Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему *Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання*

Галузь знань *12 – Інформаційні технології*
Шифр і назва галузі знань

Спеціальність *122 – Комп'ютерні науки*
Шифр і назва спеціальності

Освітня програма *Комп'ютерні науки*
Назва освітньої програми

Виконав:	*студент групи КНм-23-1*		*Олександр ЖАРНОВСЬКИЙ*
	Група виконавця	Підпис	Ім'я, ПРІЗВИЩЕ
Керівник:	*к.т.н., доц. каф. КН*		*Олександр МАЗУРЕЦЬ*
	Науковий ступінь, посада	Підпис	Ім'я, ПРІЗВИЩЕ
Нормоконтроль:	*к.т.н., доц. каф. КН*		*Руслан БАГРІЙ*
	Науковий ступінь, посада	Підпис	Ім'я, ПРІЗВИЩЕ

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор — Підпис — *Олександр БАРМАК* (Ім'я, ПРІЗВИЩЕ)
17 грудня 2024 р.

Хмельницький 2024

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет *інформаційних технологій*
Кафедра *комп'ютерних наук*
Освітній ступінь *магістр*
Галузь знань *12 – Інформаційні технології*
Спеціальність *122 – Комп'ютерні науки*

ЗАТВЕРДЖУЮ
Завідувач кафедри *комп'ютерних наук*

_____
(підпис)
д.т.н., професор Олександр БАРМАК
«*02*» *вересня* *2024* року

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА

1. Тема кваліфікаційної роботи магістра: *«Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання»*

2. Завдання видано студенту *Олександру ЖАРНОВСЬКОМУ*
(Ім'я, прізвище)

3. Керівник роботи *доцент кафедри КН Олександр МАЗУРЕЦЬ*
(посада, ім'я, прізвище)

4. Затверджені наказом університету від «*26*» *08* *2024* р. № *60*

5. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета кваліфікаційної роботи магістра – підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання, для чого слід розробити відповідний метод ідентифікації зображень, спроєктувати компоненти метода та архітектуру відповідної нейронної мережі, що використовує розроблений метод; після чого слід програмно реалізувати метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання у вигляді спроєктованої програмної системи й дослідити його ефективність.

Реферат

Кваліфікаційна робота магістра розв’язує науково-технічну задачу покращення автоматизованої ідентифікації зображень, згенерованих штучним інтелектом. Результатом роботи є метод, призначений для аналізу автентичності зображення людини завантаженого користувачем, що перетворює вхідні дані у вигляді зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані – відсоткова оцінка автентичності зображення та його походження, і програмна реалізація розробленого методу для предметної області ідентифікації зображень. З використанням розробленого методу досягається підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання, що визначає досягнення мети кваліфікаційної роботи магістра.

Актуальність теми. У сучасному світі спостерігається значне зростання використання генеративного штучного інтелекту, що пов’язано з активним розвитком технологій та доступності, простотою у використанні, швидкістю та продуктивністю.

Хоча інструменти генеративного штучного інтелекту дозволяють значно підвищити креативність, продуктивність чи частково автоматизувати види діяльності, вони також можуть бути використані у зловмисних цілях. Соціальні мережі є вразливими до сплеску використання генеративного штучного інтелекту для створення фальшивих зображень з метою маніпуляції публічної думки, монетизації та шахрайства, наклепу, підробки та інших зловживань штучним інтелектом для досягнення власних цілей. У зв'язку з цим виникає необхідність у розробці ефективних методів ідентифікації автентичності зображень.

Розроблений у кваліфікаційній роботі метод має ряд переваг у порівнянні з існуючими методами. Зокрема, він дозволяє ідентифікувати метод походження згенерованого зображення. Це дозволить краще аналізувати методи генерації штучного інтелекту для подальшого покращення ефективності.

Отже, магістерська кваліфікаційна робота володіє вагомою науковою і практичною значущістю. Її результати можуть слугувати основою для створення сучасних інформаційних систем та сервісів, зокрема орієнтованих на застосування у соціальних мережах.

Мета і задачі роботи. Метою кваліфікаційної роботи магістра є підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання. Для досягнення мети необхідно виконати наступне:

1. Дослідити сучасний стан предметної області генерацій зображень з використанням штучного інтелекту, їх методи та засоби. Виконати аналіз сучасних наукових публікацій у задачах генерації та виявлення зображень створених штучним інтелектом.

2. Розробити метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Розроблений метод має забезпечувати визначення автентичності зображення за допомогою відсоткової оцінки та визначення можливих методів використаних для генерації зображення з використанням навченої згорткової нейронної мережі.

3. Створити прикладну реалізацію методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

4. Дослідити практичну ефективність застосування методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Об’єкт дослідження – процес ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Предмет дослідження – моделі, методи та засоби ідентифікації згенерованих штучним інтелектом зображень.

Методи дослідження, що застосовані для вирішення поставлених завдань, наступні: основні положення методів аналізу даних й теорії множин, методології проектування інформаційних систем, об’єктно-орієнтований підхід.

Наукова новизна одержаних результатів. Результати виконання кваліфікаційної роботи магістра містять *інновації та наукову новизну*, зокрема було розроблено новий метод ідентифікації зображень, згенерованих штучним інтелектом, який дозволяє автоматизовано виконувати аналіз завантаженого користувачем зображення, виконуючи при цьому як і аналіз автентичності зображення, так і знаходячи можливі методи походження способу генерації засобами штучного інтелекту. Такий ефект досягається за перетворення вхідних даних у вигляді зображення, моделі для ідентифікації генерації зображення, моделі для знаходження походження генерації засобами ШІ у вихідні дані у вигляді відсоткової оцінки автентичності зображення та походження його генерації, якщо вона є.

Практичне значення одержаних результатів. Було створено інформаційну систему нейромережевого аналізу згенерованих зображень людей засобами машинного навчання, що є прикладною програмною реалізацією методу аналізу зображень людей згенерованих штучним інтелектом, із файлу завантаженим користувачем, що призначений для пошуку згенерованим штучним інтелектом зображень та вхідними даними має набір зображення, моделі для ідентифікації зображення, моделі для знаходження походження, що перетворює їх у вихідні дані у вигляді відсоткової оцінки автентичності зображення та його походження.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні наукові й практичні результати роботи доповідались у доповіді «Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning» на XLV Міжнародній науково-практичній конференції «Key Aspects of the Development of Scientific Research in Modern Conditions» (Constanta, Romania) 1 листопада 2024 року та у доповіді «Практична реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання» на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп’ютерних наук АПКН-2024»

(м. Хмельницький) 15-16 листопада 2024 року. За темою кваліфікаційної роботи автором виконано 4 наукові публікації:

1. Zharnovskyi O., Mazurets O., Sobko O. Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning. Key Aspects of the Development of Scientific Research in Modern Conditions. Proceedings of the XLV International scientific and practical conference. October 30 – November 1, 2024. Constanta, Romania. 2024. Pp. 69-73.

2. Жарновський О.В., Казмірчук Я.М., Собко О.В., Мазурець О.В. Практична реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 198-204.

3. Zharnovskyi O., Sobko O. Molchanova M. Neural Network Method for Detection of Fake Document Images for Personality Identification Systems. Black Sea Science 2024: Proceedings of the International Competition of Student Scientific Works. Odesa National University of Technology. Odesa, ONUT, 2024. Pp. 434-448.

4. Мазурець О.В., Жарновський О.В., Гладун О.В., Собко О.В. Нейромережеве виявлення фейкових зображень людей. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1 (Довідка з редакції).

Структура і обсяг роботи. Кваліфікаційна робота магістра складається з наступного: реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань з 50 найменувань й 5 додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 88 сторінок. В роботі наведено 63 зображення та 14 таблиць.

Ключові слова: штучні нейронні мережі, згорткові нейронні мережі, класифікація зображень, штучний інтелект, інформаційна система.

Зміст

Перелік скорочень

Скорочення, термін, позначення	Пояснення
МН	Машинне навчання
ІТ	Інформаційні технології
ПП	Програмний продукт
ERP	Планування ресурсів підприємства
CNN	Згорткова нейромережа
GAN	Генеративна змагальна мережа
UI	Користувацький інтерфейс
IDE	Інтегроване середовище розробки

Вступ

Кваліфікаційна робота магістра розв’язує науково-технічну задачу автоматизованої ідентифікації зображень, згенерованих штучним інтелектом. Результатом роботи є метод, призначений для аналізу автентичності зображення людини завантаженого користувачем, що перетворює вхідні дані у вигляді зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані – відсоткова оцінка автентичності зображення та його походження, і програмна реалізація розробленого методу для предметної області ідентифікації зображень. З використанням розробленого методу досягається підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання, що визначає досягнення мети кваліфікаційної роботи магістра.

Актуальність теми. У сучасному світі спостерігається значне зростання використання генеративного штучного інтелекту, що пов’язано з активним розвитком технологій та доступності, простотою у використанні, швидкістю та продуктивністю.

Хоча інструменти генеративного штучного інтелекту дозволяють значно підвищити креативність, продуктивність чи частково автоматизувати види діяльності, вони також можуть бути використані у зловмисних цілях. Соціальні мережі є вразливими до сплеску використання генеративного штучного інтелекту для створення дипфейків з метою маніпуляції публічної думки, монетизації та шахрайства, наклепу, підробки та інших зловживань штучним інтелектом для досягнення власних цілей. У зв'язку з цим виникає необхідність у розробці ефективних методів ідентифікації автентичності зображень.

Розроблений у кваліфікаційній роботі метод має ряд переваг у порівнянні з існуючими методами. Зокрема, він дозволяє ідентифікувати метод походження згенерованого зображення. Це дозволить краще аналізувати методи генерації штучного інтелекту для подальшого покращення ефективності.

Отже, магістерська кваліфікаційна робота володіє вагомою науковою і практичною значущістю. Її результати можуть слугувати основою для створення сучасних інформаційних систем та сервісів, зокрема орієнтованих на застосування у соціальних мережах.

Мета і задачі роботи. Метою кваліфікаційної роботи магістра є підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання. Для досягнення мети необхідно виконати наступне:

1. Дослідити сучасний стан предметної області генерацій зображень з використанням штучного інтелекту, їх методи та засоби. Виконати аналіз сучасних наукових публікацій у задачах генерації та виявлення зображень створених штучним інтелектом.

2. Розробити метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Розроблений метод має забезпечувати визначення автентичності зображення за допомогою відсоткової оцінки та визначення можливих методів використаних для генерації зображення з використанням навченої згорткової нейронної мережі.

3. Створити прикладну реалізацію методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

4. Дослідити практичну ефективність застосування методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Об’єкт дослідження – процес ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Предмет дослідження – моделі, методи та засоби ідентифікації згенерованих штучним інтелектом зображень.

Методи дослідження, що застосовані для вирішення поставлених завдань, наступні: основні положення методів аналізу даних й теорії множин, методології проектування інформаційних систем, об’єктно-орієнтований підхід.

Наукова новизна одержаних результатів. Результати виконання кваліфікаційної роботи магістра містять *інновації та наукову новизну*, зокрема було розроблено новий метод ідентифікації зображень, згенерованих штучним інтелектом, який дозволяє автоматизовано виконувати аналіз завантаженого користувачем зображення, виконуючи при цьому як і аналіз автентичності зображення, так і знаходячи можливі методи походження способу генерації засобами штучного інтелекту. Такий ефект досягається за перетворення вхідних даних у вигляді зображення, моделі для ідентифікації генерації зображення, моделі для знаходження походження генерації засобами ШІ у вихідні дані у вигляді відсоткової оцінки автентичності зображення та походження його генерації, якщо вона є.

Практичне значення одержаних результатів. Було створено інформаційну систему нейромережевого аналізу згенерованих зображень людей засобами машинного навчання, що є прикладною програмною реалізацією методу аналізу зображень людей згенерованих штучним інтелектом, із файлу завантаженим користувачем, що призначений для пошуку згенерованим штучним інтелектом зображень та вхідними даними має набір зображення, моделі для ідентифікації зображення, моделі для знаходження походження, що перетворює їх у вихідні дані у вигляді відсоткової оцінки автентичності зображення та його походження.

Інформаційна структура системи складається із набору зображень (датасетів) та кількох підсистем: «Підсистема взаємодії з НМ», «Підсистема розпізнавання завантажених зображень», «Підсистема інтерфейсу користувача», «Підсистема налаштувань», що дозволяють аналізувати завантажене зображення.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні наукові й практичні результати роботи доповідались на XLV Міжнародній науково-практичній конференції «Key Aspects of the Development of Scientific Research in Modern Conditions» (Constanta, Romania) 1 листопада

2024 року та на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп'ютерних наук АПКН-2024» (м. Хмельницький) 15-16 листопада 2024 року.

За темою кваліфікаційної роботи магістра автором виконано 4 наукові публікації:

1. Zharnovskyi O., Mazurets O., Sobko O. Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning. Key Aspects of the Development of Scientific Research in Modern Conditions. Proceedings of the XLV International scientific and practical conference. October 30 – November 1, 2024. Constanta, Romania. 2024. Pp. 69-73.

2. Жарновський О.В., Казмірчук Я.М., Собко О.В., Мазурець О.В. Практична реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 198-204.

3. Zharnovskyi O., Sobko O. Molchanova M. Neural Network Method for Detection of Fake Document Images for Personality Identification Systems. Black Sea Science 2024: Proceedings of the International Competition of Student Scientific Works. Odesa National University of Technology. Odesa, ONUT, 2024. Pp. 434-448.

4. Мазурець О.В., Жарновський О.В., Гладун О.В., Собко О.В. Нейромережеве виявлення фейкових зображень людей. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1 (Довідка з редакції).

Структура і обсяг роботи. Кваліфікаційна робота магістра складається з наступного: реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань з 50 найменувань й 5 додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 88 сторінок. В роботі наведено 63 зображення та 14 таблиць.

Розділ 1 Дослідження предметної області генерації зображень засобами штучного інтелекту

1.1 Особливості штучного створення зображень з використанням штучного інтелекту

Технологія генерації штучних зображень має широкий спектр застосувань, що робить її корисною в багатьох сферах людської діяльності таких як: арт та дизайн, гейм-дизайн, маркетинг, персоналізація реклами, медицина [1, 2].

Генерація зображень є корисним інструментом для художників і дизайнерів, які можуть застосовувати штучний інтелект для створення прикладів, ітерацій власних проєктів або використання згенерованих зображень як основи для подальшого редагування. Вона також дозволяє автоматизувати створення другорядних елементів, наприклад, деталей заднього фону в уже існуючих композиціях [3].

У сфері маркетингу та реклами штучний інтелект забезпечує можливість швидкого створення візуальних елементів. Зокрема, замість організації фотосесій для нового продукту, маркетологи можуть використовувати ШІ для генерації високоякісних зображень, придатних для застосування у рекламних матеріалах (рисунок 1.1).

Рисунок 1.1 – Обкладинка магазину Cosmopolitan, зроблена штучним інтелектом [3]

Також штучний інтелект може генерувати динамічну персоналізовану рекламу для користувачів базуючись на їх звичках. Amazon інтегрує ШІ в свій онлайн магазин, в особливості для сфери одягу. Компанія представили декілька функцій на базі штучного інтелекту, які допомагають клієнтам знаходити одяг що їм підходить. Штучний інтелект призваний вирішити поширену проблему повернення онлайн покупок, що виникає через те що користувачі схильні замовляти різні розміри чи кольори, повертаючи ті, що не відповідають їхнім очікуванням [4].

У галузі медицини штучній інтелект грає важливу роль у покращенні якості діагностичних зображень. Наприклад його можливо використовувати для створення більш чітких і детальних зображень тканин та органів, що допомагає в постановці діагнозів. Синтетичні дані згенеровані ШІ можуть прискорити розробку нових інструментів в сфері радіології. Також це може вирішити питання конфіденційності обміну даними між медичними установами.

Основними перевагами використання засобів штучного інтелекту для генерації зображень є [5,6]:

– простота – для генерації зображення достатньо знайти сайт з натренованою мережею та ввести текст-запит;

– швидкість – генерація зображення триває не більше кількох хвилин;

– розмірність зображення – мережі здатні генерувати зображення високого розширення та деталізації;

Недоліки та особливості використання ШІ для генерації зображень [7,8]:

– недоліки зображення;

– автентичність;

– технологічна комплексність;

– етичні дилеми.

Системам штучного інтелекту часто важко створювати зображення без недоліків, неможливість згенерувати бездоганні людські обличчя, частини тіла чи комплексні об’єкти (рисунок 1.2).

Рисунок 1.2 – Спроба DALL-E відтворити людські руки [3]

Хоча зображення, створеним штучним інтелектом можуть бути візуально вражаючими, вони бракують емоційної глибини та автентичності, якими володіють зображення створені художниками.

Якість зображень, створених за допомогою ШІ, безпосередньо залежить від обсягу, якості та доступності зображень, використаних для навчання моделі. До того ж, досягнення необхідного рівня деталізації потребує точного налаштування параметрів моделі, що є складним і трудомістким завданням, особливо в медичній сфері, де зображення мають відповідати високим стандартам точності.

Використання штучного інтелекту для генерації зображень також порушує питання авторських прав. Матеріали, залучені для навчання моделі, можуть бути захищені копірайтом, що потенційно спричиняє ризик юридичних суперечок щодо інтелектуальної власності. Окрім того, створені ШІ зображення можуть стати причиною проблем, пов’язаних із ринковою заміною оригінальних робіт.

Одним із головних недоліків вільного публічного використання ШІ для генерації зображень є створення дипфейків – через свою простоту та загальну доступність, а також якість зображення, генеративний ШІ може бути використаній в створенні зображень подій, що ніколи не мали місце для поширення дезінформації в соціальних мережах (рисунок 1.3).

Рисунок 1.3 – Згенерованим штучним інтелектом зображення Папи Римського в пуховій куртці [9]

Шахраї можуть використовувати такі зображення для поширення дезінформації з метою маніпуляції громадською думкою. Ба більше, поєднання технологій генерації зображень і аудіо дозволяє створювати реалістичні копії відомих осіб, що відкриває можливості для викрадення коштів або конфіденційної інформації у жертв [9,10].

Отже, в результаті аналізу предметної області були описані основні позитивні та негативні риси автоматичної генерації зображень з використанням штучного інтелекту, такі як простота, якість вихідного зображення, автентичність, системна комплексність та етичні проблеми, що актуалізує розробку методів для ідентифікації згенерованих зображень.

1.2 Методи та засоби генерації зображень штучним інтелектом

Перші спроби створення зображень за допомогою штучного інтелекту датуються 1970-ми роками, проте протягом наступних десятиліть прогрес у цій сфері залишався обмеженим. Це пояснювалося недостатньою обчислювальною

потужністю та примітивністю алгоритмів, які не могли обробляти реалістичні зображення.

Ситуація змінилася з появою глибокого навчання та згорткових нейронних мереж, які стали фундаментом для розробки генеративних змагальних мереж (Generative Adversarial Network, GAN). Ці технології значно підвищили якість і реалістичність створюваних зображень [11].

GAN складається з двох мереж:

– генератор – завдання мережі згенерувати дані що максимально схожі до реальних;

– дискримінатор – мережа вчиться розрізняти дані створені генератором від реальних (рисунок 1.4).

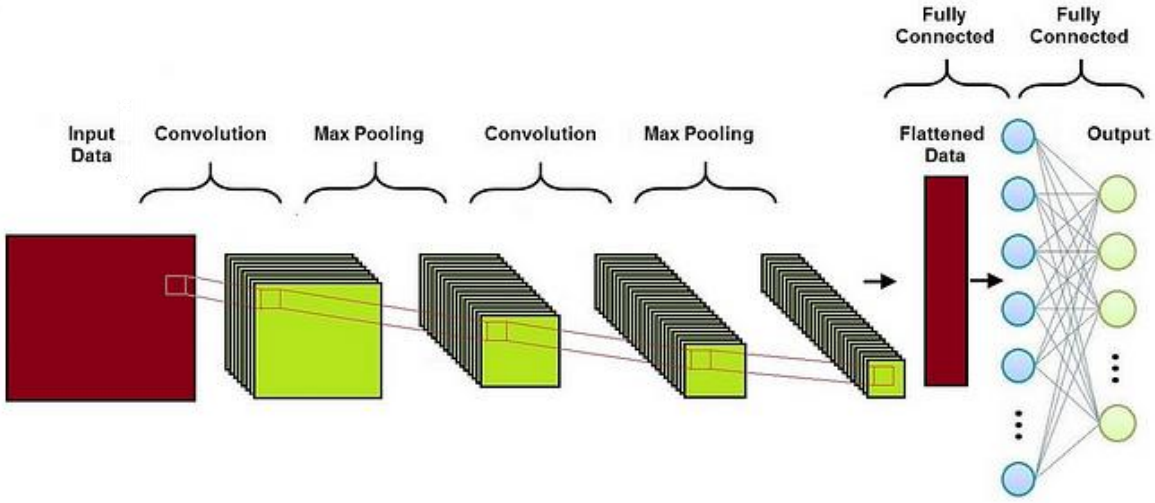


Рисунок 1.4 – Схематичне зображення мережі [11]

Процес навчання є ітеративним, де генератор намагається створити дані щоб обдурити дискримінатор, а дискримінатор покращує свою здатність розрізняти справжні та підробки (рисунок 1.5).

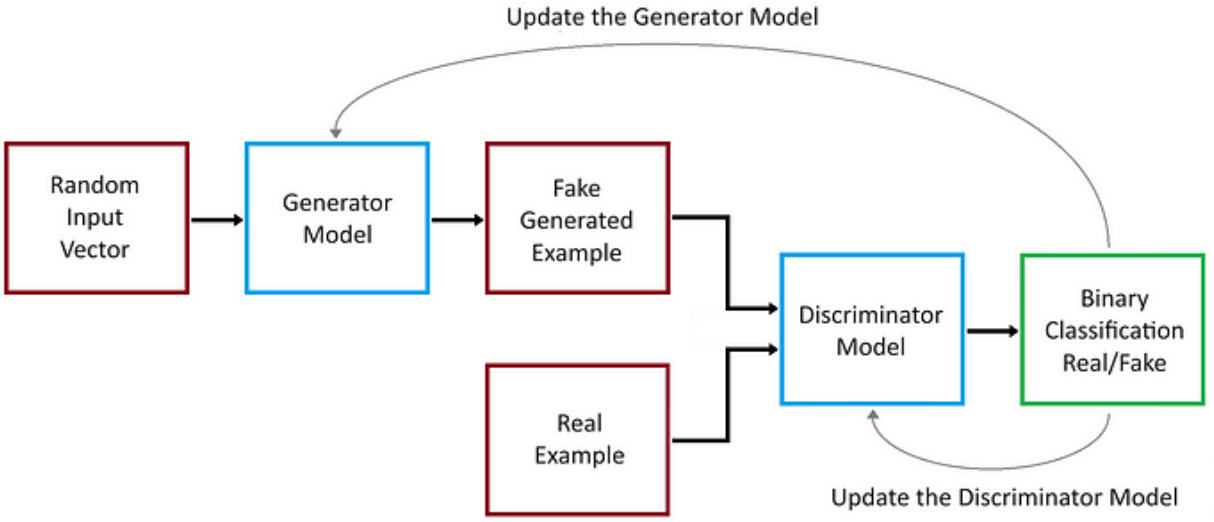


Рисунок 1.5 – Тренування GAN [11]

Процес генерації зображень базується на використанні різних типів вхідних даних, таких як RGB-зображення, відео, медичні дані або текст. На виході можуть бути отримані як статичні зображення, так і відеоматеріали.

Основними способами генерації зображень з використанням штучного інтелекту є [12]:

– Image-to-Image Translation;
– Sketch-to-image Generation;
– Text-to-Image Generation;
– Video Generation;
– Panoramic Image Generation.

Спосіб Зображення-Зображення (Image-to-Image Translation) – конверсія заданого зображення в цільове зображення зі збереженням ключових параметрів оригіналу (рисунок 1.6) [13,14].

Процес починається із встановлення цільових областей зображення, що задають тип вхідних даних для обробки системою. Мережу навчають на парних вхідних зображеннях. Далі комбінують генератор та дискримінатор використовуючи GAN. Генератор створює вихідне зображення базуючись на вхідному зображенні, в той же час дискримінатор визначає область реального та

згенерованого зображень. Функція втрат використовується для визначення різниці між вихідним та цільовим зображеннями.

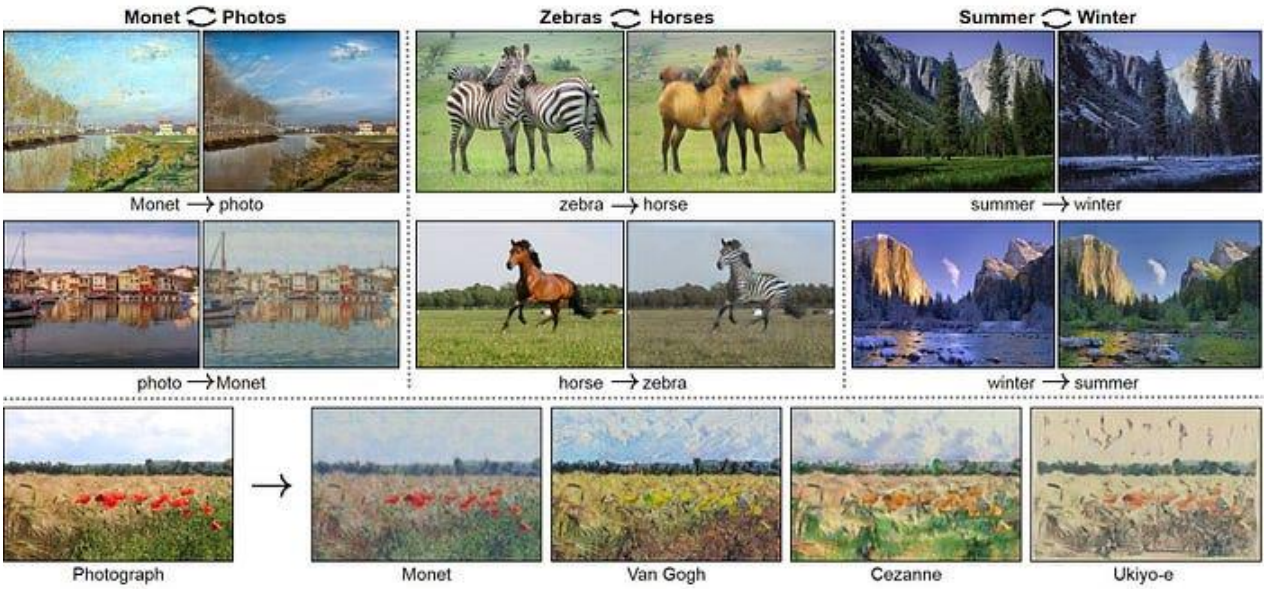


Рисунок 1.6 – Image-to-image translation використовуючи CycleGAN [13]

Спосіб Скетч-Зображення (Sketch-to-image Generation) – генерація зображення використовуючи скетч як вхідні дані (рисунок 1.7) [14,15].

Рисунок 1.7 – Sketch-to-image Generation [15]

Для тренування використовують комбінацію вхідного зображення та зображення конвертоване в скетч.

Спосіб Текст-Зображення (Text-to-Image Generation) – вхідний текст перетворюють в представлення, зрозуміле для машинної моделі за допомогою токенізації. Кожен «токен» конвертується у вектор що фіксує семантичне значення та контекстну інформацію. В свою чергу вхідний текст складає послідовність високовимірних векторів що використовується для генерації зображення [15, 16].

Тренування здійснюється за допомогою пари реального або невідповідного зображення і тексту (рисунок 1.8).

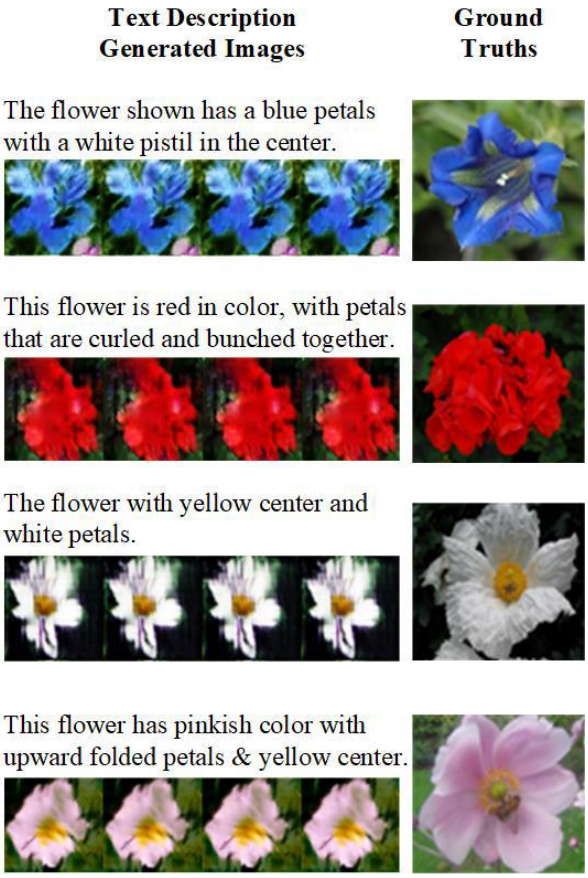


Рисунок 1.8 – Дані для тренування Text-to-Image моделі [17]

Спосіб генерації відео (Video Generation) – генерація здійснюється за допомогою токенізації вхідного тексту, генерації базового зображення та подальшої його ітерації для створення відео (рисунок 1.9) [18].

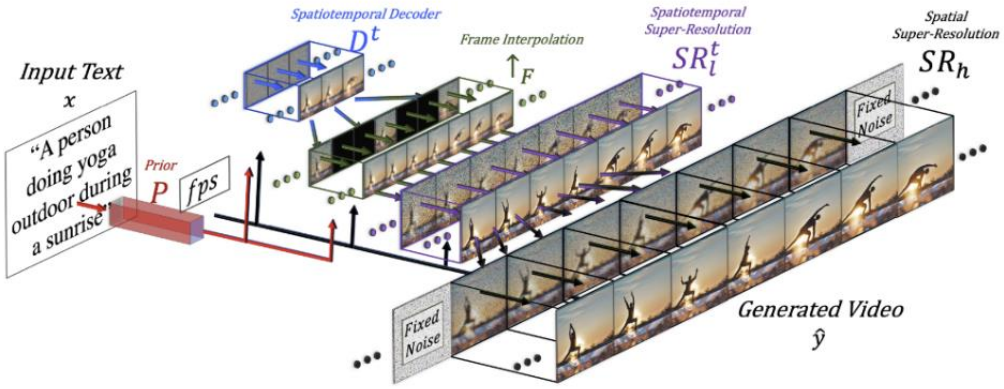


Рисунок 1.9 – Метод генерації відео [18]

Спосіб генерації панорами (Panoramic Image Generation) – вхідними даними може бути текст що слугує для генерації базового зображення, або комбінація базового зображення та тексту. Далі генерується декілька допоміжних зображень що потрібно скомбінувати із базовим та між собою (рисунок 1.10) [19].

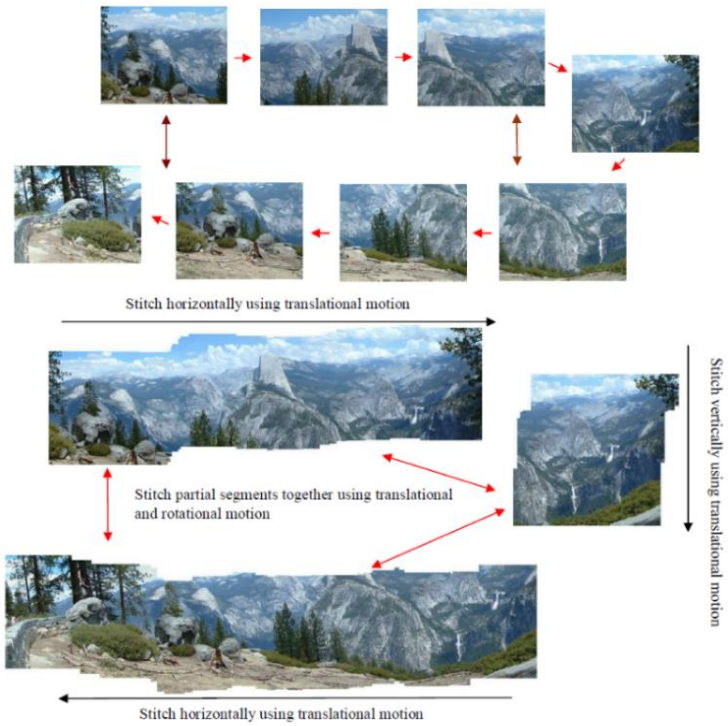


Рисунок 1.10 – Генерація зображення методом панорами [19]

Таким чином, були проаналізовані методи та засоби генерації зображень штучним інтелектом, а саме метод генерації зображення з використанням GAN та основні способи генерації зображень, а саме Image-to-Image Translation, Sketch-to-image Generation, to-Image Generation, Video Generation, Panoramic Image Generation.

1.3 Аналіз сучасних моделей генерацій зображень

На теперішній час існує величезна кількість пре-тренованих доступних моделей штучного інтелекту для генерації зображень.

DALLE-E 2 це один із найпопулярніших штучних інтелектів для створення зображень розроблений компанією OpenAI. Назва походить від злиття Dali – назви художника та WALL-E символізуючий штучний інтелект.

Модель включає в себе кодер та декодер, кожен з яких складається з кількох шарів нейронних мереж самоконтролю та прямого зв’язку. Завдяки високому рівню паралелізму, така мережа досягла передових результатів у багатьох завданнях обробки природної мови (рисунок 1.11).

Рисунок 1.11 – Зображення згенероване DALL-E «крісло у формі авокадо» [20]

CLIP є моделлю, що навчена на 400 мільйонах пар текстових заголовків і зображень, зібраних з Інтернету. Вона передбачає одночасне навчання на зображеннях і текстових даних, що дозволяє моделі зрозуміти зв'язок між двома модальностями – як текстові описи взаємодіють з візуальним змістом зображень. [21].

Одним із рішень з відкритим кодом, що використовувало CLIP, є DeepDaze, розроблений у січні 2021 року Філом Вангом. Ця модель поєднала CLIP із мережею неявного нейронного представлення під назвою Siren. DeepDaze здобула популярність завдяки здатності створювати вражаючі та сюрреалістичні зображення, часто нагадуючи фантастичні пейзажі або абстрактне мистецтво (рисунок 1.12).

Рисунок 1.12 – «Мандрівник часом» згенероване DeepDaze [21]

BigSleep – це ще одна генеративна модель глибокого навчання, розроблена тим самим дослідником, що використав моделі, опубліковані Раяном Мердоком. Модель поєднує CLIP з системою BigGAN, створеною дослідниками Google, яка використовує варіант архітектури GAN для генерації зображень високої роздільної здатності з випадкових векторів шуму. BigSleep використовує результати BigGAN для пошуку зображень, які отримують високі оцінки за

CLIP. Далі модель поступово коригує вхідний шум у генераторі BigGAN, поки створене зображення не відповідатиме заданому текстовому запиту (рисунок 1.13).

Рисунок 1.13 – «Повітряні кулі над руїнами міста» згенероване BigSleep [21]

Моделі Diffusion – у червні 2021 року автор VQGAN-CLIP представив нову роботу, поєднавши модель Contrastive Language-Image з алгоритмом дифузії для створення CLIP-керованої дифузії.

Алгоритми дифузії – це методи генерації зображень, які базуються на моделюванні поведінки частинок, що дифундують у середовищі. Зображення перетворюються на рівномірний розподіл через додавання випадкового шуму. Алгоритм дифузії поступово руйнує структуру зображення, застосовуючи шум, поки не залишається нічого, окрім випадкового шуму.

В свою чергу завдання моделей штучного інтелекту навчання алгоритму зворотної дифузії – знаходження параметрів що дозволять перетворити шум в зображення (рисунок 1.14).

Рисунок 1.14 – «Собака-робот,4к з високою деталізацією» згенероване Stable Diffusion [16]

Таким чином, був проведений аналіз сучасних моделей автоматичної генерації зображень з використанням штучного інтелекту, їх компоненти та структура.

1.4 Аналіз наукових публікацій із проблеми виявлення згенерованих штучним інтелектом зображень

Розгляд та аналіз уже існуючих рішень та статей на тему ідентифікації згенерованих штучним інтелектом зображень допоможе встановити труднощі у розробці методу та прикладної реалізації.

Автори публікації Identifying AI-Generated Art with Deep Learning [22] розглядають проблему ідентифікації згенерованих зображень з точки зору презервації людських художніх робіт та підтримки автентичності.

Метод, описаний в статті використовує кілька моделей згорткових нейронних мереж. VGG-19, обрану через свою здатність до отримання ієрархічних характеристик із вхідного зображення. Ця характеристика дозволяє мережі вловлювати комплексні шаблони. В свою чергу сама мережа характеризується однорідною структурою та простотою. ResNet-50 презентує концепт «залишкового навчання», що включає пропуск-з’єднання, які сприяють

більш плавному перебігу інформації по всій мережі та вирішують проблеми зникнення градієнта. ViT характеризується своєю винятковою продуктивністю у завданнях класифікації зображень, а також своєю масштабованістю.

Автори створили власний датасет що складається із комбінації датасетів ArtGraph, що містить ~110 тисяч зображень, розподілених по 32 стилях та 18 жанрах і Artifact, що складається з більше двох мільйонів зображень, мільйон з яких реальні на різну тематику – людські обличчя, тварини, машини та мистецтво. Отриманий датасет був розділений на 80% для тренування, та по 10% для тестування і валідації.

В результаті експериментів з використанням платформи Google Collab на базі PyTorch автори зазначають найкращі параметри мережі ViT: accuracy у 0.9758, precision 0.9752, recall 0.9759, F1 0.9755.

Автори статті Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks [23] розглядають по-піксельний аналіз зображення з використанням технологій Photo Response Non-Uniformity та ErrorLevelAnalysis.

Розроблений авторами метод використовує власну архітектуру згорткової нейронної мережі (рисунок 1.15) та згенеровані авторами зображення із використанням декількох мереж штучного інтелекту – Dall E, Stable Diffusion, OpenArt. Реальні зображення були взяті із різних джерел, таких як VISION датасет.

Зазначається що отриманий результат в 95% для PRNU та 98% точності для ELA при тренування в 100 епох. При цьому тренування для RNU було швидшим в 109 хвилин проти 167 для ELA.

Інша стаття [24], автор якої розглядає методи покращення мережі MobileNet-v2 для задач по класифікації згенерованих штучним інтелектом зображень обличь людей. Автор використав датасети Flickr-Faces-HQ для реальних зображень та 1M AI generated images 128x128 для згенерованих.

