рис.3.1.4: Інтуїтивна ілюстрація фільтру PatchShuffle

Фільтри аугментацій можуть бути використані одразу у згорткових

нейронних мережах, що може полегшити імплементацію у алгоритми DL.

Міксування зображень має багато проявів у своїй реалізації. Основна суть у комбінуванні зображень якимось алгоритмом, наприклад накладанням один на одного за допомогою усереднення значень контрастності пікселів обох зображень, отримуючи нове (Рис.3.1.5). Також можуть використовувати вирізання з першого зображення частини та додавання його до іншого. Методів безліч, як лінійних так і нелінійних [34]. Як не дивно, такі трансформації покращують результати точності класифікацій на деяких даних сетах, наприклад CIFAR-10 та CIFAR-100 [34]. Ефективність такого підходу складно пояснити чи описати, але у загальному такі нові дані дають більш різноманітну вибірку зображень з різними ознаками, хоча й маючи однакову інформацію з інших зображень. Наприклад такою інформацією є закріплення знань що до різноманітних ознак, таких як краї об'єктів.

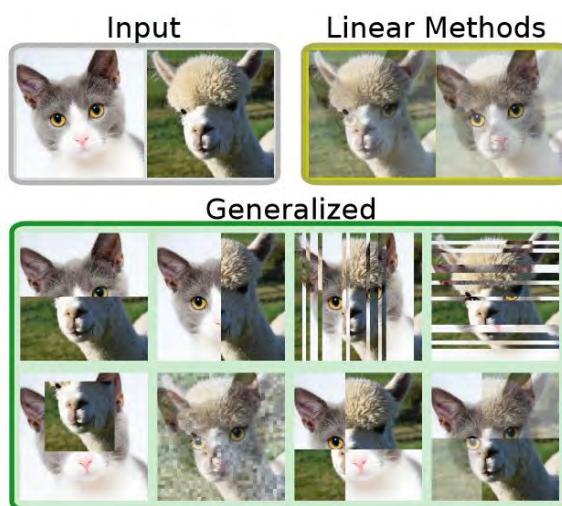


Рис.3.1.5: Аугментація міксуванням зображень

Випадкове стирання, або вирізання полягає у зміні деякої невеликої області зображення до стану, коли не зрозуміло що зображено на цій змінній області (Рис.3.1.6). Базові ознаки у такої області не можуть бути виділені саме через їх відсутність. Це дає змогу алгоритму виділяти признаки більш виразно із різних частин, так як у тих що немає ознак буде фактично оминатися. Такий підхід підвищує стійкість процесу до перенавчання [35].





Рис.3.1.6: Аугментація випадковим вирізанням

Усі методи простої аугментації є у більшому чи меншому прояві корисні при використанні за для розширення дата сету. Але деякі аугментації складно використовувати із конкретними завданнями. Наприклад задача розпізнавання рукописних цифр потребує зберігання певного нарису цифри. Використати перевертання, тобто відзеркалюючи наприклад цифру 2 буде не доречно — отриманий символ не буде відноситися до двійки. Аналогічно із обертотом на великі кути, у межах від 30-180 градусів обертання не має сенсу, семантика цифри губиться. Так само й з випадковим стиранням. Але зміщення у сторону або зашумлення буде достатньо ефективним методом аугментації, саме через те що такі перетворення є натуральними до цього типу даних. Тобто деякі геометричні перетворення не має сенсу застосовувати до даних, що прив'язані семантично до деяких геометрій. Ще одним прикладом є повздовжні та поперечні тріщини дефектів доріг. Звісно у деяких випадках коли у кадрі лише тріщина без будь-якої розмітки складно розпізнати тип дефекту. Але наприклад це можливо зрозуміти по розмітці, смугам накату, чи іншим ознакам. Це може стосуватися й фотометричних перетворень, якщо задача стоїть у розпізнаванні відтінків кольорів — не має сенсу змінювати кольоровий градієнт аугментацією.

Можливе комбінування простих аугментацій, комбінуючи обертання,

перевертання, обрізання та варіації цих методів із фотометричними можуть надати великої різноманітності нових даних. Але є межа коли велика кількість нових штучних даних почне ще більше призводити до перенавчання, тому важливо зберігати пропорції аугментованих даних до реальних.

Комплексні методи аугментації реалізуються складнішими перетвореннями, наприклад використанням алгоритмів із CNN. Розглянемо декілька найпопулярніших, а тобто генеративну змагальну мережу (Generative Adversarial Network) [36] та нейронне перенесення стилю (Neural Style Transfer) [37].

GAN це одна із технік змагального моделювання. Вона заснована на змаганні двох згорткових нейронних мереж: генератора та дискримінатора. Генератор це згорткова нейронна мережа що навчається без вчителя, на вхід якої подаються випадкові значення, наприклад гаусовий шум. Задача стоїть у визначенні таких параметрів генератора, щоб генерувати бажані зображення з поданого шуму, наприклад зображення квітки (Рис.3.1.7). Для того щоб корегувати параметри генератора використовується дискримінатор. Дискримінатор це CNN що класифікує зображення квітки як реальні та фейкові, тобто передбачає чи є достатньо схожим вхідне зображення на квітку. Якщо зображення згенероване генератором, що аналізується дискримінатором, не достатньо правдоподібне, параметри генератора змінюються. Через велику кількість ітерацій змінення генератора отримується такий генератор, що в змозі генерувати досить реальні зображення квітки, фактично “обманюючи” дискримінатор, бо вхідне зображення з генератора насправді є фейковим, згенерованим, тобто не справжнім. Також застосовується і зміна параметрів дискримінатора у разі коли він ідентифікував фейкове зображення генератора реальним. Дискримінатор навчається на реальних зображеннях квіток. Параметри двох мереж змінюються на основі стандартного алгоритму Backpropagation.

