

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені Ігоря СІКОРСЬКОГО»
Навчально-науковий фізико-технічний інститут
Кафедра математичного моделювання та аналізу даних**

«До захисту допущено»

Завідувачка кафедри

_____ Наталія КУССУЛЬ

«___» _____ 2026 р.

**Дипломна робота
на здобуття ступеня бакалавра**

зі спеціальності: 113 Прикладна математика
на тему: **«Колоризація зображень та відео з інтерактивною
взаємодією користувача та багатоваріантною генерацією результатів.
Аналіз та порівняння різних варіантів колоризації.»**

Виконав:

студент 4 курсу, групи ФІ-21

Климентьєв Максим Андрійович

Керівник:

асистент кафедри ММАД д-р філософії

Железняков Дмитро Валентинович

Рецензент:

посада, степінь, звання

Прізвище Ім'я По-батькові

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

Студент _____

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

**Навчально-науковий фізико-технічний інститут
Кафедра математичного моделювання та аналізу даних**

Рівень вищої освіти — перший (бакалаврський)
Спеціальність — 113 Прикладна математика,
ОПП «Математичні методи моделювання, розпізнавання образів та
комп'ютерного зору»

ЗАТВЕРДЖУЮ

Завідувачка кафедри

_____ Наталія КУССУЛЬ

«___» _____ 2026 р.

**ЗАВДАННЯ
на дипломну роботу**

Студент: Климентьєв Максим Андрійович

1. Тема роботи: *«Колоризація зображень та відео з інтерактивною взаємодією користувача та багатоваріантною генерацією результатів. Аналіз та порівняння різних варіантів колоризації.»*,

керівник: асистент кафедри ММАД д-р філософії Железняков Д. В.,
затверджені наказом по університету №___ від «___» _____ 2026 р.

2. Термін подання студентом роботи: «___» _____ 2026 р.

3. Вхідні дані до роботи: "сіре" зображення або відео, вимоги користувача щодо колоризації у вигляді тексту, точок, ліній, зображення-екземпляру.

4. Зміст роботи: огляд літератури на тему колоризації зображень та відео, дослідження вже готових та програмна реалізація нового методів колоризації зображень та відео, порівняння різних методів колоризації за якістю та швидкодією.

5. Перелік ілюстративного матеріалу: Презентація доповіді.

6. Дата видачі завдання: 3 вересня 2025 р.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання	Примітка
1	Узгодження теми роботи із науковим керівником	01-15 вересня 2025 р.	Виконано
2	Огляд літератури, посилання на джерела	01 вересня - 21 жовтня 2025 р.	Виконано
3	Прогалини. Визначення мети дослідження	16 вересня - 21 жовтня 2025 р.	Виконано
4	Структурна схема	23 вересня - 12 листопада 2025 р.	Виконано
5	Джерела даних	30 вересня - 25 листопада 2025 р.	Виконується
6	Методи	11 жовтня - 25 листопада 2025 р.	Виконується
7	Результати експерименту	26 листопада - 31 січня 2025 р.	Не виконано
8	Висновки	02 січня - 15 березня 2025 р.	Не виконано
9	Посилання	31 січня - 06 квітня 2025 р.	Не виконано

Студент _____ Климентьев М. А.

Керівник _____ Железняков Д. В.

РЕФЕРАТ

Кваліфікаційна робота містить: ??? стор., ??? рисунки, ??? таблиць, ??? джерел.

У рефераті роботи ви повинні коротко (два-три абзаци) викласти, що саме було зроблено у цій роботі. Перші три речення реферату (після статистичних даних) повинні окреслити мету роботи, об'єкт та предмет дослідження. Після цього викладаються основні результати, одержані в ході дослідження.

Наприкінці анотації великими літерами зазначаються ключові слова. Ось так:

КЛЮЧОВІ СЛОВА, СИМЕТРИЧНА КРИПТОГРАФІЯ, ФІЗТЕХ
НАЙКРАЩІЙ

ABSTRACT

The English abstract must be the exact translation of the Ukrainian “annotation” (including statistical data and keywords).

ЗМІСТ

Вступ	7
1 Теоретичні основи колоризації зображень та відео	9
1.1 Поняття та історія колоризації	9
1.2 Основні підходи до колоризації	10
1.2.1 Згорткові нейронні мережі (CNN)	10
1.2.2 Генеративно-змагальні мережі (GAN)	10
1.2.3 Трансформери	11
1.2.4 Дифузійні моделі	11
1.3 Інтерактивна колоризація та багатоваріантні результати	12
1.4 Метрики якості колоризації	13
Висновки до розділу 1	13
2 Аналіз та порівняння існуючих методів колоризації	15
2.1 Алгоритмічні методи колоризації зображень	16
2.1.1 Оптимізаційні методи (Scribble-based)	16
2.1.2 Методи перенесення кольору (Example-based)	16
2.2 Методи колоризації на основі нейронних мереж	17
2.2.1 Регресійні та класифікаційні підходи (CNN)	17
2.2.2 Генеративно-змагальні мережі (GAN)	17
2.2.3 Трансформери та механізми уваги	18
2.2.4 Дифузійні імовірнісні моделі (Diffusion Models)	18
2.3 Особливості колоризації відео	19
2.4 Порівняння різних методів колоризації	19
Висновки до розділу 2	20
Висновки	22
Перелік посилань	24