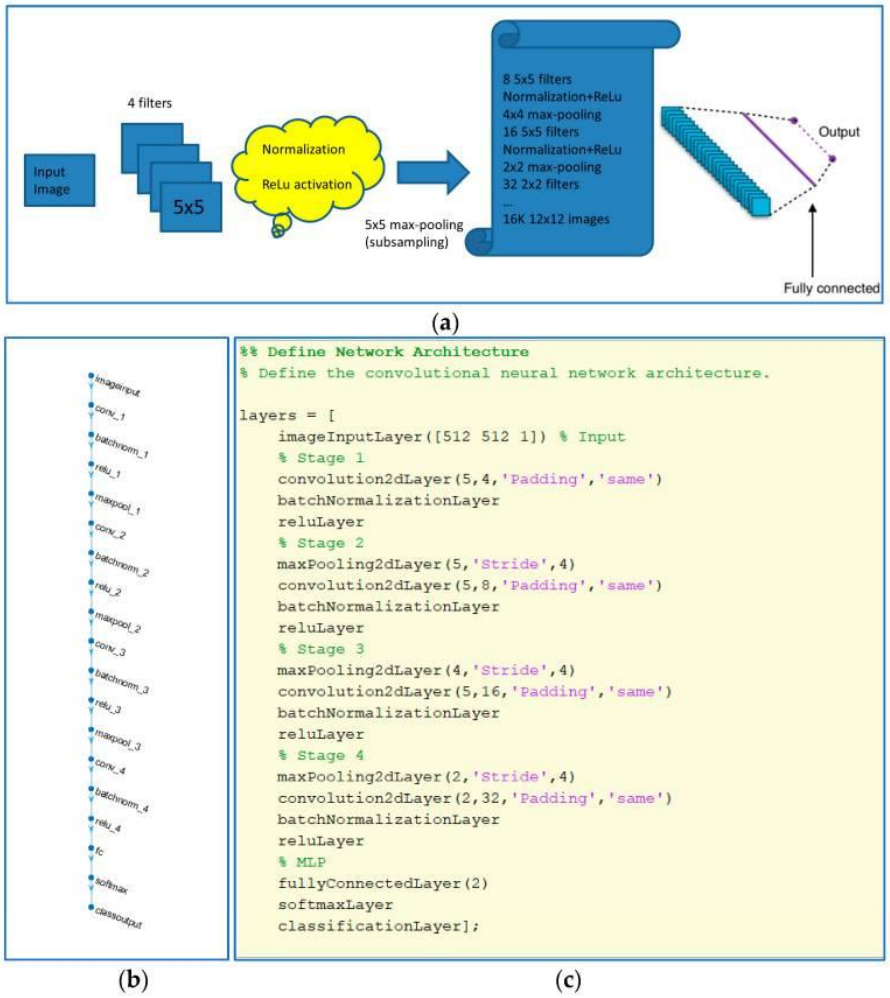


Рисунок 1.15 – Розроблена авторами мережа [23]

Для початкового тестування автор обрав 1000 зображень із датасету реальних зображень та 1422 із згенерованих із розподілом тренування-тестування в 0.2. Таким чином автор отримав 1932 зображення для тренування та 484 для валідації.

В результаті тренування 10 епох були отримані результати в 0.72% точності при тренування та 0.73% при валідації. Було зазначено що модель має труднощі із зображеннями людей зо містять аксесуари, такі як окуляри, чи погане світло.

Щоб покращити результат автор замінив Avarage Pooling Layer на Max Pooling Layer, а також додав випадкові трансформації яскравості, контрасту, відтінку та насиченості до використаних датасетів, а також гаусівський шум з параметрами 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5 та 1. Автор зазначає

що із всіх використаних параметрів тільки 0.005, 0.1 та 0.5 дали універсальне покращення точності при валідації, в свою чергу 0.0005, 0.01 та 0.05 погіршило результати. Як результат тестування найкраща отримана точність була 0.76% з параметрами тренування мережі в 15 епох, без використання шару втрат та параметром шуму в 0.1.

Як підсумок, були проаналізовані наукові публікації в сфері виявлення згенерованих штучним інтелектом зображень, використані авторами методи та отримані метрики, що потрібно враховувати під час реалізації програмного застосунку.

1.5 Мета, задачі та вимоги до реалізації інформаційної системи

Метою кваліфікаційної роботи магістра є підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання. Для досягнення мети слід вирішити наступні завдання:

1. Дослідити сучасний стан предметної області генерацій зображень з використанням штучного інтелекту, їх методи та засоби.
2. Виконати аналіз сучасних наукових публікацій у задачах генерації та виявлення зображень створених штучним інтелектом.
3. Розробити метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Розроблений метод має забезпечувати визначення автентичності зображення за допомогою відсоткової оцінки та визначення можливих методів використаних для генерації зображення з використанням навченої згорткової нейронної мережі.
4. Створити прикладну реалізацію методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.
5. Дослідити практичну ефективність застосування методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Розділ 2 Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

2.1 Схема та кроки методу ідентифікації згенерованих штучним інтелектом зображень людей

Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання передбачає розробку нейронної мережі, здатної розпізнавати та класифікувати зображення.

Для цього найефективнішими є згорткові нейронні мережі – тип глибоких нейронних мереж, який широко застосовується для аналізу зображень, аудіо та відео. CNN складається з багатошарових перцептронів, спеціально спроектованих для мінімізації потреби у попередній обробці даних.

Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання складається з наступних елементів (рисунок 2.1).

Вхідними даними методу є:

- файл зображення (png, jpg, webp, tiff, bmp);
- файл моделі для ідентифікації зображення;
- файл моделі для знаходження походження.

На першому кроці вхідне зображення потрібно трансформувати, а саме змінити розмір та перетворити у тензор. Далі завантажити натреновану мережу для ідентифікації зображення як згенерованого мережею чи реального, та проаналізувати завантажене зображення. В результаті отримуються відсоткові значення що потрібно проаналізувати та вивести. В залежності від того чи зображення реальне чи згенероване, потрібно завантажити мережу для ідентифікації походження, для випадку коли вхідне зображення згенероване, та проаналізувати і вивести результат.

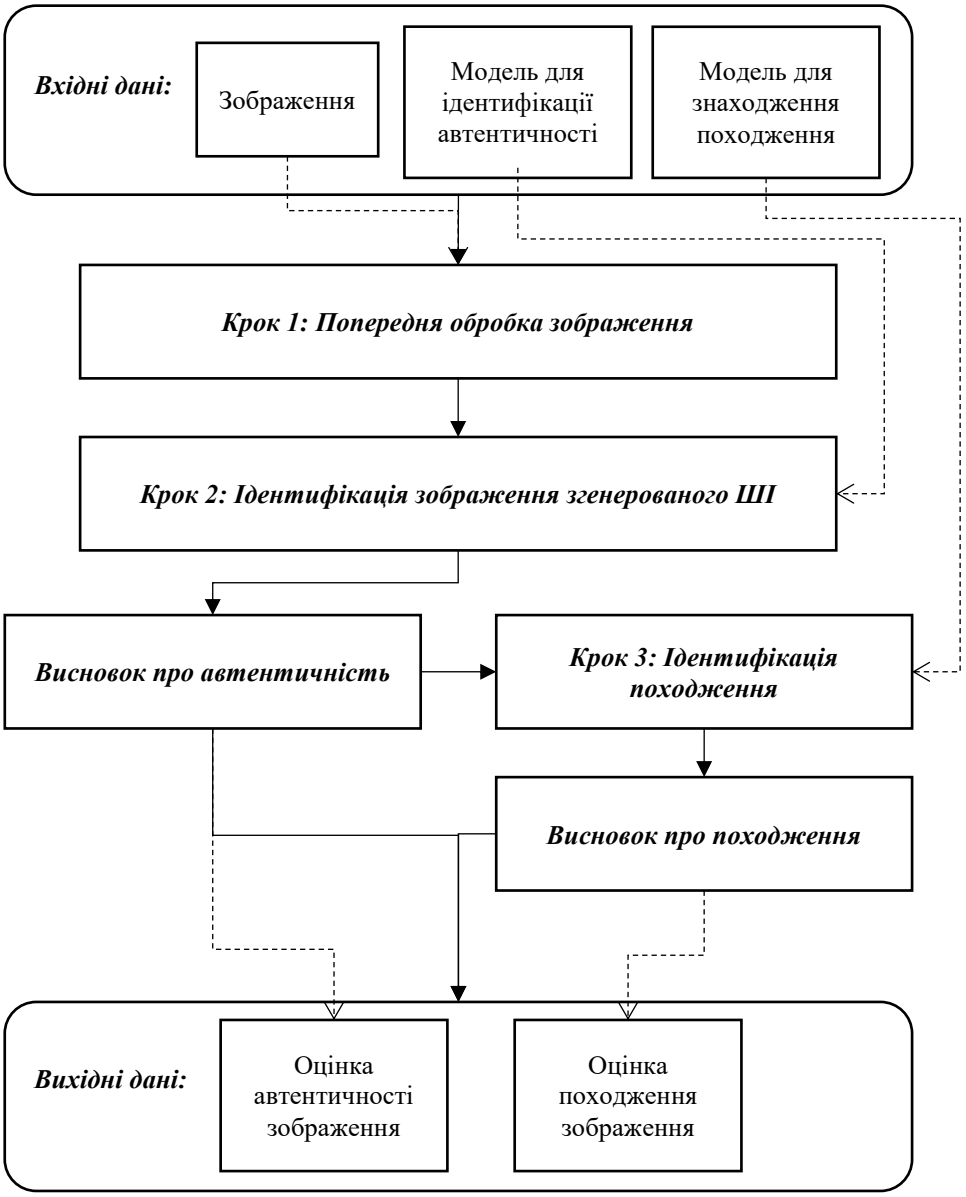


Рисунок 2.1 – Схема методу ідентифікації згенерованих штучним інтелектом зображень людей

Функціональна складова буде поділена на дві частини: для взаємодії з користувачем і для роботи нейронних мереж, а саме imageClassifier відповідає за ідентифікацію зображення як реального чи підробленого, а methodClassifier визначає відповідність зображення популярним моделям ШІ для його генерації. (рисунок 2.2).

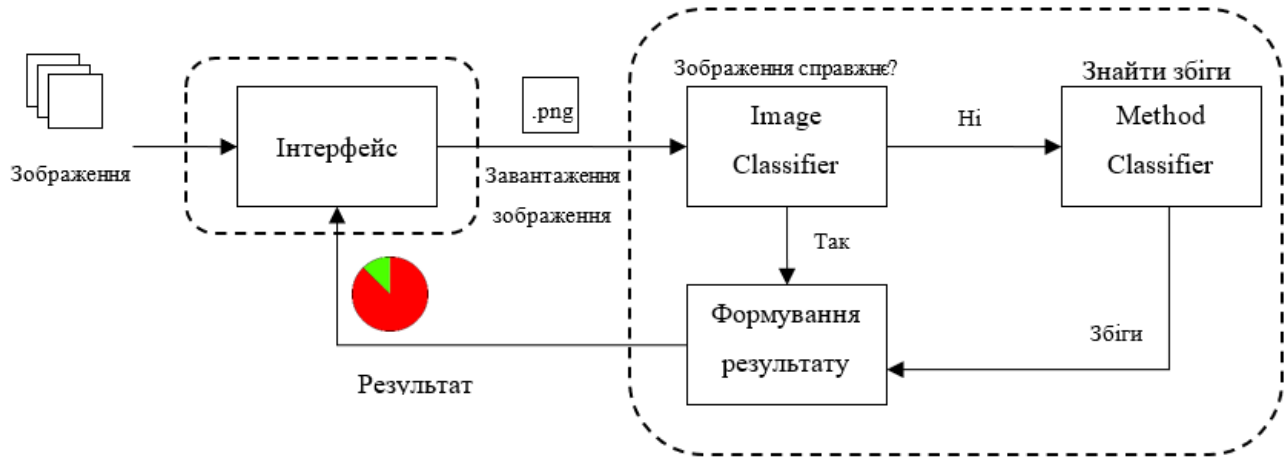


Рисунок 2.2 – Функціональні складові методу

Отже, був створений метод для ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання використовуючи комбінації двох згорткових нейронних мереж, що працює на основі перетворення вхідних даних – зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані – відсоткова оцінка автентичності зображення та його походження.

2.2 Формування датасету для ідентифікації штучно згенерованих зображень людей

Найважливішим етапом розробки мережі є підбір чи створення необхідних даних для навчання мережі, в залежності від обраних даних мережа може мати кардинально різний результат.

Датасети зазвичай складаються із трьох основних груп (рисунок 2.3) [25].

Рисунок 2.3 – Загальний поділ датасету

Навчальний набір – це частина набору даних для навчання моделі відповідності даних до класів, слугує для безпосереднього покращення параметрів. Щоб отримати добре навчену мережу навчальний набір має бути достатньо великим, щоб дати значимі результати, але не занадто великим, для запобігання перетренування.

Перетренування – модель машинного навчання є занадто спеціалізованою до навчальних даних, що погіршує здатність до узагальнення та робити правильні прогнози на нових даних. В такому випадку модель буде мати занадто хороші результати із набором для навчання, але занадто малі коли представлені інші дані.

Валідаційний набір – частина що використовується для оцінки моделі машинного навчання, слугує для оцінки продуктивності моделі та внесення корегувань.

Тестовий набір – частина яка потрібна для кінцевої оцінки продуктивності навченої моделі. Тестовий набір зберігають під час всього процесу навчання та слугує показником наскільки добре модель узагальнює дані в реальних умовах.

Оптимальні параметри розподілу датасету визначають емпірично, змінюючи параметри по необхідності.

Для навчання нейромережі по ідентифікації штучно згенерованих зображень були взяті наступні датасети:

– Flickr-Faces-HQ Dataset (Nvidia) [26].

– 1 Million Fake Faces [27];

Flickr-Faces-HQ Dataset (Nvidia) складається із 70000 реальних зображень 1024x1024 обличь людей, з різними варіаціями по віку, фону та аксесуарах (рисунок 2.4).

Рисунок 2.4 – Зразок зображень Flickr-Faces-HQ Dataset (Nvidia)

Із завантаженого датасету обрано 5000 зображень для навчання мережі, 1000 для валідації та 1000 для тестування класу «реальні».

1 Million Fake Faces складається з ~1 мільйону згенерованих зображень 1024x1024 обличь чоловік та жінок з використанням Nvidia StyleGAN (рисунок 2.5).

Рисунок 2.5 – Зразок зображень датасету 1M AI generated faces v.1.4

Використовувати весь датасет не є доцільним, тому обрано 5000 зображень для навчання мережі, 1000 для валідації та 1000 для тестування, і помістити їх у клас «згенеровані».

Як результат, отримуємо датасет із 10000 зображень для тренування, 2000 для валідації та 2000 для тестування, з двома класами та співвідношенням 1:1 (рисунок 2.6).

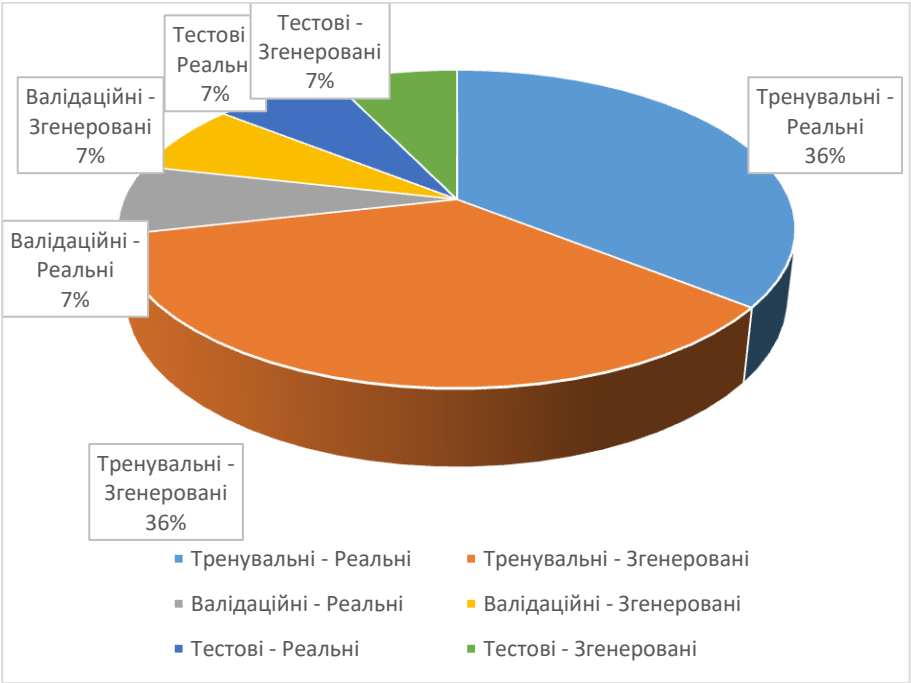


Рисунок 2.6 – Отриманий датасет

Для навчання мережі по виявленню походження штучно генерованих зображень було обрано додаткові датасети:

– 1 Million Fake Faces [26];

– Face Dataset Using Stable Diffusion v.1.4 [28];

– Real vs fake faces [29];

Із вище описаного датасету 1 Million Fake Faces було обрано 800 згенерованих зображень для тренування та по 100 для валідації і тестування для класу «StyleGan».

Face Dataset Using Stable Diffusion v.1.4 складається з ~2000 згенерованих зображень 512x512 обличь чоловік та жінок з використанням Stable Diffusion та Azure VM (рисунок 2.7).

Рисунок 2.7 – Зразок чоловічого та жіночого зображень датасету Face Dataset Using Stable Diffusion v.1.4

Датасет розділений на тренувальний та валідаційний сет по два класи – чоловічі та жіночі обличчя (рисунок 2.8). Для створеного методу було обрано 800 зображень для тренування та по 100 для валідації і тестування, що належать до класу «StableDiffusion».

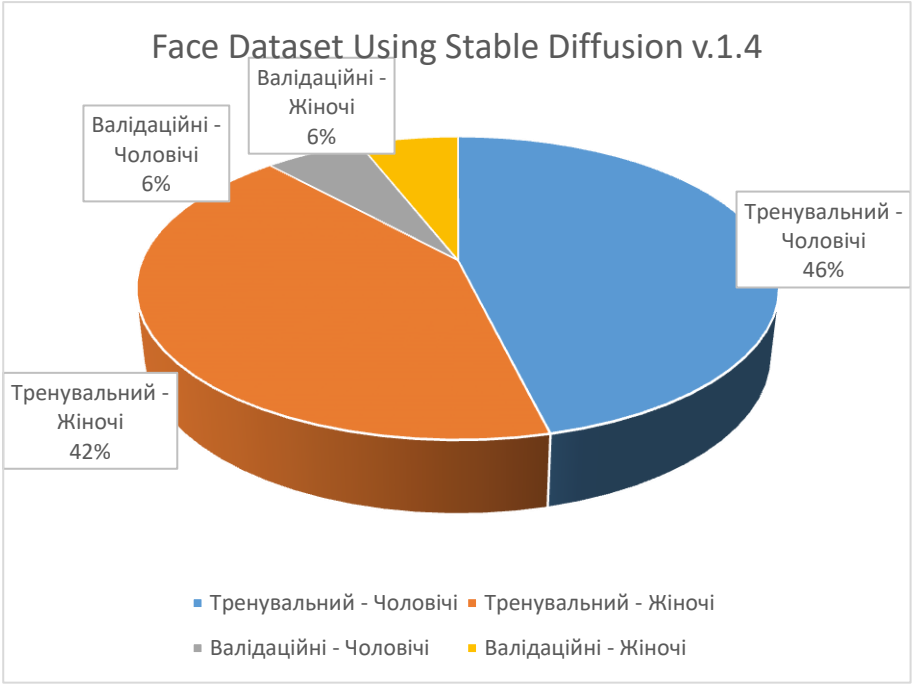


Рисунок 2.8 – Розподіл датасету

Real vs fake faces містить ~1000 згенерованих та реальних зображень зображень. Згенеровані зображення були отримані за допомогою Photoshop Experts (рисунок 2.9).

Рисунок 2.9 – Зразок згенерованих зображень датасету Real vs fake faces за допомогою Photoshop Experts

Датасет розділений відповідні класи «згенеровані» та «реальні» (рисунок 2.10). Для створеного методу були взяті згенеровані зображення 800 для тренування та по 200 для валідації і тестування, що належать до класу «PhotoshopExperts».

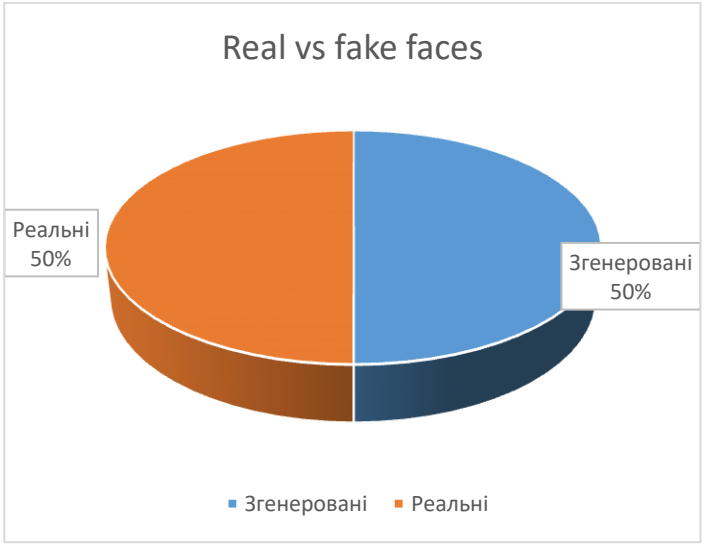


Рисунок 2.10 – Розподіл датасету

Таким чином був сформований датасет для ідентифікації штучно згенерованих зображень людей та виявлення їх походження, що складається із 10000 тренувальних, 2000 валідаційних та 2000 тестувальних зображень розподілених на два класи для мережі по ідентифікації зображень, і 2400 тренувальних, 300 валідаційних та 300 тестувальних зображень розподілених на три класи для мережі по ідентифікації походження.

2.3 Архітектура нейромережі для ідентифікації штучно згенерованих зображень людей

Одним із ключових етапів створення штучної нейромережі є розробка архітектури.

згорткова нейронна мережа має вхідний шар, шар згортки, шар активації, пул-шар, повністю-з’єднаний шар, шар нормалізації, шар відсіювання та вихідний шар.

Вхідний – шар на який надходить зображення, має аналогічну розмірність із вхідним зображенням.

Шар згортки – складається з набору K фільтрів, також відомих як кернел (kernel), із заданими параметрами висоти та ширини,. Використовуються для агрегації зображення в формат карти активації за допомогою множення значень фільтрів та оригінального зображення (рисунок 2.11), та має такі параметри як:

– глибина, відповідає за кількість каналів даних, RGB зображення має 3 канали;

– розмір, висота та ширина матриці що зазвичай є квадратною та відносно малою (3x3,5x5,7x7);

– крок, використовується для виміру кількості пікселів через які «переступає» матриця;

– нульове заповнення, тобто використання нулів по зовнішніх межах матриці зображення для збереження оригінального розміру на виході, що може бути важливим при використанні декількох шарів згортки.

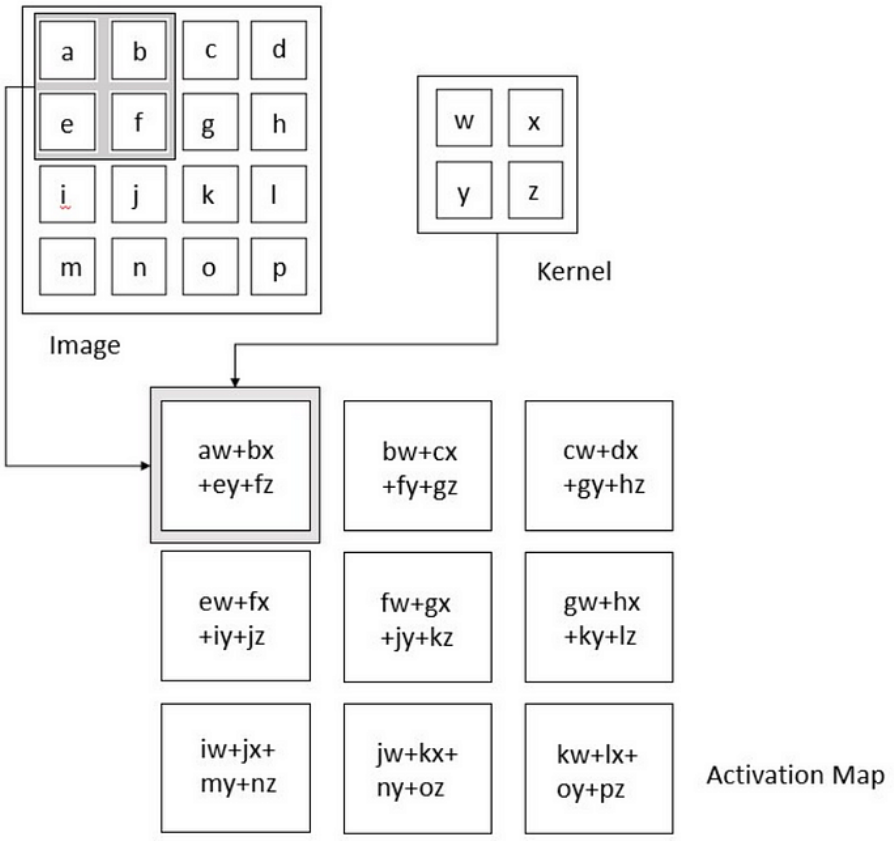


Рисунок 2.11 – Принцип роботи згортки [30]

Розмірність згортки можна визначати за допомогою формули (2.1) [30].

$$Wout = \frac{W-F+2P}{S} + 1, \tag{2.1}$$

Де W розмір вхідного зображення (якщо зображення квадратне), розмір згортки F, розмір нульового заповнення P та крок S. Кінцеве значення W_{out} має бути цілим числом.

Шар активації – містить функцію активації, такі як ReLu чи ELU. Зазвичай використовується після згортки (рисунок 2.12) .

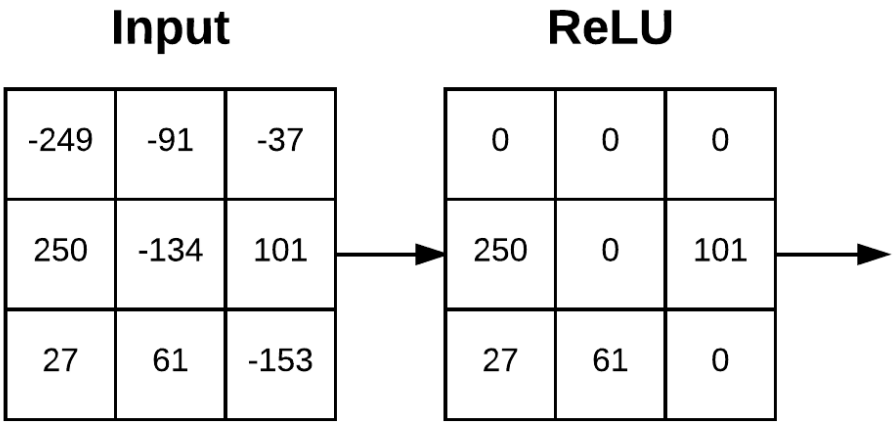


Рисунок 2.12 – Використання активаційної функції ReLu max(0,x) [31]

Пул-шар – зменшує розмір зображення без значної втрати інформації за допомогою використання функцій max чи avg, мають параметри розміру та кроку (рисунок 2.13). Зазвичай використовується в середині мережі.

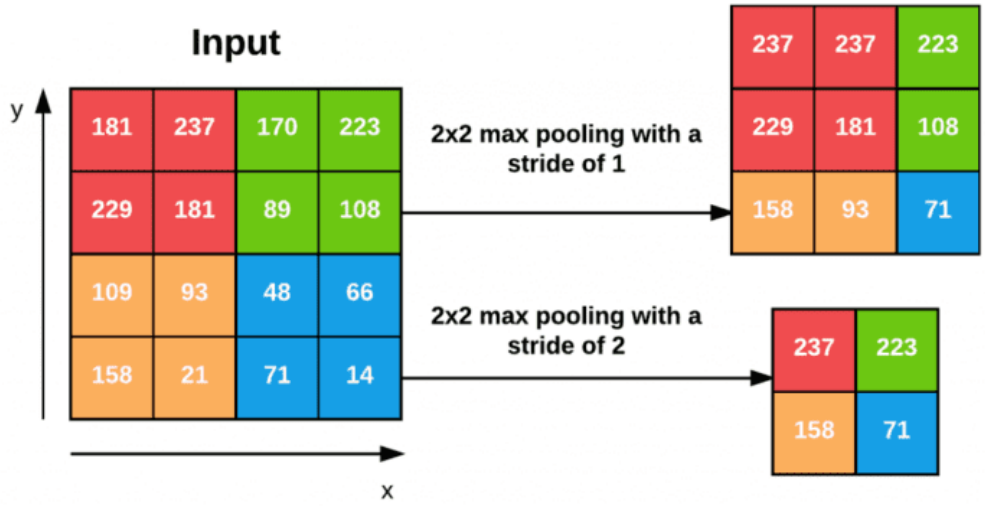


Рисунок 2.13 – Застосування пул-шару [32]

Повністю-з’єднаний шар – нейрони що виконують класифікацію на основі ознак отриманих всіх активацій шару до цього. Використовуються після згортки і пул-шару в кінці мережі.

Шар нормалізації – містить функцію нормалізації такі як мін-макс (2.2), z-score (2.3), batch-нормалізація (2.4), шар допомагає стабілізувати тренування [33].

$$x' = \frac{x - min}{x + max} \tag{2.2}$$

$$x' = \frac{x - mean}{STDEV} \tag{2.3}$$

$$x' = \frac{x - mean}{STDEV} * a + b. \tag{2.4}$$

де x вхідне значення, мінімальне значення min, максимальне значення max, середнє значення mean та стандартне відхилення STDEV. Параметри a та b це довільні навчальні параметри.

Шар відсіювання – в випадковому порядку від'єднує штучні нейрони від мережі, слугує для запобігання перенавчанню (рисунок 2.14).

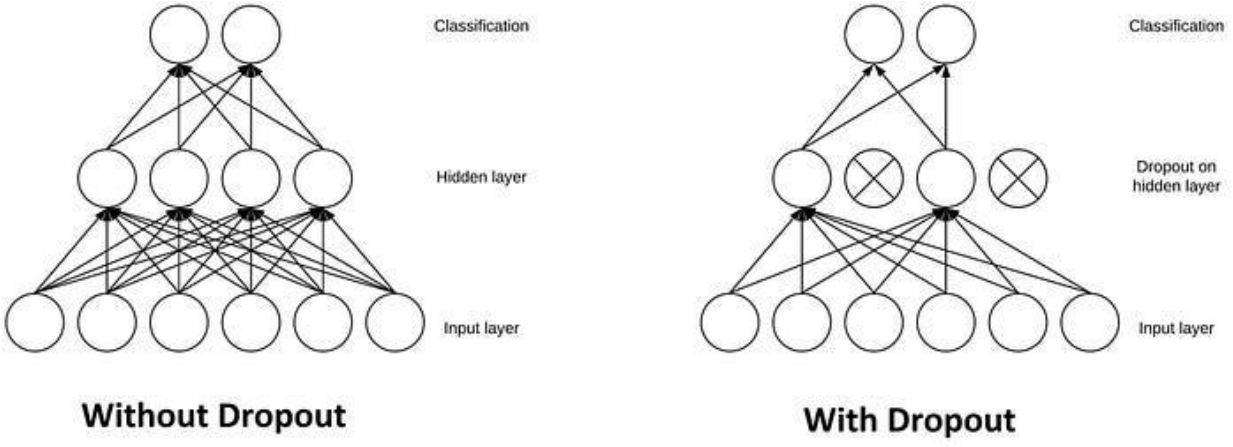


Рисунок 2.14 – Використання шару втрат [34]

Вихідний шар – класифікатор та останній шар в мережі, використовує логістичну функцію таку як sigmoid чи softmax, кількість нейронів залежить від кількості очікуваних вихідних класів мережі.

За допомогою комбінації означених шарів згорткових нейронних мереж була створена архітектура нейромережі imageClassifier для ідентифікації штучно згенерованих зображень людей (рисунок 2.15).

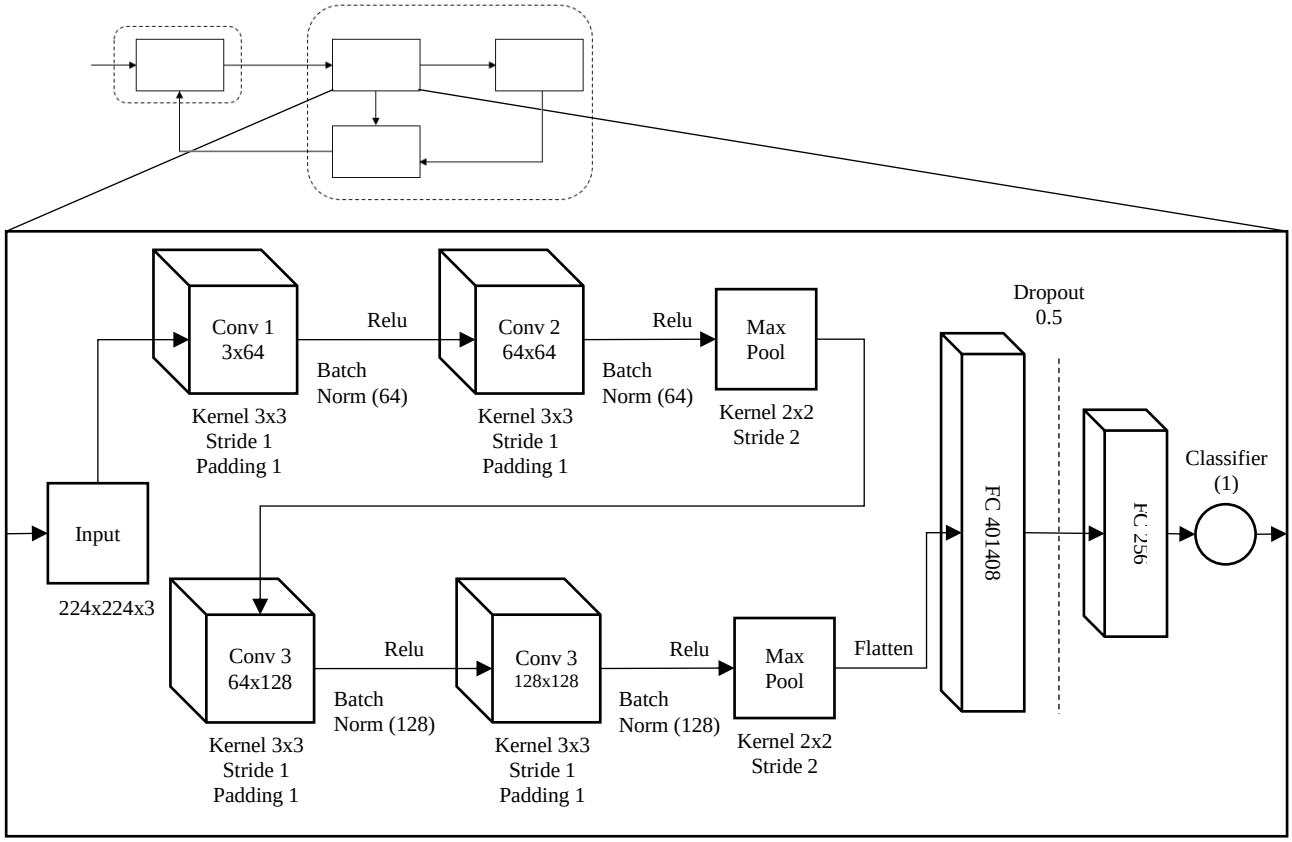


Рисунок 2.15 – Архітектура imageClassifier

Нейромережа imageClassifier використовує бінарну класифікацію відповідно до класів «реальне» та «згенероване».

Таким чином було розроблено базову архітектуру мережі imageClassifier. Архітектури штучних нейронних мереж не є статичними і можуть бути оновленими в залежності від результатів навчання під час тестування та ітерації.

2.4 Архітектура нейромережі для виявлення походження штучно згенерованих зображень людей

Аналогічним чином була розроблена структура нейромережі methodClassifier для виявлення походження штучно згенерованих зображень людей (рисунок 2.16).

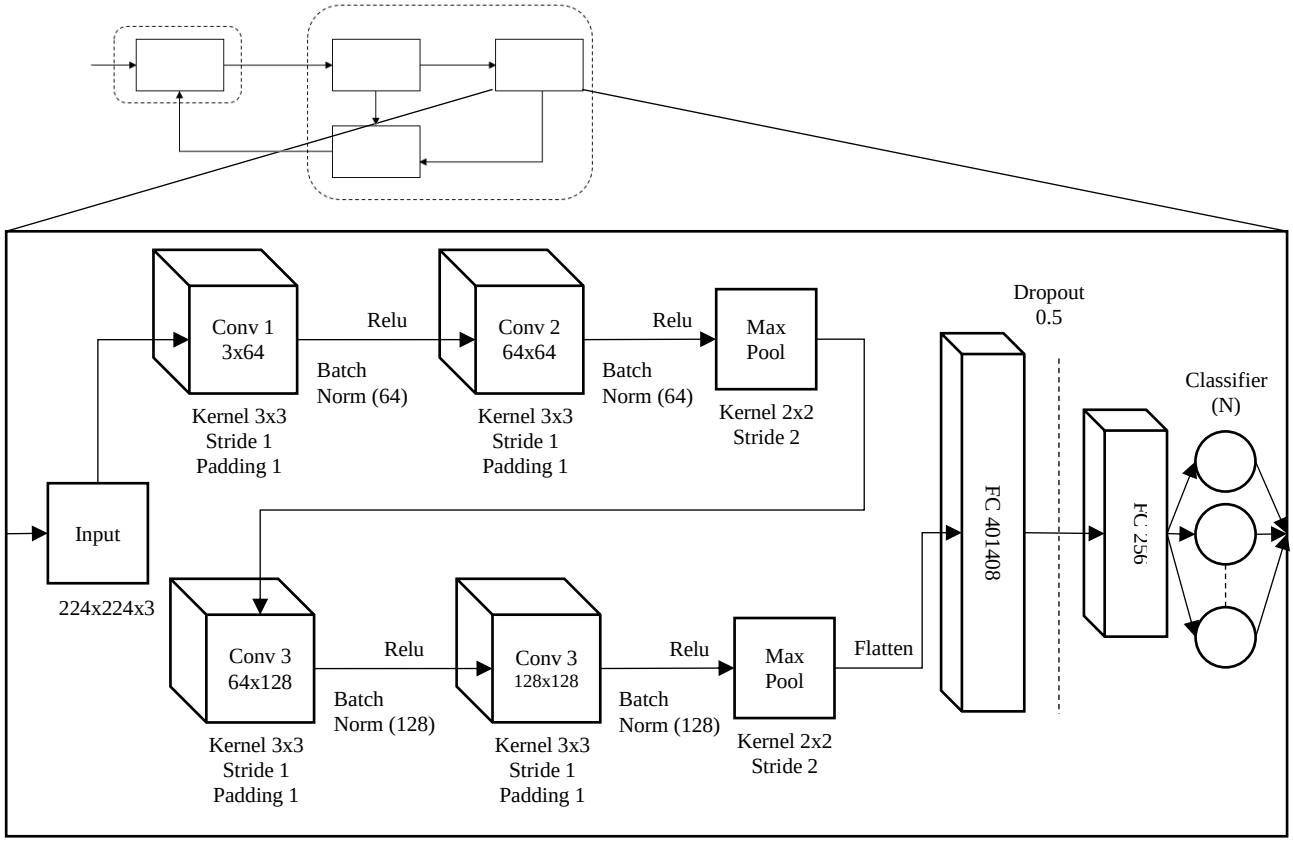


Рисунок 2.16 – Архітектура methodClassifier

Таким чином було розроблено базову архітектуру мережі methodClassifier, що базується на архітектурі нейромережі imageClassifier, але кількість вихідних шарів methodClassifier відповідає кількості класів, що відповідають кількості обраних під-датасетів, з яких складається створений датасет для мережі по ідентифікації походження.