За допомогою GAN можливо отримувати штучні зображення будь-яких об’єктів, звісно у випадку якщо є достатня кількість даних на яких можливе тренування дискримінатора.

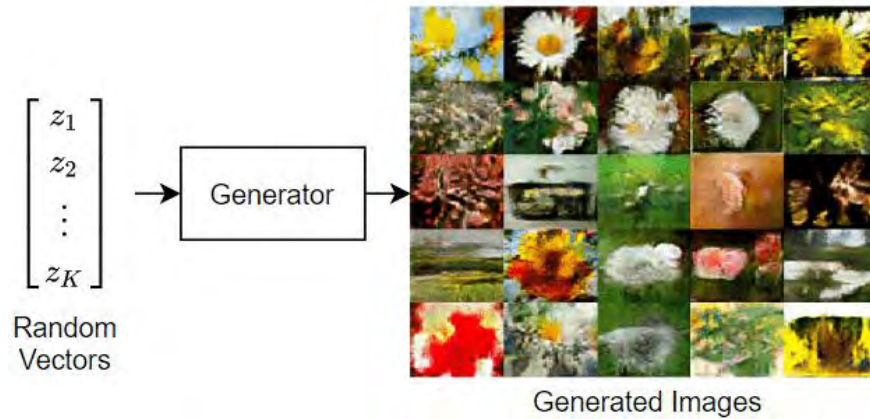


Рис.3.1.7: Ілюстрація роботи GAN

Нейронне перенесення стилю це задача надання одному зображенню стилю іншого (Рис.3.1.8). У загальному алгоритм полягає у взятті з двох зображень, основного і стилю, усіх матриць первинних ознак за допомогою будь-якої CNN. На основі об'єднання цих ознак математичним апаратом отримується функція втрат ознак основного та стильового зображення, їх сума дає функцію втрат що дає змогу генерувати з шумових зображень бажані об'єднанні стильові зображення.

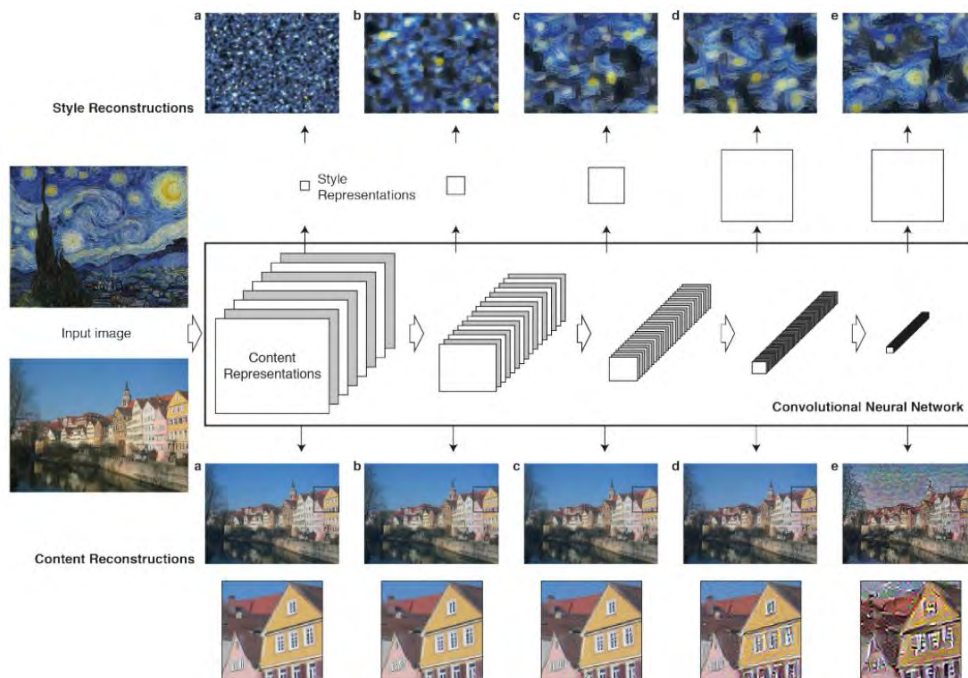


Рис.3.1.8: Ілюстрація принципу NST

Описані методи аугментацій різноманітні. Прості методи виділяються

простотою імплементації та реалізації у алгоритми. Комплексні ж є більш складним виходом з ситуації браку даних, це фактично ще одна підзадача із навчанням декількох CNN, що може бути дуже ефективним, але не гарантованим на успіх. Так само й прості методи не є гарантією підвищення точності класифікування CNN до високих меж, але їх реалізація набагато швидша та ефективніша порівняно із віддачею результативністю комплексних методів.