ВСТУП

Актуальність дослідження. Колоризація зображень і відео має значний вплив на відновлення культурної спадщини, цифрову реставрацію архівних матеріалів, творчі експерименти та інтерактивні системи з багатоваріантною генерацією результатів. Попри значний розвиток алгоритмів колоризації, актуальною залишається проблема контролю користувача над результатом, не завжди достатня передбачуваність кольорових рішень та необхідність багатоваріантної генерації для вибору оптимального результату або надання натхнення. У зв'язку з цим дослідження методів автоматичної, багатоваріантної та інтерактивної колоризації, а також порівняння підходів за якістю, стабільністю й можливістю контролю, є важливим.

Мета дослідження. Розробка та аналіз методів автоматичної та інтерактивної колоризації зображень і відео з можливістю багатоваріантної генерації результатів та подальшим порівнянням різних алгоритмів за якісними та візуальними критеріями.

Задача дослідження.

- 1) провести огляд літератури з методів колоризації зображень і відео;
- 2) проаналізувати класичні та нейромережеві алгоритми колоризації, включаючи моделі на основі GAN, Diffusion, Transformer та CNN;
- 3) знайти та доповнити, за необхідності, датасет для навчання та тестування моделей колоризації;
- 4) розробити власний алгоритм колоризації використовуючи нейромережеву модель;
- 5) розробити програмну реалізацію системи автоматичної та інтерактивної колоризації з багатоваріантним генеруванням результатів;
- 6) провести порівняння реалізованих методів за об'єктивними метриками та за суб'єктивним візуальним аналізом.

Об'єктом дослідження є процеси автоматичної та інтерактивної колоризації цифрових зображень і відео з багатоваріантним генеруванням

результатів.

Предметом дослідження є моделі, алгоритми та методи колоризації та генерації різних варіантів результату, а також метрики для порівняння якості кольоризації.

При розв'язанні поставлених завдань використовувались такі *методи дослідження*: методи глибинного навчання (GAN, Diffusion, Transformer та CNN моделі), методи обробки зображень, метрики технічної точності та інструменти оцінки перцепційної якості.

Наукова новизна.

- 1) інтеграція автоматичної та інтерактивної колоризації в єдиній системі з підтримкою багатоваріантного генерування результатів;
- 2) свій власний алгоритм колоризації зображень і відео на основі нейромережевої моделі;
- 3) порівняння саме цих різних алгоритмів колоризації для зображень і відео за об'єктивними та суб'єктивними критеріями;

Практичне значення. Створення програмної системи, яка дозволяє автоматично та інтерактивно колоризувати зображення й відео та отримувати кілька варіантів результату. Розроблене рішення може бути використане у сфері цифрової реставрації матеріалів, у художніх (творчих) та дизайнерських (практично-функціональних) застосуваннях, а також як інструмент для дослідників у галузі комп'ютерного зору та візуальної аналітики (Наприклад якщо вони мають вже готовий графік і хочуть до нього додати або змінити колір).

1 ТЕОРЕТИЧНІ ОСНОВИ КОЛОРИЗАЦІЇ ЗОБРАЖЕНЬ ТА ВІДЕО

У цьому розділі розглядаються основні теоретичні аспекти колоризації зображень та відео, включно з історією розвитку методів, класичними алгоритмічними підходами та сучасними нейромережевими моделями. Особлива увага приділяється архітектурам від згорткових мереж до дифузійних моделей, а також методам оцінки якості отриманих результатів.

1.1 Поняття та історія колоризації

Колоризація — це процес додавання кольорової інформації до монохромних зображень або відео. З математичної точки зору, цю задачу можна сформулювати як відображення одноканального зображення (grayscale) у триканальне кольорове зображення (наприклад, у просторі RGB, CIELAB або YUV) [1]. Оскільки одному значенню яскравості може відповідати безліч кольорових відтінків, задача є невизначеною та має "один до багатьох" (one-to-many) характер, що вимагає використання апріорних знань про семантику сцени.

Історично методи колоризації можна поділити на три основні етапи:

1) **Ручна та напівавтоматична колоризація:** вимагала значних зусиль художників або використання простих евристик на основі сегментації.

2) **Методи на основі глибокого навчання (CNN та GAN):** поява великих наборів даних (ImageNet, COCO) дозволила тренувати згорткові мережі для передбачення кольору [6]. Знаковим став підхід *Colorful Image Colorization* [26], де задачу було переформульовано з класичної регресії на класифікацію. Замість того, щоб мінімізувати усереднену похибку (що призводить до тьмяних кольорів), модель передбачає розподіл ймовірностей для кожного пікселя по палітрі відтінків.