2.5 Навчання нейромережевих моделей

Іншим надзвичайно важливим етапом є навчання – ітеративний процес, що дає можливість мережі класифікувати зображення згідно класів, у якому обчислення виконуються через кожен рівень мережі вперед та назад, доки не буде мінімізована функція витрат або процес не буде зупинено (рисунок 2.17) [35, 36].

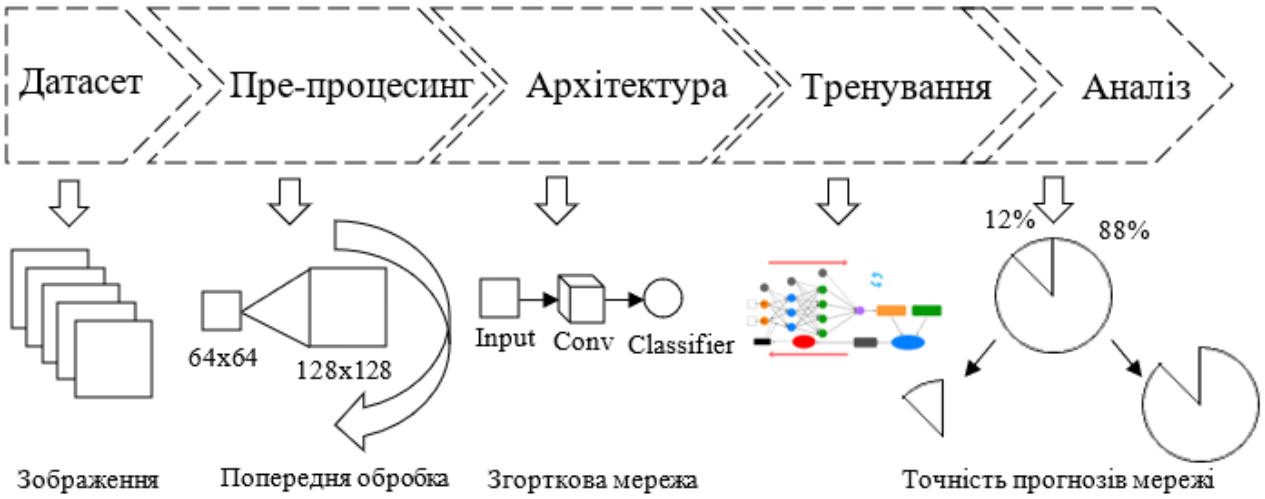


Рисунок 2.17 – Складові тренування мережі

Навчання складається із трьох основних етапів (рисунок 2.18):

– пряме поширення (forward propagation);

– обчислення функції втрат (loss function);

– зворотнє поширення (backwards propagation).

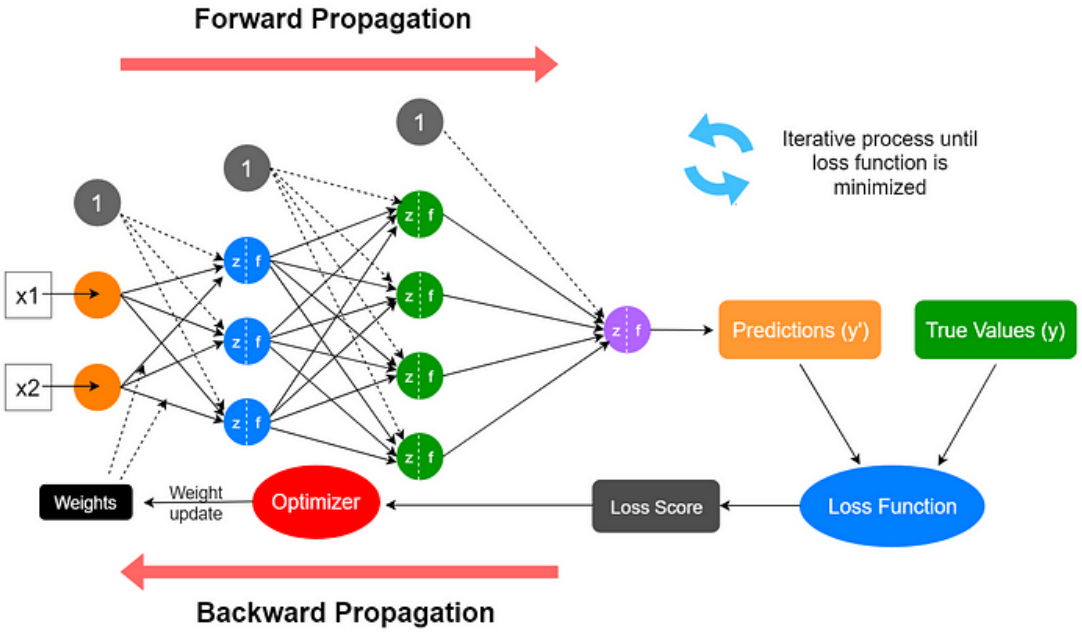


Рисунок 2.18 – Процес тренування мережі [36]

Спершу ініціалізуються параметри – нейронам присвоюється не-нульове значення ваг, далі виконується лінійна функція та функція активації по всій мережі на кожному штучному нейроні на основі вхідних та ініціалізованих

значень. Після завершення обчислень у вихідному шарі отримується результат прямого поширення – прогноз.

На наступному етапі використовується функція втрат для порівняння значень прогнозу та реальних. Функція втрат визначає наскільки добре модель працює на кожній ітерації моделі, що використовується у зворотному поширенні для оновлення параметрів мережі. Зазвичай результат подається у вигляді відсотків, де ідеальний показник є 0% помилок.

У зворотному поширенні розраховуються часткові похідні функції втрат та параметри моделі в кожному шарі і алгоритм оптимізації для корегування параметрів мережі.

Сам процес навчання поділяється на два типи:

– кероване;

– некероване.

Кероване машинне навчання потребує позначених вхідних даних під час навчання моделі машинного навчання. Ці навчальні дані маркуються розробником на етапі підготовки, перш ніж використовуватися для навчання та тестування моделі. Після того, як модель дізналася про зв’язок між вхідними та вихідними даними, її можна використовувати для класифікації нових і невідомих наборів даних і прогнозування результатів [37].

Некероване навчання (кластеризація) – навчання на необроблених і немаркованих навчальних даних. Його часто використовують для виявлення закономірностей і тенденцій у необроблених наборах даних або для кластеризації схожих даних у певну кількість груп.

Для методу ідентифікації штучно згенерованих зображень людей краще підходить кероване навчання через відповідність зображення одному із класів.

Менш важливими параметрами тренування нейронних мереж є кількість епох та групи даних. Кількість епох відповідає за цикли навчання що проходить мережа через весь датасет, а розмір групи скільки даних вона отримує до поновлення параметрів. У випадку якщо кількість груп помножити на розмір

групи більше чим взятий датасет частина даних буде використана повторно, що може призвести до перенавчання (рисунок 2.19).

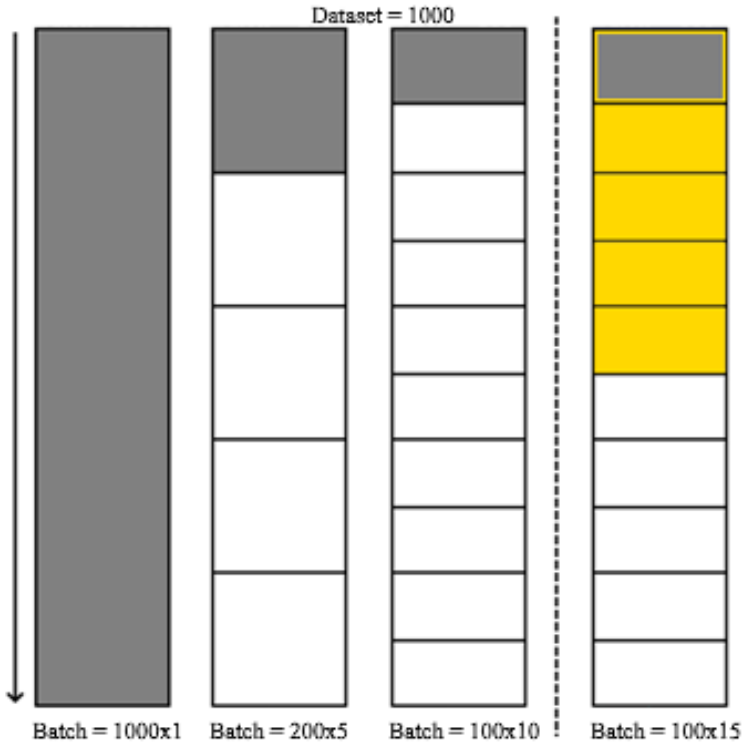


Рисунок 2.19 – Залежність кількості епох та розміру групи

Для тренування були обрані початкові дані в 10 епох та розподіл датасету на групи по 32 зображення для тренування, валідації та тестування, оптимальні параметри будуть визначатися емпірично.

Отже, для навчання розроблених нейронних мереж буде використовуватися навчання із вчителем з пре-розподіленими класами зображень відповідно до поставленого завдання.

Висновки до другого розділу

Під час написання другого розділу кваліфікаційної роботи магістра, що присвячений розробці методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання, було одержано:

1. Спроєктовано метод для ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання використовуючи комбінацію двох згорткових нейронних мереж, що працює на основі перетворення вхідних даних – зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані – відсоткова оцінка автентичності зображення та його походження.

2. Були сформовані датасети для ідентифікації штучно згенерованих зображень людей та виявлення їх походження, що складаються із 10000 тренувальних, 2000 валідаційних та 2000 тестувальних зображень обличь людей розподілених на два класи для мережі по ідентифікації зображень, і 3000 тренувальних, 600 валідаційних та 600 тестувальних зображень розподілених на три класи для мережі по ідентифікації походження.

3. Розроблені базові архітектури мережі для ідентифікації штучних зображень та мережі для ідентифікації підходу до генерації зображення, що складаються із трьох груп згорток, доповнених шарами нормалізації та функціями активації, двох пул-шарів та двох повно’єдних шарів.

4. Була створена схема навчання та обрані початкові параметри тренування мереж в 10 епох та розподіл датасетів на групи по 32 зображення, що будуть ітеративно змінюватись з метою знаходження емпірично кращих.

Розділ 3 Проєктування інформаційної системи нейромережевого аналізу згенерованих зображень людей

3.1 Схема інформаційної системи

Інформаційна система нейромережевого аналізу згенерованих зображень людей за допомогою машинного навчання є прикладною програмною реалізацією методу аналізу зображень людей згенерованих штучним інтелектом. Система призначена для обробки файлів, завантажених користувачем, з метою виявлення згенерованих зображень. Вхідними даними є файли зображень у форматах png, jpg, webp, tiff, bmp.

Інформаційна структура системи складається із набору зображень (датасетів) та кількох підсистем: «Підсистема взаємодії з НМ», «Підсистема розпізнавання завантажених зображень», «Підсистема інтерфейсу користувача», «Підсистема налаштувань» (рисунок 3.1).

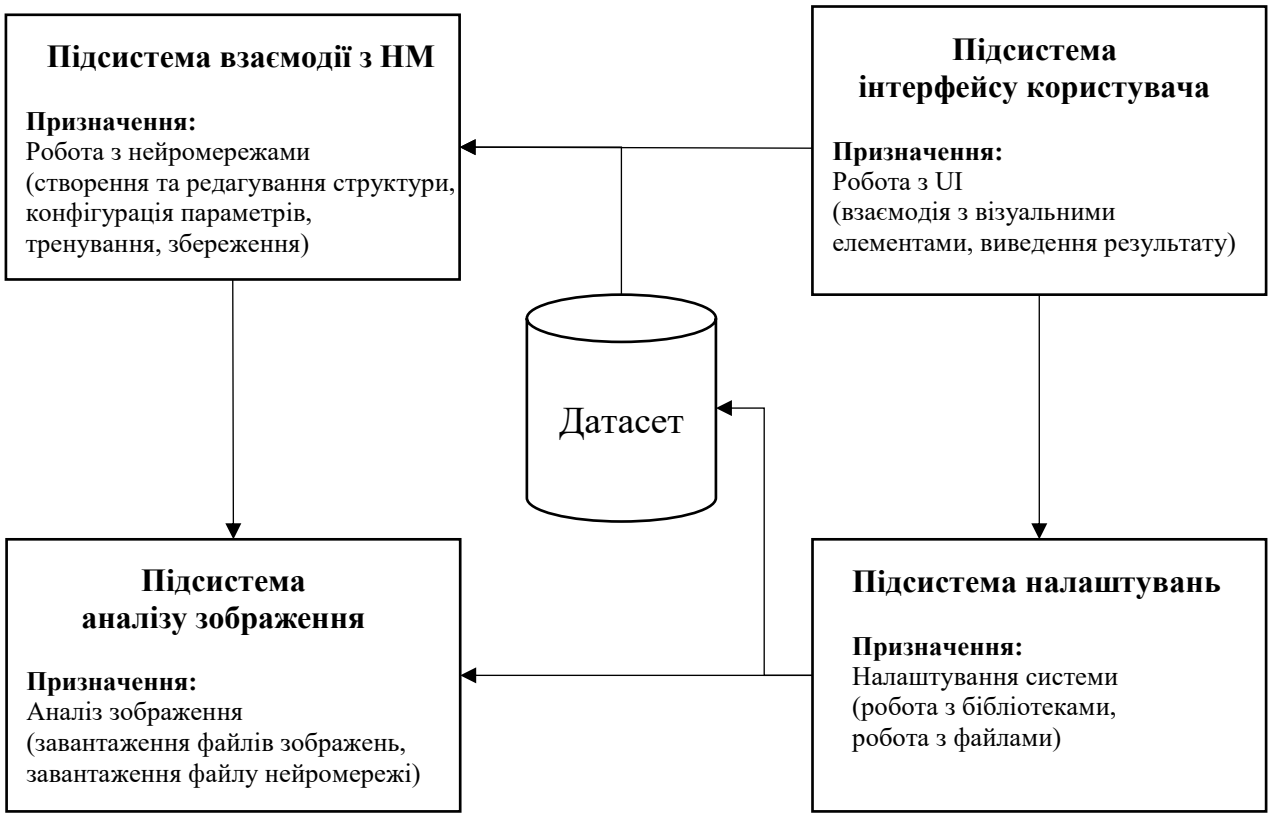


Рисунок 3.1 – Інформаційна структура системи нейромережевого аналізу зображень людей

Набір даних зображень складається з декількох датасетів, розподілених на дві відповідні частини:

– класифікація зображень – датасети, розділені на реальні та згенеровані зображення для нейронної мережі imageClassifier.

– класифікація методів створення – датасети, що містять лише згенеровані зображення, поділені за технологією генерації.

Підсистема роботи з нейронними мережами є основною і призначена для роботи з методами нейронних мереж. Вона включає функції, такі як створення та модифікація архітектури мережі, налаштування параметрів, вибір датасетів, процес тренування, а також збереження натренованої мережі у файл.

Підсистема розпізнавання зображень є допоміжною і відповідає за аналіз завантажених зображень або кількох зображень за допомогою натренованої нейронної мережі, яка завантажується з файлу. Вона має функціонал для завантаження зображень, подальшого звільнення файлів, завантаження файлів нейромережі та виведення результатів.

Підсистема інтерфейсу користувача забезпечує функціональну взаємодію користувача з іншими підсистемами через UI. Вона включає динамічну генерацію візуальних елементів відповідно до дій користувача.

Підсистема налаштувань надає користувачу можливість змінювати обрані параметри, як візуальні (наприклад, розмір вікна), так і функціональні (обрані бібліотеки та файли).

Таким чином, в даному пункті спроєктовано інформаційну структуру системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання, що за вхідним зображенням користувача визначає автентичність завантаженого зображення та визначає засоби його генерації.

3.2 Схема та функції підсистеми взаємодії з нейромережею

Підсистема взаємодії з нейронною мережею, основним призначенням якої є робота з нейромережею (рисунок 3.2).

Рисунок 3.2 – Схема та функції підсистеми взаємодії з НМ

Першою функцією підсистеми є вибір шляхів до зображень. Для коректного завантаження зображень необхідно вказати правильний шлях до основної директорії та директорій для тренування та тестування. Зображення з цих директорій будуть розподілені по класах відповідно до їх назв.

Наступною функцією є завантаження зображень у даталоадери, де можна вказати основні параметри: розмір зображення для трансформації, розмір однієї групи та потребу в перемішуванні.

Третя група функцій включає створення та редагування архітектури нейронної мережі з можливістю додавання обраної кількості шарів та вказівки їх

основних параметрів: вхідної та вихідної розмірності, розміру кернела, кроку, нульового заповнення та інших відповідних параметрів.

Кінцевою функцією є налаштування тренування з вказівкою параметрів: кількість епох, оптимізатор та його параметри. Після успішного завершення тренування можна зберегти натреновану мережу.

Отже, було спроєктовано схему та розглянуто основні функції підсистеми редагування НМ, що є основною складовою інформаційної системи нейромережевого аналізу згенерованих зображень людей.

3.3 Схема та функції підсистеми розпізнавання зображень

Підсистема аналізу зображень є вторинною підсистемою інформаційної системи нейромережевого розпізнавання згенерованих зображень людей (рисунок 3.3). Основне призначення – аналіз завантаженого зображення використовуючи вже навчені нейромережі.

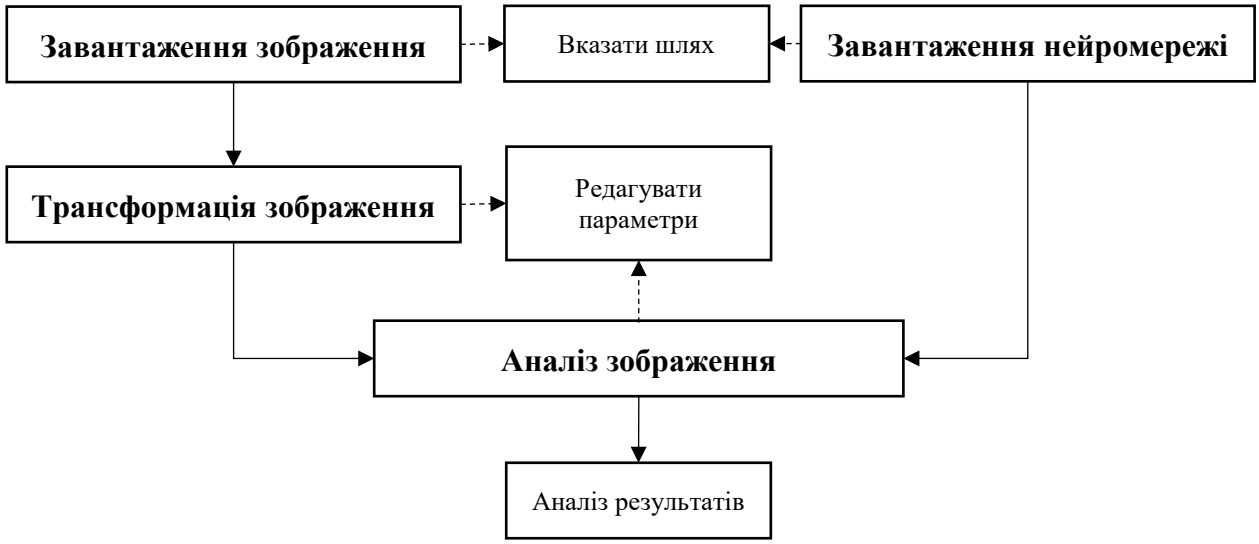


Рисунок 3.3 – Схема та функції підсистеми розпізнавання зображень

Першою парою функцій підсистеми є функції завантаження зображень та нейромережі за вказаним шляхом. Наступною функцією є трансформація зображення відповідно заданих параметрів. Останньою функцією є аналіз

завантаженого зображення з використанням нейромережі та аналіз отриманих результатів.

Отже, було спроєктована підсистема аналізу зображень, що є вторинною підсистемою проєктованої інформаційної системи.

3.4 Формування комбінації засобів розробки інформаційної системи

Для прикладної реалізації програмного забезпечення на базі методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання, необхідно обрати мову програмування, бібліотеки та середовище програмної розробки.

Python була обрана для програмної реалізації алгоритмів нейронних мереж завдяки своїй гнучкості, динамічній типізації та універсальності. Python є привабливою мовою програмування для розробки програмного забезпечення з використанням нейромереж через велику кількість спеціалізованих бібліотек, простоту у вивченні синтаксису та його чистота, здатність до інтеграції та масштабування [38].

Основні переваги Python:

– простий синтаксис – код легко зрозуміти, поширювати та підтримувати, немає багатослівності, мова легко вивчається;

– потужний інструментарій – має величезний набір сторонніх додатків, бібліотек і фреймворків що прискорюють процес розробки;

– гнучкість – Python можна використовувати в різних проектах та гнучкий у типі даних;

– портативність – програми Python є текстовими файлами, що містять інструкції для інтерпретатора та можуть бути написані в не тільки в IDE, а й текстовому редакторі.

Python – це об'єктно-орієнтована мова, яка ефективно застосовується для роботи з великими наборами даних та масштабованими системами, що є важливим фактором у розробці програмного забезпечення для нейронних мереж.

Python забезпечує ефективну обробку великих обсягів інформації завдяки використанню спеціалізованих бібліотек для обробки даних та паралельного виконання.

Також було обрано C# для розробки UI. C# це об’єктно-орієнтована мова програмування, що дозволяє створювати кросплатформенні додатки будь-якого призначення. Важливою особливістю є статична типізація [39].

Основними перевагами C# є те, що створювалась відповідно до правил об’єктно-орієнтованого програмування. Об’єктно-орієнтоване програмування поміщає дані в об’єкти, що полегшує розбиття програми на менші частини, які простіше створювати, керувати та об’єднувати. Також ООП дозволяє керувати об’єктами без безпосередньої роботи з їхніми внутрішніми властивостями;

C# все ще вдосконалюється та підтримується корпорацією Microsoft. Також Microsoft підтримує велику кількість документації по C# і .NET, включаючи докладні пояснення поширених проблем.

В якості IDE були обрані Visual Studio Code для розробки Python та Microsoft Visual Studio для розробки C#.

Visual Studio Code або VS code це гнучкий текстовий редактор що надає розробникам набір настроюваних функцій через розширення. Створений для веб-розробників і тих та невеликих проектів VS Code здобув є популярним середовищем для розробки завдяки зручному інтерфейсу та вражаючій функціональності [40,41].

Основні переваги:

– гнучкість – середовище можна налаштувати за допомогою плагінів додаючи лише потрібний функціонал;

– спрощений редактор – починається як базовий текстовий редактор що робить середовище швидким та простим до використання;

– підтримка мов – налаштувати для мов, таких як C, C++ і Python, за допомогою довантаження плагінів;

– універсальність – редактор можна використовувати для широкого кола завдань, таких як редагування та перегляд файлів розмітки

Microsoft Visual Studio – це інтегроване середовище розробки, призначене для створення настільних програм, GUI, консольних додатків, веб-програм та мобільних додатків. Visual Studio підтримує різні платформи для розробки програмного забезпечення Microsoft та дозволяє писати код на 36 мовах програмування, включаючи C#, C++, Visual Basic, JavaScript та інші. [42].

– на відміну від VS Code, Visual Studio не вимагає ручного налаштування для компіляторів і функцій специфічних для мови.

– Visual Studio краще підходить для C# - налаштування середовища VS для C# відповідно до можливостей Visual Studio може бути складним завданням.

Для розробки інформаційної системи ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання, було обрано комбінацію мов програмування Python та C#, середовища розробки Visual Studio Code та Visual Studio. Python є ідеальним вибором для розробки програмного застосунку завдяки своїй гнучкості та потужному інструментарії бібліотек машинного навчання, C# в свою чергу надає розширені можливості для створення UI. В свою чергу обрані середовища розробки дозволяють ефективно розробляти та налагоджувати код.

3.5 Вибір спеціалізованих програмних розширень

Для реалізації інформаційної системи нейромережевого аналізу згенерованих зображень людей, важливим є вибір відповідного інструментарію бібліотек для обробки зображень, методів створення та тренування нейронних мереж.

Були обрані наступні бібліотеки:

– TorchVision;

– Numpy;

– Pathlib;

– Pillow;

– Mathplotlib;

– PythonNet.

TorchVision – це бібліотека для комп’ютерного зору, яка йде рука об руку з PyTorch та містить методи для ефективного перетворення зображень і відео, а також деякі попередньо навчені моделі та датасети [43].

Основні переваги TorchVision:

– автоматично постачається з підтримкою GPU що значно прискорює роботу;

– компактність та простота реалізації;

– поставляється із значним набором готових нейромереж, датасетів та зразків коду

– активно розробляється та підтримується командою FacebookAI

TorchVision постачається з можливістю легкого завантаження деяких із часто-використовуваних наборів даних таких як CIFAR, CelebA, COCO, Omniglot, VOC, Flickr, FashionMNIST.

Для спрощення навчання TorchVision має попередньо підготовлені моделі для класифікації зображень, виявлення об’єктів, сегментації екземплярів та класифікації відео що можуть бути окремо локально завантажені. Одні з таких мереж: ResNet 3D 18, MASK R-CNN, AlexNet, VGG, ResNet, Inception.

Підсумовуючи Torchvision для розробки нейромереж в контексті програмного застосунку на базі методу нейромережевого аналізу згенерованих зображень людей документів є обґрунтованим та ефективним, оскільки бібліотека значно сприяє спрощує розробку архітектури та тренування нейромереж.

Для обробки масивів даних використовується бібліотека NumPy. NumPy – це бібліотека Python, яка надає об'єкти багатовимірних масивів та їх похідні (такі як замасковані масиви та матриці), а також набір процедур для ефективних операцій над масивами. До таких операцій належать математичні, логічні обчислення, маніпуляції формою, сортування, вибір, введення/виведення, дискретне перетворення Фур'є, базова лінійна алгебра, основні статистичні операції та випадкове моделювання [44].

В основі NumPy лежить об'єкт ndarray, який інкапсулює n-вимірні масиви однорідних типів даних. Багато операцій над такими масивами виконуються за допомогою скомпільованого коду, що значно підвищує швидкість обробки даних.

Основні відмінності NumPy від стандартних засобів Python:

– масиви мають статичний розмір на відміну від Python списків, для зміни розмірів потрібно створити новий масив та видалити оригінальний;

– елементи мають мати однаковий тип даних, що дозволяє їх займати однакову кількість пам’яті;

– масиви мають розширений функціонал для математичних та інших видів операцій над великою кількістю даних, що краще оптимізовані порівняно з вбудованим функціоналом Python;

– користувацька підтримка.

Іншою обраною бібліотекою є Pathlib що забезпечує елегантне рішення для обробки шляхів файлової системи використовуючи об’єктно-орієнтований підхід, а також забезпечує незалежну від платформи поведінку [45].

До Python 3.4 для обробки шляхів до файлів було більш традиційно використовувати модуль OS. Але з часом модуль OS почав втрачати свою ефективність. Основні з недоліків OS.

– довгий і нечитабельний код для відносно простої операції;

– передбачає знання про розуміння списків;

– включає рядкові операції що є схильними до помилок та не дуже стислими.

В свою чергу ідентична операція по завантаженню зображення із використанням Pathlib є набагато простішою. Об’єктно-орієнтований підхід Pathlib організовує код навколо об’єктів шляху та їх взаємодії, що дозволяє створювати більш модульний та платформонезалежний код, який простіший у використанні та підтримці.

Функціонал Pathlib дозволяє створювати об’єкти шляху використовуючи інші об’єкти, поточного та домашнього робочого каталогу, а також роботу з кореневим та похідним каталогом і назвами файлів.

Для роботи з зображеннями була обрана бібліотека Pillow – форк бібліотеки PIL. Бібліотека містить легкі інструменти обробки зображень, які допомагають редагувати, створювати та зберігати зображення, але її підтримку було припинено в 2011 році. Проєкт Pillow розділив оригінальний PIL і додав до нього підтримку Python3.x. Бібліотека підтримує велику кількість форматів файлів зображень таких як: png, jpeg, bmp, tiff [46].

PythonNet це пакет бібліотек, що надає програмістам Python майже повну інтеграцію з .NET Framework, .NET Core і середовищем виконання Mono у Windows, Linux і macOS. Python.NET надає потужний інструмент створення сценаріїв додатків для розробників .NET. Пакет PythonNet дозволяє створювати сценарії додатків .NET або створювати цілі додатки на Python, використовуючи служби та компоненти .NET, написані будь-якою мовою, націленою на CLR (C#, VB.NET, F#, C++/CLI) [47].

Отже, було визначено бібліотеки для реалізації методу нейромережевого аналізу згенерованих зображень людей, а саме TorchVision для методів нейронних мереж, Pathlib для роботи з файлами, Pillow для обробки зображень, Numpy та Mathplotlib для обробки та візуалізації даних під час тестування, PythonNet для взаємодії між Python та C#.

3.6 Програмна архітектура інформаційної системи нейромережевого аналізу згенерованих зображень людей

Була розроблена схема запропонованої програмної архітектури для інформаційної системи нейромережевого аналізу згенерованих зображень людей (рисунок 3.4).

Програмна архітектура системи є об'єктно-орієнтованою та складається з таких модулів:

– Menu – внутрішній модуль, який реалізує перехід між іншими класами;

– ModelTrain – модуль, що реалізує методи тренування мережі;

– TrainWindow – внутрішній модуль, що відповідає за роботу з нейромережами;

– ImageTest – зовнішній модуль, що реалізує методи завантаження та аналізу зображень;

– TestWindow – внутрішній модуль, що забезпечує аналіз завантажених зображень;

– SettingWindow – внутрішній модуль, що дозволяє змінювати налаштування програми.

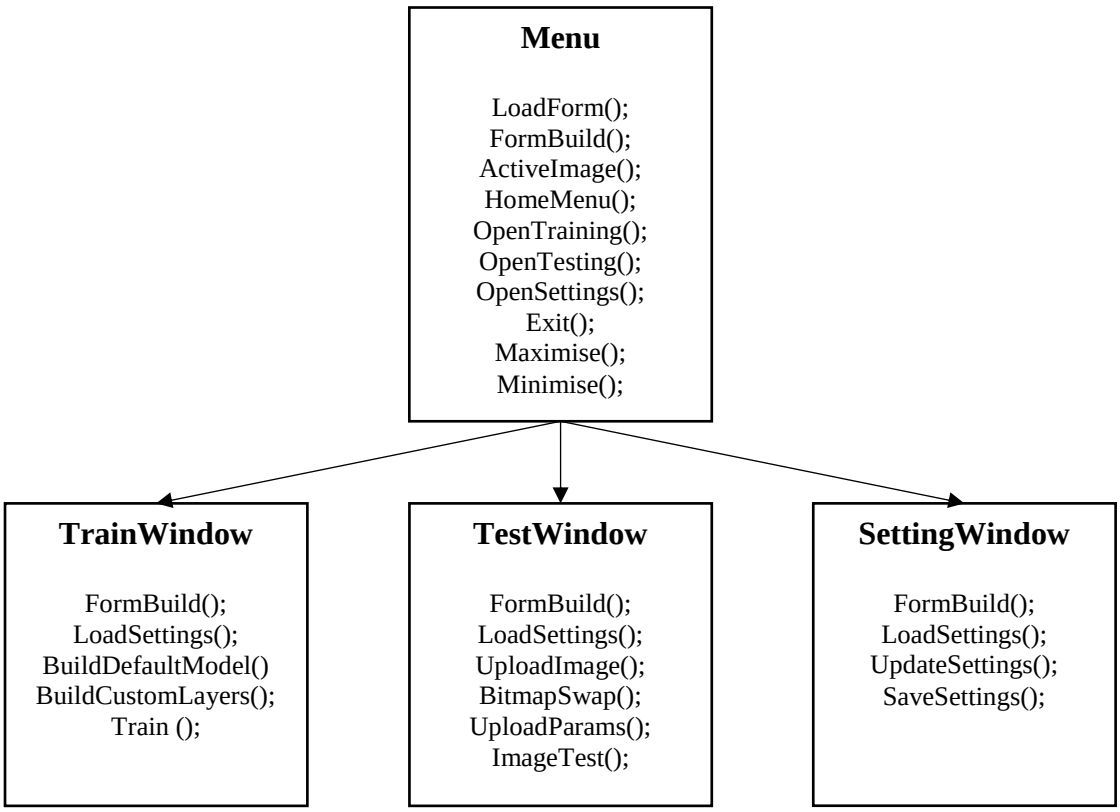


Рисунок 3.4 – Програмна архітектура системи

Модуль «TrainWindow» має основне призначення — створення та редагування архітектури нейромережі, а також її тренування. Методи InitialiseComponent() та FormBuild() відповідають за ініціалізацію та створення форми за допомогою C#. Метод LoadSettings() завантажує налаштування програми з відповідного файлу. Метод TrainButtonClick() завантажує параметри

нейромережі з відповідних полів та запускає процес тренування, використовуючи завантажений клас ModelTrain.py на Python. Метод CreateCustomLayers() динамічно додає нові елементи UI для створення власної архітектури мережі.

Модуль «ModelTrain.py» використовує методи бібліотеки Torchvision мови Python для тренування нейромережі та має такі основні методи: SetImagePaths(), який завантажує шляхи до зображень; LoadImages(), що завантажує файли зображень, які потім перетворюються на формат даних для навчання мережі в методі ToDataloaders(). Метод BuildDefaultModel() завантажує стандартну архітектуру, а BuildCustomModel() створює архітектуру динамічно. Метод Run() запускає процес навчання та зберігає натреновану мережу.

Модуль «TestWindow» використовує навчену нейромережу для аналізу завантаженого зображення користувачем та відображення результатів. Метод LoadSettings() працює аналогічно методу цього ж імені в модулі TrainWindow і завантажує файл налаштувань. UploadImage() викликає інтерфейс для вибору та завантаження зображення в додаток і звільняє його для паралельної обробки, використовуючи метод BitmapSwap. Метод ImageTest() завантажує клас ImageTest.py на Python для аналізу зображення вибраною нейронною мережею та виводить результат в інтерфейсі користувача.

Модуль «SettingWindow» здійснює взаємодію з елементами інтерфейсу для завантаження та редагування налаштувань програми. Метод LoadSettings() завантажує файл налаштувань та заповнює елементи UI для подальшого редагування. Метод SaveSettings() зберігає або перезаписує налаштування в відповідний файл для подальшого використання іншими модулями.

Таким чином, була спроєктована архітектура інформаційної системи для реалізації методу нейромережевого аналізу згенерованих зображень людей.

Висновки до третього розділу

В ході виконання третього розділу було виконано проєктування інформаційної системи нейромережевого аналізу згенерованих зображень людей, зокрема:

1. Було спроєктовано інформаційну структуру системи нейромережевого аналізу згенерованих зображень людей, яка на основі вхідних даних з датасету зображень та відповідних класів навчає та зберігає нейромережу, а також завантажує її для аналізу обраного зображення з виведенням результатів.

2. Наведено основні схеми та функції головної підсистеми тренування мережі та вторинної підсистеми аналізу зображень спроєктованої інформаційної системи.

3. Виконано вибір комбінації засобів розробки інформаційної системи, що включає мови програмування Python та C#, а також середовища розробки Visual Studio Code та Microsoft Visual Studio. Python було обрано для реалізації алгоритмів нейронних мереж завдяки гнучкості, динамічній типізації та універсальності, а C# – для створення елементів інтерфейсу користувача. Як IDE обрано Visual Studio Code для розробки на Python та Microsoft Visual Studio для розробки на C#.

4.Вибрано спеціалізовані програмні розширення для реалізації прикладного застосунку на базі методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Серед них бібліотека PyTorch для розробки нейронної мережі, бібліотека Pathlib для роботи з файлами, Numpy для обробки зображень та PythonNet для взаємодії між мовами програмування.

5. Спроєктовано програмну архітектуру інформаційної системи для методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Архітектура включає 4 внутрішні та 2 зовнішні класи з визначеним функціоналом.

Розділ 4 Дослідження методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

4.1 Особливості розробки прикладних компонентів експериментальної інформаційної системи

Відповідно до розробленої програмної архітектури інформаційної системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання, що складається із 4 внутрішніх та 2 зовнішніх модулів, були створені прикладні компоненти що реалізовують заявлений функціонал.

Клас ModelTrain своїм основним призначенням якого є тренування нейромережі. Головним методом класу є метод run(), що призначений для покрокової ініціалізації та виконання усіх елементів мережі, а також сам процес тренування. Метод складається з кількох етапів, які схематично представлені на рисунку 4.1.

Вхідними даними є зчитаний з форми шлях до основної директорії датасету, назв для тренування та валідації, що автоматично розподіляється по класам відповідно назв директорій, крім того з форми зчитуються параметри трансформації датасету, такі як вихідний розмір, та параметри тренування мережі, а саме кількість епох, коефіцієнт навчання, оптимізатор функція втрат.

Першим кроком методу є завантаження та перевірка усіх файлів зображень на цілісність. Далі ці зображення трансформують відповідно до вказаних параметрів, таких як в розмір, нормалізація та перетворюють у тензор. Наступним кроком завантажені зображення перетворюють у даталоадери – об’єкт що спрощує розподіл датасету у групи. Останніми кроками методу є ініціалізація архітектури та тренування відповідно до вхідних параметрів. Вихідними даними є об’єкт натренованої мережі який можливо зберегти для подальшого використання.

