

ВПЛИВ ГІПЕРПАРАМЕТРІВ ДИФУЗІЙНИХ МОДЕЛЕЙ НА ЯКІСТЬ ГЕНЕРУВАННЯ

Вінницький національний технічний університет

Анотація

У роботі розглянута проблема налаштування гіперпараметрів дифузійних моделей для покращення результатів створення контенту. Показано приклади того як гіперпараметри, такі як ключові слова в описі або `cfg_scale`, впливають на результат. Відзначено методи та показано актуальність проблеми якісного оцінювання штучних даних.

Ключові слова: дифузійні моделі, Stable Diffusion, глибоке навчання, гіперпараметри, оцінювання, штучний контент.

Abstract

The article presents the problem of setting the hyperparameters of diffusion models to improve the results of content creation. It also shows the examples of how hyperparameters such as keywords in the prompt or `cfg_scale` affect the result. The methods are noted and the relevance of the problem of qualitative assessment of artificial data is shown.

Keywords: diffusion models, Stable Diffusion, deep learning, hyperparameters, evaluation, artificial content.

Проблема швидкого створення якісного штучного контенту стає все більш актуальною у сучасному світі, оскільки все більше компаній і організацій стикаються з потребою в автоматизації та оптимізації своїх процесів, включаючи створення візуального контенту.

Завдяки розширенню технологій штучного інтелекту а саме глибокого навчання, створення штучного контенту стає все більш доступним та ефективним. Наприклад, сучасні системи генерування тексту можуть створювати великі текстові матеріали (узагальнюючі описові та пояснювальні статті, огляди продуктів, маркетингові матеріали тощо) з меншими зусиллями та витратами часу, ніж люди. А часто, абстрагуючись від певних фактологічних проблем, і кращої якості з точки зору повноти, граматики та стилю.

Дифузійні моделі генерування зображень [1], такі як Stable Diffusion, останнім часом набирають все більшої популярності. Основна задача таких моделей – це створення деталізованого зображення на основі текстового опису-підказки (prompt). Але їх можна застосовувати і для інших завдань, наприклад, для замальовування, домальовування, перетворювання та редагування зображення.

'A painting of the last supper by Picasso.'



Рисунок 1 – Приклад генерування зображення моделлю Stable Diffusion за описом-підказкою.

Окрім текстового опису, модель використовує ще ряд інших параметрів, які суттєво впливають на результат. Щодо самого опису-підказки, то використання різних ключових слів в описі одного і того ж об'єкту також часто призводить до зовсім різних результатів генерування. На рисунку 2 показано як впливає додавання слів “photorealistic”, “highly detailed”, “4K”, “realistic” на створення зображень kota при фіксованих значеннях решти параметрів SD моделі [2].



Рисунок 2 – Приклад зміни результату генерування Stable Diffusion при зміні опису-підказки.

Іншим важливим параметром у моделі Stable Diffusion є `cfg_scale`, при малих значеннях якого вхідний опис майже не впливає на результат, а при великих – результат стає більш нереалістичним. На рисунку 3 показано результат генерування з вхідним описом “cat” та значенням `cfg_scale` від одного до дев'ятнадцяти з кроком 2. Видно, що зображення зі значеннями від 7 до 11 мають найбільш реалістичний вигляд, а 15-19 – набувають рис твору образотворчого мистецтва.



Рисунок 3 – Результати генерування Stable Diffusion для різних значень параметра `cfg_scale`.

Висновок

При створенні штучного контенту часто важливим фактором є його схожість з реальними даними (наприклад, це можуть бути реальні фотографії котів). І не завжди, навіть маючи лише найсучаснішу (state-of-the-art) модель, можна отримати ідеальні результати. Це потребуватиме додаткових зусиль щодо налаштування відповідних гіперпараметрів генеративної моделі з урахуванням багатьох факторів, у тому числі – предметної галузі. Тому актуальною є задача автоматизації контролю якості генерування. Наприклад, для певних видів задач генерування можна використати автоматичні (Inception Score, Fréchet Inception Distance) та експертні (Human eYe Perceptual Evaluation) оціночні метрики [3].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models,” arXiv: 2112.10752 [cs.CV], Apr. 2022.
2. Ashley Chow, inversion, Will Cukierski, “Stable Diffusion - Image to Prompts,” Kaggle, 2023. [Online]. Available: <https://kaggle.com/competitions/stable-diffusion-image-to-prompts>. Accessed on: Apr. 21, 2023.
3. Sharon Zhou, Mitchell L. Gordon, Ranjay Krishna, Austin Narcomey, Li Fei-Fei, Michael S. Bernstein, “HYPE: A Benchmark for Human eYe Perceptual Evaluation of Generative Models,” arXiv:1904.01121 [cs.CV], Oct. 2019.

Ісаєнков Ярослав Олександрович — аспірант, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: yisaienkov@gmail.com

Мокін Олександр Борисович — д-р техн. наук, професор, професор кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: abmokin@gmail.com

Isaienkov Yaroslav — graduate student, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: oiuygl@gmail.com

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