У цій роботі пропонується дослідити точність класифікації CNN на основі RoadDamageDetection2022 [42] дата сету з 3741 зображень та додаванням до нього у відсотковому відношенні аугментованих зображень, отриманими одними з трьох описаних методів.

Дата сет поділяється на п'ять класів (у дужках зазначені короткі умовні маркування класів):

- лінійна тріщина (DLinear);
- сітка тріщин (D20);
- вибоїна (D40).

Кожен клас має 1247 екземплярів зображень, що репрезентують дефекти.

Відсоткові відношення пропонується обрати у 10, 20 та 30 відсотків що до кількості зображень основного дата сету.

Виходячи з поставленої умови обирається наступні 3 прості геометричні аугментації:

- відзеркалення;
- обрізання;
- обертання.

Усі три методи є натуральною трансформацією із обраним типом даних. Такі методи є найбільш цікавими саме через простоту їх застосування.

### 3.2. Застосування методів аугментації для збільшення бази даних

Маючи 1247 зображень для кожного з класів потрібно сформувати дата сети із відсотковим вмістом зображень.

Усі перетворення зображень реалізовані за допомогою мови програмування Python та бібліотек PIL, Tensorflow, Numpy й інших допоміжних.

Відзеркалення пропонується обрати за горизонтальною та вертикальною осями (Рис.3.2.1).



Рис.3.2.1: Відзеркалення. а) D20, б) D40, в) DLinear

За рахунок різноманітного розташування дефектів на зображеннях є велика вірогідність що обрізання випадкової області не буде містити в собі бажаний дефект. Отже пропонується вирізати центральну частину зображення, що називається зумуванням. (Рис.3.2.2).



Рис.3.2.2: Зумування. а) D20, б) D40, в) DLinear

Обертання пропонується обрати випадковим у межах від -15 до 15 градусів (Рис.3.2.3).

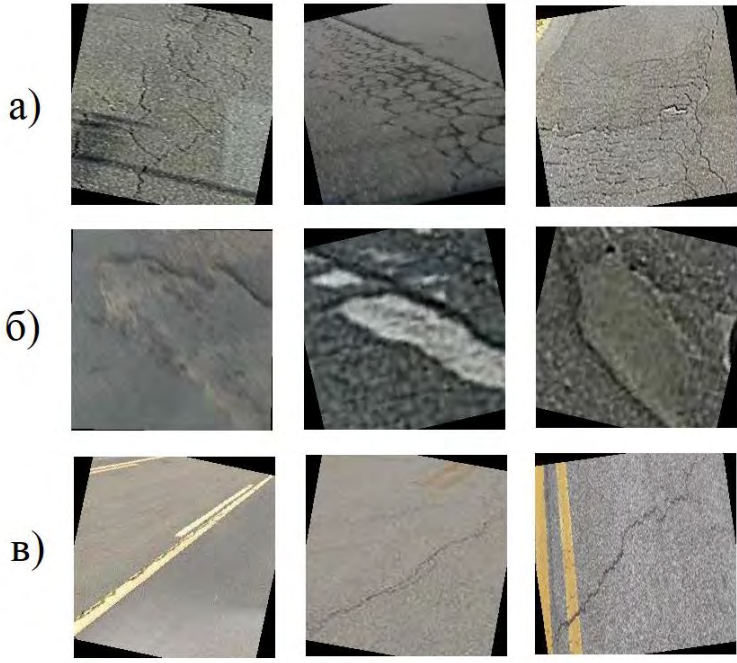


Рис.3.2.3: Обертання. а) D20, б) D40, в) DLinear

### 3.3. Вибір архітектури згорткової нейронної мережі

Вибір архітектури CNN для поставленої задачі може бути будь-який, але цікаво дослідити ту архітектуру, що найбільш вживана у відповідних задачах. З таких можливо виділити AlexNet, VGGNet, ResNet, MobileNet.

AlexNet [38] це одна з найперших стабільних та результативних згорткових нейронних мереж, вона складається з 5-ти згорткових шарів та 3-х повністю з'єднаних (Рис.3.3.1). Має у собі 58,3 мільйонів тренуваних параметрів та розрахована на вхід кольорового зображення розмірністю  $224 \times 224$ . Є простою реалізацією CNN що отримала нагороду у застосуванні до задачі конкурсу розпізнавання тисячі класів даних ImageNet у 2012 році, за що й стала популярною. Також була першою мережею із використанням активаційної функції ReLU, принципом випадкового обнулення навчальних параметрів Dropout, та підвищеною глибиною накладання згорткових шарів.