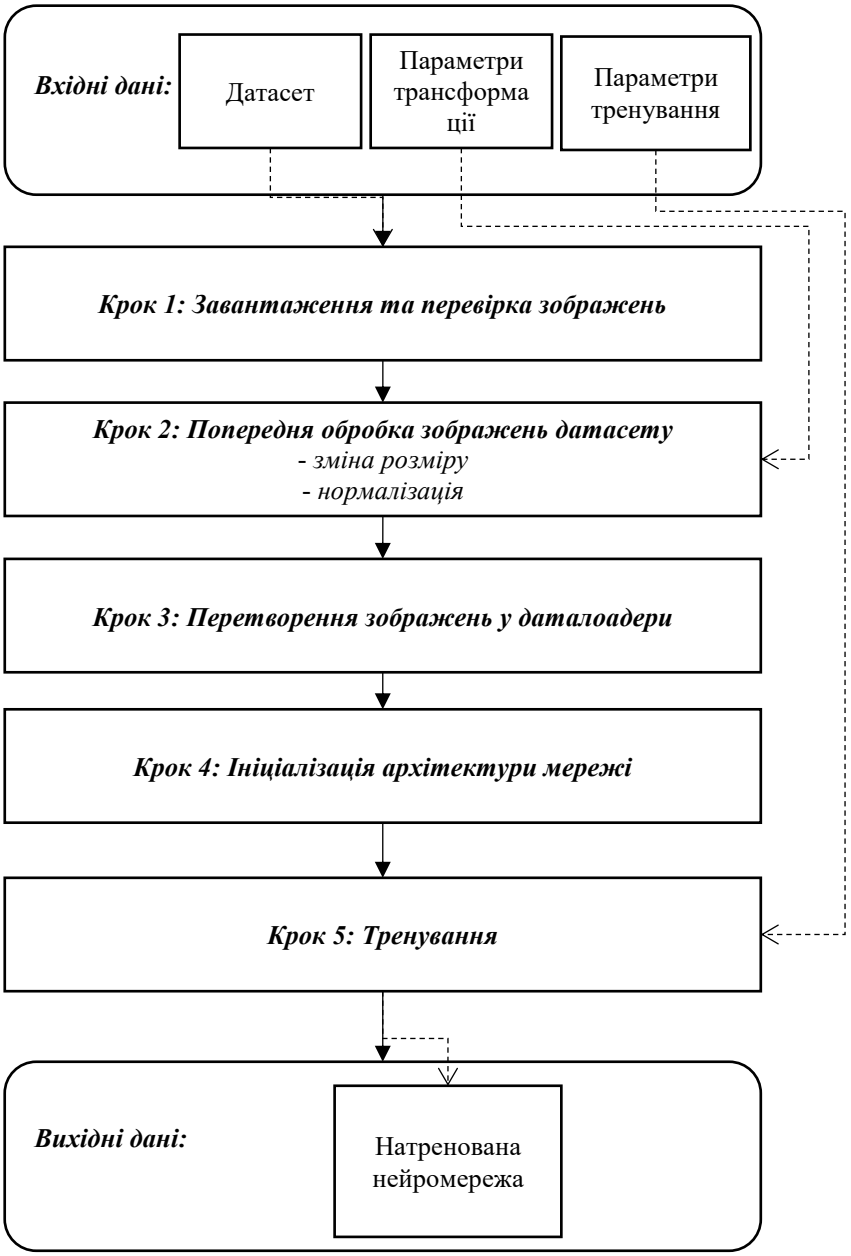


Рисунок 4.1 – Схема роботи методу run()

Іншим важливим методом класу ModelTrain є buildCustomModel(), що дозволяє створювати та редагувати архітектуру мережі для тренування. Метод buildCustomModel() розширює метод run() (рисунок 4.2).

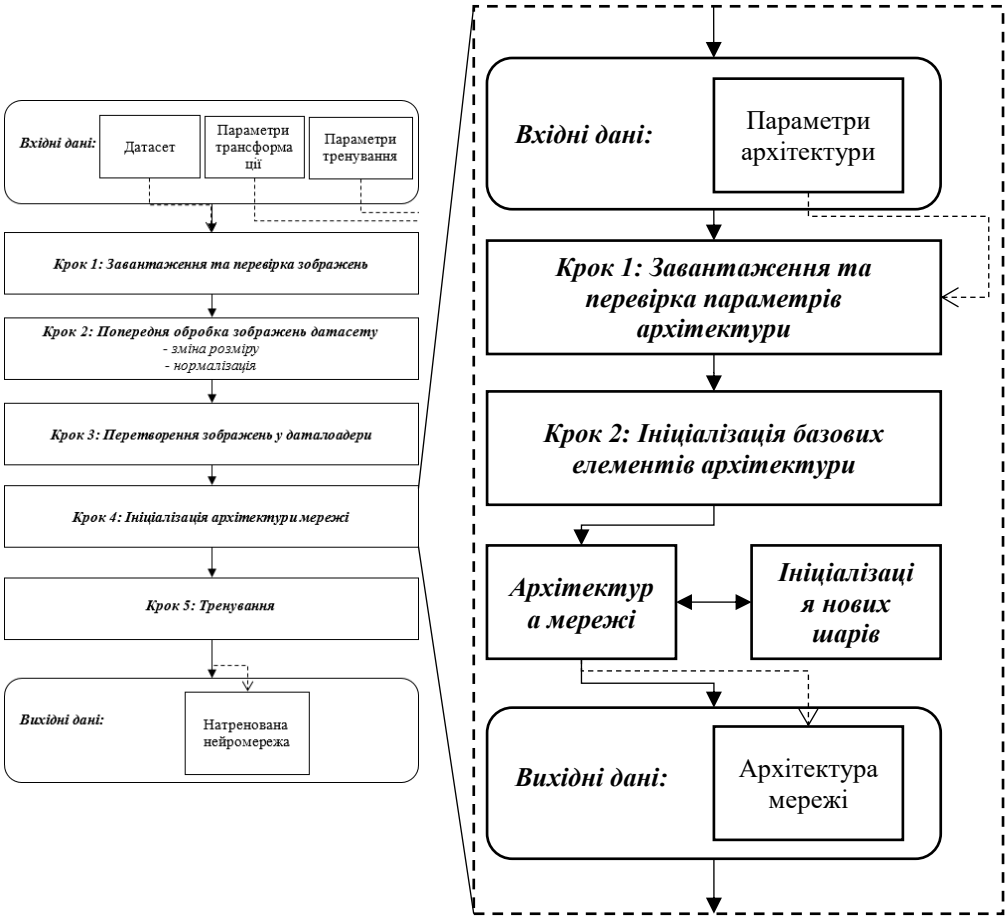


Рисунок 4.2 – Схема методу buildCustomModel()

Першим кроком методу є завантаження та перевірка параметрів архітектури. Другим кроком є ініціалізація базових шарів мережі – вхідного та вихідного. Наступним кроком є циклічно додавати шари в об’єкт архітектури відповідно до заданих параметрів. Вихідними даними є об’єкт архітектури для використання в методі run().

Клас TrainWindow слугує для завантаження елементів користувацького інтерфейсу та взаємодії з функціональним класом ModelTrain.

Клас ImageTest, що реалізовує однойменний метод ImageTest() для завантаження користувацього зображення, двох нейронних мереж для ідентифікації зображення та походження та подальшого аналізу. З форми зчитується шлях до зображення, двох нейронних мереж для аналізу зображення та його походження, що є вхідними даними. Вихідними даними є відсоткова оцінка зображення по наведених категоріях.

Схема методу наведена на рисунку 4.3.

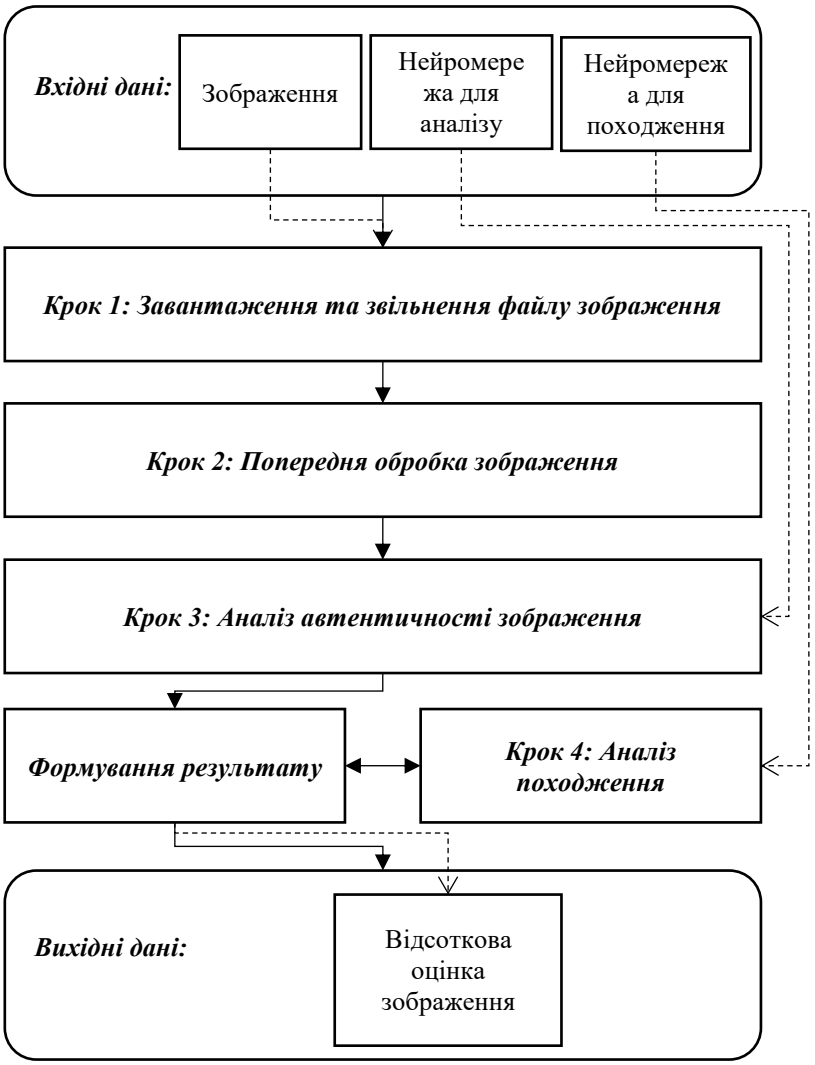


Рисунок 4.3 – Схема методу ImageTest()

Першим кроком методу є завантаження файлу зображення для подальшої обробки. На другому кроці методу зображення трансформують відповідно до вказаного розміру та перетворюють у тензор. Наступним кроком зображення аналізує перша мережа та формується результат. Якщо вердикт мережі що зображення згенероване його додатково аналізує друга мережа та додає свій результат. Вихідними даними є відсоткова оцінка зображення, чи воно згенероване та яким методом.

Під час завантаження зображення метод ImageTest() використовує метод BitmapSwap() що дозволяє завантажити файл в пам’ять та звільнити його для подальшого використання за допомогою підміни бітмапів (рисунок 4.4).

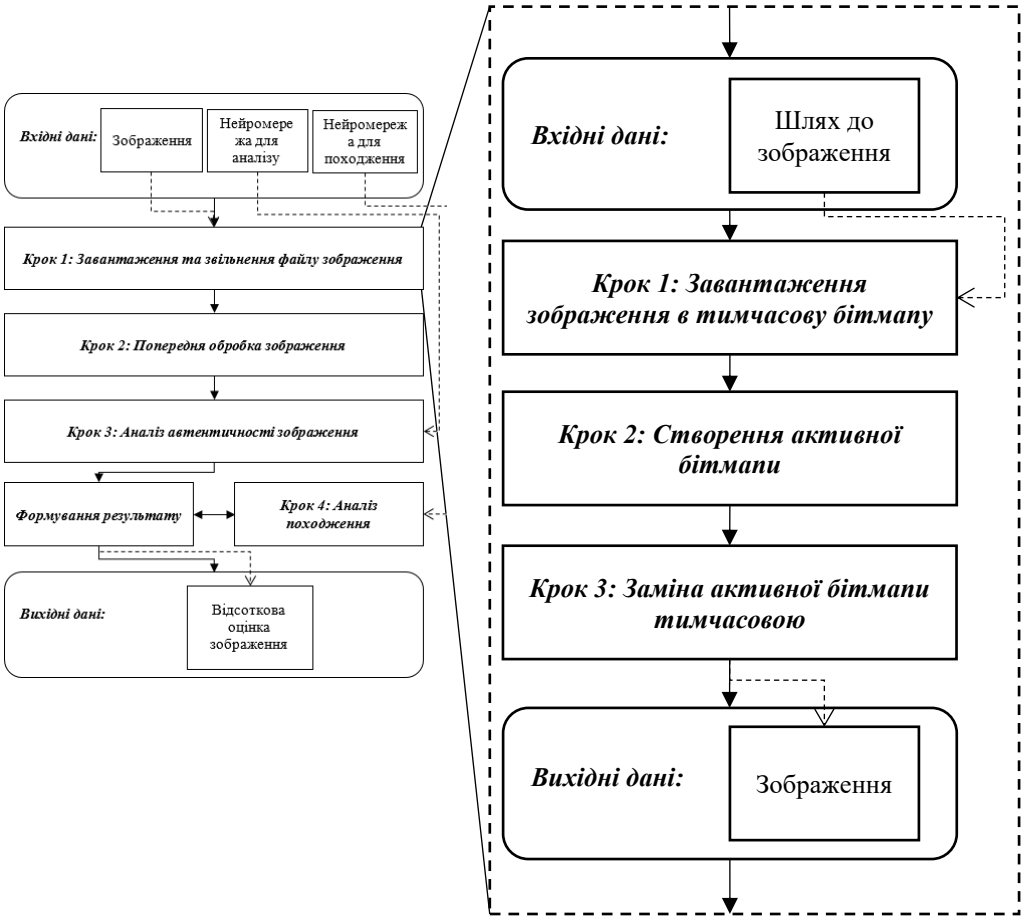


Рисунок 4.4 – Метод BitmapSwap()

Це допомагає уникнути помилки коли активне зображення неможливо перемістити чи відкрити (рисунок 4.5)

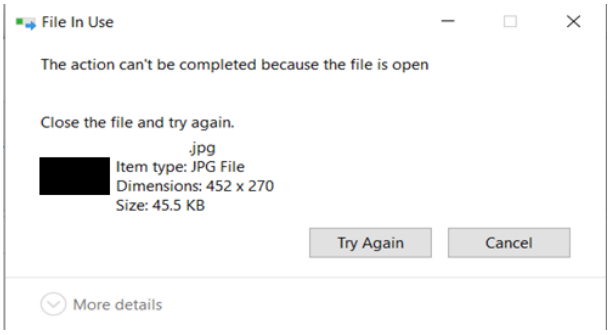


Рисунок 4.5 – Помилка «File in use»

Клас TestWindow слугує для завантаження елементів користувацького інтерфейсу та взаємодії з функціональним класом ImageTest.

Клас Settings дозволяє користувачеві налаштовувати параметри програми та змінювати шлях до зовнішніх класів та бібліотек використовуючи метод SaveSettings() (рисунок 4.6).

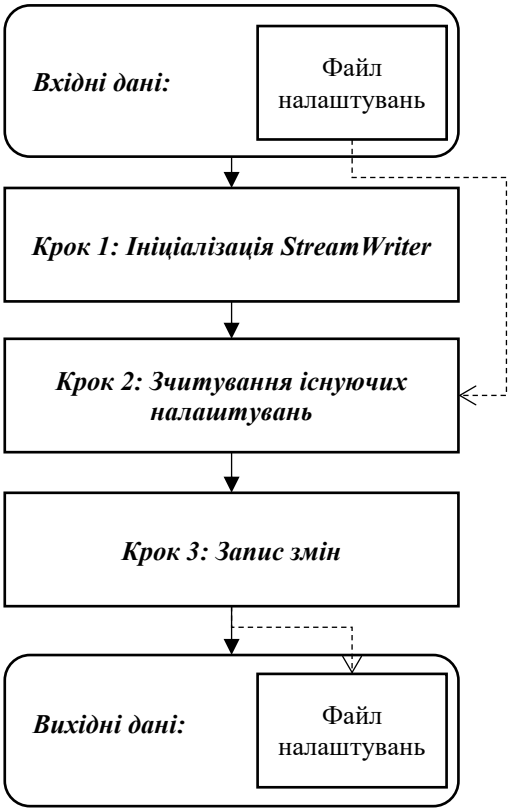


Рисунок 4.6 – Метод SaveSettings()

Файл налаштувань використовується класами TrainWindow та TestWindow для завантаження параметрів шляхів до необхідних класів ModelTrain та ImageTest() та необхідної бібліотеки.

Клас Menu призначений для переходу між модулями TrainWindow, TestWindow і Settings, використовуючи UI та має набір методів що відповідають елементам інтерфейсу для завантаження відповіднох форми.

Таким чином, було описано прикладні особливості розробки компонентів інформаційної системи, що включає наведенні класи, які реалізують заявлений

функціонал системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання.

4.2 Прикладне тестування інформаційної системи

Для перевірки коректності роботи прикладного застосунку інформаційної системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання було проведено тестування усіх модулів системи.

Першим тестовим кейсом є перевірка існування файлу налаштувань та коректного його завантаження, в випадку відсутності цього файлу потрібно використовувати налаштування по замовчуванням. Кроки тест-кейса наведено у таблиці 4.1.

Таблиця 4.1 – Тест-кейс 001

Тест-кейс ID: 001	**Пріоритет: 1**	Жарновський О.В.
Назва: Перевірка коректного завантаження файлу налаштувань **Вхідні дані:** Нічого		
Кроки		**Очікуваний результат**
1. Видалити файл налаштувань; 2. Відкрити застосунок; 3. Обрати опцію Settings із стартового меню.		Відкрився застосунок Відкрилась підсистема налаштувань та повідомлення про відсутність файлу налаштувань і його генерації по-замовчуванню
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу К001 наведений на рисунку 4.7.

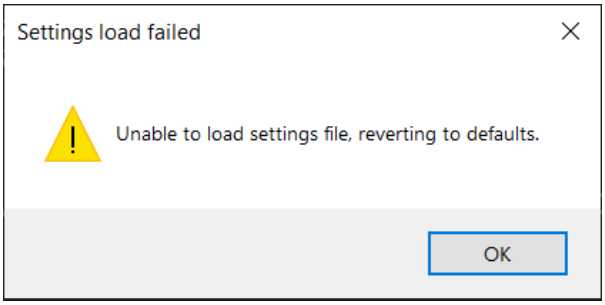


Рисунок 4.7 – Результат виконання тест-кейсу 001

Наступним тестовим кейсом буде перевірка коректного завантаження файлу моделі класу TestWindow. Кроки тест-кейса подано у таблиці 4.2.

Таблиця 4.2 – Тест-кейс 002

Тест-кейс ID: 002	**Пріоритет: 1**	Жарновський О.В.
Назва: Перевірка коректного завантаження файлу налаштувань **Вхідні дані:** Нічого		
Кроки	**Очікуваний результат**	
1. Відкрити застосунок; 2. Перейти на підсистему тестування зображень; 3. Натиснути на завантаження моделі; 4. Вийти із інтерфейсу завантаження без обраного файлу.	Відкрився застосунок Відкрилась підсистема тестування зображень, з’явився відповідний інтерфейс завантаження файлу, з’явилось повідомлення про помилку завантаження моделі	
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу 002 наведено на рисунку 4.8.

Наступним тестовим випадком буде перевірка коректності завантаженого файлу моделі класу TestWindow. Кроки тест-кейса наведено у таблиці 4.3.

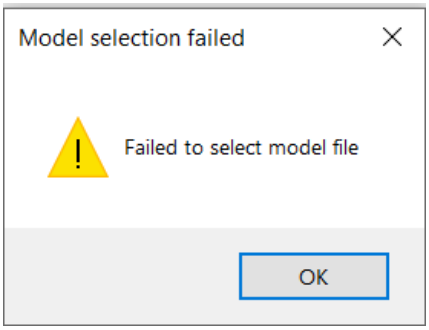


Рисунок 4.8 – Результат виконання тест-кейсу К002

Таблиця 4.3 – Тест-кейс 003

Тест-кейс ID: 003	**Пріоритет: 1**	Жарновський О.В.
Назва: Перевірка достовірності завантаженої моделі **Вхідні дані:** Тестовий файл зображення, пустий txt файл		
Кроки		**Очікуваний результат**
1. Відкрити застосунок; 2. Перейти на підсистему тестування зображень; 3. Завантажити файл зображення; 4. Завантажити пустий txt файл.		Відкрився застосунок Відкрилась підсистема тестування зображень, з'явився відповідний інтерфейс завантаження файлу, з'явилось повідомлення про помилку завантаження моделі
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу 003 наведено на рисунку 4.9.

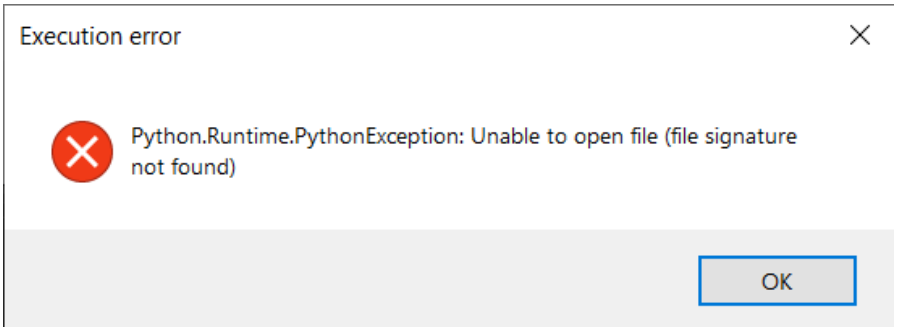


Рисунок 4.9 – Результат виконання тест-кейсу 003

Наступним тестовим випадком буде перевірка коректності параметрів конвертування зображення класу TestWindow. Кроки тест-кейса наведено у таблиці 4.4.

Таблиця 4.4 – Тест-кейс 004

Тест-кейс ID: 004	**Пріоритет: 1**	Жарновський О.В.
Назва: Перевірка достовірності параметрів трансформації **Вхідні дані:** Тестовий файл зображення, тестовий файл моделі		
Кроки	**Очікуваний результат**	
1. Відкрити застосунок; 2. Перейти на підсистему тестування зображень; 3. Вказати параметр розміру зображення 10000.	Відкрився застосунок Відкрилась підсистема тестування зображень, параметри автоматично зменшилися до максимального значення та з’явилося відповідне повідомлення про помилку.	
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу К004 наведено на рисунку 4.10.

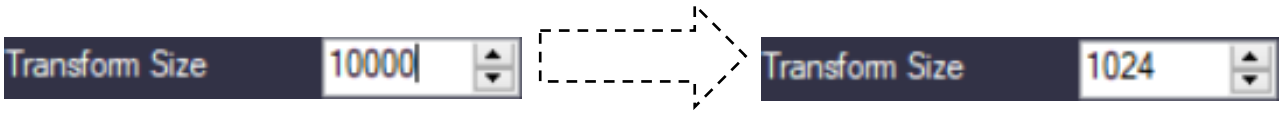


Рисунок 4.10 – Результат виконання тест-кейсу 004

Наступним тестовим випадком буде перевірка розширення інтерфейсу для створення користувацької архітектури моделі класу TrainWindow. Кроки тест-кейса наведено у таблиці 4.5.

Таблиця 4.5 – Тест-кейс 005

Тест-кейс ID: 005	**Пріоритет: 2**	Жарновський О.В.
Назва: Перевірка розширення інтерфейсу **Вхідні дані:** Тестовий файл зображення, тестовий файл моделі		
Кроки		**Очікуваний результат**
1. Відкрити застосунок; 2. Перейти на підсистему тренування моделей; 3. Обрати кількість шарів 11; 4. Натиснути на «Змінити».		Відкрився застосунок Відкрилась підсистема тренування моделей, інтерфейс був успішно розширений
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу 005 наведено на рисунку 4.11.

Рисунок 4.11 – Результат виконання тест-кейсу 005

Також потрібно перевірити динамічне редагування архітектури нейромережі класу TrainWindow. Кроки тест-кейса наведено у таблиці 4.6.

Таблиця 4.6 – Тест-кейс 006

Тест-кейс ID: 006	**Пріоритет: 2**	Жарновський О.В.
Назва: Перевірка динамічної зміни інтерфейсу **Вхідні дані:** Тестовий файл зображення, тестовий файл моделі		
Кроки		**Очікуваний результат**
1. Відкрити застосунок; 2. Перейти на підсистему тренування зображень; 3. Обрати кількість шарів 11; 4. Натиснути на «Змінити»; 5. Обрати кількість шарів 5; 6. Натиснути на «Змінити».		Відкрився застосунок Відкрилась підсистема тренування моделей, інтерфейс був успішно розширений, інтерфейс був успішно змінений
Результат виконання тест-кейсу: пройдено успішно		

Результат успішного виконання тест-кейсу 006 наведено на рисунку 4.12.

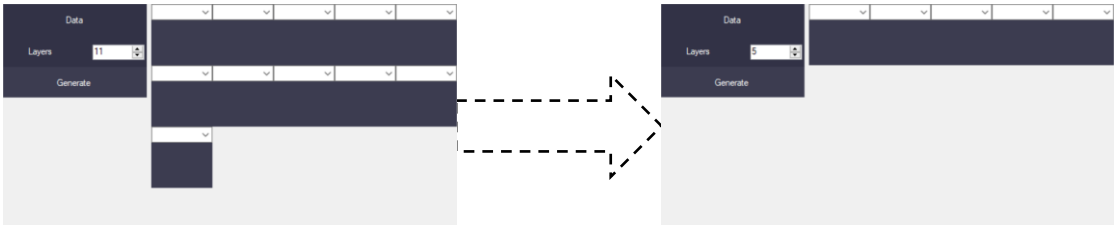


Рисунок 4.12 Результат виконання тест-кейсу 006

Таким чином, було проведено тестування розробленого програмного застосунку, що реалізує метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Під час тестування був перевірений функціонал, що працює згідно поставлених завдань без випадків некоректної роботи.

4.3 Особливості використання інформаційної системи

Інформаційна система нейромережевого аналізу згенерованих зображень людей складається із чотирьох підсистем – двох основних та двох допоміжних. По замовчуванню завантажується підсистема меню.

Допоміжна підсистема меню (рисунок 4.13), що слугує для переходу між іншими підсистемами, для чого користувачеві потрібно натиснути на відповідний елемент інтерфейсу до бажаної підсистеми, що розширить вікно та додасть нові елементи керування.

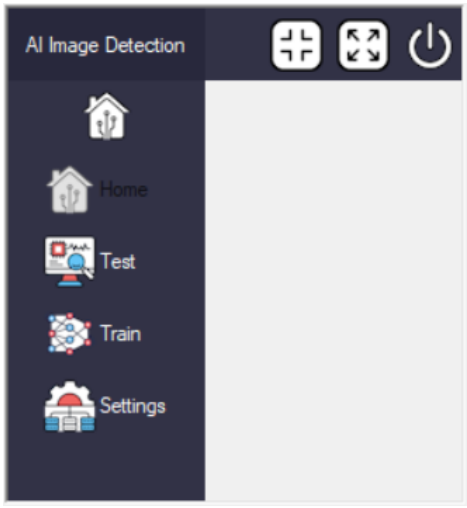


Рисунок 4.13 – Інтерфейс підсистеми меню

Основна підсистема аналізу зображень (рисунок 4.14) дозволяє користувачеві завантажити обране зображення обличчя людини та використовуючи вже навчену мережу отримати прогноз щодо його достовірності.

Для цього користувачеві потрібно завантажити зображення, натиснувши відповідну кнопку «Upload», що викликає контексте меню навігації де потрібно обрати файл зображення. Далі потрібно заповнити поля з параметрами, де:

– transform size: розмір трансформації зображення, залежіть від архітектури мережі;

– normalize: чи використовувати нормалізацію зображення;

– path: шлях до зображення, обраного раніше;

– modelpath: шлях до файлу моделі, що буде використовуватися для аналізу;

– datapath: шлях до класів, на які будуть класифікувати зображення;

– use method mode: крім аналізу на достовірність користувач може додатково отримати можливе походження зображення.

– model path, datapath: у випадку якщо був обраний параметр use method model, потрібно додатково вказати аналогічні параметри мережі для класифікації методів походження.

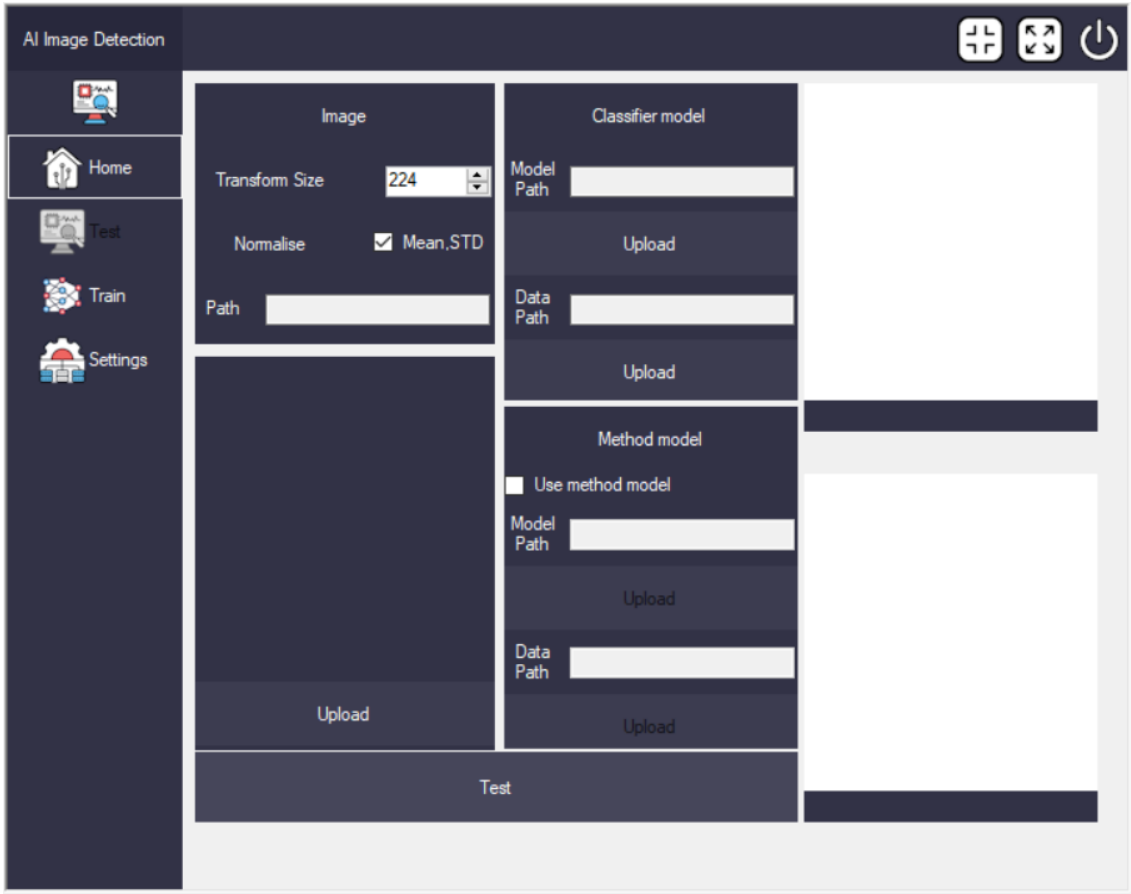


Рисунок 4.14 – Підсистема аналізу зображень

Після заповнення наведених полів користувачеві потрібно натиснути кнопку «Test» що виведе результат у вигляді пари класу та графіку для кожної мережі (рисунок 4.15). На рисунку продемонстрований результат аналізу

валідаційного зображення з результатом в 60% що зображення є згенерованим та 0.99% що це зроблено використовуючи StyleGan.

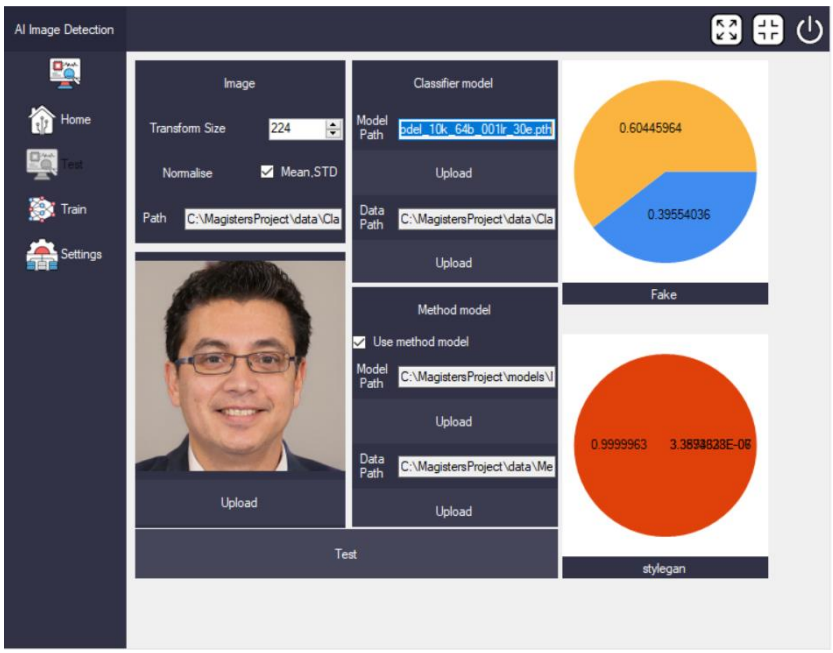


Рисунок 4.15 – Результат виконання

Основна підсистема взаємодії з НМ (рисунок 4.16) дозволяє користувачеві натренувати власну мережу використовуючи введені параметри, включаючи зміну шарів.

Рисунок 4.16 – Підсистема взаємодії з НМ

Для тренування мережі користувачеві потрібно ввести наступні параметри:

– path: шлях до папки з датасетом, що має містити папки «train», «val» та «test» з зображеннями, розподіленими на відповідні класи;

– transform size: розмір до якого будуть змінені усі завантажені зображення;

– normalize: чи використовувати нормалізацію зображення;

– batch size: розмір груп, на які буде розподілений датасет;

– hidden units: параметр що відповідає за кількість прихованих шарів;

– epochs: скільки епох буде навчатися мережа;

– learning rate: коефіцієнт навчання мережі;

– name: назва файлу, що буде збережена мережа, має мати розширення pt чи pth.

У випадку якщо користувач хоче використовувати власну архітектуру мережі, потрібно обрати параметр «Architecture:custom» що розширить форму та дозволить додавати шари з встановленими параметрами.

По завершенню налаштування, користувачеві потрібно натиснути кнопку «Train», що почне процес навчання. Створений системою файл «Log.txt» дозволяє відслідковувати прогрес тренування мережі з проміжними параметрами точності та втрат, також по його завершенню файл буде містити результати тестування мережі.

Допоміжна підсистема налаштувань (рисунок 4.17) дозволяє користувачеві змінювати шляхи до зовнішніх модулів та бібліотеки.

Для того щоб змінити шлях користувачеві потрібно натиснути відповідну кнопку «Upload», що викликає контекстне меню для вибору файлу. Після завершення змін потрібно натиснути кнопку «Save», що створить або замінить файл «Settings.txt» з змереженими даними, та перезавантажити додаток.

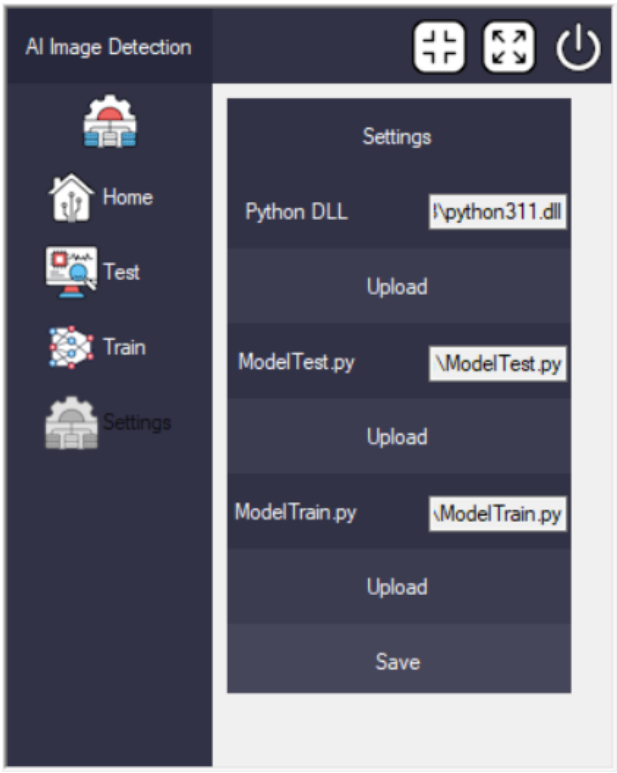


Рисунок 4.17 – Підсистема налаштувань

Таким чином були описані особливості використання системи нейромережевого аналізу згенерованих зображень людей складається із двох основних та двох допоміжних підсистем, та описаний їх основний функціонал.

4.4 Дослідження ефективності та інтерпретація отриманих результатів

Для дослідження ефективності інформаційної системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання було використано два набори тестових даних, що складаються із 2000 зображень розділених на «згенеровані» та «реальні», та 600 зображень розділених на методи походження.

Під час дослідження було натреновані три нейромережі ImageClassifier для ідентифікації зображень та MethodClassifier для ідентифікації походження з параметрами наведеними на таблиці 4.7. В якості вихідних метрик були використані параметри точності та втрат під час навчання та під час тренування.

Таблиця 4.7 – Параметри навчання нейромереж

		Parameters				
		Epochs	Learning Rate	Batch size	Dropout	Normalization
ImageClassifier	Model_1_0	10	0.001	32	0.5	None
	Model_1_1	10	0.001	64	0.5	Mean, STD
	Model_1_2	10	0.005	64	0.5	Mean, STD
	Model_1_3	30	0.001	64	0.3	Mean, STD
MethodClassifier	Model_2_0	10	0.001	32	0.5	None
	Model_2_1	10	0.005	64	0.5	Mean, STD
	Model_2_2	20	0.001	64	0.5	Mean, STD

Тестування моделі Model_1_0 класу ImageClassifier з наступними вхідними параметрами:

– кількість епох: 10;

– коефіцієнт навчання: 0.001;

– розмір групи: 32;

– відсоток втрат на повноз’єдному шарі: 0.5;

В результаті тренування були отримані графіки точності та втрат (рисунок 4.18). Кінцеві параметри точності мережі становлять 0.57 та 0.64 для валідації.