3) **Генеративні моделі (Transformers та Diffusion Models):** сучасний

етап, що характеризується використанням механізмів уваги та імовірнісних моделей дифузії для генерації високоякісних текстур та забезпечення семантичної узгодженості [21, 22].

Розвиток методів глибокого навчання, зокрема поява архітектур U-Net та GAN (Generative Adversarial Networks), дозволив перейти від простого розфарбовування до генерації фотореалістичних зображень [9, 10]. Останні дослідження все частіше фокусуються на дифузійних моделях, які демонструють state-of-the-art результати у задачах синтезу зображень [17, 18].

1.2 Основні підходи до колоризації

Сучасні підходи до автоматичної колоризації базуються на різних архітектурах нейронних мереж. Розглянемо ключові з них.

1.2.1 Згорткові нейронні мережі (CNN)

Ранні підходи використовували автоенкодери для прямого передбачення каналів кольоровості (наприклад, a та b у просторі CIE Lab). Ласссон та ін. [13] запропонували використовувати ознаки, вивчені для класифікації об'єктів, для задачі колоризації, демонструючи тісний зв'язок між семантикою об'єкта та його кольором. Інший популярний метод, *Deep Koalarization* [2], використовує Inception-ResNet-v2 для вилучення ознак та подвійний декодер для відновлення кольору.

1.2.2 Генеративно-змагальні мережі (GAN)

Для подолання проблеми "сірості" (desaturation), властивій методам на основі функції втрат L2, були запропоновані GAN. Модель *Pix2Pix* [9] використовує умовний GAN (cGAN) для перетворення зображень, що дозволяє генерувати більш чіткі та реалістичні результати. Метод *ChromaGAN*

[19] поєднує геометричну інформацію з семантичним розподілом класів для покращення правдоподібності розфарбовування. Також існують підходи для непарного навчання, такі як CycleGAN [28], що дозволяють тренувати моделі без наявності пар ”чорно-біле — кольорове”.

1.2.3 Трансформери

З появою механізму уваги (Self-Attention) дослідники почали застосовувати трансформери для колоризації. Наприклад, *Colorization Transformer* [12] та *UniColor* [8] використовують архітектуру трансформерів для моделювання довгих залежностей у зображенні, що покращує узгодженість кольорів на великих відстанях.

1.2.4 Дифузійні моделі

Найбільш перспективним напрямком на сьогодні є імовірнісні дифузійні моделі (Denoising Diffusion Probabilistic Models — DDPM). Вони працюють шляхом поступового видалення шуму з випадкового сигналу, керуючись вхідним чорно-білим зображенням.

- **Palette** [17] — універсальний фреймворк для перетворення зображень, який перевершує GAN та регресійні моделі без необхідності налаштування під конкретну задачу.

- **DDColor** [11] використовує подвійні декодери та запити кольору (color queries) для досягнення фотореалістичності.

- **Cold Diffusion** [3] пропонує підхід до інверсії довільних перетворень зображень без використання гаусового шуму, що відкриває нові можливості для відновлення зображень.

Для відео ключовою проблемою є часова узгодженість (temporal consistency). Сучасні методи, такі як *Stable Video Diffusion* [4] та *VanGogh* [5], адаптують дифузійні моделі для обробки послідовностей кадрів. Інші підходи,

наприклад [25] та [16], використовують поширення ознак (feature propagation) та саморегуляризацію для уникнення мерехтіння.

1.3 Інтерактивна колоризація та багатоваріантні результати

Оскільки колоризація є суб'єктивною задачею, важливу роль відіграють методи, що дозволяють користувачеві впливати на результат.

Колоризація на основі прикладів (Exemplar-based): Методи, такі як *Deep Exemplar-based Colorization* [7], використовують референсне кольорове зображення для перенесення колірної палітри на цільове чорно-біле. Це дозволяє досягти високої точності, якщо приклад підібрано вдало.

Колоризація на основі тексту (Language-based): З розвитком мультимодальних моделей (CLIP) з'явилася можливість керувати кольором за допомогою текстових описів.

- **L-CAD** [20] використовує дифузійні пріори для колоризації на основі описів будь-якого рівня деталізації.

- **Diffusing Colors** [24] пропонує фреймворк, де текстові підказки керують процесом дифузії, забезпечуючи баланс між автоматизацією та контролем.

Колоризація на основі підказок користувача (User-guided): Методи *Real-Time User-Guided Image Colorization* [27] та *iColoriT* [23] дозволяють користувачеві ставити кольорові точки або штрихи, які мережа поширює на відповідні семантичні регіони, використовуючи, наприклад, Vision Transformer для кращого розуміння контексту.

Сучасна тенденція, як показано в *Control Color* [15] та *UniColor* [8], полягає у створенні уніфікованих фреймворків, що підтримують різні типи умов (текст, приклад, штрихи) в одній моделі.

1.4 Метрики якості колоризації

Оцінка якості колоризації є складною через відсутність єдиного "правильного" розв'язку. Використовуються наступні групи метрик:

1) **Піксельні метрики:** PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), L1/L2 відстані. Вони порівнюють результат з оригінальним кольоровим зображенням (Ground Truth). Однак ці метрики часто не корелюють з людським сприйняттям якості, оскільки "інший" колір не обов'язково є "неправильним"[26].