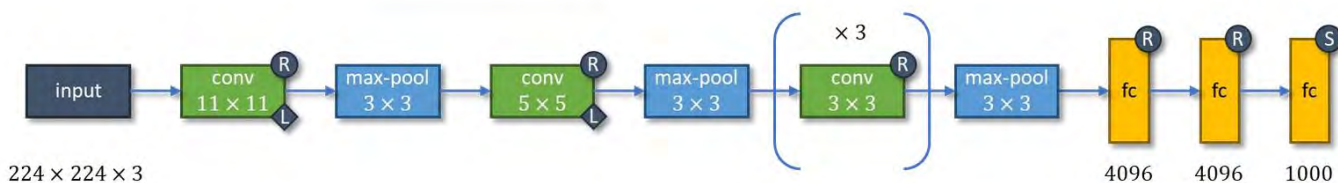


Рис.3.3.1: Архітектура AlexNet

VGGNet [39] є також достатньо простою мережею саме через лінійність шарів згортки. Але на відміну від AlexNet має більшу глибину за рахунок від 13 згорткових шарів (Рис.3.3.2). VGGNet є першою мережею на якій досліджували залежність глибини мережі на покращення ефективності та точності класифікації. Достатньо гарні результати дали архітектури з глибиною 16 та 19 згорткових шарів, отримавши друге місце у задачі ImageNet, і фактично підтвердили деякий поріг кореляції між глибиною та ефективністю, бо з підвищенням глибини мережам притаманно деградувати. На жаль мережа є досить повільною за рахунок великої кількості навальних параметрів — 143 мільйонів у VGG19, час тренування такої

мережі лічився у днях. Недоліком VGGNet є повільна швидкість та потреба у великій кількості даних. Але згодом навіть із такими недоліками стала основою більшості алгоритмів та систем AI.

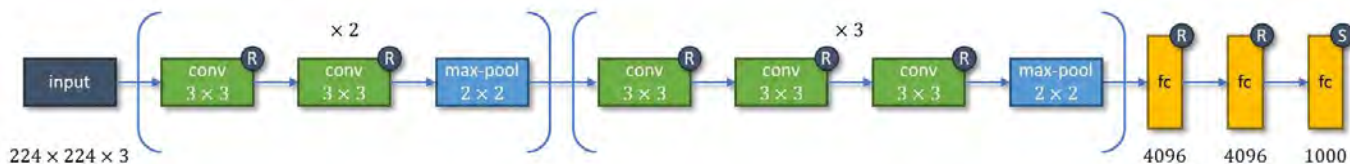


Рис.3.3.2: Архітектура VGGNet

ResNet [40] це CNN у якій вперше використовувався концепт залишкових блоків (Residual Blocks) для розбиття більш глибокої CNN. Принцип такого блоку полягає у проходженні вхідними даними через деяку кількість згорткових шарів, так само як і у попередньо розглянутих мережах, але результат цих шарів опісля додається до вхідних значень цього блоку. Розбиття великої лінійної мережі на малі частини дасть можливість цим частинам навчатися більш ефективно. Додавання частіше за всього реалізується згортковим фільтром  $1 \times 1$ , що фактично зрівнює розмірності матриць входу та виходу згорткових шарів. За допомогою накладання таких блоків можливо спростити обчислення параметрів навчання, даючи можливість алгоритму Backpropagation вираховувати градієнти для кожного блоку окремо, не залежачи від інших. Такий підхід дає змогу збільшити глибину мережі нелінійно. Стандартними реалізаціями ResNet є мережі із 50, 101 та 152 кількістю блоків (Рис.3.3.3). На відміну від VGG19 кількість параметрів такої мережі менша — 60,4 мільйони, але швидкість навчання помірно менша. Також ResNet використовує шари Batch normalisation що тільки посилює мережу від перенавчання.

На Рис.3.3.3 ілюстрований приклад архітектури ResNet. У даному випадку застосовуються два подібні блоки, де блоки ідентичності (Identity block) та згортки (Conv block) виконують одну й ту ж функцію, з відмінністю в тому що після кожної серії блоків ідентичності, що виділяють основні ознаки, застосовується блок згортки, який прирівнює розмірності матриць первинних ознак, даючи можливість їх



об'єднати.