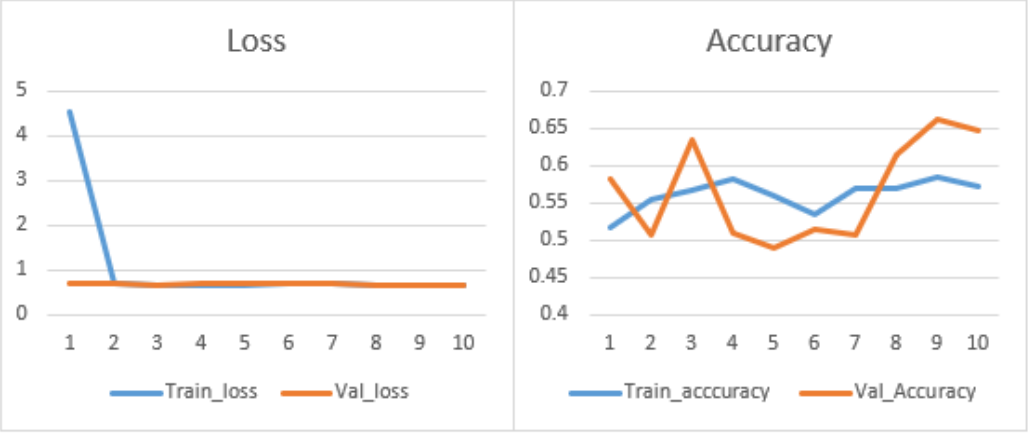


Рисунок 4.18 – Результат тренування Model_1_0

В результаті тестування була отримана матриця сплутаності (таблиця 4.8) та наступні метрики Accuracy: 0.628, Precision: 0.324, recall: 0.839, f1:0.467.

Таблиця 4.8 – Матриця сплутаності мережі

True label	0	324	676
	1	62	922
		0	1
		Predicted label	

Помічене мінімальне відхилення параметрів, мережа була недотренована. Для наступної мережі був подвоєний розмір груп та додана нормалізація з метою покращення результату.

Тестування моделі Model_1_1 класу ImageClassifier з наступними вхідними параметрами:

– кількість епох: 10;

– коефіцієнт навчання: 0.001;

– розмір групи: 64;

– відсоток втрат на повноз’єдному шарі: 0.5;

– нормалізація mean[0.485,0456,0.406], std[0.229,0.224,0.225].

В результаті тренування були отримані графіки точності та втрат (рисунок 4.19). Кінцеві параметри точності становлять 0.718 і 0.72 для валідації.

Рисунок 4.19 – Результат тренування Model_1_1

В результаті тестування була отримана матриця сплутаності (таблиця 4.9) та наступні метрики Accuracy: 0.741, Precision: 0.751, recall: 0.739, f1:0.745.

Таблиця 4.9 – Матриця сплутаності мережі

True label	0	751	249
	1	264	720
		0	1
		Predicted label	

Дана ітерація моделі має значно більше відхилення від початкових значень та в загалом кращі параметри при тренуванні та тестуванні. Наступним кроком було значно збільшено коефіцієнт навчання.

Тестування моделі Model_1_2 класу ImageClassifier з наступними вхідними параметрами:

– кількість епох 10;

– коефіцієнт навчання 0.005;

– розмір групи 64;

– відсоток втрат на повноз’єдному шарі: 0.5;

– нормалізація mean[0.485,0456,0.406], std[0.229,0.224,0.225].

В результаті тренування були отримані графіки точності та втрат (рисунок 4.20). Кінцеві параметри точності мережі становлять 0.53 та 0.54 для валідації.

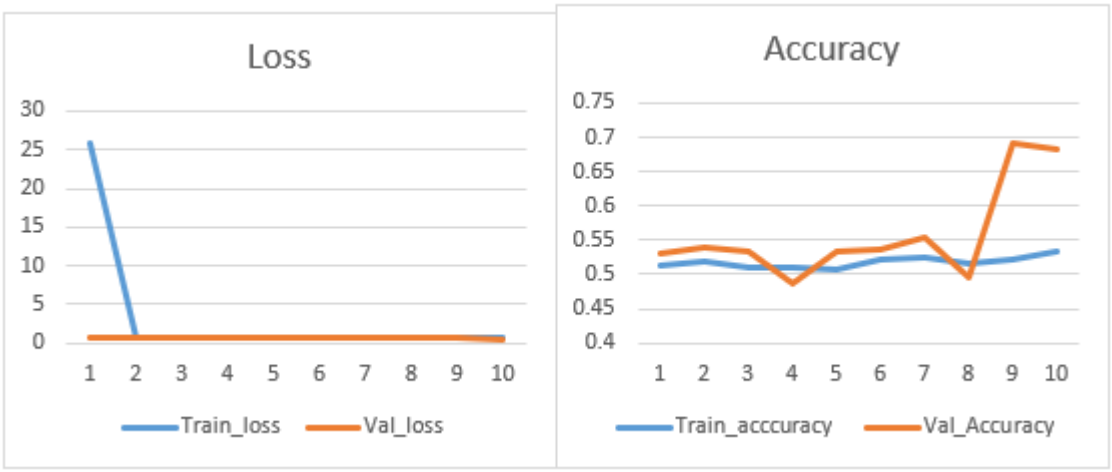


Рисунок 4.20 – Результат тренування та тестування Model_1_2

В результаті тестування була отримана матриця сплутаності (таблиця 4.10) та наступні метрики Accuracy: 0.553, Precision: 0.702, recall: 0.544, f1:0.613.

Таблиця 4.10 – Матриця сплутаності мережі

True label		0	1
	0	702	298
	1	588	396
		0	1
		Predicted label	

Дана ітерація моделі має значно гірше відхилення та погані метрики при тренуванні та тестування. Наступна ітерація буде використовувати попередній коефіцієнт навчання, але буде зменшений відсоток втрат на повно’єдному шарі та збільшена кількість епох.

Тестування моделі Model_1_3 класу ImageClassifier з наступними вхідними параметрами:

– кількість епох 30;

– коефіцієнт навчання 0.005;

– розмір групи 64;

– відсоток втрат на повноз’єдному шарі: 0.3;

– нормалізація mean[0.485,0456,0.406], std[0.229,0.224,0.225].

В результаті тренування були отримані графіки точності та втрат (рисунок 4.21). Кінцеві параметри точності становлять 0.94 та 0.92 для валідації.

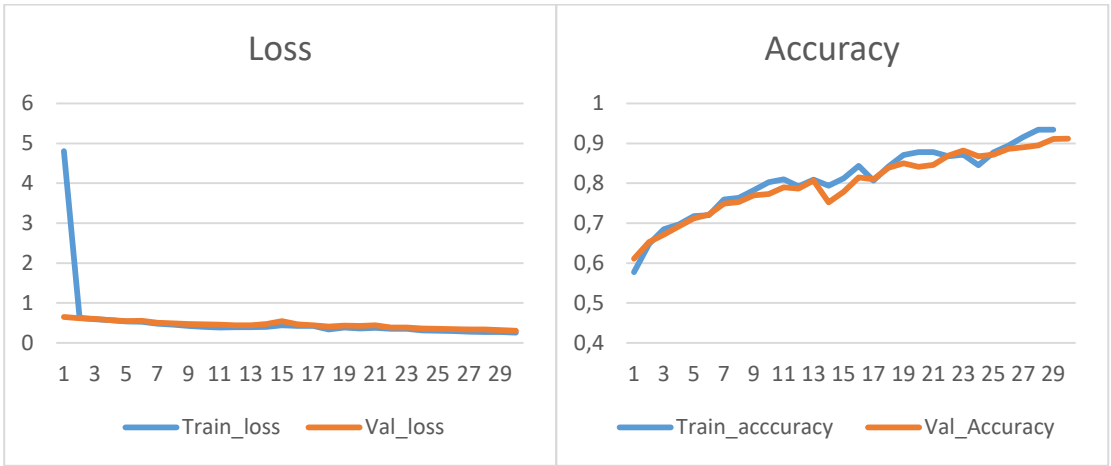


Рисунок 4.21 – Результат тренування та тестування Model_1_3

В результаті тестування була отримана матриця сплутаності (таблиця 4.11) та наступні метрики Accuracy: 0.923, Precision: 0.927, recall: 0.923, f1: 0.295.

Таблиця 4.11 – Матриця сплутаності мережі

True label		0	1
	0	937	73
	1	78	889
		0	1
		Predicted label	

Ця модель має задовільні метрики точності під час тренування та валідації, а також задовільні характеристики під час тестування. Збільшення кількості епох та зменшення відсотку втрат позитивно сприяли на результат, але подальшому збільшенні кількості епох мережа не дала видимих покращень при значних витратах у часі, тому на цьому ітерування ImageClassifier було завершено.

Наступною мережею для тренування та тестування була взята мережа класу MethodClasifier під назвою Model_2_0 з наступними параметрами:

- кількість епох 10;
- коефіцієнт навчання 0.001;
- розмір групи 32;
- відсоток втрат на повноз’єднаному шарі: 0.5.

В результаті тренування були отримані графіки точності та втрат (рисунок 4.22). Кінцеві параметри точності мережі становлять 0.94 та 0.833 для валідації.

В результаті тестування була отримана матриця сплутаності (таблиця 4.12) та наступні метрики Accuracy: 0.859, Precision: 0.966, recall: 0.637, f1:0.767.

На відміну від Model_1_0, мережа Model_2_0 не вгадує правильну відповідь під час тренування, але метрики все ще не є оптимальними. Наступна версія, схоже до Model_1_2, має збільшений розмір групи та коефіцієнт навчання для перевірки закономірності.

Таблиця 4.12 – Матриця сплутаності мережі

True label	0	58	0	2
	1	29	70	1
	2	4	0	92
		0	1	2
		Predicted label		

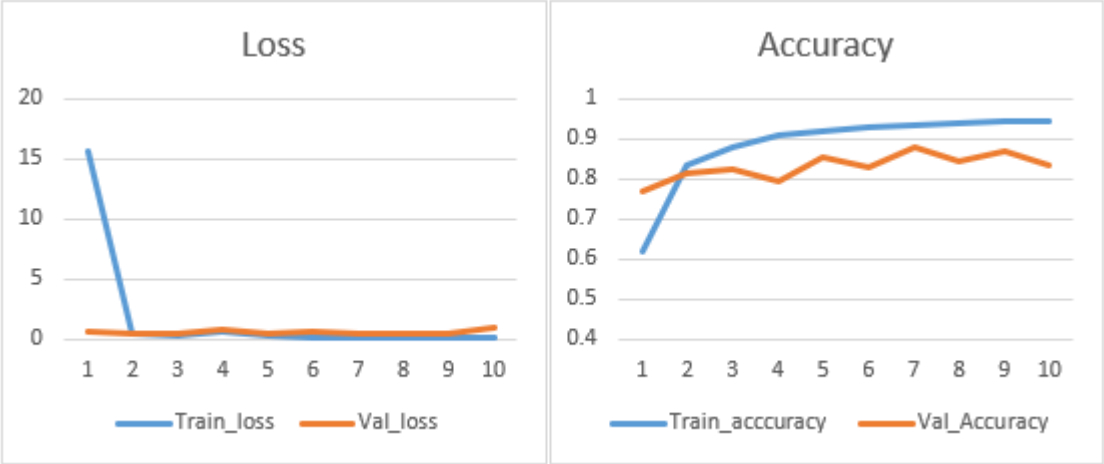


Рисунок 4.22 – Результат тренування та тестування Model_2_0

Тестування мережі Model_2_1 класу MethodClasifier з наступними параметрами:

– кількість епох 10;

– коефіцієнт навчання 0.005;

– розмір групи 64;

– відсоток втрат на повноз’єдному шарі: 0.5;

– нормалізація mean[0.485,0456,0.406], std[0.229,0.224,0.225].

В результаті тренування були отримані графіки точності та втрат (рисунок 4.23). Кінцеві параметри точності мережі становлять 0.83 та 0.75 для валідації.

В результаті тестування була отримана матриця сплутаності (таблиця 4.13) та наступні метрики Accuracy: 0.82, Precision: 0.833, recall: 0.632, f1:0.718.

Таблиця 4.13 – Матриця сплутаності мережі

True label	0	50	1	9
	1	21	73	6
	2	8	1	87
		0	1	2
		Predicted label		

Рисунок 4.23 – Результат тренування та тестування Model_2_1

Схоже до Model_1_2, модель Model_2_1 демонструє гірші та нестабільні показники під час тренування та тестування. Для наступної ітерації вирішено використати 20 епох, через досить високі показники початкової моделі, також на відміну від Model_1_3, шар втрат залишається незмінним.

Тестування моделі Model_1_2 класу MethodClassifier з наступними вхідними параметрами:

– кількість епох 20;

– коефіцієнт навчання 0.001;

– розмір групи 64;

– нормалізація mean[0.485,0456,0.406], std[0.229,0.224,0.225].

В результаті тренування кінцева точність мережі під час тренування становить 0.98 та під час тестування 0.95 (рисунок 4.24).

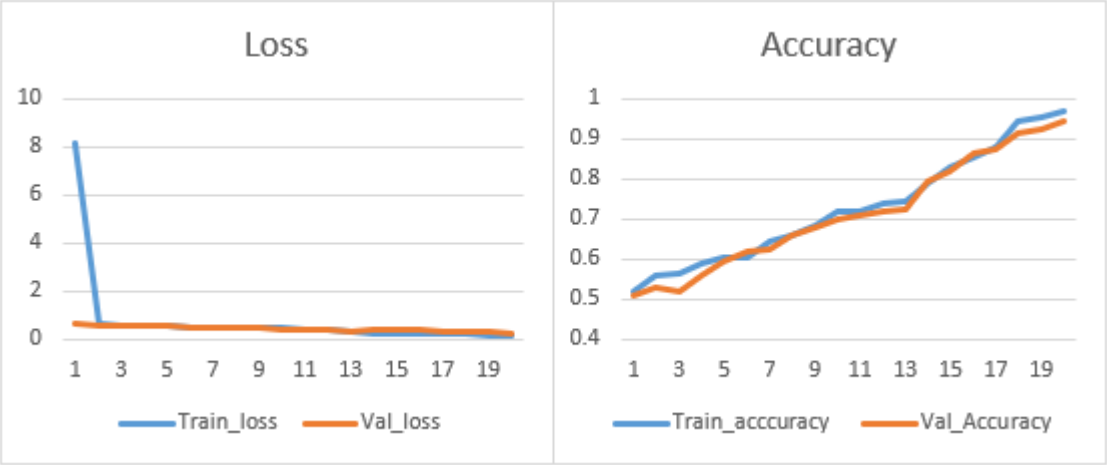


Рисунок 4.24 – Результат тренування та тестування Model_2_2

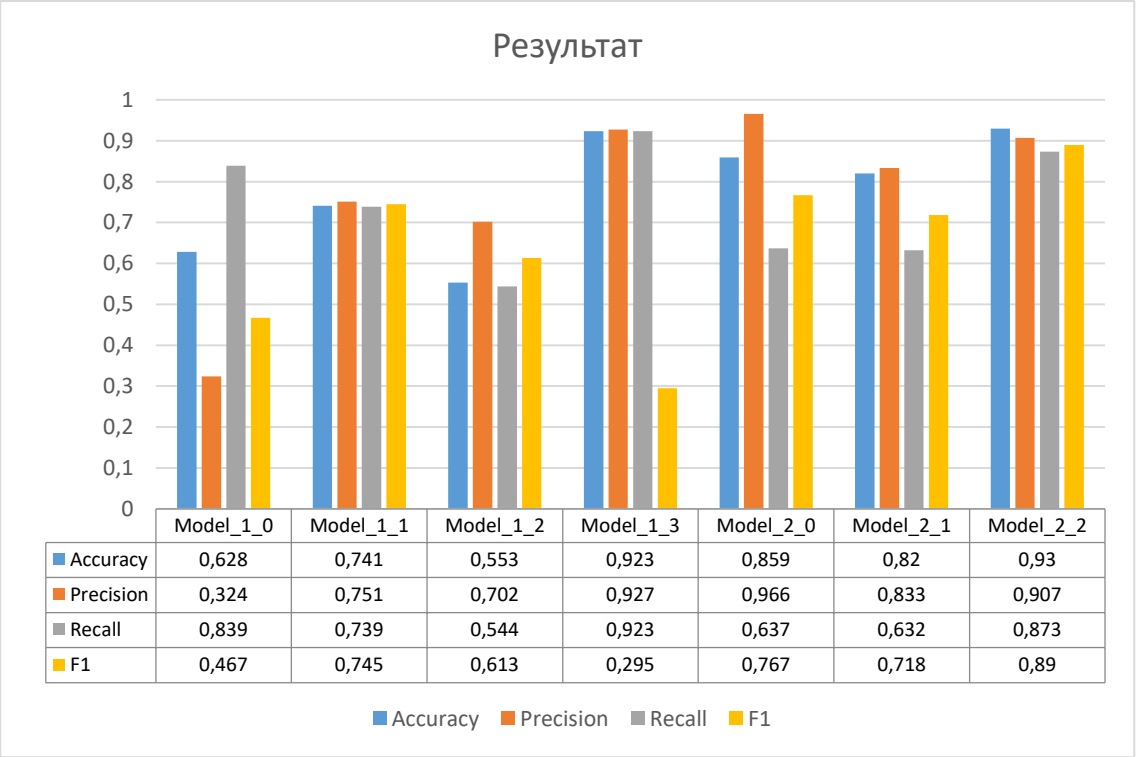
В результаті тестування була отримана матриця сплутаності (таблиця 4.14) та наступні метрики Accuracy: 0.93, Precision: 0.907, recall: 0.873, f1:0.89.

Таблиця 4.14 – Матриця сплутаності мережі

True label	0	69	2	5
	1	6	78	4
	2	4	1	89
		0	1	2
		Predicted Label		

Третя ітерація моделі має задовільні відсотки точності та втрат під час тестування, збільшення кількості епох та додавання нормалізації позитивно сприяли на результат, тому на цьому ітерування MethodClassifier було завершено.

В результаті тренування були отримані результати наведені на рисунку 4.25.



	Model_1_0	Model_1_1	Model_1_2	Model_1_3	Model_2_0	Model_2_1	Model_2_2
Accuracy	0,628	0,741	0,553	0,923	0,859	0,82	0,93
Precision	0,324	0,751	0,702	0,927	0,966	0,833	0,907
Recall	0,839	0,739	0,544	0,923	0,637	0,632	0,873
F1	0,467	0,745	0,613	0,295	0,767	0,718	0,89

Рисунок 4.25 – Результати дослідження

Отже, було проведено дослідження ефективності та інтерпретації отриманих результатів, в ході чого продемонстровано що моделі класифікації методу мають загалом кращий результат порівняно із ідентифікацією зображення. Також, під час тестування виявлено покращення результатів при збільшенні кількості зображень для тренування мережі, збільшення коефіцієнтів та використанні нормалізації. Із представлених мереж найкращі результати мають мережі Model_1_3 та Model_2_2 з результатом accuracy в 92% та 93% при тестуванні з експериментальною вибіркою в 2000 та 600 зображень.

Висновки до четвертого розділу

Під час написання четвертого розділу було виконано дослідження ефективності інформаційної системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання, в рамкая якого:

1. Описано прикладні особливості розробки компонентів інформаційної системи, що складається із чотирьох підсистем та шести класів, які реалізофують

заявлений фукціонал інформаційної системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання.

2. Проведено прикладне тестування інформаційної системи нейромережевого аналізу згенерованих зображень людей засобами машинного навчання використовуючи тест-кейси. Під час тестування був перевірений функціонал що працює згідно поставлених завдань без випадків некоректної роботи.

3. Виконано дослідження ефективності, яке показало, що моделі класифікації методу мають загалом кращий результат порівняно з ідентифікацією зображення. Також, під час тестування спостережено покращення результатів при збільшенні кількості зображень для тренування мережі, збільшення коефіцієнів та використанні нормалізації. Із представлених мереж найкращі результати мають мережі Model_1_3 та Model_2_2 з результатом accuracy в 92% та 93% при тестуванні з експериментальною вибіркою в 2000 та 600 зображень.

Загальні висновки

Кваліфікаційна робота магістра розв’язує задачу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Результатом роботи є метод та його програмна реалізація, призначена для ідентифікації зображень завантажених користувачем, що використовує комбінацію двох згорткових нейронних мереж, та працює на основі перетворення вхідних даних – зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані – відсоткова оцінка автентичності зображення та його походження. З практичним використанням розробленого методу досягається підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання, що визначає досягнення мети кваліфікаційної роботи магістра.

Для досягнення мети дослідження було виконано:

1. Досліджено сучасний стан предметної області генерацій зображень з використанням штучного інтелекту, їх методи та засоби. Виконано аналіз сучасних наукових публікацій у задачах генерації та виявлення зображень створених штучним інтелектом.

2. Розроблено метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Розроблений метод забезпечує визначення автентичності зображення за допомогою відсоткової оцінки та визначення можливих методів використаних для генерації зображення з використанням навченої згорткової нейронної мережі.

3. Створено прикладну реалізацію методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

4. Досліджено практичну ефективність застосування методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Результати виконання кваліфікаційної роботи магістра містять інновації та наукову новизну, зокрема було створено новий метод ідентифікації

згенерованих штучним інтелектом зображень людей засобами машинного навчання, що дозволяє автоматизовано аналізувати завантажене зображення за запитом користувачів, виконуючи при цьому як і аналіз автентичності зображення, так і знаходячи можливі методи його генерації, перетворюючи вхідні дані у вигляді зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані – відсоткова оцінка автентичності зображення та його походження.

Розроблений у кваліфікаційній роботі метод має ряд переваг у порівнянні з існуючими методами, дозволяючи ідентифікувати не тільки зображення, а й методи його генерації. Це дозволяє краще аналізувати методи генерації штучного інтелекту для подальшого покращення ефективності. Із представлених нейромереж найкращі результати мають мережі з результатом accuracy в 92% та 93% при тестуванні з експериментальною вибіркою в 2000 та 600 зображень.

За темою кваліфікаційної роботи магістра автором виконано 4 наукові публікації. Основні наукові й практичні результати роботи доповідались у доповіді «Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning» на XLV Міжнародній науково-практичній конференції «Key Aspects of the Development of Scientific Research in Modern Conditions» (Constanta, Romania) 1 листопада 2024 року та у доповіді «Практична реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання» на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп’ютерних наук АПКН-2024» (м. Хмельницький) 15-16 листопада 2024 року [48, 49, 50].

Також наукова робота в межах кваліфікаційної роботи магістра на тему «Neural Network Method for Detection of Fake Document Images for Personality Identification Systems» зайняла призове ІІІ місце на фінальному етапі Міжнародного конкурсу студентських наукових робіт «Black Sea Science 2024» з роботою, січень – березень 2024 року (Додаток Д).

Перелік посилань

1. Tess. Good Use Cases for AI Image Generators. URL: https://www.tess.design/blog/good-use-cases-for-generative-ai

2. Softblues. How Generative AI Is Changing Creative Work. URL: https://softblues.io/blog/how-does-ai-image-generation-work/

3. Altexsoft. AI Image Generation Explained: Techniques, Applications, and Limitations. URL: https://www.altexsoft.com/blog/ai-image-generation/

4. HBR. How Generative AI Is Changing Creative Work. URL: https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work

5. Aidocmaker. Use-Cases of Text to Image for AI Image Generation. URL: https://www.aidocmaker.com/blog/4-use-cases-of-text-to-image-for-ai-image-generation

6. Conroy Creative Counsel. The Pros and Cons of AI Images. URL: https://conroycreativecounsel.com/the-pros-and-cons-of-ai-images/

7. Medium. The limitations of AI-generated photorealistic images. URL: https://medium.com/@tracyyxchen/the-limitations-of-ai-generated-photorealistic-images-23635c0186ce

8. Tech4Future. Image Recognition: The Limitations of AI Compared to Human Vision. URL: https://tech4future.info/en/image-recognition-limitations/

9. NPR. AI-generated images are everywhere. Here's how to spot them. URL: https://www.npr.org/2023/06/07/1180768459/how-to-identify-ai-generated-deepfake-images

10. Medium. The Future of AI-Generated Images: Challenges and Possibilities. URL: https://medium.com/@mukundvyas0/the-future-of-ai-generated-images-challenges-and-possibilities-5ee533d48f3e

11. Medium. Generative Adversarial Networks. URL: https://medium.com/@marcodelpra/generative-adversarial-networks-dba10e1b4424

12. Linkedin. A Look Inside of Generative AI Visual Use Cases. URL: https://www.linkedin.com/pulse/look-inside-generative-ai-visual-use-cases-stu-tek-ybbzf

13. Ramzan S., Iqbal M. M., Kalsum T. Text-to-Image generation using deep learning. Engineering Proceedings, 20(1), 2022, pp. 16.

14. TechTarget Network. What is image-to-image translation?. URL: https://www.techtarget.com/searchenterpriseai/definition/image-to-image-translation

15. Ahmed R., Nahla hosny B. Sketch to Image Using Generative Adversarial Networks (GAN). URL: https://www.researchgate.net/publication/362481944_Sketch_to_Image_Using_Generative_Adversarial_Networks_GAN

16. Medium. Sketch-guided Image Generation with Stable Diffusion. URL: https://medium.com/@geronimo7/sketch-guided-stable-diffusion-a-tutorial-fb25bc69ddb5

17. Ramzan S., Iqbal M. M., Kalsum T. Text-to-Image generation using deep learning. Engineering Proceedings, 20(1), 2022, pp. 16.

18. DeterminedAI. How does Video Generation work?. URL: https://www.determined.ai/blog/how-does-video-gen-work

19. Bonny M. Z., Uddin M. S. A technique for panorama-creation using multiple images. International Journal of Advanced Computer Science and Applications, 11(2), 2020.

20. Medium. DALL·E: Creating Images from Text (by OpenAI). URL: https://farzanaanjum.medium.com/dall-e-creating-images-from-text-by-openai-b9da6d785f6

21. Sii. A brief history of AI-powered image generation. URL: https://sii.ua/blog/en/a-brief-history-of-ai-powered-image-generation/

22. Bianco, T., Castellano, G., Scaringi, R., Vessio, G. Identifying AI-Generated Art with Deep Learning. In CREAI@ 2023, pp. 16-25.

23. Martin-Rodriguez F., Garcia-Mojon R., Fernandez-Barciela M. Detection of AI-created images using pixel-wise feature extraction and convolutional neural networks. Sensors, 23(22), 2023, pp. 9037.

24. Rewatkar S. Analyzing and Improving Existing Neural Network-Based Approaches to Identify AI Generated Images. Journal of Student Research, 2024, 13(1).

25. Medium. Understanding the Difference between Training, Test, and Validation Sets in Machine Learning. URL: https://medium.com/syntaxerrorpub/understanding-the-difference-between-training-test-and-validation-sets-in-machine-learning-c59feec6483b

26. Kaggle. Flickr-Faces-HQ Dataset (Nvidia) - Part 1. URL: https://www.kaggle.com/datasets/xhlulu/flickrfaceshq-dataset-nvidia-part-1

27. Kaggle. 1 Million Fake Faces – 1 dataset. URL: https://www.kaggle.com/datasets/tunguz/1-million-fake-faces

28. Kaggle. ace Dataset Using Stable Diffusion v.1.4. URL: https://www.kaggle.com/datasets/bwandowando/faces-dataset-using-stable-diffusion-v14

29. Kaggle. Real vs Fake Faces dataset. URL: https://www.kaggle.com/datasets/uditsharma72/real-vs-fake-faces

30. Medium. Convolutional Neural Networks, Explained. URL: https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

31. Byers B., Sheta A. Design of Convolutional Neural Networks for Fish Recognition and Tracking. Artificial Intelligence and Machine Learning AIML, 22(1), 2023, pp. 1-9.

32. PyImageSearch. Convolutional Neural Networks (CNNs) and Layer Types. URL: https://pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/

33. DigitalOcean. Batch Normalization in Convolutional Neural Networks. URL: https://www.digitalocean.com/community/tutorials/batch-normalization-in-convolutional-neural-networks

34. Baeldung. How ReLU and Dropout Layers Work in CNNs. URL: https://www.baeldung.com/cs/ml-relu-dropout-layers

35. Iosifidis A., Tefas A. Deep learning for robot perception and cognition. Academic Press 2022, pp. 17-34

36. Medium. Overview of a Neural Network's Learning Process. URL: https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa

37. GeekForGeeks. Supervised and Unsupervised learning. URL: https://www.geeksforgeeks.org/supervised-unsupervised-learning/

38. Python. URL: https://www.python.org/

39. C# .NET. URL: https://learn.microsoft.com/en-us/dotnet/csharp/

40. VS Code. URL: https://code.visualstudio.com/

41. DistantJob. Visual Studio vs Visual Studio Code: What's the Key Difference? URL: https://distantjob.com/blog/visual-studio-vs-visual-studio-code/

42. Microsoft Visual Studio. URL: https://visualstudio.microsoft.com/

43. Pytorch. URL: https://pytorch.org/vision/stable/index.html

44. Numpy. URL: https://numpy.org/

45. Pathlib. URL: https://docs.python.org/3/library/pathlib.html

46. PIL. URL: https://pypi.org/project/pillow/

47. PythonNet. URl: https://github.com/pythonnet/pythonnet

48. Zharnovskyi O., Mazurets O., Sobko O. Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning. Key Aspects of the Development of Scientific Research in Modern Conditions. Proceedings of the XLV International scientific and practical conference. October 30 – November 1, 2024. Constanta, Romania. 2024. Pp. 69-73.

49. Жарновський О.В., Казмірчук Я.М., Собко О.В., Мазурець О.В. Практична реалізація методу ідентифікації згенерованих штучним інтелектом

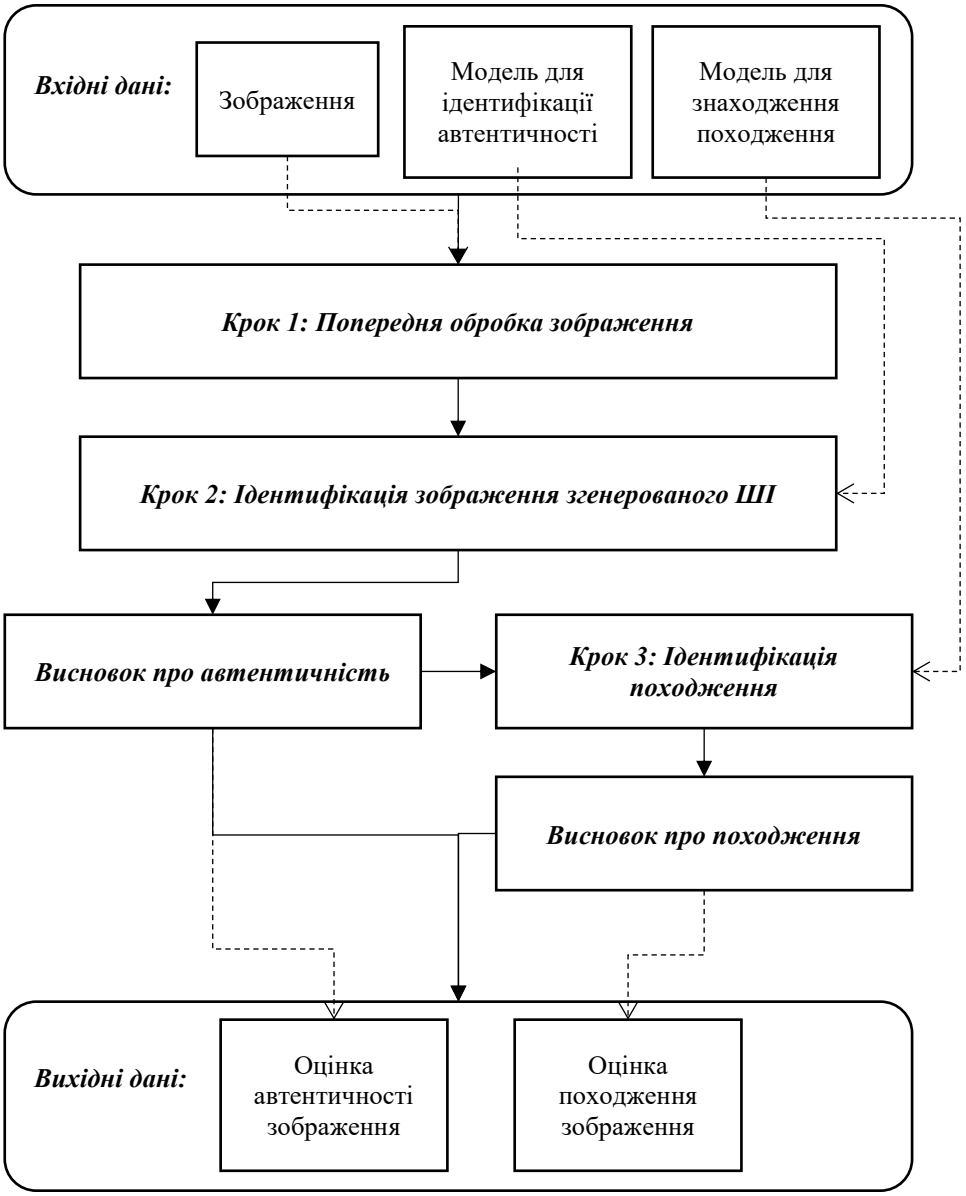
зображень людей засобами машинного навчання. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 198-204.

50. Zharnovskyi O., Sobko O. Molchanova M. Neural Network Method for Detection of Fake Document Images for Personality Identification Systems. Black Sea Science 2024: Proceedings of the International Competition of Student Scientific Works. Odesa National University of Technology. Odesa, ONUT, 2024. Pp. 434-448.

ДОДАТКИ

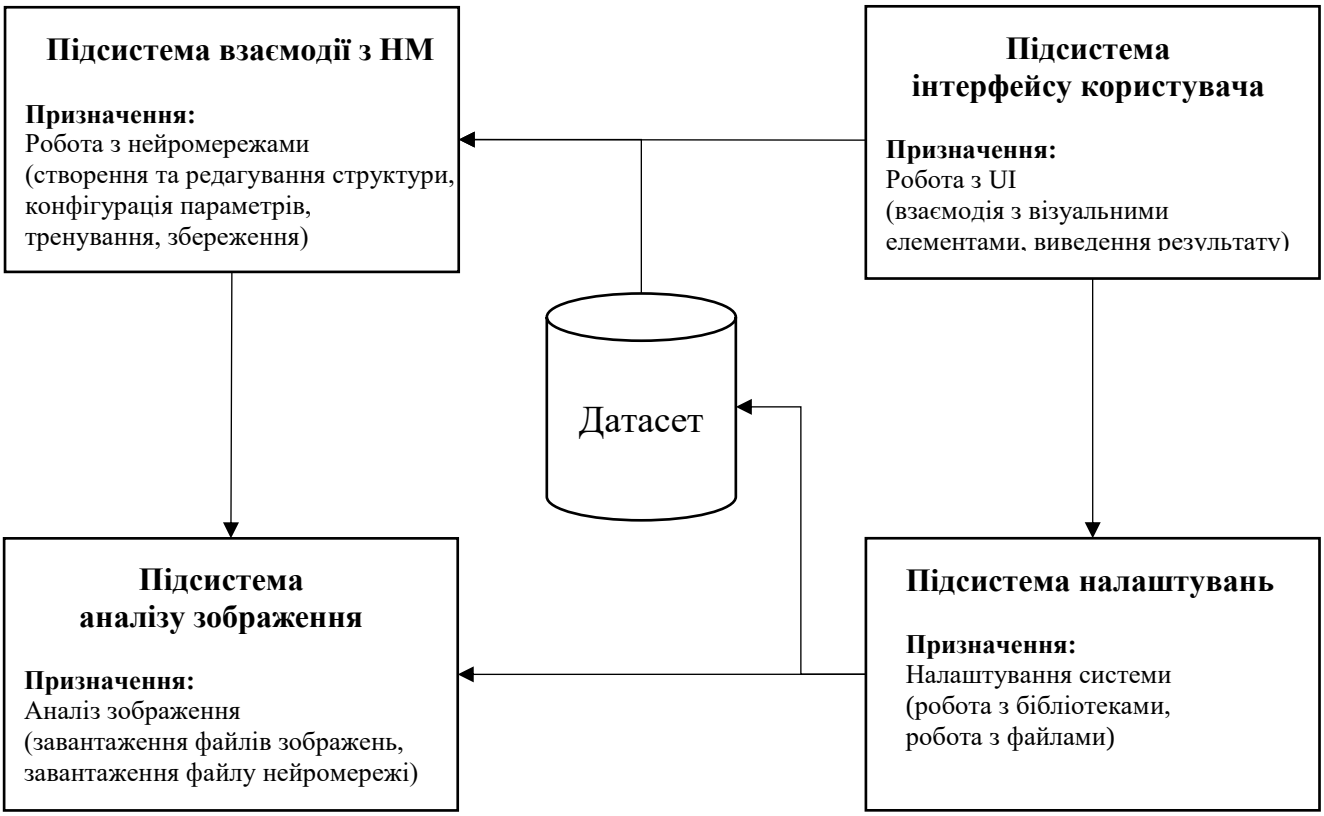
Додаток А

Схема методу ідентифікації згенерованих штучним інтелектом зображень людей



Додаток Б

Інформаційна структура системи ідентифікації згенерованих штучним інтелектом зображень людей



Додаток В

Світлини наукових публікацій, виконаних при роботі над кваліфікаційною роботою магістра

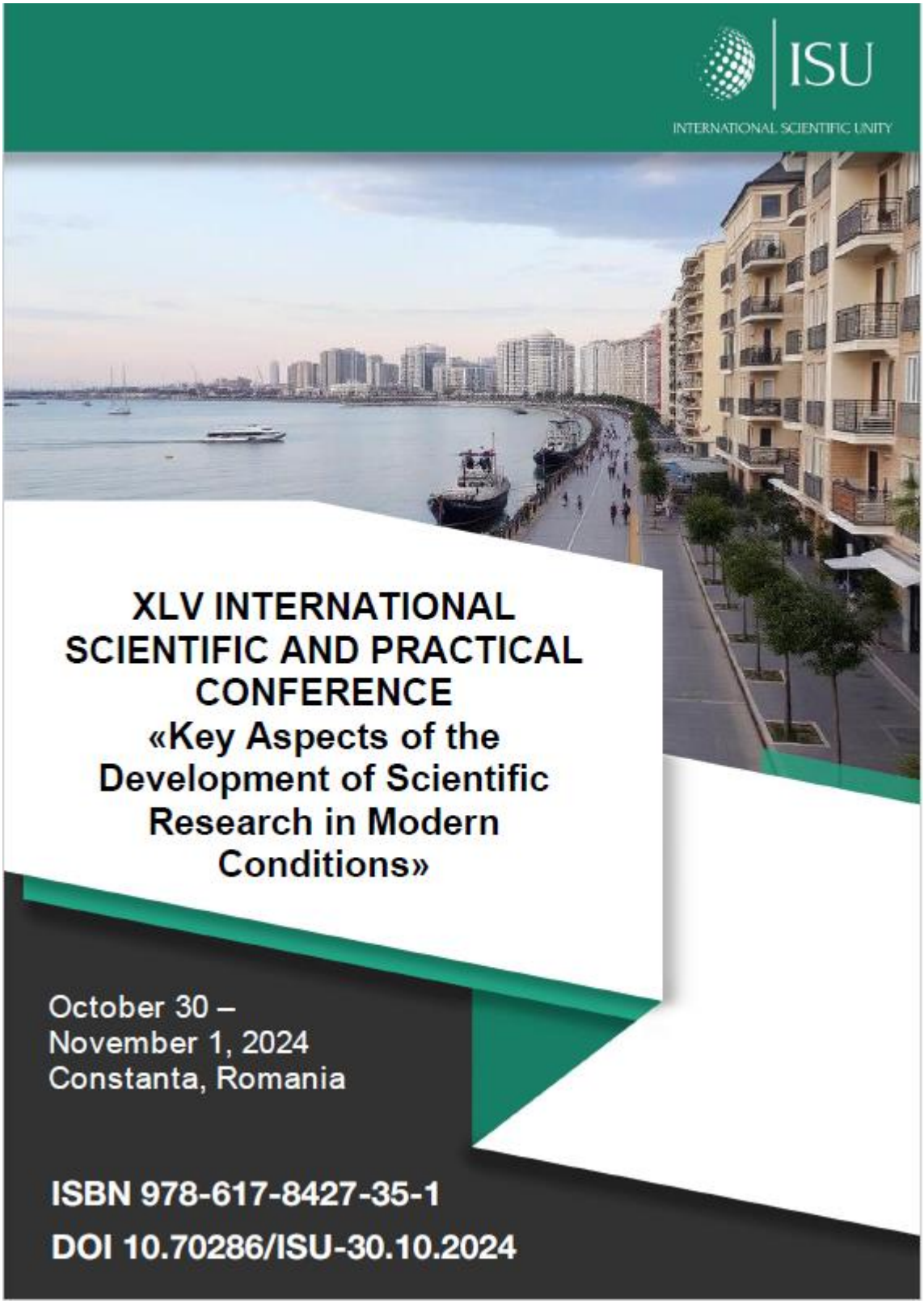
Наукові публікації:

1. Zharnovskyi O., Mazurets O., Sobko O. Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning. Key Aspects of the Development of Scientific Research in Modern Conditions. Proceedings of the XLV International scientific and practical conference. October 30 – November 1, 2024. Constanta, Romania. 2024. Pp. 69-73.