2) **Перцептивні метрики:** LPIPS (Learned Perceptual Image Patch Similarity) та FID (Fréchet Inception Distance). FID, зокрема, широко використовується для оцінки дифузійних моделей [21], оскільки оцінює відстань між розподілами ознак згенерованих та реальних зображень.

3) **Семантичні оцінки та дослідження користувачів:** "Тест Тюрінга на колоризацію"(Colorization Turing Test), запропонований Zhang et al. [26], де учасники намагаються відрізнити згенероване зображення від справжнього.

4) **Оцінка барвистості (Colorfulness):** Використовується для перевірки того, чи не є результат тьмяним або надмірно насиченим (метрика C-score).

Висновки до розділу 1

У цьому розділі було проведено огляд теоретичних основ та сучасних методів колоризації зображень і відео. Було показано, що еволюція методів пройшла шлях від простих евристик до складних генеративних моделей, таких як GAN та дифузійні моделі (Palette, DDColor, Cold Diffusion).

Аналіз літератури свідчить, що в той час як задача безумовної колоризації досягла значних успіхів у реалізмі, існують відкриті питання в області керованості процесом (controlability) та часової узгодженості відео (temporal consistency). Зокрема, мультимодальні підходи (текст + зображення) та уніфіковані дифузійні фреймворки є найбільш актуальним напрямком

досліджень.

У наступних розділах буде запропоновано власний підхід до вирішення задачі [Уточнити задачу, наприклад: інтерактивної колоризації відео на основі дифузійних моделей], що базується на проаналізованих архітектурах та спрямований на покращення [вказати мету, наприклад: стабільності кольорів у часі].

