УДК 004.89

Л. Р. Кулик[1]
О. Б. Мокін[1]

ПРОБЛЕМА КОНСИСТЕНТНОСТІ ТРАНСФОРМАЦІЇ СТИЛЮ ДЛЯ ДИФУЗІЙНИХ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

[1] Вінницький національний технічний університет

Анотація

Показано актуальність проблеми забезпечення консистентності трансформації стилю зображень, згенерованих за допомогою дифузійних моделей глибокого навчання. Наведено приклади існуючих підходів забезпечення консистентної генерування та трансформації стилю, їх переваги та недоліки. Запропоновано комбінований підхід, що дозволяє досягти кращої консистентності трансформації стилю під час генерування зображення, посилити контроль над його стилем та зменшити кількість спотворень (артефактів).

Ключові слова: глибоке навчання, дифузійні моделі, консистентність генерування, трансформація стилю.

Abstract

This paper highlights the relevance of the problem of ensuring consistency in the style transformation of images generated using deep-learning diffusion models. It presents examples of existing approaches to achieving consistent generation and style transformation, along with their advantages and disadvantages. The proposed combined approach allows for better consistency in style transformation during image generation, enhances control over its style, and reduces the number of image distortions (artifacts).

Keywords: deep learning, diffusion generative models, generation consistency, style transformation.

Дифузійні моделі глибокого навчання [1] – це нове покоління генеративних моделей, які показали відмінні результати в генеруванні зображень, у тому числі фотореалістичних. Але попри високу якість генерування, все більшої актуальності набуває проблема забезпечення єдиного консистентного стилю між генераціями. Для вирішення цієї проблеми дослідники та інженери штучного інтелекту активно застосовують різноманітні підходи, які можна узагальнити, як трансформацію стилю.

Трансформація стилю спрямована на зміну візуального стилю зображення на бажаний, при цьому зберігаючи оригінальний зміст цього вхідного зображення. До прикладів прикладного застосування даного підходу можна віднести:

- художню стилізація: наприклад, перетворення фотографій у картини в стилі відомих художників;
- редагування певних ознак зображення: зміна кольорів, освітлення, текстури тощо;
- перенесення стилю: з одного зображення на інше.

Зупинимося детальніше на найбільш відомих та ефективних моделях і методах забезпечення консистентності генерування за допомогою трансформації стилю:

1. DreamBooth [2]. Це підхід базується на донавчанні («файнтюнингу») дифузійних моделей на додаткових даних для вивчення та відтворення нових стилів, об’єктів та концептів. Даний підхід можна використовувати сумісно з такими оптимізаційними механізмами тренування, як LoRA (Low-Rank Adaptation) [3].

2. ControlNet [4]. Дана нейронна мережа, застосована до дифузійної моделі, дозволяє врахувати під час генерування додаткові умови, наприклад, границі об’єктів на зображенні, пози людей, глибину об’єктів та сцен, сегментаційні маски тощо.

3. Noise Inversion [5]. Цей метод пропонує альтернативний простір шуму для дифузійної моделі, в якому можна легко маніпулювати зображеннями за допомогою текстових описів. Він створює зручні для редагування карти шуму, які дозволяють відновлювати структуру та семантику вхідного зображення та виконувати наступні маніпуляції: зміщення, редагування кольорів, зміна стилю, зміна освітлення тощо. Цей підхід також можна використовувати для покращення якості та підвищення різно-

манітності інших методів.

4. Zero-shot Image-to-Image Translation [6]. Метод pix2pix-zero вирішує проблему редагування реальних зображень без спеціальних запитів. В його підході закладено автоматичне визначення напрямків редагування, що відображають бажані зміни у просторі текстового опису. Для збереження загальної структури вмісту після редагування використовується карта перехресної уваги вхідного зображення, яка використовується протягом усього процесу дифузії.

5. Style Aligned Image Generation via Shared Attention [7]. Метод StyleAligned використовує так зване мінімальне спільне використання уваги під час процесу дифузії для декількох одночасних генерацій, завдяки чому зберігається стильова або концептуальна узгодженість між зображеннями в дифузійних моделях. Це дозволяє створювати узгоджені в стилі зображення за допомогою зразкового стилю за допомогою простої операції інверсії дифузії для вхідного зображення.

Кожен з вказаних підходів має свою область застосування і продемонстрував якісні результати для певних (вузьких) задач. Але для вирішення проблеми трансформації стилю за наявності лише одного прикладу стилю та одного зображення, для якого цей стиль має бути застосований (тобто зображення концепту), жоден з них не підходить. Тому для вирішення цієї проблеми розроблено комбінований підхід, який ефективно поєднує такі моделі та методи: Shared Attention, ControlNet, LoRA та Noise Inversion:

1. Noise Inversion: використовується для інверсії дифузії вхідного зображення-джерела стилю.

2. Shared Attention: використовується для отримання «стильових уваг» (attentions) зображення-джерела стилю за допомогою моделі Stable Diffusion XL [8] з Canny ControlNet [4] та стилізованої моделі LoRA.

3. ControlNet та LoRA: застосування до зображення-джерела стилю моделі Stable Diffusion XL з Canny ControlNet та стилізованою моделлю LoRA для зображення, яке потрібно стилізувати.

Схема розробленого підходу зображена на рисунку 1.