2. Жарновський О.В., Казмірчук Я.М., Собко О.В., Мазурець О.В. Практична реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 198-204.

3. Zharnovskyi O., Mazurets O., Sobko O. Neural network method for detection of fake document images for personality identification systems. Black Sea Science 2024: Proceedings of the International Competition of Student Scientific Works. Odesa National University of Technology. Odesa, ONUT, 2024. Pp. 434-448.

4. Мазурець О.В., Жарновський О.В., Гладун О.В., Собко О.В. Нейромережеве виявлення фейкових зображень людей. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1 (Довідка з редакції).


ISU
INTERNATIONAL SCIENTIFIC UNITY
XLV INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE
«Key Aspects of the Development of Scientific Research in Modern Conditions»
October 30 – November 1, 2024
Constanta, Romania
ISBN 978-617-8427-35-1
DOI 10.70286/ISU-30.10.2024

прогнозування кібератак має значний потенціал для підвищення безпеки мережевих систем та своєчасного запобігання інцидентам.

References

1. Столяр А. Л. Аналіз сучасних методів виявлення аномалій в комп'ютерних мережах. 2023, URL: https://doi.org/10.18372/2073-4751.74.17888.
2. Sunanda Gamage, Jagath Samarabandu. Deep learning methods in network intrusion detection: A survey and an objective comparison. 2020, URL: https://doi.org/10.1016/j.jnca.2020.102767
3. Mujaheed Abdullahi, Yahia Baashar, Hitham Alhussian. Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review. 2022, URL: https://doi.org/10.3390/electronics11020198
4. Nachaat Mohamed. Current trends in AI and ML for cybersecurity: A state-of-the-art survey". 2023, URL: https://doi.org/10.1080/23311916.2023.2272358

APPROACH TO IDENTIFICATION OF ARTIFICIAL INTELLIGENCE-GENERATED PEOPLE IMAGES BY MEANS OF MACHINE LEARNING

Zharnovskyi Oleksandr
Postgraduate student
Mazurets Oleksandr
Ph.D in Engineering Science, Associate Professor
Sobko Olena
Teacher
Computer Science Department
Khmelnytskyi National University, Ukraine

Artificial image generation technology has a wide range of applications, artificial image generators such as Midjourney, StableDiffusion, Adobe FIrefly, FLUX, Runaway can greatly simplify a large number of areas of human activity [1].

Some of the most popular fields of activity using artificial image generation are marketing – where artificial images can replace a photo shoot for a new product and create personalized advertising, medicine – which allows improving image diagnostics by creating clearer images, art and design – artists can use generative AI to create references, base image or less important background elements.

As images become more and more realistic, the creation of deep fakes becomes one of the biggest problems and threats to the existence of open tools [2, 3], which is why the development of methods for identifying generated images is relevant [4, 5].

The proposed method uses a combination of two convolutional neural networks that are excellent for the task of image analysis (Figure 1). The input image is

classified into real and generated, in case the generated image the neural network tries to find matches with known image generators (Figure 2, 3).

Images of people's faces combined with images created by five popular image generators, namely Midjourney, Stable diffusion, Dalle-3, Dreamstudio, Craiyon, will be used as datasets [6].

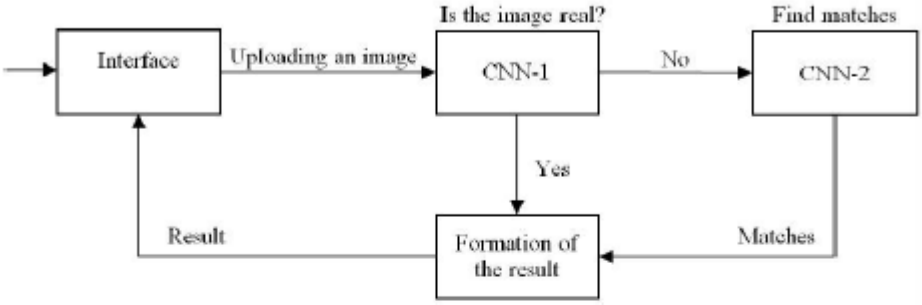


Figure 1. Approach to identification of artificial intelligence-generated people images by means of machine learning.

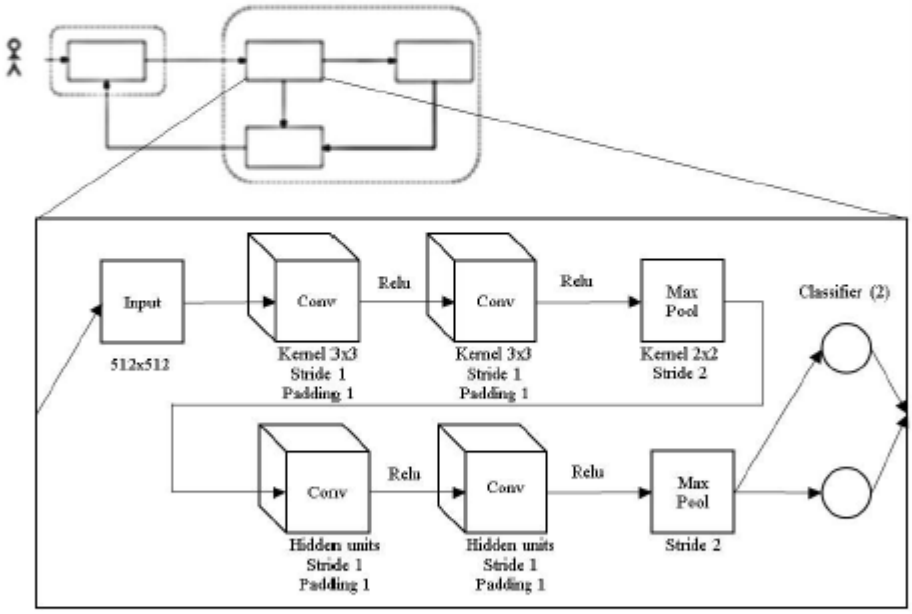


Figure 2. Architecture of CNN neural network model for basic classification of generated images.

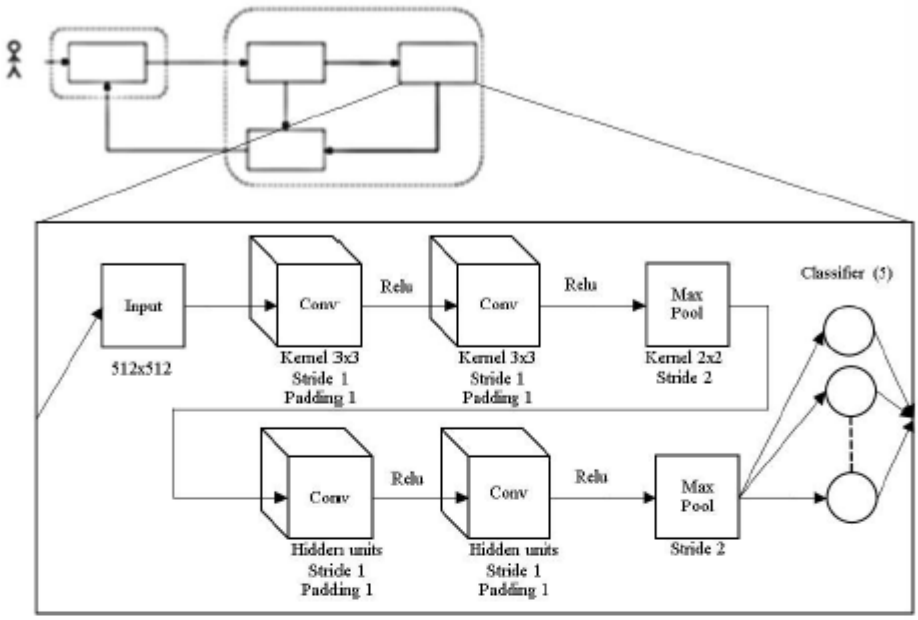


Figure 3. Architecture of CNN neural network model for detailed classification of generated images.

The architecture of convolutional neural networks consists of the following layers:

– activation layer – contains the activation function, usually used after convolution;

– pool layer – reduces image size without significant loss of information, usually used in the middle of the network;

- fully connected layer – perform classification based on features obtained by the previous layer, usually used after convolution and pool layer;

– normalization layer – contains the normalization function, stabilizes training, is used after the activation function;

– screening layer – randomly disconnects artificial neurons from the network, serves to prevent retraining;

– loss layer – determines the level of error between the original result and the expected one;

– output layer – the last layer in the network, the number depends on the number of expected output classes of the network.

The combination of these layers allows creating a functional network architecture for the proposed method [7].

The ready-made TorchVision library was used for the software implementation of the CNN architecture.

TorchVision is a library consisting of popular datasets, model architectures, and image transformation functions for computer vision tasks. It consists of: learning

methods for object detection, image classification, instance segmentation, video classification and semantic segmentation.

Supervised machine learning requires labeled input when training a machine learning model. This training data is labeled by the developer in the training phase before being used to train and test the model. Once the model has learned the relationship between input and output data, it can be used to classify new and unknown data sets and predict outcomes. Unsupervised learning (clustering) – learning on raw and unlabeled training data. It is often used to identify patterns and trends in raw data sets or to cluster similar data into a certain number of groups. Less important parameters for training neural networks are the number of epochs and the group size. The number of epochs is responsible for the training cycles that the network goes through, and the size of the group is how much data it receives in one cycle. If the number of epochs is multiplied by the size of the group more than the taken dataset, part of the data will be reused, which can lead to retraining (Figure 4). Optimal values are determined during training.

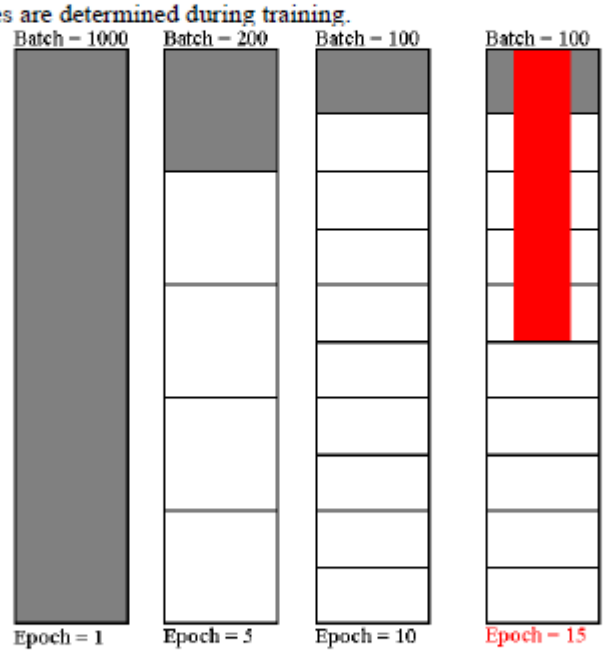


Figure 4. Dependence on the number of epochs and the size of the group.

For the CNN-1 network, whose task is to classify images into real and generated ones, the data needs to be grouped into two classes. In turn, CNN-2 will try to assign the input image to one of them.

So, the result of the work is the development of a method of identification of images of people generated by artificial intelligence by means of machine learning. The developed method allows for efficient image identification and can be integrated into mass media and social networks for automatic verification of image authenticity.

In addition, the network can be constantly improved and adapted to new methods of image generation to prevent the spread of false information.

References

1. Zharnovskyi O., Sobko O., Klimenko V. Intelligent System for Neural Network Detection of Fake Document Images for Automated Personality Identification. Proceedings of IV International Scientific and Practical Conference «Innovative research and perspectives of the development of science and technology». January 29-31, 2024. Stockholm, Sweden. 2024. Pp. 337-343.
2. Mazurets O. V., Klimenko V. I., Molchanova M. O., Sultanov A. V. Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones. Global Science: Prospects and Innovations. Proceedings of the 10th International scientific and practical conference. Cognum Publishing House. Liverpool, United Kingdom. 2024. Pp. 198-207.
3. Mazurets O., Molchanova M., Klimenko V., Klopotivskyi D. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services. Proceedings of XXII International Scientific and Practical Conference «Modern Scientific Research: Theoretical and Practical Aspects». May 8-10, 2024. Oslo, Norway. 2024. Pp. 64-68.
4. Molchanova M., Mazurets O., Klimenko V., Kuflevsky Ev. Object-oriented model for neural network damage detection of mail packages. Proceedings of XIV International Scientific and Practical Conference «Solving Scientific Problems Using Innovative Concepts». March 13-15, 2024. Copenhagen, Denmark. 2024. Pp. 58-62.
5. Novak Y., Mazurets O. Practical Application of Method of Automated Personal Identification by Fingerprints Using Convolution Neural Networks. Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». December 27-29, 2023. Stockholm, Sweden, International Scientific Unity. 2023. Pp. 136-140.
6. Mazurets O., Sobko O., Vit R., Pasternak V. Practical Approach for Detection by Deep Learning of Target Objects of Subject Area Based on Semantic Connectivity Indicators in Audio Database. Proceedings of XXIV International Scientific and Practical Conference «Modern Scientific Challenges are the Driving Force of the Development of Scientific Research». May 22-24, 2024. Bruges, Belgium. International Scientific Unity. 2024. Pp. 91-96.
7. Mazurets O., Zalutska O., Tyschenko O., Bohdanova A. An Approach to Using MobileNet CNN-model for Gesture Recognition. Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions». May 15-17, 2024. Munich, Germany. 2024. Pp. 59-64.

Міністерство освіти і науки України
Хмельницький національний університет

ЗБІРНИК НАУКОВИХ ПРАЦЬ

за матеріалами XVI Всеукраїнської науково-практичної конференції
«Актуальні проблеми комп'ютерних наук АПКН-2024»

15-16 листопада 2024

Хмельницький 2024

УДК 004.8

Жарновський О.В., Казмірчук Я.М., Собко О.В., Мазурець О.В.

Хмельницький національний університет

ПРАКТИЧНА РЕАЛІЗАЦІЯ МЕТОДУ ІДЕНТИФІКАЦІЇ ЗГЕНЕРОВАНИХ ШТУЧНИМ ІНТЕЛЕКТОМ ЗОБРАЖЕНЬ ЛЮДЕЙ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

Розроблено метод для ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання на основі використання комбінацій згорткових нейронних мереж. Спроєктовано інформаційну структуру системи нейромережевого аналізу згенерованих зображень облич людей засобами машинного навчання, що за вхідним зображенням дозволяє визначити автентичність зображення та визначити можливі засоби його генерації. Також було спроєктовано програмну архітектуру інформаційної системи для прикладної програмної реалізації розробленого методу нейромережевого аналізу згенерованих зображень облич людей.

Method for identifying human images generated by artificial intelligence by means of machine learning based on the use of combinations of convolutional neural networks has been developed. The information structure of the system of neural network analysis of generated images of people's faces by means of machine learning was designed, which allows to determine the authenticity of the image based on the input image and to determine the possible means of its generation. The software architecture of the information system was also designed for the applied software implementation of the developed method of neural network analysis of generated images of people's faces.

Технологія генерації штучних зображень має широкий спектр застосувань, що робить її корисною в багатьох сферах людської діяльності. Генерація зображень корисна для арту та дизайну – дизайнери та художники можуть використовувати штучний інтелект для генерації референсів та ітерації власних робіт, використовувати штучне зображення як базове для подальшого редагування чи генерувати менш важливі деталі для вже існуючого зображення, такі як об'єкти заднього фону.

В сфері маркетингу та реклами штучний інтелект здатен швидко генерувати візуал. Наприклад, замість того, щоб організовувати фотосесію для нового продукту, маркетологи можуть використовувати ШІ для створення високоякісних зображень, для використання в рекламних матеріалах.

Якість зображень згенерованим ШІ прямим чином залежить від кількості та якості зображень, що використовували для тренування моделі, а також їх доступності.

Складність налаштування – досягнення бажаного рівня деталізації вимагає ретельного налаштування параметрів моделі, що є складнім та трудомістким процесом, особливо для сфери медицини, де зображення повинні мати високу точність.

Проблеми копірайту – крім того що самі зображення, використані для тренування мережі, можуть бути захищені авторським правом, отримане зображення, може призвести до юридичних проблем маркет-заміни та інтелектуальної власності.

Створення дипфейків – через свою простоту та загальну доступність, а також якість зображення генеративний ШІ може бути використаній в створенні зображень подій, що ніколи не мали місце для поширення дезінформації в соціальних мережах. Це може бути використано шахраями для розповсюдження дезінформації з метою впливу на громадську думку.

Хоча найперші спроби генерувати зображення з використанням штучного інтелекту відносяться до 1970-х років, протягом десятиліть прогрес у цій галузі був незначним. Доступні обчислювальні потужності були обмежені, а алгоритми занадто прості щоб працювати із реалістичними зображеннями.

Однак це змінилося з розвитком глибокого навчання та згорткових нейронних мереж [1, 2], що в свою чергу забезпечили основу для створення генеративних змагальних мереж [3, 4].

Генерація зображень здійснюється за допомогою різних форм вхідних даних включаючи RGB зображення, відео, медичні дані чи текст де на виході отримують зображення чи відео.

Метою роботи є прикладне вирішення задачі ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання вимагає розробки нейронної мережі здатної до розпізнаванню та класифікації зображень.

Для цього найкраще підходить згорткова нейронна мережа – тип глибоких нейронних мереж, що активно використовується для аналізу зображень, аудіо та відео [5]. CNN складається із різновиду багатошарових перцептронів, розроблених так, щоб вимагати мінімальний обсяг попередньої обробки [6, 7].

Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання складається з наступних кроків (рисунок 1):

- завантаження датасетів;
- пре-процесинг зображень;
- створення нової чи редагування вже існуючої архітектури нейромережі;
- тренування;
- аналіз результатів та корегування мережі.

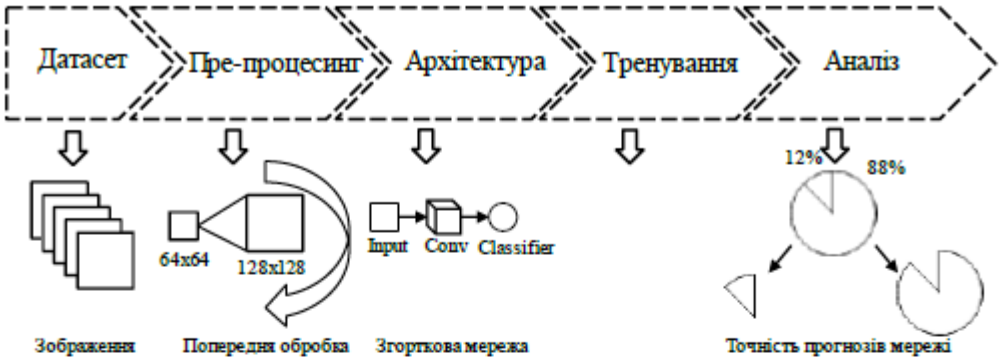


Рисунок 1 – Структура методу

Функціональна складова буде розділена на дві окремих нейронних мережі: imageClassifier, що відповідає за ідентифікацію зображення як реального чи підробки, та methodClassifier, що визначає збіг з популярними моделями ШІ для генерації зображень (рисунок 2).

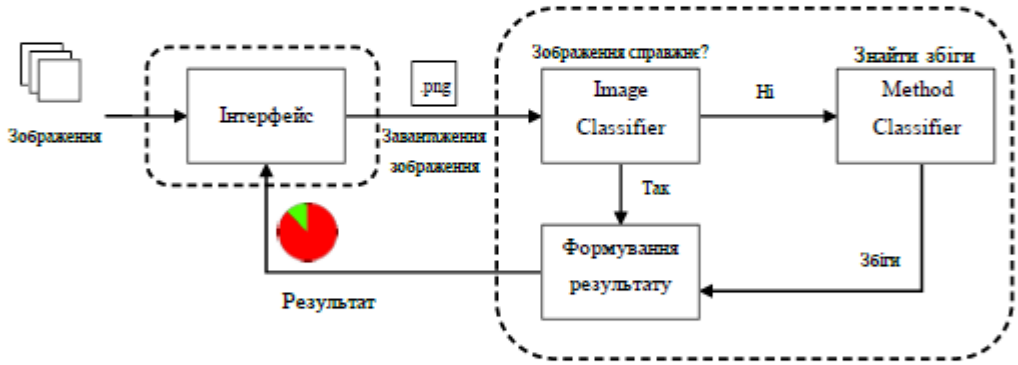


Рисунок 2 – Схема роботи методу

Вхідними данними є файли зображення, що користувач завантажує через інтерфейс, кожне зображення обробляє нейронна мережа imageClasifier, в залежності від результату додатково використовується methodClasifier, а результатом функціонального виконання є класифікація зображення як реального чи згенерованого, та віднесення до можливих методів генерації.

Інформаційна система нейромережевого аналізу згенерованих зображень облич людей засобами машинного навчання є прикладною програмною реалізацією методу аналізу зображень обличь людей, із файлу завантаженим користувачем, що призначений для пошуку згенерованим штучним інтелектом зображень та вхідними даними має. Вхідними є файл(и) зображень, такі як: png, .jpg, webp, tiff, bmp.

Інформаційна структура системи складається із набору зображень (датасетів) та кількох підсистем: «Підсистема взаємодії з НМ», «Підсистема розпізнавання завантажених зображень», «Підсистема інтерфейсу користувача», «Підсистема налаштувань».

Набір даних зображень складається із наведених датасетів, що були розподілені дві відповідні частини:

– класифікація зображень – датасети розділені на реальні та згенеровані зображення для нейромережі imageClasifier;

– класифікація методів створення – датасети з виключно згенерованих зображень поділені відповідно до технології генерації.

Підсистема роботи з НМ є головною, що призначена для роботи з методами нейронних мереж [8, 9]. Включає в себе ряд функцій, таких як: створення та зміна архітектури нейромережі, конфігурація параметрів нейромережі, вибір датасетів та процес тренування, а також збереження натренованої мережі у відповідний файл.

Підсистема розпізнавання зображень є вторинною, вона призначена для аналізу завантаженого зображення чи декількох зображень обраною натренованою нейромережею, завантаженою з файлу. Має наступний функціонал: завантаження зображення з подальшим звільненням файлу, завантаження файлу нейромережі та виведення отриманих результатів.

Підсистема інтерфейсу користувача забезпечує функціональну взаємодію користувача з іншими підсистемами за допомогою UI. Включає в себе динамічну генерацію візуальних елементів відповідно до дій користувача.

Підсистема налаштувань дає можливість користувачу змінювати обрані параметри, як і візуальні, як розмір вікна, так і функціональні, обрані бібліотеки та файли.

Підсистема взаємодії з НМ, що основним призначенням має роботу з нейромережею (рисунку 3).

Першою функцією підсистеми є вибір шляхів до зображень. Для успішного завантаження зображень потрібно коректно вказати шлях до основної директорії та директорії для тренування та тестування. Із вказаних директорій будуть взяті зображення та розподілені по класам відповідним їх назв.

Наступною функцією є завантаження зображень у даталоадери, де є можливість вказати основні параметри: розмір зображення для трансформації, розмір однієї групи та потреба в перемішуванні.

Наступною групою функцій є створення та редагування архітектури нейромережі з можливістю додатки обрану кількість шарів з вказанням їх основних параметрів: вхідна та вихідна розмірність, розмір кернела, крок, нульове заповнення та інші відповідні пареметри.

Кінцевою функцією є налаштування тренування із вказанням параметрів: кількість епох, оптимізатор та його параметри. Після успішного завершення тренування є можливість зберегти отриману мережу.

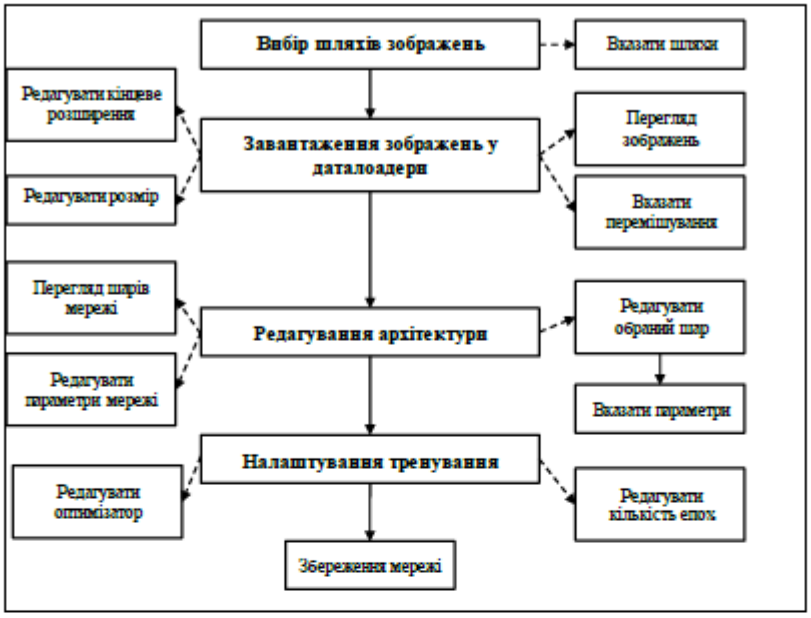


Рисунок 3 – Схема та функції підсистеми взаємодії з НМ

Була розроблена схема пропонованої програмної архітектури для інформаційної системи нейромережевого аналізу згенерованих зображень облич людей (рисунок 4).

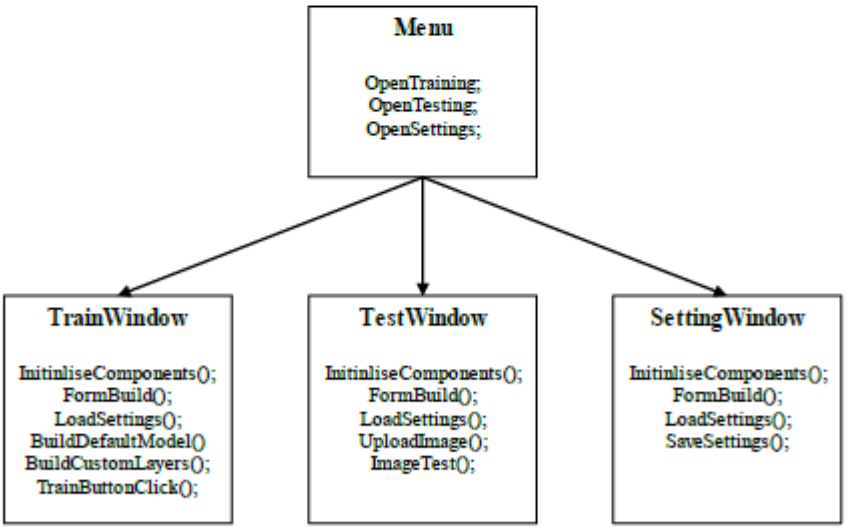


Рисунок 4 – Програмна архітектура системи

Програмна архітектура системи є об'єктно-орієнтованою та складається з наступних класів:

– Menu – внутрішній клас що реалізує перехід між іншими класами;
– ModelTrain.py – зовнішній клас що реалізує методи тренування мережі;
– TrainWindow – внутрішній клас що реалізує роботу з нейромережами;
– ImageTest.py – зовнішній клас що реалізує методи завантаження та аналізу зображення;
– TestWindow – внутрішній клас що реалізує аналіз завантажених зображень;
– SettingWindow – внутрішній клас що реалізує можливість зміни налаштувань програми.

Клас «TrainWindow», основним призначенням якого є створення та редагування архітектури, тренування нейромережі. Методи InitialiseComponent() та FormBuild() призначені для ініціалізації та створення форми використовуючи C#. Метод LoadSettings() завантажує налаштування програми із відповідного файлу. Метод TrainButtonClick() завантажує параметри нейромережі із відповідних полів та запускає процес тренування використовуючи завантажений ModelTrain.py Python клас. Метод CreateCustomLayers() динамічно довантажує нові елементи UI для створення власної архітектури мережі.

Клас «ModelTrain.py» використовує методи Torchvision мови Python для тренування мережі та має наступні методи: SetImagePaths() для завантаження шляхів зображень, LoadImages() завантажує файли зображень що далі використовуються в ToDataloaders() для перетворення завантажених зображень у формат даних для навчання мережі, BuildDefaultModel() завантажує спроєктовану по замовчуванню архітектуру, в свою чергу BuidCustomModel() створіє її динамічно, а метод Run() запускає процес навчання з подільшим збереженням мережі.

Клас «TestWindow» використовує навчену нейромережу для аналізу завантаженого користувачем зображення та виводу результатів. Метод LoadSettings() аналогічно відповідному методу класа TrainWindow ззавантажує файл налаштувань. Метод UploadImage() викликає інтерфейс для вибору та завантаження зображення в додаток та його звільнення для паралельної обробки використовуючи метод BitmapSwap. Метод ImageTest() завантажує ImageTest.py Python клас для аналізу зображення обраною нейронною мережею та виводить результат в UI.

Клас «SettingWindow» здійснює роботу з елементами інтерфейсу для завантаження та редагування налаштувань додатку. Метод LoadSettings() призначений для завантаження файлу налаштувань та заповнення елементів UI для подальшого редагування. Метод SaveSettings() зберігає чи перезаписує налаштування у відповідний файл для подальшого використання іншими класами.

Таким чином, був розроблений метод для ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання на основі використання комбінацій згорткових нейронних мереж. Спроєктовано

інформаційну структуру системи нейромережевого аналізу згенерованих зображень облич людей засобами машинного навчання, що за вхідним зображенням дозволяє визначити автентичність зображення та визначити можливі засоби його генерації. Також було спроєктовано програмну архітектуру інформаційної системи для прикладної програмної реалізації розробленого методу нейромережевого аналізу згенерованих зображень облич людей.

Перелік посилань

1. Мазурець О.В., Скрипник Т.К., Ізотов А.В. Фасетковий метод перетворення зображень за допомогою нейромережевого розпізнавання. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2020, №1 (281). – С.119-125.
2. Mazurets O., Uspenska K., Vit R., Tyschenko O. Intelligent System for Determining the Object Attributes Values by Neural Networks Means by Graphic Images in Databases. Current Trends in the Development of Scientific Research in Today's Conditions. Proceedings of XXV International scientific and practical conference. May 29-31, 2024. International Scientific Unity. Florence, Italy. 2024. Pp. 86-91.
3. Kharysh I., Sobko O., Mazurets O. Designing CNN Neural Network Model for Detecting Fractures of Lower Extremities by X-ray Images. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. October 23-25, 2024. Dubrovnik, Croatia. 2024. Pp. 91-96.
4. Mazurets O. V., Klimenko V. I., Molchanova M. O., Sultanov A. V. Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones. Global Science: Prospects and Innovations. Proceedings of the 10th International scientific and practical conference. Cognum Publishing House. Liverpool, United Kingdom. 2024. Pp. 198-207.
5. Mazurets O., Zalutska O., Tyschenko O., Bohdanova A. An Approach to Using MobileNet CNN-model for Gesture Recognition. Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions». May 15-17, 2024. Munich, Germany. 2024. Pp. 59-64.
6. Pokhytun A., Mazurets O., Molchanova M., Tyschenko O. Method for Neural Network Detecting Changed Images of People's Faces Using CNN. New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International scientific and practical conference. October 21-23, 2024. Marseille, France. 2024. Pp. 35-40.
7. Mazurets O., Molchanova M., Klimenko V., Klopotivskyi D. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services. Proceedings of XXII International Scientific and Practical Conference «Modern Scientific Research: Theoretical and Practical Aspects». May 8-10, 2024. Oslo, Norway. 2024. Pp. 64-68.
8. Novak Y., Mazurets O. Practical Application of Method of Automated Personal Identification by Fingerprints Using Convolution Neural Networks. Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». December 27-29, 2023. Stockholm, Sweden, International Scientific Unity. 2023. Pp. 136-140.
9. Мазурець О.В., Петровський С.С., Дидо Р.А. Нейромережева модель для ідентифікації особистості за зображенням обличчя у реальному часі Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції. 31 жовтня – 1 листопада 2024 р. Одеса, ОНТУ. 2024. С.655-658.

Ministry of Education and Science of Ukraine

ODESA NATIONAL UNIVERSITY OF TECHNOLOGY

International Competition of Student Scientific Works

BLACK SEA SCIENCE 2024

PROCEEDINGS

ODESA, ONUT 2024

NEURAL NETWORK METHOD FOR DETECTION OF FAKE DOCUMENT IMAGES FOR PERSONALITY IDENTIFICATION SYSTEMS

Author: Oleksandr Zharnovskyi
Advisors: Olena Sobko, Maryna Molchanova
Khmelnytskyi National University (Ukraine)

Abstract. *Forged document images can be used for various malicious purposes such as identity theft, credit card fraud or illegal immigration. This makes it a very important task to develop a method for the detection of fake document images. One of the promising ways of detecting such images is the use of artificial intelligence, convolutional neural networks in particular. Convolutional neural networks are capable of learning typical characteristics of fake document images, such as light mismatch, unnatural colors and artificial or modified elements made by using graphic editors.*

This work aims to develop and implement a method for the detection of fake document images for personal identification systems. This method is based on using convolutional neural networks to detect distinctive signs of image forgery. The study showed that the method has an accuracy of 95,16% and the resulting system can determine levels of image authenticity.

Keywords: *personal identification, identity documents, document image falsification, CNN.*

I. INTRODUCTION

In modern digital world personal identification using document photographs is becoming more and more important. Ever increasing number of online operations creates new possibilities for fraud. Personal identification helps protect users from such threats.

For example, personal identification by document photographs helps banks and credit companies prevent fraudulent activities involving credit cards, bank accounts and loans. It also helps online stores prevent stealing user data. By using document photographs citizens can get access to social benefits and be able to use public services.

One of many ways of personal identification is identity documents, they contain a set of key parameters that help establish a person's identity, such as name, surname and photo card. Such documents are split into two main categories [1]:

- confirm identity and verify citizenship;
- confirm identity and confirm their special status.

These include but are not limited to:

- passport of the citizen;
- passport of the citizen for traveling abroad;
- diplomatic passport;
- service (government) passport;
- driving license;
- refugee identity certificate.

Document circulation of any form creates potential risks that can vary from different copies of the same document being used in different official institutions to attempts of complete forgery.

The simplest examples of falsification are erasement (removing some parts of a document), reprinting (adding new words or printing over) and paper replacement (in case a document consists of multiple parts). It's easy enough to detect signs of erasing or reprinting valuable spots without full expertise, but in cases of complete replacements or falsification, it would require a more detailed examination. There are a wide variety of protection types used to prevent counterfeiting [2]. Any kind of document with at least some bureaucratic value will be using a combination of such methods.

Passport, a mandatory document for every citizen of Ukraine is one of the best examples of widely used document with a huge amount of various security measures (Fig. 1).

Fig. 1. Security measures used on ID card of citizen of Ukraine [3]

Some of those are special fonts, barcodes, elements that are not visible to the naked human eye (such as minified text, text only visible to untarred or ultraviolet), touch elements and more. However, even all of these measures cannot completely prevent image forgery through the use of complex graphic editors.

II. LITERATURE ANALYSIS

To deal with massive quantity of image data flowing through various organizations the system needs to be automated by using neural networks.

Neural networks (NN) are complex machine learning methods that use interconnected nodes of artificial neurons in layers, a structure similar to the human brain. Generally, such networks consist of input and output layers and varying amounts of layers in-between.

The two broad types of neural networks are feedforward and recurrent.

Feedforward Neural Network (FNN) consists of neurons that feed data only in one direction (direction of output), often used for face or speech recognition and can deal with data "noise", easy to maintain thanks to relatively simple structure.

Recurrent Neural Network (RNN) on the other hand "memorizes" some of the

data after feeding it in the next layer which allows it to make better conclusion in case of errors in the next layer. Because of its structure, it is often used in automated text-related systems (such as search engines or grammar checkers) where data is split into easily definable segments – words.

There are a lot of specialized types of neural networks, one of which is convolutional neural network (CNN) – a network type that excels at pattern recognition of audio or video signal [4]. In the case of an image, layers can split a given image by edges, faces or contours highlighting key features. With every new layer, the complexity of a fragment, object or class increases along algorithmic complexity (Fig. 2).

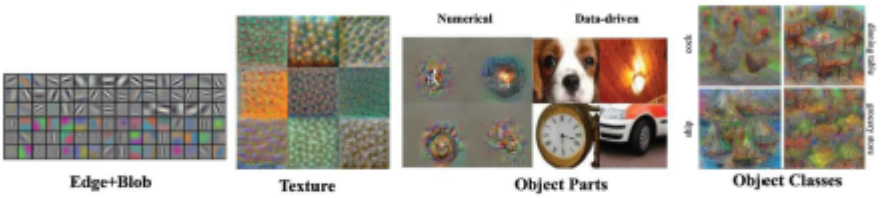


Fig. 2. Examples of image processing with neural network [4]

Artificial intelligence and neural networks have grown in popularity for automated identity detection and image recognition due to their advantages over more traditional identification methods.