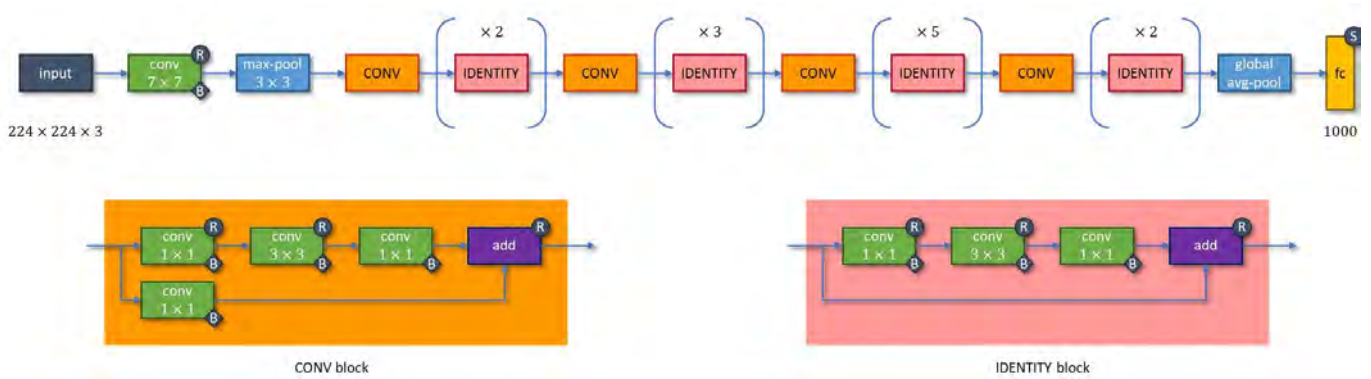


Рис.3.3.3: Архітектура ResNet

MobileNet [41] це CNN що створена за для використання на пристроях з малими обчислювальними потужностями. Принцип полягає у зменшенні кількостей операцій що здійснюється під час навчання. Це досягається зміною механізму згортки. Замість стандартного фільтру згортки застосовується дві більш прості: згортка по глибині (Depthwise convolution) та поелементна згортка (Pointwise convolution). За допомогою застосування цих двох згорток послідовно можливо відтворити роботу стандартного механізму згортки, значно зменшуючи кількість операцій у процесі навчання, що напряму залежить від потужністю обчислювання системи. Звісно за рахунок оптимізації втрачається й точність класифікації такої мережі. Але фактично можливо знизити кількість операцій у 10 разів порівняно із стандартним методом, та втратити лише 1-1.5% точності.

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
5× Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool $7 \times 7$	$7 \times 7 \times 1024$
FC / s1	$1024 \times 1000$	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Рис.3.3.3: Архітектура MobileNet

Існують варіації названих мереж, наприклад ResNetV2 та MobileNetV2, але принцип їх дії незмінний.

У цій роботі пропонується дослідити архітектуру MobileNet що має глибину у 50 шарів та 4,3 мільйони навчальних параметрів. Це є одна з найбільш простих, розповсюджених, стабільних та швидкодіючих згорткових нейронних мереж, тому й може бути цікавою у поставленій задачі, наприклад у системі MMS із використанням не дуже коштовного обладнання.

## Розділ 4. Експериментальна перевірка ефективності отриманих результатів

### 4.1. Застосування отриманих результатів для перевірки дієздатності вибраної

#### НМ

Усі операції та алгоритми реалізовані на фреймворку Keras, що оснований на пакетному забезпеченні TensorFlow.

У роботі використовується нейронна мережа MobileNet із застосуванням Transfer Learning.

Із CNN MobileNet використовується Feature extraction, параметри якого навчені на дата сету ImageNet. Ці параметри заморожуються, тобто вони не тренуються із алгоритмом навчання. Причому Classifier CNN MobileNet відкидається, бо він розрахований на 1000 класів дата сету ImageNet. Тому створюється новий класифікатор, що буде навчатися. Отже пропонується наступна архітектура нейронної мережі, представлена на Рис.4.1.1.

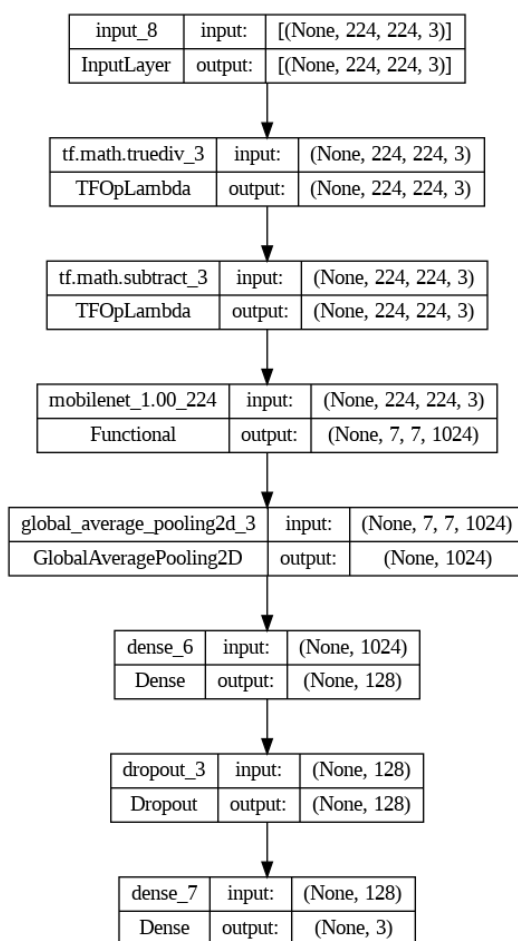


Рис.4.1.1: Імплементована архітектура MobileNet

У якості класифікатору використовуються шари:

- Global Average Pooling 2D;
- Dense зі 128 нейронами;
- Dropout із шансом 20%;
- Dense із 3 нейронами.

Global Average Pooling 2D – це алгоритм вирівнення значень з трьохмірної матриці у одновимірний масив, за для подальшої обробки для класифікатору, що є масивом нейронів.

Dense – це звичайний шар ANN, що складається з якоїсь кількості нейронів. У нашому випадку 128 та 3, останній — це кінцевий шар що передбачає клас даних.

Dropout – це підхід викидання деяких значень нейронів із певним шансом. Такий підхід зменшує перенавчання моделі, заважаючи ефекту переповненню параметрів (вагів) впливати на результати, фактично Dropout нормалізує значення параметрів вагів.

Важливими також є другий та третій шари з верху мережі на Рис.4.1.1. Другим є шар перемасштабування пікселів кольорових каналів. Так як усі значення пікселів ранжуються у межах 0-255, нам зручно зменшити ці значення, фактично нормалізуючи їх діленням на 255. У результаті виходять значення від нуля до одиниці, з такими числами алгоритму працювати легше. Третій шар віднімає певне значення від усіх значень пікселів, що трохи затьмарить зображення. Ці два шари є шарами препроцесингу зображення на вхід нейронної мережі, та передбачені імплементацією MobileNet. Розмір вхідного зображення 244x244x3.