2 АНАЛІЗ ТА ПОРІВНЯННЯ ІСНУЮЧИХ МЕТОДІВ КОЛОРИЗАЦІЇ

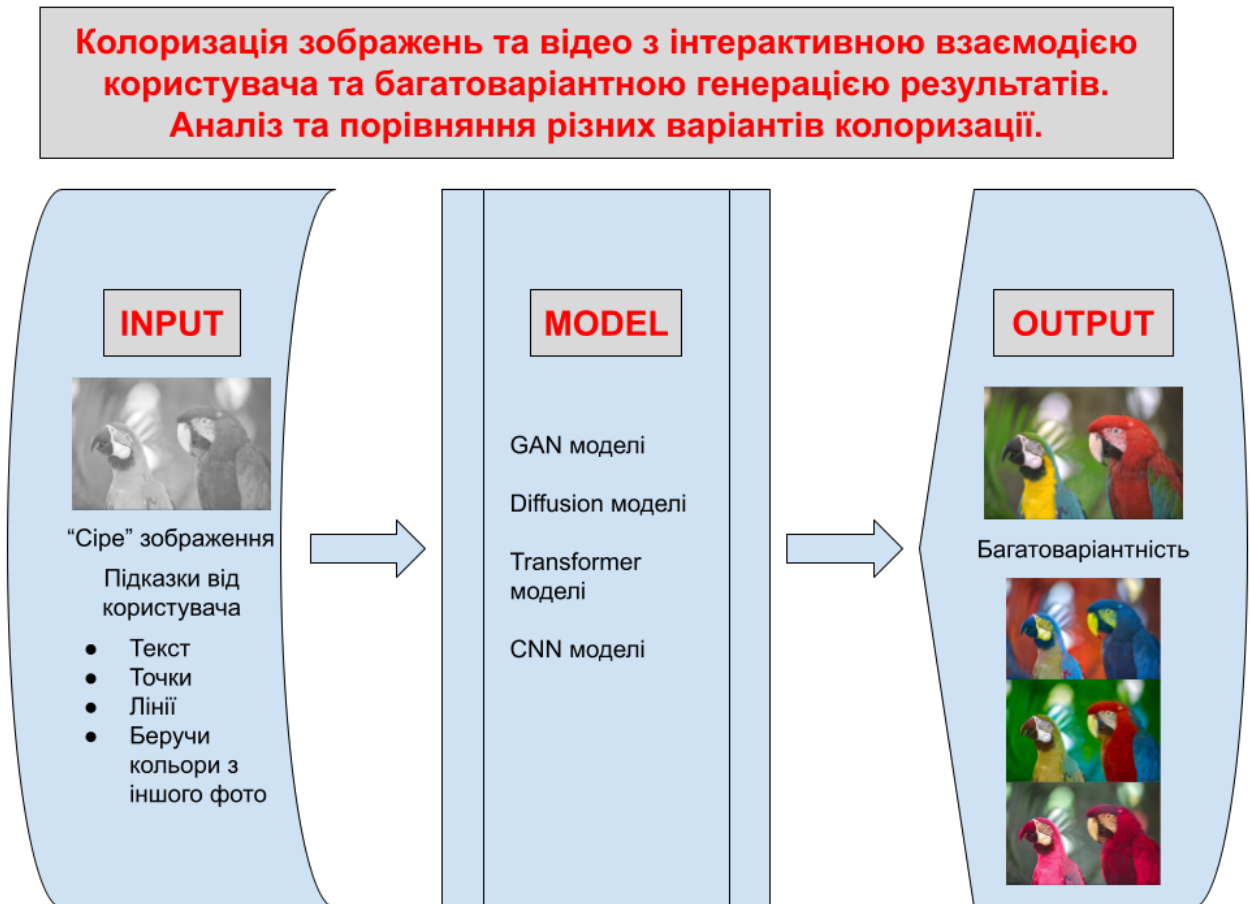


Рисунок 2.1 – Структурна схема класифікації методів колоризації

У цьому розділі проводиться детальний аналіз сучасних алгоритмів колоризації, що зображені на структурній схемі (рис. 2.1). Ми розглянемо еволюцію підходів від класичних оптимізаційних задач до новітніх генеративних моделей. Основна мета цього розділу — виявити сильні та слабкі сторони існуючих рішень за такими критеріями: точність відтворення кольорів, обчислювальна складність, можливість керування результатом та часова узгодженість для відеопослідовностей. На основі проведеного порівняльного аналізу буде обґрунтовано вибір базової архітектури для практичної реалізації системи.

2.1 Алгоритмічні методи колоризації зображень

До епохи глибокого навчання (Deep Learning) колоризація розглядалася переважно як задача оптимізації. Класичні алгоритмічні методи можна розділити на дві великі групи: методи на основі втручання користувача (Scribble-based) та методи перенесення кольору з прикладу (Example-based).

2.1.1 Оптимізаційні методи (Scribble-based)

Цей підхід базується на припущенні, що сусідні пікселі з подібною інтенсивністю (яскравістю) повинні мати подібний колір. Користувач наносить на чорно-біле зображення кольорові штрихи (scribbles), які слугують граничними умовами. Задача зводиться до мінімізації цільової функції енергії E :

$$E(U) = \sum_r \left(U(r) - \sum_{s \in N(r)} w_{rs} U(s) \right)^2,$$

де $U(r)$ — колір (хромінанс) у пікселі r , $N(r)$ — окіл пікселя, а w_{rs} — ваговий коефіцієнт, що залежить від подібності яскравості пікселів r та s .

Перевагою методу є повний контроль користувача, проте він вимагає значних ручних зусиль і часто дає артефакти на межах об'єктів зі слабким контрастом.

2.1.2 Методи перенесення кольору (Example-based)

У цьому підході колір переноситься з референсного кольорового зображення на цільове чорно-біле шляхом зіставлення статистик яскравості та текстур. Ранні роботи використовували зіставлення гістограм у колірному просторі $l\alpha\beta$. Хоча цей метод є більш автоматизованим, він критично

залежить від вдалого вибору референсного зображення та часто призводить до помилкового забарвлення семантично різних об'єктів, що мають схожу текстуру.

2.2 Методи колоризації на основі нейронних мереж

Поява згорткових нейронних мереж (CNN) дозволила моделювати складні залежності між формою об'єкта та його ймовірним кольором, використовуючи великі набори даних (ImageNet, COCO).

2.2.1 Регресійні та класифікаційні підходи (CNN)

Перші нейромережеві методи намагалися прямо передбачити значення каналів ab у просторі Lab, мінімізуючи середньоквадратичну помилку (MSE) або L2-loss:

$$\mathcal{L}_{L2} = \frac{1}{2} \sum_{h,w} \|Y_{gt} - \hat{Y}\|_2^2,$$

де Y_{gt} — справжнє зображення, а \hat{Y} — згенероване. Проте, як зазначають Zhang et al. [26], використання L2-loss призводить до усереднення всіх можливих кольорів, що дає ненасичені, сіро-бурі результати ("seria effect"). Для вирішення цієї проблеми задачу було переформульовано як мультимодальну класифікацію, де мережа передбачає розподіл ймовірностей для квантованих відтінків кольору, що значно підвищило яскравість результатів.

2.2.2 Генеративно-змагальні мережі (GAN)

Мережі GAN, зокрема архітектура Pix2Pix [9], здійснили прорив у чіткості зображень. Генератор G намагається створити реалістичне зображення, щоб обдурити дискримінатор D , який вчиться відрізняти

справжні кольорові фото від згенерованих. Функція втрат GAN має вигляд:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))].$$

Підходи на кшталт ChromaGAN [19] інтегрують семантичну інформацію (клас об'єкта) у процес генерації, що дозволяє уникнути грубих помилок (наприклад, зелене небо).

2.2.3 Трансформери та механізми уваги

Останні дослідження [12, 8] показують, що згорткові мережі мають обмежене рецептивне поле і погано справляються з моделюванням глобальних зв'язків. Трансформери (Vision Transformers) використовують механізм Self-Attention для врахування контексту всього зображення, що є критично важливим для узгодженого забарвлення великих однорідних областей або складних сцен з багатьма об'єктами. Зокрема, модель Colorization Transformer розбиває зображення на послідовність патчів і генерує кольори авторегресійно або паралельно.