One of the main advantages of artificial intelligence and neural networks is their ability to learn on massive data sets which allows them to spot details a human wouldn't. For example, neural networks can be used to detect fake images based on blur, light mismatch or unnatural color channels.

III. OBJECT, SUBJECT, AND METHODS OF RESEARCH

The object of study – the process of detecting fake document images for personality identification systems.

The subject of study – models, methods, algorithms and means to verify the authenticity of scanned and photographed images of identity documents.

3.1. Description of the method of detecting fake document images with neural networks

Overall methods of creating forged document images do not differ much from creating regular fake images with graphic editors or other similar software.

Some of the most common ways to modify an image are:

- image retouching – changing some image characteristics;
- copy-move – copying one part of the image over the other (Fig. 3. a);
- splicing – creating one image from two or more source images (Fig. 3. b);
- processing – image goes through color pattern changes and filters.

a) b)

Fig. 3. Common ways to modify images: a – copy-move, b – splicing

When tempering an image with an editor such as Paint or Photoshop, usually there is going to be an error around edited spots or metadata.

Metadata is additional information in an image file containing stuff like color pattern, camera type, resolution, location date and more. Often times its lost during editing.

Such errors can be used for image analysis through a combination of ELA and CNN in functional methods of fake document image identification (Fig. 4).

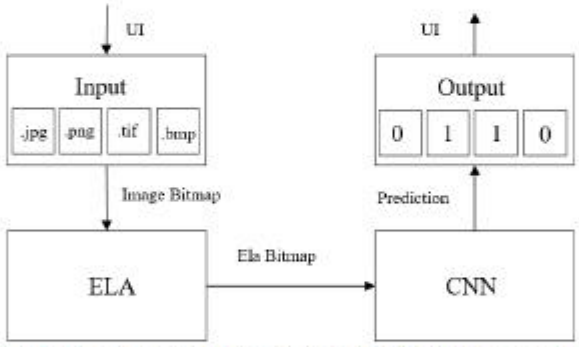


Fig. 4. Simplified diagram for method of detecting fake images using neural network and error level analysis

Input – user interface that allows uploading a selected image with chosen settings.

Error Level Analysis (ELA) – one of the methods of identifying changes to an image through compression on a given brightness and comparing it with the original. If the image was not tempered the error should be the same across all images, otherwise modified spot will be standing out (Fig. 5).

Fig. 5. Using ELA on a spliced image

The functional task of the ELA module is to save the input image with selected brightness and compare it with the original to create ELA which will then be used in a neural network (Fig. 6).

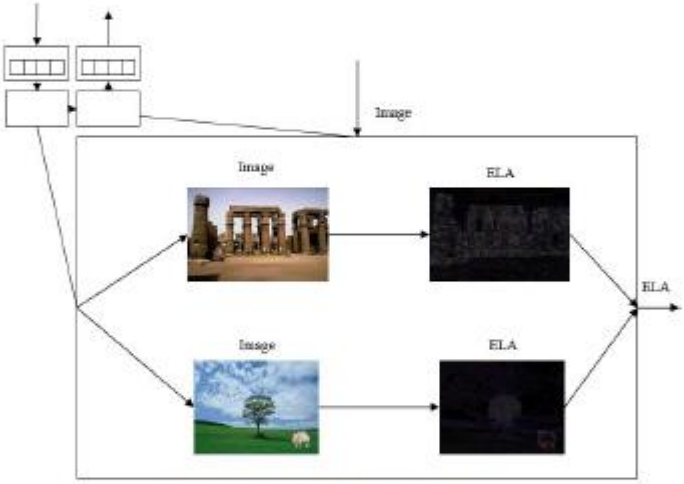


Fig. 6. ELA module functionality

ConvNet/CNN – convolutional neural network module containing input, output and hidden layers. Hidden layers can be convolutional, aggregative, fully-connected or normalization layers.

The main task of the CNN module is to learn how to classify real and edited images during the learning phase and then classify input images.

Output is an interpretation and UI module that converts CNN results in a user-friendly format.

This concludes describing a method of using neural networks to identify fake document images for a personality identification system.

3.2. Neural network architecture for identifying fake document images

For neural networks to function properly it's mandatory and extremely important to correctly set up its training. Training allows the network to classify input images according to "fed" data samples during the learning process and the accuracy of such a conclusion entirely depends on how you set it up and which data you feed [5]. The two main training algorithms for neural networks are supervised and unsupervised learning.

Unsupervised learning or "clustering" means that the network doesn't know the correct labels and groups data in clusters on its own. This method has substantial accuracy loss and doesn't fit for the given task.

Supervised learning means that input data needs to be classified, in our case fake or real. After the learning, algorithm calculates accuracy using the loss function, weights get shifted and the process repeats until loss gets minimized or accuracy reaches optimal values (Fig. 7).

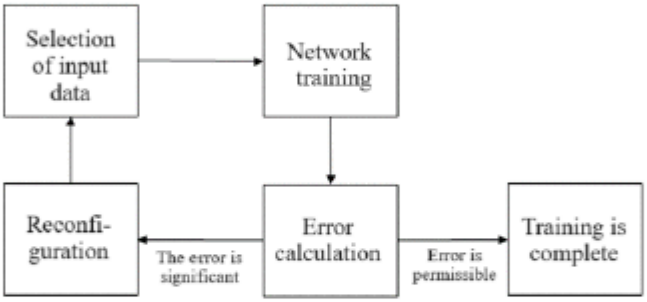


Fig. 7. Learning process

The neural network structure is divided into main functional layers:

– convolution layer – the first layer that is used for feature extraction and outputs feature map for further network layers;

– pooling layer – often used after convolution, the main function is to reduce feature map to reduce algorithmic complexity which speeds up the process and reduces computer load;

– fully connected layer – used for connecting different layers of the network and flattening feature map, often used before the output layer;

– dropout – removes some of the neurons to prevent overstudy and reduce the overall size of the resulting network.

As a result of combining such layers, the network architecture was created for identifying fake document images (Fig. 8).

Another important parameter for neural network learning besides the structure is the iteration count. This value is defined by the amount of batch count and volume taken from the dataset (Fig. 9)

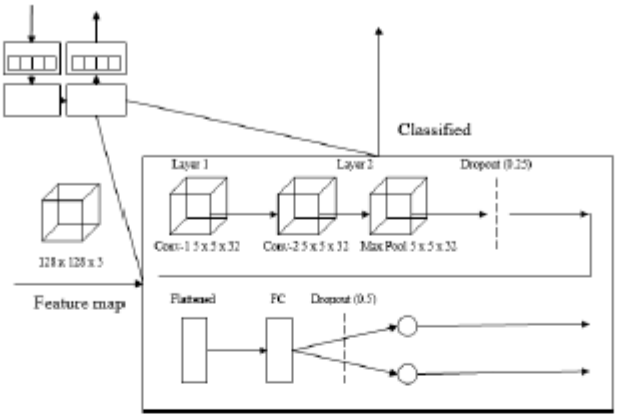


Fig. 8. Neural network architecture

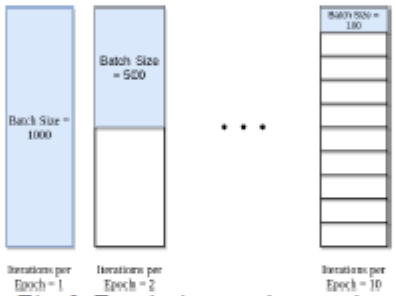


Fig. 9. Epochs in neural networks

In cases where the batch size is not evenly distributed relative to the number of images, some of the images will be used multiple times. An optimal number of epochs will be determined by testing. Also, it is taken into account the possibility of customizing the neural network structure with the user interface.

This concludes neural network architecture for identifying falsified document images.

3.3. Preparing input data for the neural network

For the neural network to function properly it's an important development step to prepare a dataset – a set of data (images) the network will be learning on. For training and testing CASSIA2 [6] and MIDV [7] datasets have been selected.

CASSIA2 consists of more than 12000 different authentic and fake images (Fig. 10).

a) b)

Fig. 10. Cassia dataset sample: a – authentic, b – fake

MIDV consists of 500 photographs or clips of different documents for personal identification (Fig. 11).

Fig. 11. MIDV dataset sample

The selected data needs to be preprocessed – all the files have different resolution, brightness and quality which needs to be standardized and classified for further use by the neural network.

Classification is a process of splitting images into classes which will then be fed into neural network inputs, in this case, the classes are "fake" and "real".

It's also important to randomly split the dataset with every model iteration into three sets – training, validation and testing (Fig. 12).

Fig. 12. Dataset splitting

Training set – a main set of data that has the most images (~70-80%) and will be used for model training.

The validation set – second set by size (~20-10%) used for testing each model iteration and shifting its parameters.

To prevent overfitting and properly evaluate results the third set can be used – the test set. Containing the least data and servers as a final test. If the testing set is not being used, the model could display good validation results, but these results will be

specific to the selected dataset and might heavily vary for other data.

If the testing accuracy on validation and testing sets is within the margin of error, the model can be considered trained, if the results differ by a substantial amount it needs to be retrained accounting for possible errors.

IV. RESULTS

As a result of research, an information system was developed that uses a created method and allows to automatically determine the authenticity of photographed or scanned images of personal identification documents with the use of a convolutional neural network.

For practical implementation of the method for identifying fake document images class diagram was created according to expected functionality (Fig. 12).

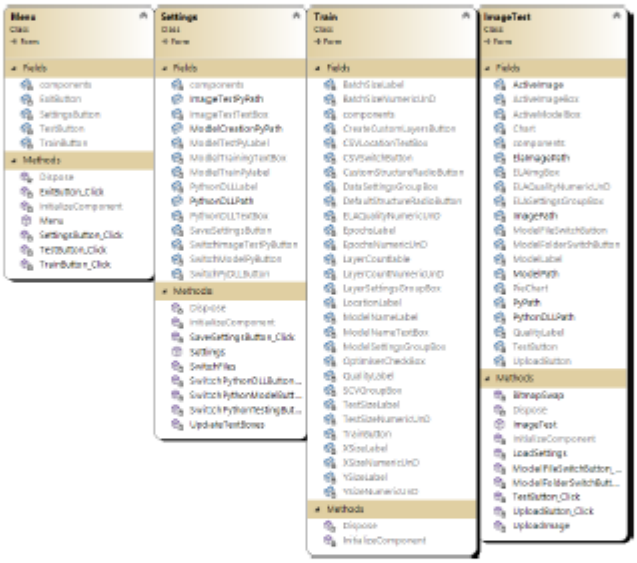


Fig. 12. Class Diagram

SettingsButton_Click, TestButton_Click and TrainButton_Click methods allow to switch active form with a press of a corresponding button.

Settings class contains methods for switching paths to Python classes and a path to a Python DLL file which are required for proper function. Changes are saved in a Txt file that will be used by other modules and will be saved for further use (Fig. 13).

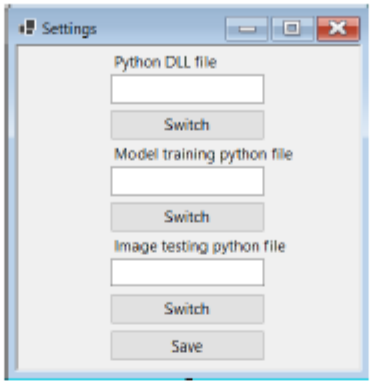


Fig. 13. Settings class interface

SwitchPythonDLLButton_Click, SwitchPythonModelButton_Click and SwitchPythonTestingButton_Click methods invoke user interface for switching paths to corresponding files. SaveSettingsButton_Click saves those changes into the Settings.txt file.

PythonNet library was used for the functional interaction of modules written in C# and Python. This third-party extension allows C# to load .py files (and vice versa) containing methods for working with CNN written in Python with its own set of libraries, invoking those methods with selected parameters and returning possible results.

Train class contains all the logic for creating, setting up and training CNN as well as saving it for further use (Fig. 14).

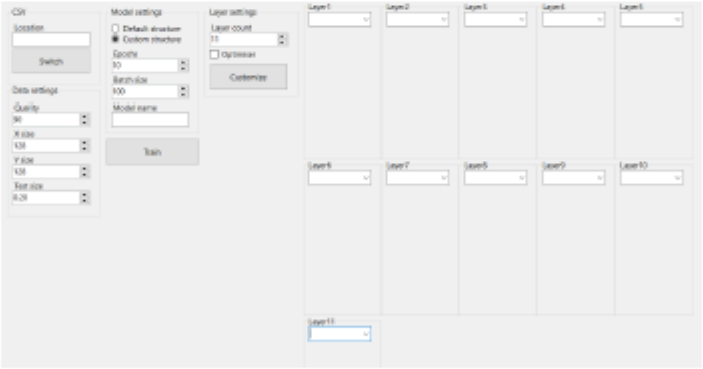


Fig. 14. Train class interface

Load settings method loads setting from a file, if it exists, or uses paths set up by default. Before training user needs to set up neural network parameters into corresponding fields and paths to a CSV file that contains image paths and classes (fake or real). Train loads .py file written in Python and invokes methods it contains with selected parameters. After the model is done training it can be saved in the

SavedModels folder as a file or folder with a given name, and a new model can be trained with different parameters.

ImageTest class serves to check the authenticity of an uploaded document image with a selected trained neural network (Fig. 15).

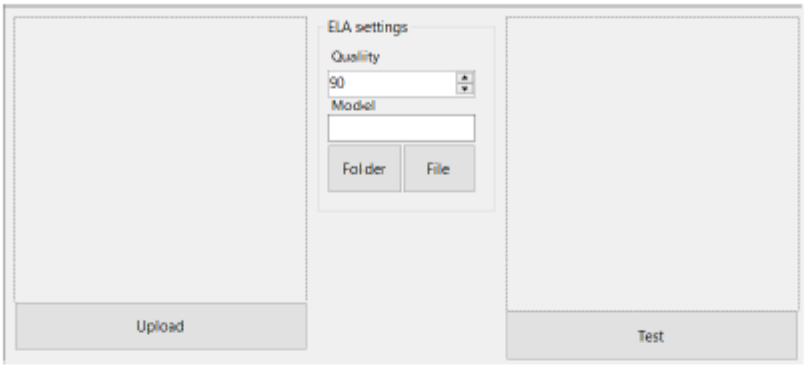


Fig. 15. ImageTest class interface

LoadSettings, similarly to the method in the Train class, loads paths to the required files. UploadImageButton_Click invokes the UploadImage method which in turn shows a user interface to upload an image for testing. BitmapSwap allows simultaneously use, move, delete or alter the image without it being locked by or affecting the system. ModelFileSwitchButton_Click and ModelFolderSwitchButton_Click methods allow the user to select a folder or file with trained model data which will be used to check the image. ImageTestButton_Click loads a .py file with the ImageTest method in Python to test image authenticity with neural network and selected settings, and then displays the results in a form of a pie chart.

A neural network was trained and tested to study the effectiveness of the developed method for detecting fake document images. The network had a set structure with altered key parameters before training and had the following results (Table 1) та (Fig. 16).

Table 1. – Neural network training parameters and its results

		Network training settings			Results		
		Epochs	Batch size	Train-Test Ratio	Accuracy	Val Accuracy	Diff
	CNN1	10	100	0.2	0.9638	0.9534	0.0104
	CNN2	20	125	0.2	0.9803	0.9515	0.0288
	CNN3	30	150	0.2	0.9941	0.9598	0.0343
	CNN4	10	100	0.35	0.9603	0.9430	0.0173
	CNN5	20	125	0.35	0.9841	0.9446	0.0395
	CNN6	30	150	0.35	0.9960	0.9516	0.0444

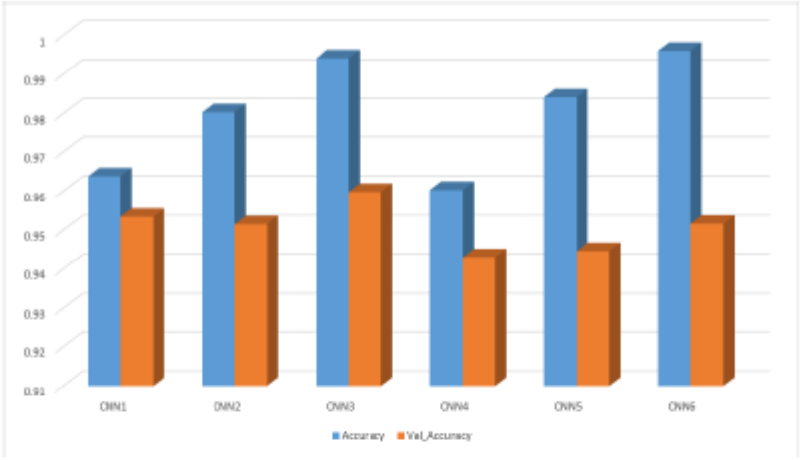


Fig. 16. Resulting accuracy

Training CNN1 model with the following parameters:

- Epochs – 10.
- Batch size – 100.
- Train-test ratio – 80:20.

With the following result (Fig. 17).

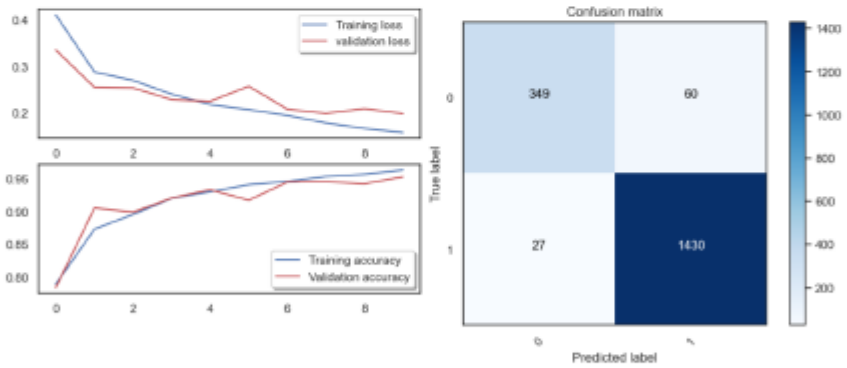


Fig. 17. Accuracy graphs and confusion matrix for CNN1

As a result of training, final training accuracy for the model is 96,38% and 95,34% for validation set.

The accuracy graph illustrates a slight deviation within the error margin.

The confusion matrix shows a low count of false negatives (fake images passed as real) and a bigger margin of false positives (real images passed as fakes).

Training CNN2 model with the following parameters:

- Epochs – 20.
- Batch size – 125.

– Train-test ratio – 80:20.

With the following result (Fig. 18).

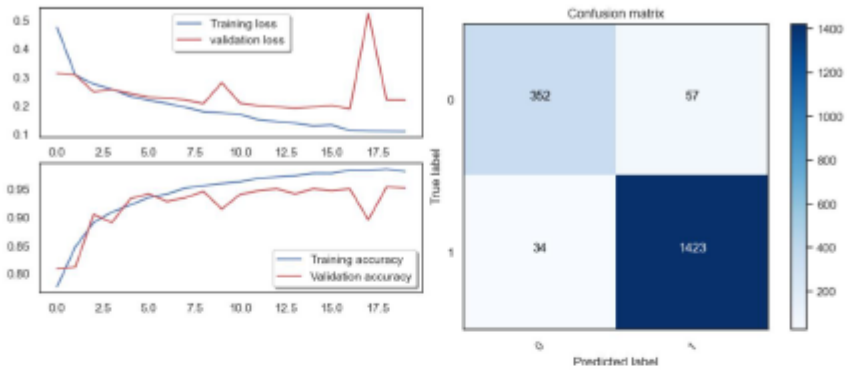


Fig. 18. Accuracy graphs and confusion matrix for CNN2

Final accuracy:

– During training – 98,03%;
– During validation – 95,15%.

The accuracy graph illustrates a substantial local deviation between training and validation accuracy.

Increasing epoch count and batch size had a positive impact on training accuracy in comparison to the previous version but didn't impact validation accuracy by much.

Training CNN3 model with the following parameters:

– Epochs – 30.
– Batch size – 150.
– Train-test ratio – 80:20.

With a following result (Fig. 19).

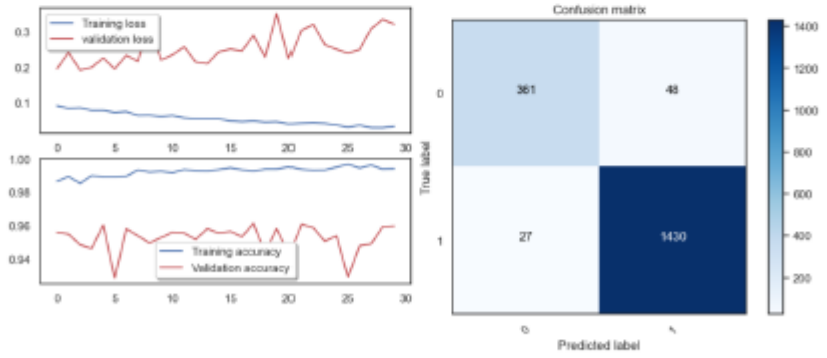


Fig. 19. Accuracy graphs and confusion matrix for CNN3

Final accuracy:

- During training – 99.41%.
- During validation – 95.98%.

The accuracy graph illustrates a substantial global deviation between training and validation accuracy.

Further increase to epochs count and batch size significantly increased training accuracy but didn't impact validation accuracy by much in comparison to previous iterations.

Because of substantial deviation and low increase in validation accuracy next model iteration had other parameters adjusted.

Notably, increasing the testing ratio from 20% to 35% resulted in a severe decrease in overall accuracy.

During the research, various CNN models have been trained with iterating key parameters such as epoch count, batch size, train-test split or adjusting datasets and more. As a result of training and testing, the overall best result was achieved by the CNN3 model with an accuracy of 96%, notably, the deviation is 3 times bigger in comparison to CNN1 when the result increased only by 0.0005.

V. CONCLUSIONS

The purpose of the work was to develop and implement a method for detecting fake document images using a neural network. The developed system allows to automatically determine the authenticity of a photographed or scanned image of a personal identification document.

The final system has the following functionality:

– testing uploaded document images on a selected trained neural network model with an ability to customize input settings and visualize results

– training model on a custom dataset with the ability to change key model parameters, saving the model as a file or folder;

– customizing functional modules.

The developed system for detecting fake document images can be an effective way of reducing fraud in personal identification systems. It can help in the automatic detection of fake personal identification documents which can be used for identity theft, credit card fraud or other fraudulent activity.

Author published theses in proceedings of International scientific and practical conference on the scientific work topic [8].

VI. REFERENCES

1. Legalclinics. Personal identification documents in Ukraine. https://legalclinics.in.ua/consult/consultation-26-06-2020-5/

2. Core.ac.uk. Document types and protection methods. https://core.ac.uk/download/pdf/12241561.pdf

3. Dmsu. Ukraine ID card protection layers. https://dmsu.gov.ua/faq/biometrichni-dokumenti-v-ukrajni/yaka-stupin-zaxistu-u-biometrichnix-dokumentiv.html

4. Evergreens. CNN basics and usage https://evergreens.com.ua/ua/articles/cnn.html

5. Slobodzian V., Molchanova M., Kovalchuk O., Sobko O., Mazurets O., Barmak O., Krak I. (2020). An Approach Based on the Visualization Model for the Ukrainian Web Content Classification. Ieeexplore. https://ieeexplore.ieee.org/document/9913162.

6. Kaggle. Cassia2 dataset. https://www.kaggle.com/datasets/sophatvathana/casia-dataset

7. PapersWithCode. MIDV-500 dataset. https://paperswithcode.com/dataset/midv-500

8. Zharnovskyi O., Sobko O., Klimenko V. Intelligent system for neural network detection of fake document images for automated personality identification. Proceedings of IV International Scientific and Practical Conference. «Innovative research and perspectives of the development of science and technology». January 29-31, 2024. Stockholm, Sweden. Pp. 337-343. https://eu-conf.com/wp-content/uploads/2024/01/INNOVATIVE-RESEARCH-AND-PERSPECTIVES-OF-THE-DEVELOPMENT-OF-SCIENCE-AND-TECHNOLOGY.pdf

Довідка: ВХНУ ТН 4-12/2024

Видання: Herald of Khmelnytskyi National University. Technical Sciences (Вісник Хмельницького національного університету. Технічні науки)

Категорія фаховості видання: затверджено як наукове фахове видання України, у якому можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії, категорії «Б» (наказ МОН №1643 від 28.12.2019, наказ МОН №409 від 17.03.2020).

Напрям – технічні науки за спеціальностями – 101, 121, 122, 123, 124, 125, 141, 151, 161, 172, 181, 182 (28.12.2019), спеціальності – 131, 132, 133 (17.03.2020).

Назва статті: НЕЙРОМЕРЕЖЕВЕ ВИЯВЛЕННЯ ФЕЙКОВИХ ЗОБРАЖЕНЬ ЛЮДЕЙ

Автори: Мазурець О.В., Жарновський О.В., Гладун О.В., Собко О.В.
Хмельницький національний університет

Номер, у який прийнято статтю: №1 за 2025 рік.

04.12.2024

Начальника відділу
інтелектуальної власності та трансферу технологій Ю.В.Кравчик

ВІДДІЛ КАДРІВ

УДК 004.8

МАЗУРЕЦЬ ОЛЕКСАНДР
Хмельницький національний університет
https://orcid.org/0000-0002-8900-0650
e-mail: exe.chong@gmail.com
ЖАРНОВСЬКИЙ ОЛЕКСАНДР
Хмельницький національний університет
e-mail: stalkar2v2@gmail.com
ГЛАДУН ОЛЕКСАНДР
Хмельницький національний університет
e-mail: olexandrgladun@gmail.com
СОБКО ОЛЕНА
Хмельницький національний університет
https://orcid.org/0000-0001-5371-5788
e-mail: olena.sobko.ua@gmail

НЕЙРОМЕРЕЖЕВЕ ВИЯВЛЕННЯ ФЕЙКОВИХ ЗОБРАЖЕНЬ ЛЮДЕЙ

Запропоновано метод ідентифікації згенерованих штучним інтелектом фейкових зображень людей засобами машинного навчання. Метод грунтується на використанні комбінації двох згорткових нейронних мереж для автоматизованого аналізу зображень з метою ідентифікації автентичності та походження. Особливістю запропонованого методу є те, що він дозволяє не тільки ідентифікувати згенеровані штучним інтелектом зображення, але і метод їх генерації, що може значно покращити ідентифікацію згенерованих штучним інтелектом зображень. Особливістю запропонованого методу є те, що від дозволяє не тільки ідентифікувати згенеровані штучним інтелектом фейкові зображення, але і метод їх генерації, що може значно покращити ідентифікацію згенерованих штучним інтелектом зображень. Метод працює на основі перетворення вхідних даних у вигляді зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані у вигляді відсоткової оцінки автентичності зображення та його походження. На першому кроці зображення проходить через попередню обробку, що включає в себе перетворення, зміна розміру та створення тензору. Другим кроком потрібно завантажити натреновану мережу для ідентифікації зображення як згенерованого мережею чи реального, та проаналізувати завантажене зображення; в результаті отримуються відсоткові значення що потрібно проаналізувати та вивести. Якщо вхідне зображення класифіковано як згенероване, то виконується третій крок і потрібно завантажити мережу для ідентифікації походження та додатково проаналізувати зображення. Архітектура нейронної мережі використовує розширену кількість вихідних шарів відповідно до методів генерації.

Для навчання мережі було підготовлено 17000 зображень людей та розроблений відповідний програмний додаток для дослідження ефективності. Встановлено, що використання тренувальної вибірки у відношенні класі згенеровані та реальні 1 до 1, збільшення розмірів груп, застосування випадкової аугментації та нормалізації сприяло значному покращенню кінцевих показників при менших витратах у часі. Це дозволило покращити максимальну точність класифікаторів до 92% для ідентифікації автентичності та 95% для встановлення методу генерації. В той же час збільшення коефіцієнту навчання дало негативний результат, а збільшення кількості епох не дало видимих результатів при значних збільшеннях витрат у часі. Одержані результати свідчать про спроможність запропонованого методу ефективно ідентифікувати згенеровані штучним інтелектом фейкові зображення засобами машинного навчання.

Ключові слова: згорткові нейронні мережі, класифікація зображень, фейкові зображення, штучний інтелект.

MAZURETS OLEKSANDR, ZHARNOVSKYI OLEKSANDR, HLADUN OLEKSANDR, SOBKO OLENA
Khmelnytskyi National University

NEURAL NETWORK DETECTION OF FAKE IMAGES OF PEOPLE

This work proposes method for identifying artificial intelligence-generated images of people using machine. The method is based on a combination of two convolutional neural networks for automated image analysis to identify authenticity and origin. The peculiarity of the proposed method is that it allows not only to identify artificial intelligence-generated images, but also the method of their generation, which can significantly improve the identification of artificial intelligence-generated images. The peculiarity of the proposed method is that it allows not only to identify artificial intelligence-generated images, but also the method of their generation, which can significantly improve the identification of artificial intelligence-generated images. The method works on the basis of converting input data in the form of an image, a model for image identification, a model for finding the origin into output data in the form of a percentage assessment of the authenticity of the image and its origin. In the first step, the image undergoes preprocessing, which includes transformation, resizing and tensor creation. In the second step, it is necessary to load a trained network to identify the image as generated by the network or real, and analyze the

loaded image; as a result, percentage values are obtained that need to be analyzed and output. If the input image is classified as generated, then the third step is performed and it is necessary to load a network for identifying the origin and additionally analyze the image. The neural network architecture uses an extended number of output layers according to the generation methods.

Dataset that consist of 17,000 images of people was prepared for training the network and a corresponding software application was developed to study its effectiveness. It was established that the use of a training sample in the ratio of generated and real classes 1 to 1, increasing the size of groups, applying random augmentation and normalization contributed to a significant improvement in the final metrics with less time. This allowed improving the maximum accuracy of classifiers to 92% for identifying authenticity and 95% for establishing the generation method. At the same time, increasing the learning rate gave a negative result, and increasing the number of epochs did not give visible improvements with significant increases in time. The results obtained indicates the ability of the proposed method to effectively identify images generated by artificial intelligence using machine learning.

Keywords: convolutional neural networks, image classification, fake images, artificial intelligence.

Аналіз предметної області та постановка задачі

Технологія генерації штучних зображень має широкий спектр застосувань, що робить її корисною в багатьох сферах людської діяльності для покращення креативності, продуктивності та часткової автоматизації [1], але в той же час, використання штучного інтелекту для генерації зображень дозволяє людям зловживати цією технологією для особистої вигоди.

Створення дипфейків є однією із головних проблем простого та загальнодоступного ШІ для генерації якісних зображень. Згенеровані штучним інтелектом зображення та неправдиві твердженням можуть швидко поширитися у соціальних мережах, охопивши тисячі людей. Цю проблему ускладнюють алгоритми, які віддають перевагу сильному емоціональному вмісту, що є часто суперечливим або сенсаційний матеріал, що призводить до дезінформації [2].

У статті пропонується метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання що грунтується на використанні комбінації двох згорткових нейронних мереж для автоматизованого аналізу зображень з метою ідентифікації автентичності та походження.

Останні публікації

Класифікація згенерованих штучним інтелектом зображень є досить складною задачею через комплексні методи генерації та варіативність. Щоб вирішити цю проблему у [3] розглядають покращення існуючих методів ідентифікації згенерованих штучним інтелектом зображень. Метод використовує розроблену згорткову нейронну мережу MobileNet-v2 для ідентифікації зображень обличь людей. В результаті тестування було отримано точність у 72% та зазначено що модель має значні труднощі із зображеннями що містять аксесуари. Для покращення результатів були використані випадкові трансформації, зміна яскравості та контрасту, а також зміна параметрів мережі що дозволили покращити результат до 76%.

Метод розглянутий у [4] зосереджується на аналізі згенерованих штучним інтелектом робіт мистецтва. Розроблений метод використовує існуючі згорткові нейронні мережі VGG-19, ResNet-50 та ViT та створений датасет із зображень людей. В результаті тестування ефективності моделей отримані результати демонструють найкращий результат у мережі ViT з точністю в 97% на 50000 зображень розміром 200x200.

Інший підхід було розглянуто у [5], що використовують по-піксельний аналіз зображень використовуючи PRNN та ELA. Розроблений метод використовує власну архітектуру згорткової нейронної мережі та створений датасету із реальних зображень, та зображень згенерованих Dall E, Stable Diffusion та Open Art. Ефективність розробленої моделі визначали за допомогою метрик Accuracy, Precision, Recall, F1. Зазначається що використання ELA мали незначні відмінності метрик в порівнянні з PRNN, але значно швидше тренування на епоху.

Метою роботи є підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання.

Основна частина

Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання призначений для автоматизованої ідентифікації зображень людей згенерованих штучним інтелектом. Особливістю запропонованого методу є те, що від дозволяє не тільки ідентифікувати згенеровані штучним інтелектом зображення, але і метод їх генерації, що може значно покращити ідентифікацію згенерованих штучним інтелектом зображень.

Схема методу наведена на рис.1. Метод працює на основі перетворення вхідних даних у вигляді зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані у вигляді відсоткової оцінки автентичності зображення та його походження

На першому кроці зображення проходить через попередню обробку, що включає в себе перетворення, зміна розміру та створення тензору. Попередньо оброблене зображення буде використане навченою нейронною мережею для оцінки.

Рис. 1. Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

Другим кроком потрібно завантажити натреновану мережу для ідентифікації зображення як згенерованого мережею чи реального, та проаналізувати завантажене зображення. В результаті отримуються відсоткові значення що потрібно проаналізувати та вивести. Використана нейромережа для аналізу автентичності зображена схематично на рис.2. Архітектура мережі складається із вхідного шару, трьох шарів згортки, шару втрат да двох повноз'єдних шарів. Використання шарів згортки добре підходить для задач класифікації зображень, проте значно збільшує витрати у часі.

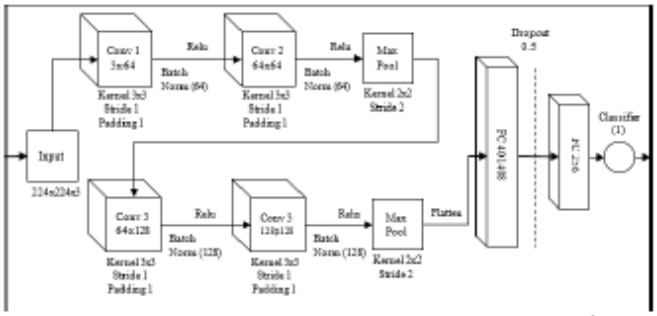


Рис. 2. Архітектура згорткової нейронної мережі для ідентифікації зображень людей

В залежності від того чи зображення реальне чи згенероване, виконується третій крок. Якщо вхідне зображення класифіковано як згенероване потрібно завантажити мережу для ідентифікації походження та додатково проаналізувати зображення. Аріхектура мережі наведена на рис.3. Архітектура мережі використовує розширену кількість вихідних шарів відповідно до методів генерації.

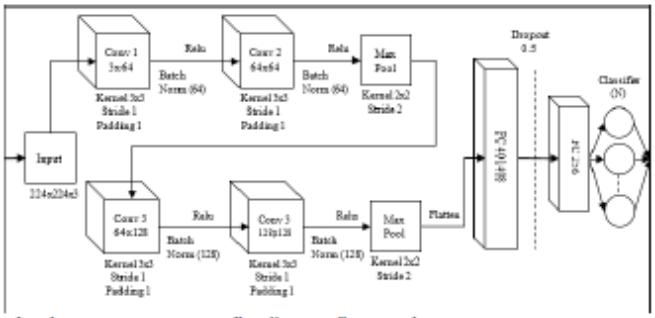


Рис. 3. Архітектура згорткової нейронної мережі для виявлення походження

Результатом аналізу зображення є відсоткова оцінка та віднесення його до одного з двох класів «реальне» чи «згенероване» для мережі по ідентифікації штучних зображень людей, та «Stable Diffusion», «StyleGan», «PhotoshopExperts» для мережі виявлення походження. Мережа може бути додатково налаштована для використання нових методів генерації.

Отже, запропонований метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання дозволяє виявляти штучні зображення, ґрунтуючись на використанні комбінації двох згорткових нейронних мереж для автоматизованого аналізу зображень з метою ідентифікації автентичності та походження

Дані дослідження

Для навчання розроблених нейромереж було сформовано набір даних із 17000 зображень. Зображення були взяті із Flickr-Faces-HQ Dataset (Nvidia) [6], 1 Million Fake Faces [7], Face Dataset Using Stable Diffusion v.1.4 [8], Real vs Fake Faces dataset. [9]. Сформований набір розподілений на два датасети.

Датасет для ідентифікації зображення, розподілений на «реальні» та «згенеровані» та складається із 14000 зображень, по 5000 на клас, а також 2000 для валідації та 2000 для тестування.

Датасет для виявлення походження розподілених на три класи «Stable Diffusion», «StyleGan», «PhotoshopExperts» та складається із 3000 зображень, по 800 зображень на клас для тренування, 300 для валідації та 300 для тестування. Зображення стандартизовано до розміру 224x224, нормалізовано та перетворено у тензор.

Таким чином був описаний набір даних, що буде використаний для навчання та тестування розроблених нейромереж.

Дослідження ефективності методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

Для оцінки ефективності розробленого методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання був розроблений програмний додаток що складається із 4 модулів для взаємодії із користувачем написаних на мові C#, та двох функціональних модулів мови Python. Приклад роботи застосунку наведений на рис. 4.

Рис. 4. Програмна реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

В подальшому використовуючи функціональні модулі Python було виконано дослідження ефективності розробленого методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Результати експерименту з дослідження ефективності розробленого методу

Експеримент було проведено використовуючи розроблені нейромережі та отриманий набір даних [10, 11]. Для оцінки ефективності методу були натреновані нейромережі з параметрами наведеними на таблиці 1.

Таблиця 1

Параметри навчання нейромереж

		Parameters				
		Epochs	Learning Rate	Batch size	Dropout	Normalization
ImageClassifier	Model_1_0	10	0.001	32	0.5	None
	Model_1_1	10	0.001	64	0.5	Mean, STD
	Model_1_2	10	0.005	64	0.5	Mean, STD
	Model_1_3	30	0.001	64	0.3	Mean, STD
MethodClassifier	Model_2_0	10	0.001	32	0.5	None
	Model_2_1	10	0.005	64	0.5	Mean, STD
	Model_2_2	20	0.001	64	0.5	Mean, STD

Для оцінки під час тренування та валідації були використані метрики Accuracy та Loss [12, 13]. Завдяки збільшенню розмірів групи, використанню нормалізації та відношенню кількості зображень класів 1 до 1, вдалося значно покращити показники точності з 60% до 70% (рис. 5). В той же час збільшення коефіцієнту навчання дало негативний результат в 53%. При збільшенні кількості епох до 30 для мережі ідентифікації зображення та до 20 для знаходження походження результат покращився до 95%. При послідуючому збільшенні кількості епох не було отримано значних покращень при значних витратах у часі.