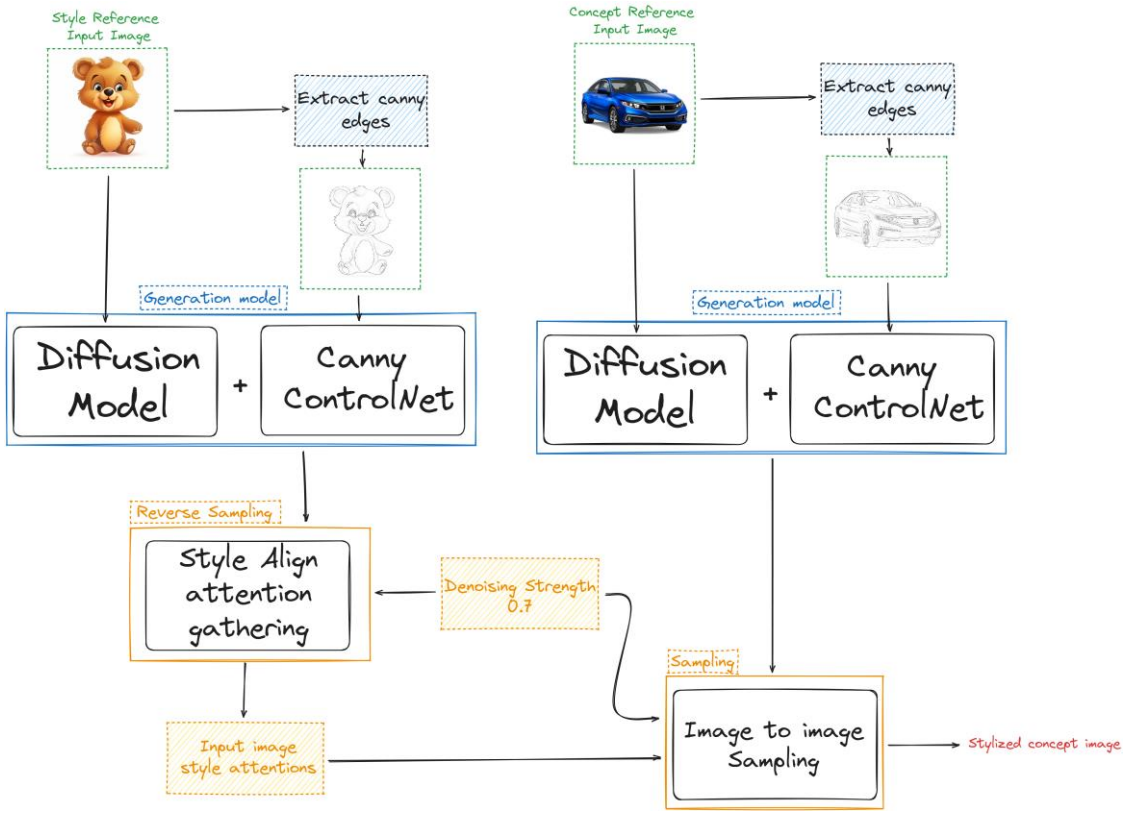

Рисунок 1 – Розроблений підхід для трансформації стилю

До переваги розробленого підходу можна віднести:

- краща консистентність: стиль генерується послідовно, уникаючи різких змін протягом трансформації;

- посилений контроль: користувач може налаштовувати параметри стилю за допомогою Canny ControlNet і LoRA;
- менше артефактів: завдяки Canny ControlNet зберігається структура зображення, що мінімізує так звані артефакти (небажані викривлення зображення).
- гнучкість: підхід працює з різними стилями та зображеннями.

Висновки

Показано, що проблема забезпечення консистентності генерування зображень та трансформації зображень за допомогою дифузійних моделей глибокого навчання є вкрай актуальною для галузі штучного інтелекту та глибокого навчання зокрема. Наведено приклади, переваги та недоліки найбільш ефективних і перспективних моделей та методів забезпечення консистентності стилю під час генерування зображень. Запропоновано комбінований підхід, що дозволяє досягти кращої консистентності трансформації стилю, посилити контроль над стилем згенерованих зображень та зменшити кількість спотворень зображення, що виникають при трансформації.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Jonathan Ho, Ajay Jain, Pieter Abbeel “Denoising Diffusion Probabilistic Models” arXiv:2006.11239 [cs.LG], Jun. 2020.
2. Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, Kfir Aberman “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation” arXiv:2208.12242 [cs.CV], Aug. 2022.
3. Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, Luke Zettlemoyer “QLoRA: Efficient Finetuning of Quantized LLMs” arXiv:2305.14314 [cs.LG], May. 2023.
4. Lvmin Zhang, Maneesh Agrawala “Adding Conditional Control to Text-to-Image Diffusion Models” arXiv:2302.05543 [cs.CV], Feb. 2023.
5. Inbar Huberman-Spiegelglas, Tomer Michaeli “An Edit Friendly DDPM Noise Space: Inversion and Manipulations” arXiv: 2304.06140 [cs.CV], Apr. 2023.
6. Gaurav Parmar, Krishna Kumar Singh “Zero-shot Image-to-Image Translation” arXiv:2302.03027 [cs.CV], Feb. 2023.
7. Amir Hertz, Andrey Voynov “Style Aligned Image Generation via Shared Attention” arXiv:2312.02133 [cs.CV], Jan. 2024.
8. Dustin Podell, Zion English “SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis” arXiv: 2307.01952 [cs.CV], Jul. 2023.

Кулик Леонід Русланович – аспірант факультету інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: leonidkulik2707@gmail.com

Мокін Олександр Борисович – доктор технічних наук, професор, професор кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: abmokin@gmail.com

Kulyk Leonid – a graduate student, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, email : leonidkulik2707@gmail.com

Mokin Oleksandr – Dr. Sc. (Eng.), Professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: abmokin@gmail.com