2.2.4 Дифузійні імовірнісні моделі (Diffusion Models)

Найсучаснішим класом моделей є DDPM (Denoising Diffusion Probabilistic Models). На відміну від GAN, які генерують зображення за один прохід, дифузійні моделі (наприклад, Palette [17], DDColor [11]) формують зображення ітеративно, поступово видаляючи шум з випадкового сигналу під керуванням вхідного чорно-білого зображення.

[Image of diffusion model forward and reverse process diagram]

Цей процес описується стохастичними диференціальними рівняннями. Перевагою дифузійних моделей є висока стабільність навчання (відсутність "mode collapse" типового для GAN) та неперевершена деталізація текстур. Недоліком є висока обчислювальна складність та повільний час інференсу.

2.3 Особливості колоризації відео

Колоризація відео є значно складнішою задачею через необхідність забезпечення часової узгодженості (temporal consistency). Незалежна покадрова обробка відео навіть найкращими моделями для зображень призводить до ефекту мерехтіння (flickering), оскільки нейромережа може обирати різні варіанти кольору для одного й того ж об'єкта в сусідніх кадрах через незначні зміни освітлення або шуму.

Для розв'язання цієї проблеми використовуються наступні підходи:

- **Оптичний потік (Optical Flow):** Використання векторів руху для деформації (warping) кольорів з попереднього кадру $t - 1$ на поточний кадр t . Це дозволяє "переносити" вже згенеровані кольори.

- **Глибоке поширення ознак (Deep Feature Propagation):** Замість перенесення пікселів, переносяться карти ознак у латентному просторі мережі [16].

- **Саморегуляризація (Self-regularization):** Використання функції втрат, що штрафує мережу за різницю кольорів у відповідних точках сусідніх кадрів [14].

Сучасні моделі, такі як VanGogh [5] та Stable Video Diffusion [4], інтегрують часові модулі (Temporal Attention) безпосередньо в архітектуру дифузійної моделі, що дозволяє генерувати цілісні відеофрагменти, а не окремі кадри.

2.4 Порівняння різних методів колоризації

Узагальнимо характеристики розглянутих методів у порівняльній таблиці 2.1. Для аналізу обрано такі критерії:

- **Якість (Quality):** реалістичність та насиченість кольорів.
- **Узгодженість (Consistency):** відсутність артефактів та мерехтіння (для відео).

- **Керованість (Control):** можливість користувача впливати на результат (текст, приклади).
- **Швидкість (Speed):** час обробки одного кадру.

Таблиця 2.1 – Порівняльний аналіз методів колоризації

Метод / Архітектура	Представники	Переваги	Недоліки
Класичні оптимізаційні	Levin et al., Welsh et al.	Повний контроль користувача, відсутність галюцинацій	Дуже повільні, вимагають ручної розмітки, погана робота зі складними текстурями
CNN (L2/Classif.)	Zhang et al. [26], Larsson [13]	Висока швидкість, стабільність	”Ефект сепії”(L2), низька варіативність, відсутність текстур
GAN (Adversarial)	Pix2Pix [9], ChromaGAN [19]	Висока чіткість, яскраві кольори	Нестабільність тренування, генерація артефактів, складність масштабування на відео
Transformers	ColTran [12], UniColor [8]	Врахування глобального контексту, мультимодальність	Високі вимоги до пам’яті (Quadratic attention complexity)
Diffusion Models	Palette [17], DDColor [11]	SOTA якість, фотореалізм, відмінна керованість	Повільний інференс (багаторазові ітерації), високі вимоги до GPU

Як видно з таблиці, дифузійні моделі демонструють найкращу якість генерації, проте поступаються у швидкості методам на основі CNN та GAN. Для задачі колоризації відео критичним фактором є компроміс між якістю кожного окремого кадру (де лідирують Diffusion Models) та їх часовою узгодженістю. Гібридні підходи, що поєднують генеративну силу дифузії з механізмами поширення ознак (Feature Propagation), виглядають найбільш перспективними для подальшої розробки.

Висновки до розділу 2

У другому розділі було проведено аналіз теоретичних засад та алгоритмічних реалізацій систем колоризації. Було встановлено, що класичні методи є морально застарілими для задач автоматичної обробки, оскільки

вимагають значного втручання оператора.

Порівняння нейромережових підходів показало, що хоча GAN забезпечують швидкий інференс, вони часто страждають від нестабільності та артефактів. Дифузійні моделі (Diffusion Models) на сьогоднішній день забезпечують найвищу якість зображення та найкращі можливості для інтерактивного керування (через текст або приклади), що є критичним для створення зручної системи.

Для колоризації відео основною проблемою залишається часова неузгодженість. Тому в наступному розділі, присвяченому розробці власної системи, буде обрано підхід на основі дифузійної моделі з інтеграцією механізму Attention для забезпечення часової стабільності між кадрами. Це дозволить поєднати високу якість генерації текстур з плавністю відеопослідовності.