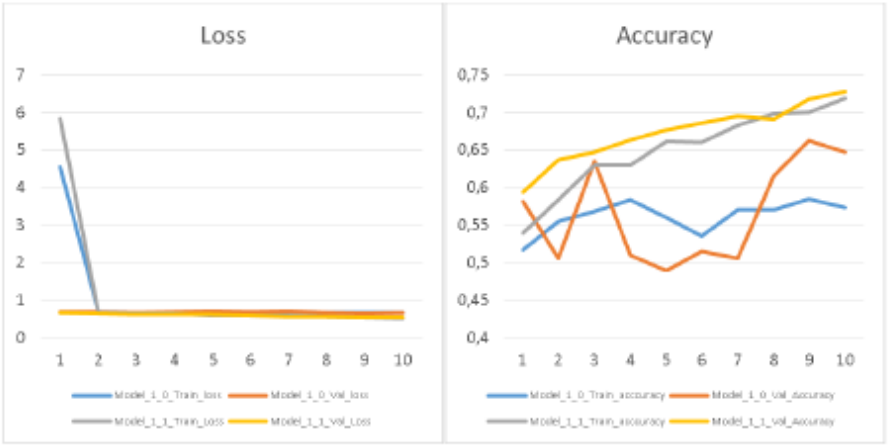
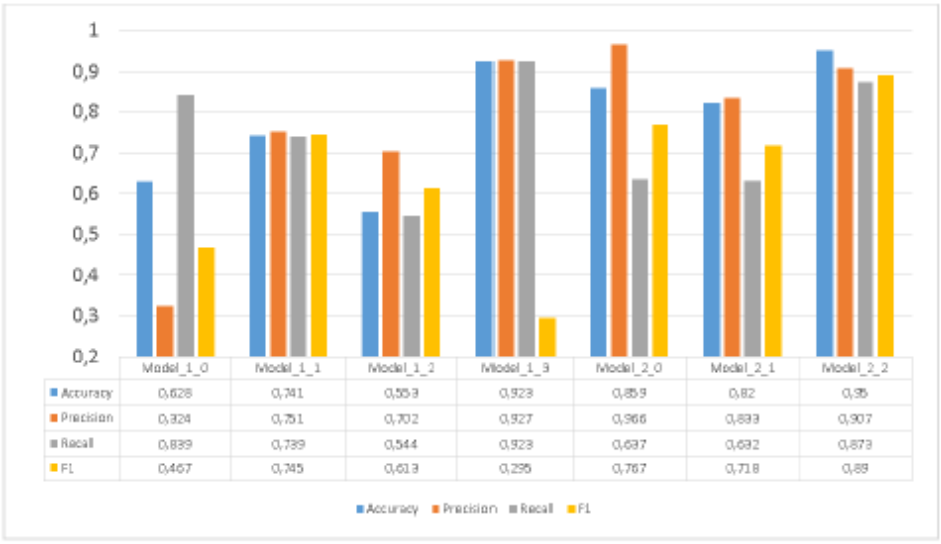


Рис. 5. Порівняння метрик мереж під час тренування

Для оцінки під час тестування були використані метрики Accuracy, Precision, Recall, F1. Результати тестування наведені на рис. 6.



	Model_1_0	Model_1_1	Model_1_2	Model_1_3	Model_2_0	Model_2_1	Model_2_2
Accuracy	0,628	0,741	0,559	0,923	0,859	0,82	0,95
Precision	0,324	0,751	0,702	0,927	0,966	0,893	0,907
Recall	0,839	0,739	0,544	0,923	0,637	0,692	0,873
F1	0,467	0,745	0,613	0,295	0,767	0,718	0,89

Рис. 6. Результати тестування

Як видно із рис. 6, збільшення кількості групи, застосування трансформацій, нормалізації та збільшення кількості епох мали позитивний ефект на результати під час тестування, а в свою чергу збільшення коефіцієнту навчання негативний. Найкращий отриманий параметр точності в 92% для мережі по ідентифікації зображення та 95% для виявлення походження.

Одержані результати свідчать про спроможність запропонованого методу ефективно ідентифікувати згенеровані штучним інтелектом зображення засобами машинного навчання. Ітерація параметрів навчання мережі дозволили значно покращити ефективність методу. Подальші дослідження будуть спрямовані на покращення архітектури нейромережі, а також на алгоритми тренування та тестування.

Висновки

У статті проведено аналіз сучасного стану сфери ідентифікації зображень людей згенерованих штучним інтелектом, виконаний аналіз та окреслено задачі для покращення та запропонований метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання, що грунтується на використанні комбінації двох згорткових нейронних мереж для автоматизованого аналізу зображень з метою ідентифікації автентичності та походження та дозволяє не тільки ідентифікувати згенеровані штучним інтелектом зображення, але і метод їх генерації, що може значно покращити ідентифікацію згенерованих штучним інтелектом зображень

Для навчання розроблених нейромереж було сформовано набір даних із 17000 зображень. Сформований набір складається із двох датасетів, для ідентифікації зображення, розподілений на «реальні» та «згенеровані», та датасет для виявлення походження розподілених на три класи «Stable Diffusion», «StyleGan», «PhotoshopExperts» .

Для оцінки ефективності запропонованого методу було створено програмну реалізацію, яка складається із набору ноутбуків, реалізованих у хмарному сервісі «Google Colab» (для навчання нейромережевої моделі гібридної архітектури та для розширення отриманого набору даних методом аугментації тексту), та застосунку з графічним інтерфейсом користувача на мові Python.

Для оцінки ефективності розробленого методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання був розроблений програмний додаток що складається із 4 C# модулів для взаємодії із користувачем та двох функціональних модулів Python.

Із проведеного дослідження видно, що збільшення кількості групи, застосування трансформацій, нормалізації та збільшення кількості епох мали позитивний ефект на результати під час тестування, а в свою чергу збільшення коефіцієнту навчання негативний. Найкращий отриманий параметр точності в 92% для мережі по ідентифікації зображення та 95% для виявлення походження. Отримані результати свідчать про спроможність запропонованого методу ефективно ідентифікувати згенеровані штучним інтелектом зображення засобами машинного навчання. Ітерація параметрів навчання мережі дозволили значно покращити ефективність методу.

Література

1. How AI Image Generators Can Benefit You [Електронний ресурс] – 2024 – Режим доступу до ресурсу: https://www.linkedin.com/pulse/how-ai-image-generators-can-benefit-you-imran-butt-frzff.
2. Mapping the misuse of generative AI [Електронний ресурс] – 2024 – режим доступу до ресурсу: https://deepmind.google/discover/blog/mapping-the-misuse-of-generative-ai/.
3. Analyzing and Improving Existing Neural Network-Based Approaches to Identify AI Generated Images [Електронний ресурс] – 2024 – Режим доступу до ресурсу: https://www.jsr.org/hs/index.php/path/article/view/5931.
4. Identifying AI-Generated Art with Deep Learning [Електронний ресурс] – 2023 – Режим доступу до ресурсу: https://www.researchgate.net/publication/375116319_Identifying_AI-Generated_Art_with_Deep_Learning.
5. Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks [Електронний ресурс] – 2023 – Режим доступу до ресурсу: https://pmc.ncbi.nlm.nih.gov/articles/PMC10674908/.
6. Flickr-Faces-HQ Dataset (Nvidia) - Part 1 [Електронний ресурс] – Режим доступу до ресурсу: https://www.kaggle.com/datasets/xhlulu/flickrfaceshq-dataset-nvidia-part-1
7. 1 Million Fake Faces – 1 dataset [Електронний ресурс] – режим доступу до ресурсу: https://www.kaggle.com/datasets/tunguz/1-million-fake-faces
8. Face Dataset Using Stable Diffusion v.1.4 [Електронний ресурс] – Режим доступу до ресурсу: https://www.kaggle.com/datasets/bwandowando/faces-dataset-using-stable-diffusion-v14
9. Real vs Fake Faces dataset [Електронний ресурс] – Режим доступу до ресурсу: https://www.kaggle.com/datasets/uditsharma72/real-vs-fake-faces
10. Zharnovskyi O. Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning / O. Zharnovskyi, O. Mazurets, O. Sobko // Key Aspects of the Development of Scientific Research in Modern Conditions: Proceedings of the XLV International Scientific and Practical Conference. – 2024. – Constanta, Romania. – Pp. 69–73.
11. Жарновський О.В. Практична реалізація методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання / О.В. Жарновський, Я.М. Казмірчук, О.В. Собко, О.В. Мазурець // Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». – 2024. – Хмельницький. – С. 198–204.
12. Pokhytun A. Method for Neural Network Detecting Changed Images of People's Faces Using CNN / A. Pokhytun, O. Mazurets, M. Molchanova, O. Tyschenko // New Horizons in Scientific Research: Challenges and Solutions: Proceedings of the 1st International Scientific and Practical Conference. – 2024. – Marseille, France. – Pp. 35–40.
13. Zharnovskyi O. Neural Network Method for Detection of Fake Document Images for Personality Identification Systems / O. Zharnovskyi, O. Mazurets, O. Sobko // Black Sea Science 2024: Proceedings of the International Competition of Student Scientific Works. – 2024. – Odesa: ONUT. – Pp. 434–448.

References

1. How AI Image Generators Can Benefit You [Elektronnyi resurs] – 2024 – Rezhym dostupu do resursu: https://www.linkedin.com/pulse/how-ai-image-generators-can-benefit-you-imran-butt-frzff.
2. Mapping the misuse of generative AI [Elektronnyi resurs] – 2024 – rezhym dostupu do resursu: https://deepmind.google/discover/blog/mapping-the-misuse-of-generative-ai/.
3. Analyzing and Improving Existing Neural Network-Based Approaches to Identify AI Generated Images [Elektronnyi resurs] – 2024 – Rezhym dostupu do resursu: https://www.jsr.org/hs/index.php/path/article/view/5931.
4. Identifying AI-Generated Art with Deep Learning [Elektronnyi resurs] – 2023 – Rezhym dostupu do resursu: https://www.researchgate.net/publication/375116319_Identifying_AI-Generated_Art_with_Deep_Learning.
5. Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks [Elektronnyi resurs] – 2023 – Rezhym dostupu do resursu: https://pmc.ncbi.nlm.nih.gov/articles/PMC10674908/.
6. Flickr-Faces-HQ Dataset (Nvidia) - Part 1 [Elektronnyi resurs] – Rezhym dostupu do resursu: https://www.kaggle.com/datasets/xhlulu/flickrfaceshq-dataset-nvidia-part-1
7. 1 Million Fake Faces – 1 dataset [Elektronnyi resurs] – rezhym dostupu do resursu: https://www.kaggle.com/datasets/tunguz/1-million-fake-faces
8. Face Dataset Using Stable Diffusion v.1.4 [Elektronnyi resurs] – Rezhym dostupu do resursu: https://www.kaggle.com/datasets/bwandowando/faces-dataset-using-stable-diffusion-v14

9. Real vs Fake Faces dataset [Elektronnyi resurs] – Rezhym dostupu do resursu: https://www.kaggle.com/datasets/uditsharma72/real-vs-fake-faces

10. Zharnovskyi O. Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning / O. Zharnovskyi, O. Mazurets, O. Sobko // Key Aspects of the Development of Scientific Research in Modern Conditions: Proceedings of the XLV International Scientific and Practical Conference. – 2024. – Constanta, Romania. – Pp. 69–73.

11. Zharnovskyi O.V. Praktychna realizatsiia metodu identyfikatsii zghenerovanykh shtuchnym intelektom zobrazhen liudei zasobamy mashynnoho navchannia / O.V. Zharnovskyi, Ya.M. Kazmirchuk, O.V. Sobko, O.V. Mazurets // Zbirnyk naukovykh prats za materialamy XVI Vseukrainskoi naukovo-praktychnoi konferentsii «Aktualni problemy kompiuternykh nauk APKN-2024». – 2024. – Khmelnytskyi. – S. 198–204.

12. Pokhytun A. Method for Neural Network Detecting Changed Images of Peoples Faces Using CNN / A. Pokhytun, O. Mazurets, M. Molchanova, O. Tyschenko // New Horizons in Scientific Research: Challenges and Solutions: Proceedings of the 1st International Scientific and Practical Conference. – 2024. – Marseille, France. – Pp. 35–40.

13. Zharnovskyi O. Neural Network Method for Detection of Fake Document Images for Personality Identification Systems / O. Zharnovskyi, O. Mazurets, O. Sobko // Black Sea Science 2024: Proceedings of the International Competition of Student Scientific Works. – 2024. – Odesa: ONUT. – Pp. 434–448.

Додаток Г

Презентаційний матеріал

Кваліфікаційна робота магістра

Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

Виконав
Студент групи КНм-23-1
Жарновський
Олександр Володимирович

Керівник
к. т. н. доцент кафедри КН
Мазурець Олександр Вікторович

Актуальність

У сучасному світі спостерігається значне зростання використання генеративного штучного інтелекту, що пов'язано з активним розвитком технологій та доступності, простотою у використанні, швидкістю та продуктивністю.

Хоча інструменти генеративного штучного інтелекту дозволяють значно підвищити креативність, продуктивність чи частково автоматизувати види діяльності, вони також можуть бути використані у зловмисних цілях. Соціальні мережі є вразливими до сплеску використання генеративного штучного інтелекту для створення дипфейків з метою маніпуляції публічної думки, монетизації та шахрайства, наклепу, підробки та інших зловживань штучним інтелектом для досягнення власних цілей. У зв'язку з цим виникає необхідність у розробці ефективних методів ідентифікації автентичності зображень.

Розроблений у кваліфікаційній роботі метод має ряд переваг у порівнянні з існуючими методами. Зокрема, він дозволяє ідентифікувати метод походження згенерованого зображення. Це дозволить краще аналізувати методи генерації штучного інтелекту для подальшого покращення ефективності.

2

Мета і задачі роботи

Мета кваліфікаційної роботи магістра

Підвищення точності ідентифікації згенерованих штучним інтелектом зображень людей методами машинного навчання.

Виконати

- Дослідити сучасний стан предметної області генерацій зображень з використанням штучного інтелекту, їх методи та засоби. Виконати аналіз сучасних наукових публікацій у задачах генерації та виявлення зображень створених штучним інтелектом.
- Розробити метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Розроблений метод має забезпечувати визначення автентичності зображення за допомогою відсоткової оцінки та визначення можливих методів використаних для генерації зображення з використанням навченої згорткової нейронної мережі.
- Створити прикладну реалізацію методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.
- Дослідити практичну ефективність застосування методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

3

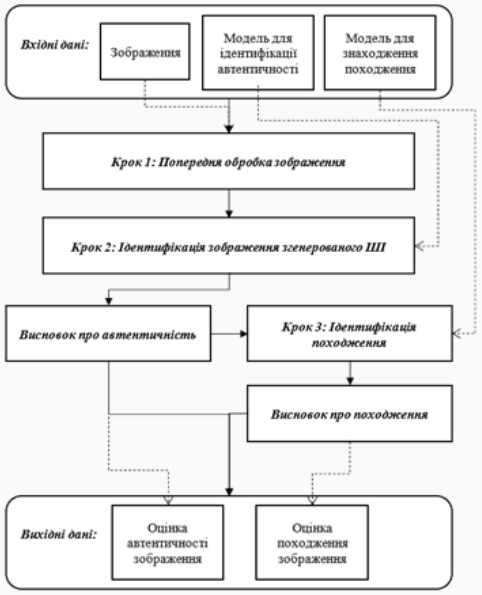


Схема методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

4

Набори даних

Flickr-Faces-HQ Dataset

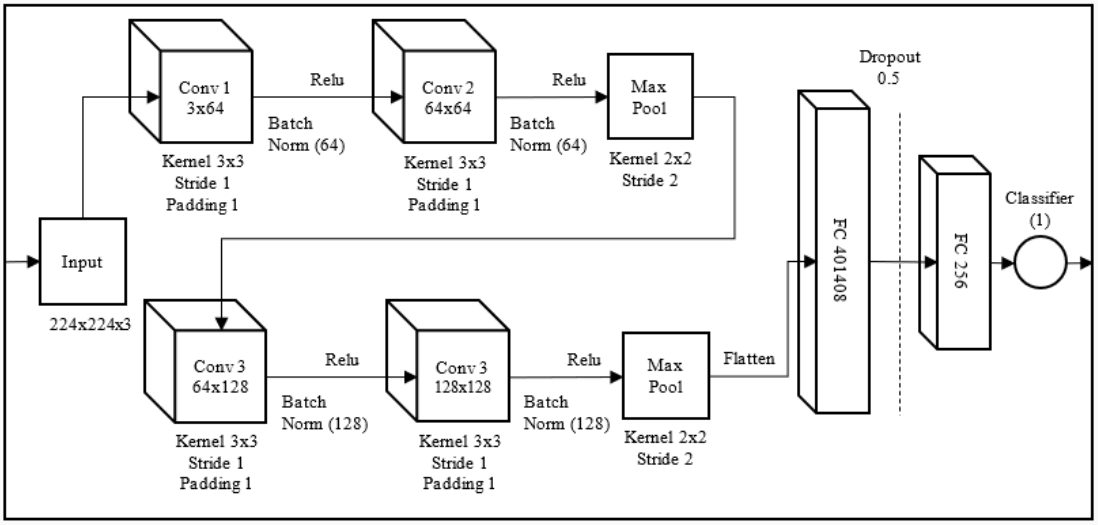
1M AI generated faces

Face Dataset Using Stable Diffusion v.1.4

Real vs fake faces

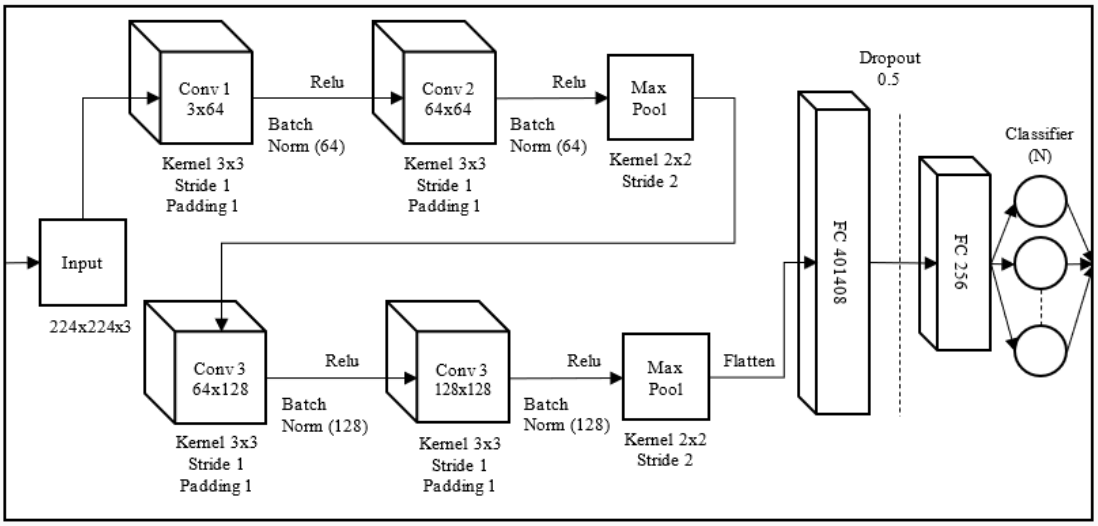
5

Архітектура мережі ImageClassifier



6

Архітектура мережі MethodClassifier

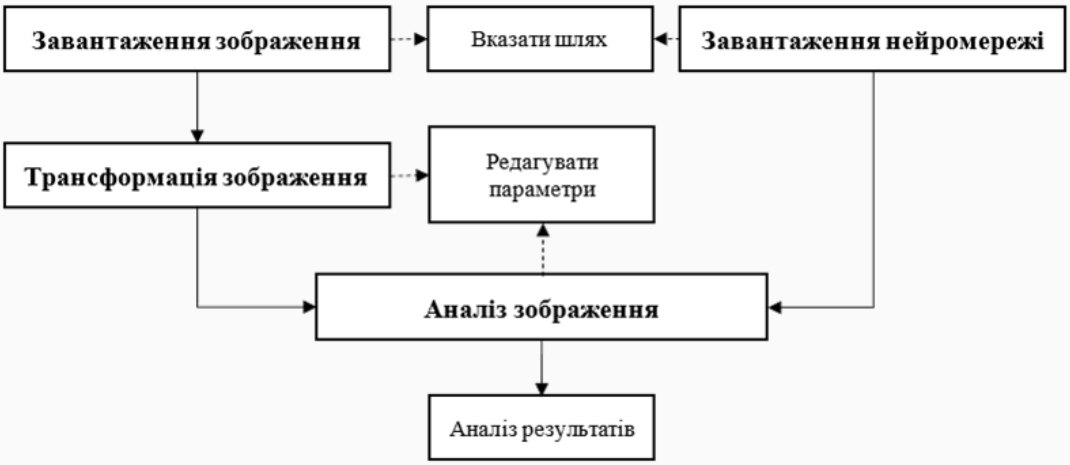


7

Інформаційна структура системи

8

Схема та функції підсистеми розпізнавання зображень

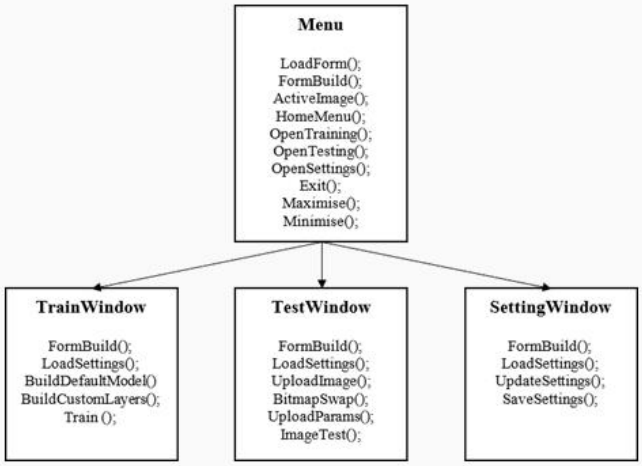


9

Схема та функції підсистеми взаємодії з нм

10

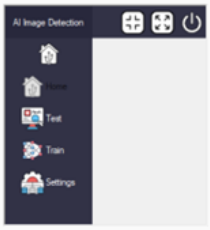
Програмна архітектура системи



11

Програмна реалізація системи

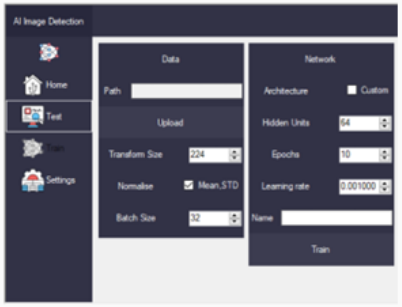
Menu



Settings

Test Window

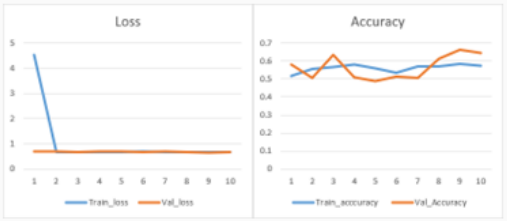
Train Window



12

Дослідження ефективності методу

Метрики під час тренування



Метрики під час тестування

True Label		0	1
	0	324	676
	1	62	922

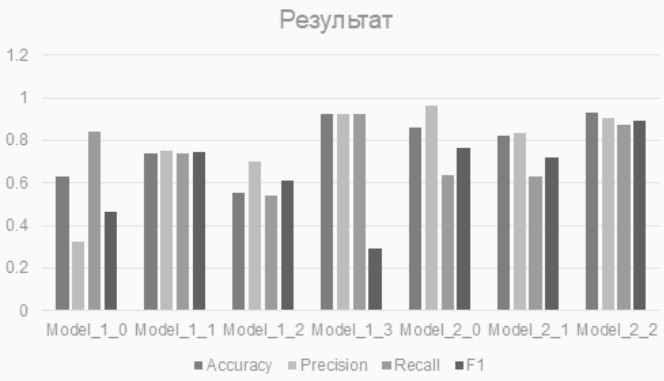
Predicted

- $Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{324+922}{324+922+62+676} = 0.628$
- $Precision = \frac{TP}{TP+FP} = \frac{324}{324+676} = 0.324$
- $Recall = \frac{TP}{TP+FN} = \frac{324}{324+62} = 0.839$
- $F1 = 2 * \frac{Precision*Recall}{Precision+Recall} = 2 * \frac{0.324*0.839}{0.324+0.839} = 0.467$

13

Дослідження ефективності - результат

		Parameters				
		Epochs	Learning Rate	Batch size	Dropout	Normalization
Image Classifier	Model_1_0	10	0.001	32	0.5	None
	Model_1_1	10	0.001	64	0.5	Mean, STD
	Model_1_2	10	0.005	64	0.5	Mean, STD
	Model_1_3	30	0.001	64	0.3	Mean, STD
MethodClassifier	Model_2_0	10	0.001	32	0.5	None
	Model_2_1	10	0.005	64	0.5	Mean, STD
	Model_2_2	20	0.001	64	0.5	Mean, STD



Під час тестування із представлених мереж найкращі результати мають мережі Model_1_3 та Model_2_2 з результатом accuracy в 92% та 93% при тестуванні з експериментальною вибіркою в 2000 та 600 зображень.

14

Висновки

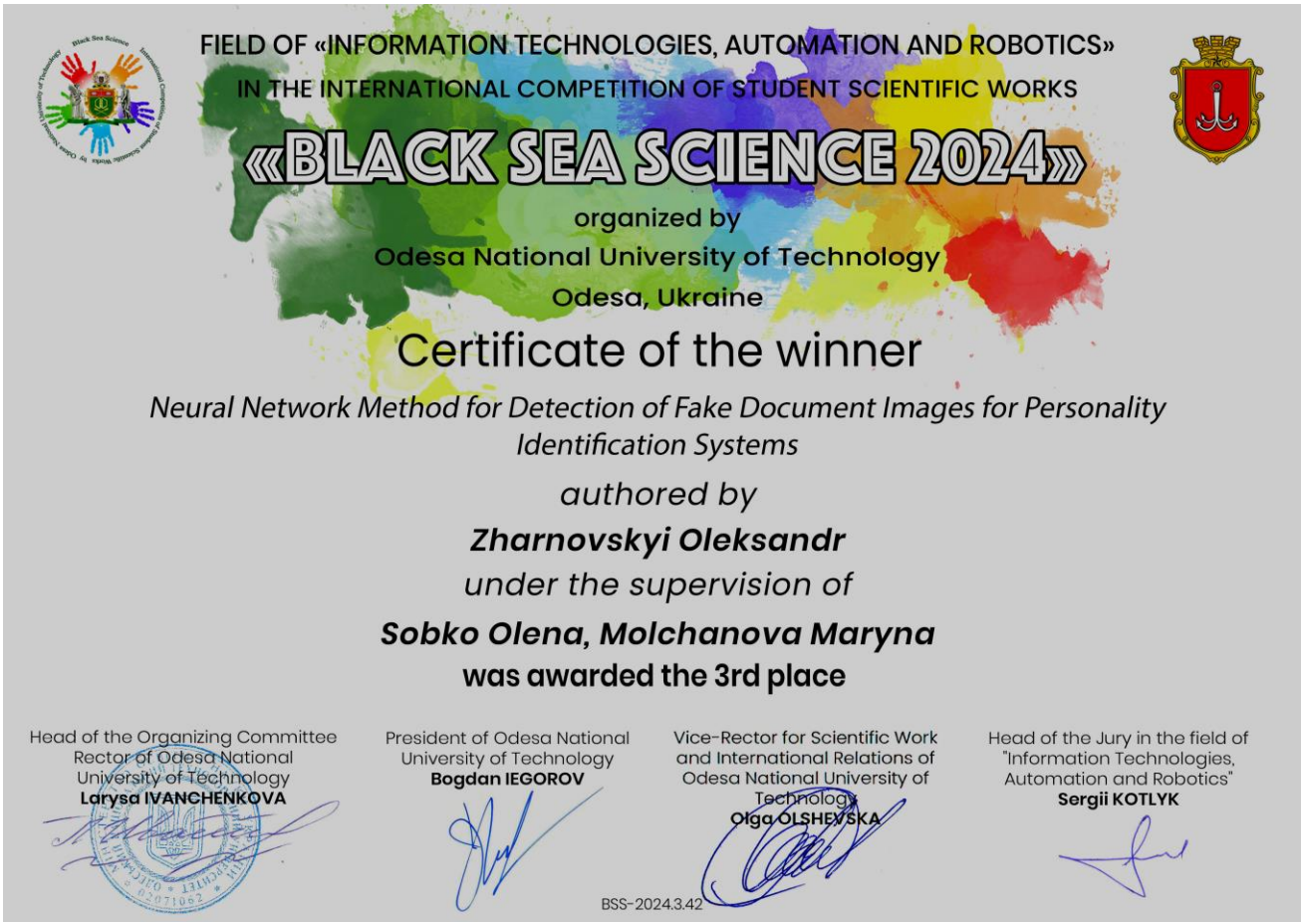
Кваліфікаційна робота магістра розв'язує задачу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Результатом роботи є метод та його програмна реалізація, призначена для ідентифікації зображень завантажених користувачем, що використовує комбінацію двох згорткових нейронних мереж, та працює на основі перетворення вхідних даних – зображення, моделі для ідентифікації зображення, моделі для знаходження походження у вихідні дані – відсоткова оцінка автентичності зображення та його походження.

Для досягнення мети дослідження було виконано:

- Досліджено сучасний стан предметної області генерацій зображень з використанням штучного інтелекту, їх методи та засоби. Виконати аналіз сучасних наукових публікацій у задачах генерації та виявлення зображень створених штучним інтелектом.
- Розроблено метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. Розроблений метод має забезпечувати визначення автентичності зображення за допомогою відсоткової оцінки та визначення можливих методів використаних для генерації зображення з використанням навченої згорткової нейронної мережі.
- Створено прикладну реалізацію методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.
- Досліджено практичну ефективність застосування методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

Додаток Д

Сертифікат про призове місце на фінальному етапі Міжнародного конкурсу студентських наукових робіт «Black Sea Science 2024»



Mon Dec 16 11:18:44 EET 2024, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 8.0%

Словники перевірки: en_US, ru_RU, ua_UA. **Помилок в документах: 13%**

	Документ		Сумарний збіг по Базі Даних	
ID: 159991 Назва: КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА на тему Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання Додано в БД: 2024-12-16 Автора: Олександр ЖАРНОВСЬКИЙ Керівники: Олександр МАЗУРЕЦЬ Консультанти: Опоненти:	Символи	Лексеми	Символи	Лексеми
	78798	1114	11226 (14%)	161 (14%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Олександр ЖАРНОВСЬКИЙ

Співавтор:

Назва: Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання

Науковий керівник: Олександр МАЗУРЕЦЬ, к.т.н., доцент

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1:16.1%

Коефіцієнт подібності 2:2.8%

Мікропробіли: 0

Заміна букв: 0

Інтервали: 0

Білі знаки: 135

Дата створення звіту: 2024-12-16 21:32:03.0

Після аналізу Звіту подібності констатую наступне:

☑ **Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.**

☐ **Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.**

☐ **Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.**

Обгрунтування:

Дата 16.12.2024 експерт Петровський С.Р.

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ
КАФЕДРИ КОМП'ЮТЕРНИХ НАУК
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ МАГІСТРА ДО ЗАХИСТУ
ЗА РЕЗУЛЬТАТАМИ АНАЛІЗУ ЗВІТУ ПОДІБНОСТІ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: *Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання*

Автор: *Олександр ЖАРНОВСЬКИЙ*

Спеціальність: *122 – Комп'ютерні науки*

Освітня програма: *освітньо-професійна*

Науковий керівник: *к.т.н., доц. каф. КН Олександр МАЗУРЕЦЬ*

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	–
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	–
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	–

Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

1) за програмою Anti-Plagiarism виявлені 8%;
2) за програмою StrikePlagiarism КП1 16,1%, КЦ 2,8%;

які містять матеріали огляду предметної області; інші схожості є фрагментарними – містять поширені конструкції, загальновідомі терміни, скорочення та визначення, що, з урахуванням наведених обґрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи. Запозичення, виявлені в роботі є законними і не є плагіатом.

Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Керівник роботи _____ Олександр МАЗУРЕЦЬ

Гарант ОП _____ Руслан БАГРІЙ

Завідувач кафедри КН _____ Олександр БАРМАК

ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу магістра

гр. КНм-23-1 Олександра ЖАРНОВСЬКОГО за темою: *Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання*

1. Актуальність обраної теми

У сучасному світі стрімкий розвиток генеративного штучного інтелекту, зумовлений його доступністю, простотою використання, швидкістю та ефективністю, суттєво змінює різні сфери діяльності. Генеративні інструменти підвищують креативність, продуктивність і дозволяють частково автоматизувати робочі процеси, проте водночас несуть серйозні ризики зловживань. Соціальні мережі стають мішенню для поширення фейкових зображень, створених генеративним штучним інтелектом, що загрожує маніпуляцією суспільною думкою, шахрайством, наклепом, підробками та монетизацією сумнівного контенту. Це підкреслює гостру необхідність у створенні ефективних методів для перевірки автентичності зображень і виявлення фальсифікацій, що робить обрану тему магістерської роботи надзвичайно актуальною.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Кваліфікаційна робота магістра Олександра Жарновського на тему «Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання» повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи.

3. Професійні та особистісні якості магістранта

Під час роботи над кваліфікаційною роботою магістра Олександр Жарновський проявив високий професіоналізм, відповідальність і цілеспрямованість. Його підхід до завдань вирізнявся старанністю, ефективністю та прагненням досягти максимально якісних результатів. Олександр продемонстрував вміння планувати роботу, дотримуватися термінів і стандартів, а також виявляв ініціативу та наполегливість у виконанні поставлених завдань.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Результати, отримані в результаті виконання кваліфікаційної роботи магістра, є результатом самостійної діяльності студента. Отримані положення наукової новизни та

інновації, описані в роботі, дозволили покращити існуючі методи в галузі ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

5. Наукова новизна та оригінальність запропонованих підходів

Результати виконання кваліфікаційної роботи магістра містять інновації та наукову новизну, зокрема було розроблено новий метод, який дозволяє автоматизовано виконувати аналіз завантаженого користувачем зображення, виконуючи при цьому як і аналіз автентичності зображення.

6. Ступінь оволодіння методами дослідження

Магістрант виявив високий ступінь оволодіння необхідними методами дослідження.

7. Повнота та якість розкриття теми роботи

Тема роботи в повній мірі обґрунтована й розкрита, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання у роботі виконані, а також проведено аналіз результатів прикладного застосування запропонованих засобів методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

8. Логічність, послідовність, аргументованість, літературна грамотність викладу матеріалу

Структура роботи й послідовність викладення логічні та відповідні поставленій меті. Викладення матеріалу грамотне та виявляє високий ступінь відповідності стилю.

9. Можливість практичного застосування кваліфікаційної роботи, окремих її частин

Було створено інформаційну систему нейромережевого аналізу згенерованих зображень людей засобами машинного навчання, що є прикладною програмною реалізацією методу аналізу зображень людей згенерованих штучним інтелектом. Інформаційна структура системи складається із набору зображень (датасетів) та кількох підсистем: «Підсистема взаємодії з НМ», «Підсистема розпізнавання завантажених зображень», «Підсистема інтерфейсу користувача», «Підсистема налаштувань», що дозволяють аналізувати завантажене зображення.

10. Висновок про можливість допуску кваліфікаційної роботи до захисту, на яку оцінку заслуговує робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Науковий керівник _____ к.т.н., доц. каф. КН Олександр МАЗУРЕЦЬ

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ
кафедра комп'ютерних наук

ВІДГУК ОПОНЕНТА

на кваліфікаційну роботу магістра

гр. КНм-23-1 Олександра ЖАРНОВСЬКОГО за темою: *Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання*

1. Актуальність обраної теми

Популярність генеративного штучного інтелекту стрімко зростає завдяки його швидкому розвитку, доступності, зручності, високій швидкості та ефективності. Ці інструменти значно підвищують продуктивність, креативність та дозволяють автоматизувати частину робочих процесів, але водночас створюють серйозні ризики зловживань. Зокрема, соціальні мережі стають вразливими до поширення фейкових зображень, створених генеративним штучним інтелектом. Це може призводити до шахрайства, маніпуляції громадською думкою, наклепів, підробок і монетизації недостовірного контенту. Тому є доцільною розробка ефективних рішень для перевірки автентичності візуальної інформації та виявлення фальсифікацій, що підтверджує актуальність теми магістерської роботи.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Обрана тема «Метод ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання», в межах якої виконані поставлені задачі, повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи магістра.

3. Повнота розкриття мети та завдань дослідження

В роботі автор цілком розкриває мету дослідження та поставленні в межах теми завдання.

4. Наявність наукової новизни

Результати кваліфікаційної роботи магістра містять елементи наукової новизни. Зокрема, було розроблено новий метод для ідентифікації зображень, створених за допомогою штучного інтелекту. Цей метод дозволяє автоматизовано аналізувати завантажене користувачем зображення, визначаючи його автентичність та виявляючи можливі способи генерації засобами штучного інтелекту. Досягнення такого результату стало можливим завдяки перетворенню вхідних даних у вигляді зображення на вихідні дані,

що включають відсоткову оцінку автентичності та походження генерації зображення, якщо таке має місце.

5. Зміст кожного розділу роботи

Робота містить чотири розділи: у першому розділі виконано дослідження предметної області генерації зображень засобами штучного інтелекту. Другий розділ присвячено розробці методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання. У третьому розділі виконано проєктування інформаційної системи нейромережевого аналізу згенерованих зображень людей. У четвертому розділі виконано дослідження ефективності методу ідентифікації згенерованих штучним інтелектом зображень людей засобами машинного навчання.

6. Ступінь розкриття теми роботи

Тема кваліфікаційної роботи повною мірою розкрита та обґрунтована, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання у роботі виконані, та проведено аналіз результатів прикладного застосування запропонованих методу і засобів.

7. Якість оформлення кваліфікаційної роботи

Оформлення роботи відповідає необхідним нормам та вимогам, які ставляться до оформлення кваліфікаційних робіт.

8. Недоліки кваліфікаційної роботи

Деякі з джерел, що наведені в переліку посилань не цілком відповідають вимогам оформлення. Суттєвих недоліків не виявлено, робота виконана на високому рівні.

9. Загальний висновок (допускається чи не допускається до захисту), якої оцінки заслуговує кваліфікаційна робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Опонент (прізвище, ім'я, по батькові, посада, місце роботи)

Мартинюк Валерій Володимирович, зав.каф. АКІТ

«17» 12 2024 р

підпис