За завданням сформовано 10 дата сетів. Один з них є базовим, тобто вхідним, з якого формувалися всі інші, він налічує для трьох класів по 1247 зображень, як наведено у Таблиці 4.1.1. Інші дев'ять є дата сети із додаванням аугментацій до вхідного, базового. Аугментовані дата сети описані у Таблиці 4.1.2.

Таблиці 4.1.1. Опис основного дата сету

<b>Основний дата сет</b>	
<b>Клас</b>	<b>Кількість зображень</b>
Алігатор, D20	1247
Вибоїна, D40	1247
Лінійна тріщина, DLinear	1247
Сума	3741

Таблиці 4.1.2. Опис аугментованих дата сетів.

<b>Кількість зображень</b>	<b>10%</b>	<b>20%</b>	<b>30%</b>
<b>Відзеркалення</b>	4116	4491	4866
<b>Зумування</b>	4116	4491	4866
<b>Обертання</b>	4116	4491	4866
<b>Кількість аугментованих зображень</b>	375	750	1125

Кожен дата сет формувався окремо, тобто аугментація є “офлайновою”. Перед аугментацією розміри зображення перемасштабовували до 244x244 пікселів.

Навчання на кожному із дата сетів проводилося п’ять разів за для забезпечення мінімальної вибірки точностей класифікування на різних вибірках. При навчанні використовувався розмір батчу у 128 зображень, у 6 епох. Дата сети ділилися на тренувальну та валідаційну підвбірки, за відношенням 80/20.

Також для перевірки точності моделі використовується метрика точності що розраховується за допомогою формуванням матриці невідповідностей (Confusion Matrix) (Рис.4.1.2). За допомогою матриці можливо розрахувати точності класифікації для кожного класу окремо, а також їх середнє значення, що дає фактичну, або загальну точність. Її будують на основі передбачень моделі, та співвставлення значень справжньої класифікації із передбачуваними:

$$Accuracy\ score = A_s = \frac{TP + TN}{TP + TN + FP + FN}$$

де TP (TruePositives) – істинно позитивні передбачення, TN (TrueNegatives) – істинно негативні передбачення, FP (FalsePositives) – хибно позитивне передбачення, FN (FalsePositives) – хибно істинне передбачення.

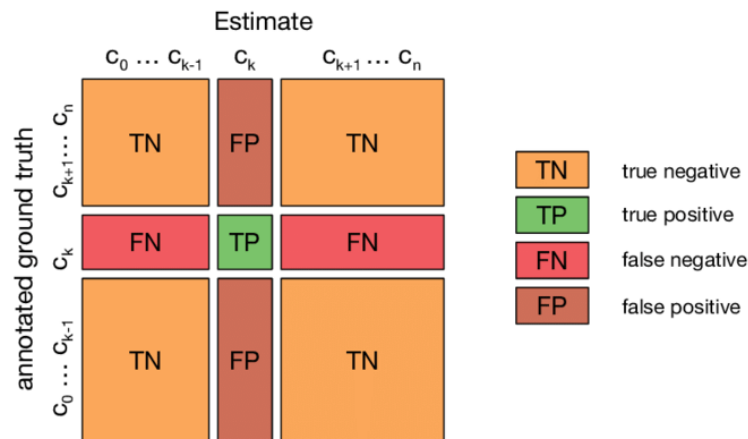


Рис.4.1.2: Confusion matrix

У цій роботі для кожного дата сету будується Confusion matrix на основі співставлення реальних класифікацій із передбачуваними основного дата сету.

У Таблиці 4.1.3 наведені результати навчання на всіх дата сетах по трьом метрикам: точність на тренувальній вибірці ( $P_t$ ), точність на валідаційній вибірці ( $P_v$ ), загальна оцінка точність класифікації за усіма класами за допомогою CM ( $A_s$ ).

Таблиці 4.1.3. Точності класифікацій

Тип	%	$\sum P_t$ 5,%	$\sum P_v$ 5,%	$\sum A_s$ 5,%
Базовий	0	94,684	95,180	94,790
Відзеркалення	10	<b>94,884</b>	94,582	<b>94,842</b>
	20	94,184	94,150	94,216
	30	94,558	<b>94,956</b>	94,474
Зумування	10	<b>94,964</b>	95,076	94,606

	20	94,717	94,587	<b>94,922</b>
	30	94,886	<b>94,866</b>	94,606
Обертання	10	94,48	<b>94,478</b>	94,544
	20	<b>94,338</b>	93,926	<b>94,804</b>
	30	94,222	94,352	94,646

## 4.2. Аналіз результатів дослідження

Із Таблиці 4.1.3. можна побачити що точності передбачення фактично не змінюються. Це може бути пов'язано із малою кількістю дослідів, кількості якої не достатньо для виявлення такого малого статистичного прояву відсоткових вмістів аугментацій та їх типів. Це свідчить про те, що малі вибірки аугментацій малодієві, тобто відсоток аугментованих зображень повинен бути як мінімум більше 30. Ефективність конкретних методів аугментацій яскраво не виражається.