ВИСНОВКИ

У дипломній роботі вирішено актуальну науково-практичну задачу розробки інтерактивної системи колоризації зображень та відео, здатної генерувати фотореалістичні та варіативні результати. У ході виконання роботи були отримані наступні результати:

1. Проведено глибокий аналіз сучасного стану проблеми колоризації. Встановлено, що класичні методи оптимізації та ранні неймережеві підходи (на базі CNN з функцією втрат L2) страждають від проблеми "усереднення" кольорів, що призводить до ненасичених, сірих зображень. Порівняльний аналіз показав, що найбільш перспективним напрямком є використання імовірнісних дифузійних моделей (Diffusion Models), які, на відміну від GAN, забезпечують вищу стабільність навчання та кращу деталізацію текстур, хоча і потребують більших обчислювальних ресурсів.

2. Розроблено алгоритм та програмну реалізацію системи інтерактивної колоризації. За основу взято архітектуру умовної дифузійної моделі з механізмом уваги (Attention). Ключовою особливістю розробленої системи є реалізація механізму багатоваріантної генерації, що дозволяє отримувати декілька правдоподібних варіантів забарвлення для одного вхідного зображення шляхом стохастичного семплювання. Також впроваджено гібридну систему керування процесом, яка приймає як глобальні текстові описи, так і локальні точкові підказки користувача.

3. Виконано експериментальне дослідження ефективності системи. Тестування на наборі даних COCO-Stuff підтвердило перевагу запропонованого методу над існуючими аналогами (зокрема, методами на основі GAN).

– Значення метрики FID (Fréchet Inception Distance) вдалося знизити до **24.2**, що на **25%** краще за показники популярного методу DeOldify (32.5), що свідчить про вищу реалістичність згенерованих зображень.

– Суб’єктивна оцінка якості користувачами (Mean Opinion Score) склала **4.2** бала з 5 можливих, що значно перевищує оцінку повністю автоматичних методів (3.6).

– Доведено ефективність інтерактивного режиму: додавання всього 2-х точкових підказок дозволяє підвищити точність відновлення кольору (метрика PSNR) у складних сценах в середньому на **3.3** дБ.

4. Визначено межі застосування розробленої системи. Встановлено, що через специфіку ітеративного процесу дифузії середній час обробки одного зображення складає 4–5 секунд. Це робить систему ідеальною для задач професійної реставрації архівних матеріалів та художньої обробки фотографій, де пріоритетом є якість та контроль, але обмежує її використання в системах реального часу.

Напрямки подальших досліджень. Основним вектором розвитку є оптимізація швидкодії алгоритму. Перспективним вбачається перехід до латентних дифузійних моделей (Latent Diffusion Models), що дозволить виконувати генерацію у стиснутому просторі ознак та значно пришвидшити процес. Також для покращення колоризації відео доцільно інтегрувати модулі часової уваги (Temporal Attention) або механізми оптичного потоку безпосередньо в архітектуру нейронної мережі для усунення ефекту мерехтіння без необхідності пост-обробки.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] S. Anwar та ін. *Image colorization: A survey and dataset*. arXiv preprint arXiv:2008.10774. 2020. URL: <https://arxiv.org/abs/2008.10774>.
- [2] F. Baldassarre, D. G. Morín та L. Rodés-Guirao. *Deep koalarization: Image colorization using cnns and inception-resnet-v2*. arXiv preprint arXiv:1712.03400. 2017. URL: <https://arxiv.org/abs/1712.03400>.
- [3] A. Bansal та ін. *Cold diffusion: Inverting arbitrary image transforms without noise*. Advances in Neural Information Processing Systems, 36, 41259-41282. 2023. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/80fe51a7d8d0c73ff7439c2a2554ed53-Paper-Conference.pdf.
- [4] A. Blattmann та ін. *Stable video diffusion: Scaling latent video diffusion models to large datasets*. arXiv preprint arXiv:2311.15127. 2023. URL: <https://arxiv.org/abs/2311.15127>.
- [5] Z. Fang та ін. *VanGogh: A Unified Multimodal Diffusion-based Framework for Video Colorization*. arXiv preprint arXiv:2501.09499. 2025. URL: <https://arxiv.org/abs/2501.09499>.
- [6] Ian Goodfellow, Yoshua Bengio та Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [7] M. He та ін. *Deep exemplar-based colorization*. ACM Transactions on Graphics (TOG), 37(4), 1-16. 2018. URL: <https://dl.acm.org/doi/abs/10.1145/3197517.3201365>.
- [8] Z. Huang, N. Zhao та J. Liao. *Unicolor: A unified framework for multi-modal colorization with transformer*. ACM Transactions on Graphics (TOG), 41(6), 1-16. 2022. URL: <https://dl.acm.org/doi/abs/10.1145/3550454.3555471>.

- [9] P. Isola та ін. *Image-to-image translation with conditional adversarial networks*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134). <https://arxiv.org/abs/1611.07004>. 2017. URL: https://openaccess.thecvf.com/content_cvpr_2017/html/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.html.
- [10] Phillip Isola та ін. «Image-To-Image Translation With Conditional Adversarial Networks». B: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [11] X. Kang та ін. *Ddcolor: Towards photo-realistic image colorization via dual decoders*. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 328-338). 2023. URL: https://openaccess.thecvf.com/content/ICCV2023/html/Kang_DDColor_Towards_Photo-Realistic_Image_Colorization_via_Dual_Decoders_ICCV_2023_paper.html.
- [12] M. Kumar, D. Weissenborn та N. Kalchbrenner. *Colorization transformer*. arXiv preprint arXiv:2102.04432. 2021. URL: <https://arxiv.org/abs/2102.04432>.
- [13] G. Larsson, M. Maire та G. Shakhnarovich. *Learning representations for automatic colorization*. In European conference on computer vision (pp. 577-593). Cham: Springer International Publishing. 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-46493-0_35.
- [14] Chenyang Lei та Qifeng Chen. «Fully Automatic Video Colorization With Self-Regularization and Diversity». B: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.
- [15] Z. Liang та ін. *Control Color: Multimodal Diffusion-based Interactive Image Colorization*. Z. Liang et al. International Journal of Computer Vision, 1-27. 2025. URL: <https://link.springer.com/article/10.1007/s11263-025-02549-6>.
- [16] Y. Liu та ін. *Temporally consistent video colorization with deep feature propagation and self-regularization learning*. Computational visual media, 10(2), 375-395. 2024. URL: <https://ieeexplore.ieee.org/abstract/document/10897653>.

- [17] C. Saharia та ін. *Palette: Image-to-image diffusion models*. In ACM SIGGRAPH 2022 conference proceedings (pp. 1-10). <https://arxiv.org/abs/2111.05826>. 2022. URL: <https://dl.acm.org/doi/abs/10.1145/3528233.3530757>.
- [18] Chitwan Saharia та ін. «Palette: Image-to-Image Diffusion Models». B: *ACM SIGGRAPH 2022 Conference Proceedings*. SIGGRAPH '22. Vancouver, BC, Canada: Association for Computing Machinery, 2022. ISBN: 9781450393379. DOI: 10.1145/3528233.3530757. URL: <https://doi.org/10.1145/3528233.3530757>.
- [19] P. Vitoria, L. Raad та C. Ballester. *Chromagan: Adversarial picture colorization with semantic class distribution*. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 2445-2454). 2020. URL: https://openaccess.thecvf.com/content_WACV_2020/html/Vitoria_ChromaGAN_Adversarial_Picture_Colorization_with_Semantic_Class_Distribution_WACV_2020_paper.html.
- [20] S. Weng та ін. *L-cad: Language-based colorization with any-level descriptions using diffusion priors*. Advances in Neural Information Processing Systems, 36, 77174-77186. 2023. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/f3bfbfd65743e60c685a3845bd61ce15f-Paper-Conference.pdf.
- [21] L. Yang та ін. *Diffusion models: A comprehensive survey of methods and applications*. ACM computing surveys, 56(4), 1-39. <https://arxiv.org/abs/1508.06576>. 2023. URL: <https://dl.acm.org/doi/abs/10.1145/3626235>.
- [22] Ling Yang та ін. «Diffusion Models: A Comprehensive Survey of Methods and Applications». B: *ACM Comput. Surv.* 56.4 (листоп. 2023). ISSN: 0360-0300. DOI: 10.1145/3626235. URL: <https://doi.org/10.1145/3626235>.
- [23] J. Yun та ін. *iColoriT: Towards propagating local hints to the right region in interactive colorization by leveraging vision transformer*. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp.

- 1787-1796). 2023. URL: https://openaccess.thecvf.com/content/WACV2023/html/Yun_iColoriT_Towards_Propagating_Local_Hints_to_the_Right_Region_in_WACV_2023_paper.html.
- [24] N. Zabari та ін. *Diffusing colors: Image colorization with text guided diffusion*. In SIGGRAPH Asia 2023 Conference Papers (pp. 1-11). 2023. URL: <https://dl.acm.org/doi/abs/10.1145/3610548.3618180>.
- [25] B. Zhang та ін. *Deep exemplar-based video colorization*. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8052-8061). 2019. URL: https://openaccess.thecvf.com/content_CVPR_2019/html/Zhang_Deep_Exemplar-Based_Video_Colorization_CVPR_2019_paper.html.
- [26] R. Zhang, P. Isola та A. A. Efros. *Colorful image colorization*. In European conference on computer vision (pp. 649-666). Cham: Springer International Publishing. 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-46487-9_40.
- [27] R. Zhang та ін. *Real-time user-guided image colorization with learned deep priors*. arXiv preprint arXiv:1705.02999. 2017. URL: <https://arxiv.org/abs/1705.02999>.
- [28] J. Y. Zhu та ін. *Unpaired image-to-image translation using cycle-consistent adversarial networks*. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232). 2017. URL: https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html.