Можлива проблема локальних мінімумів алгоритму градієнтного спуску, що має випадковий характер, через що споглядається такий розкид значень вірогідностей.

Важливою є перед тренувана модель CNN, а тобто її згортковий шар. Згортковий шар може містити занадто багато асоціацій відносно поданих типів об'єктів, тому й так швидко, не залежачи від додавання малої кількості даних, надає алгоритму швидкий приріс точності під час навчання, що спотворює бажані результати приростів.

## 4.3. Пропонована подальша робота.

У реалізації поставленої задачі можливо покращити деякі реалізації. Перше — це кількість аугментованих зображень. Їх кількість може напряму вплинути на генералізацію даних. Є також сенс дослідити й шкідливий вплив кількості аугментованих зображень. Тобто дослідити проміжок, встановивши порігові рівні ефективного та неефективного відсоткового відношення аугментацій.

Змінити дані на більш стійкі до аугментації, дефекти дорожніх доріг достатньо чутливі до деяких перетворень, а також таким чином збільшити кількість основних даних, співвіднести відношення на різних розмірах датасетів. Реальні дані не завжди оптимальні для дослідження характеристик чи поведінок алгоритмів ML та DL.

Комбінувати різні аугментації, отримуючи набагато більше інформації про синергію таких методів. Це може бути як корисним методом, так і шкідливий для загального алгоритму та даних.

Відмовитися від великих архітектур CNN, кількість їх параметрів занадто багато для такої малої кількості даних.

Відмовитися від підходу Transfer Learning, навчати невеликі моделі з нуля, наділяючи усю мережу первинними ознаками бажаного типу дефекту.



## Висновки

У роботі були застосовані деякі методи аугментації що дуже слабо вплинули на ефективність класифікації мережею трьох дефектів дорожнього покриття.

Запропонований метод підвищення ефективності нейронної мережі методами аугментаціями розширенням малої вибірки малим відсотковим відношенням аугментованих даних не є ефективним. Це пов'язано із кількістю аугментованих даних опісля додавання — цих даних занадто мало порівняно із вхідним дата сетом. Використання багато-параметрових CNN не є доцільним із малою вибіркою, це призводить до перенавчання, тому є сенс використовувати невеликі, із меншою кількістю шарів та параметрів загалом. Такі моделі дуже чутливі до кількості даних, тому можуть бути цікавим об'єктом досліджень у поставленій проблемі.

Є доцільнішим використання розширення дата сету що перевищують 30%, у випадку якщо використовуються багато методів аугментації до всього масиву розширення.

## Список використаної літератури

1. Строительство автомобильных дорог : учебник / коллектив авторов ; Ушаков В. В., Ольховикова В. М. – М., 2013. - 576 с.
2. Ремонт и содержание автомобильных дорог. Том II : справочная энциклопедия / коллектив авторов ; А.П. Васильев — М., 2004 — 1129с.
3. Distress Identification Manual for the Long-term Pavement Performance Project, (1993).
4. ASTM D6433-20, (2010). Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys.
5. ДСТУ 8954:2019. Автомобільні дороги. Оцінювання рівня дефектності дорожнього одягу.
6. Tao, C. V., & Li, J. (Eds.). (2007). *Advances in mobile mapping technology*. (Vol. 4). CRC Press. Burningham, S., & Stankevich, N. (2005).
7. Chiang, Kai-Wei & Tsai, Guang-Je & Zeng, Jih. (2021). Mobile Mapping Technologies. 10.1007/978-981-15-8983-6\_25.
8. Direct Sensor Orientation in Airborne and Land-based Mapping Applications. (2001).
9. K. -W. Chiang, G. -J. Tsai, Y. -H. Li and N. El-Sheimy, "Development of LiDAR-Based UAV System for Environment Reconstruction," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1790-1794, Oct. 2017, doi: 10.1109/LGRS.2017.2736013.
10. Giudici, Henri & Mocialov, Boris & Myklatun, A.. (2021). ViaPPS: A Mobile Pavement Profiling System.
11. Vung, Pham & Nguyen, Du & Donan, Christopher. (2022). Road Damages

Detection and Classification with YOLOv7. 10.48550/arXiv.2211.00091.

12. Muhammad Haroon Asad, Saran Khaliq, Muhammad Haroon Yousaf, Muhammad Obaid Ullah, Afaq Ahmad, "Pothole Detection Using Deep Learning: A Real-Time and AI-on-the-Edge Perspective", *Advances in Civil Engineering*, vol. 2022, Article ID 9221211, 13 pages, 2022. <https://doi.org/10.1155/2022/9221211>.

13. Maeda, Hiroya & Sekimoto, Yoshihide & Seto, Toshikazu & Kashiya, Takehiro & Omata, Hiroshi. (2018). Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images: Road damage detection and classification. *Computer-Aided Civil and Infrastructure Engineering*. 33. 10.1111/mice.12387.

14. Hongxia Li, Weixing Wang, Mengfei Wang, Limin Li, Vivian Vimlund, A review of deep learning methods for pixel-level crack detection, *Journal of Traffic and Transportation Engineering (English Edition)*, Volume 9, Issue 6, 2022, Pages 945-968, ISSN 2095-7564, <https://doi.org/10.1016/j.jtte.2022.11.003>.

15. Silva, Wilson & Schwerz de Lucena, Diogo. (2018). Concrete Cracks Detection Based on Deep Learning Image Classification. *Proceedings*. 2. 5387. 10.3390/ICEM18-05387.

16. Majidifard, Hamed & Adu-Gyamfi, Yaw & Buttlar, William. (2020). Pavement Image Datasets: A New Benchmark Dataset to Classify and Densify Pavement Distresses. *Transportation Research Record: Journal of the Transportation Research Board*. 2674. 10.1177/0361198120907283.

17. Girshick, Ross & Donahue, Jeff & Darrell, Trevor & Malik, Jitendra. (2013). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 10.1109/CVPR.2014.81.

18. Long, Jonathan & Shelhamer, Evan & Darrell, Trevor. (2015). Fully convolutional networks for semantic segmentation. 3431-3440.

10.1109/CVPR.2015.7298965.

19. Feng, C.; Zhang, H.; Wang, H.; Wang, S.; Li, Y. Automatic Pixel-Level Crack Detection on Dam Surface Using Deep Convolutional Network. *Sensors* 2020, 20, 2069. <https://doi.org/10.3390/s20072069>.

20. Kang, Dong H., and Young-Jin Cha. "Efficient attention-based deep encoder and decoder for automatic crack segmentation." *Structural Health Monitoring* 21.5 (2022): 2190-2205.

21. Qiao, W.; Liu, Q.; Wu, X.; Ma, B.; Li, G. Automatic Pixel-Level Pavement Crack Recognition Using a Deep Feature Aggregation Segmentation Network with a scSE Attention Mechanism Module. *Sensors* 2021, 21, 2902. <https://doi.org/10.3390/s21092902>.

22. Q. Yang and X. Ji, "Automatic Pixel-Level Crack Detection for Civil Infrastructure Using Unet++ and Deep Transfer Learning," in *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19165-19175, 1 Sept.1, 2021, doi: 10.1109/JSEN.2021.3089718.

23. H. Liu, X. Miao, C. Mertz, C. Xu and H. Kong, "CrackFormer: Transformer Network for Fine-Grained Crack Detection," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 3763-3772, doi: 10.1109/ICCV48922.2021.00376.

24. Di Benedetto, A.; Fiani, M.; Gujski, L.M. U-Net-Based CNN Architecture for Road Crack Segmentation. *Infrastructures* 2023, 8, 90. <https://doi.org/10.3390/infrastructures8050090>.

25. Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. 10.1007/978-3-319-24574-4\_28.

26. Simard, Patrice Y. et al. "Best practices for convolutional neural networks applied to visual document analysis." Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. (2003): 958-963.

27. Lazaric, Alessandro. (2011). Transfer in Reinforcement Learning: A

- Framework and a Survey. 10.1007/978-3-642-27645-3\_5.
28. Штучна нейронна мережа [Електронний ресурс] – Режим доступу до ресурсу: [https://uk.wikipedia.org/wiki/Штучна\\_нейронна\\_мережа](https://uk.wikipedia.org/wiki/Штучна_нейронна_мережа).
29. Chain rule [Електронний ресурс] – Режим доступу до ресурсу: [https://en.wikipedia.org/wiki/Chain\\_rule](https://en.wikipedia.org/wiki/Chain_rule).
30. Convolutional neural networks [Електронний ресурс] – Режим доступу до ресурсу: <http://cs231n.github.io/convolutional-networks/>.
31. Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* **6**, 60 (2019).  
<https://doi.org/10.1186/s40537-019-0197-0>.
32. Kang, Guoliang & Dong, Xuanyi & Zheng, Liang & Yang, Yi. (2017). PatchShuffle Regularization.
33. Taylor, Luke & Nitschke, Geoff. (2018). Improving Deep Learning with Generic Data Augmentation. 1542-1547. 10.1109/SSCI.2018.8628742.
34. C. Summers and M. J. Dinneen, "Improved Mixed-Example Data Augmentation," *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2019, pp. 1262-1270, doi: 10.1109/WACV.2019.00139. 10.1109/WACV.2019.00139.
35. Zhun Z, Liang Z, Guoliang K, Shaozi L, Yi Y. Random erasing data augmentation. ArXiv e-prints. 2017.
36. Ian JG, Jean PA, Mehdi M, Bing X, David WF, Sherjil O, Aaron C, Yoshua B. Generative adversarial nets. NIPS. 2014.
37. Leon AG, Alexander SE, Matthias B. A neural algorithm of artistic style. ArXiv. 2015.
38. Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E. Advances in Neural Information Processing Systems. Curran Associates, Inc. 2012
39. Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv e-prints. 2015.
40. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual

Learning for Image Recognition. ArXiv e-prints. 2015.

41. Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

42. Arya, Deeksha & Maeda, Hiroya & Ghosh, Sanjay & Toshniwal, Durga & Sekimoto, Yoshihide. (2022). RDD2022: A multi-national image dataset for automatic Road Damage Detection.

**Додатки**

**Додаток А**

код надається за запитом до авторів