

Генерация персонализированных изображений

Степанов Илья

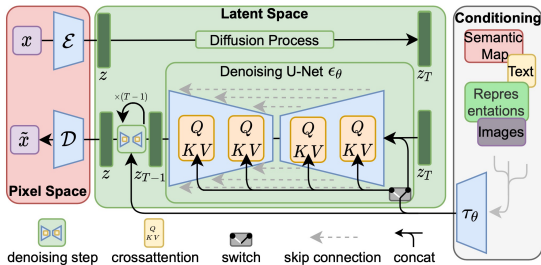
Московский физико-технический институт

Курс: Научный трек Иннпрака ФПМИ

Эксперт: Филатов Андрей Викторович

2024

Введение



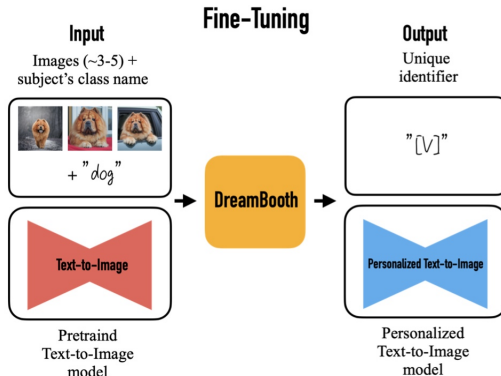
В существующей диффузионной модели решается задача минимизации данной функции потерь:

$$\mathcal{L} = \mathbb{E}_{\epsilon \sim N(0, I), t, z_t, c} \|\epsilon - \epsilon_\theta(t, z_t, c)\|^2$$

где $t \in [0, T]$ — временной шаг диффузионного процесса;
 $z_t = \alpha_t z + \sigma_t \epsilon$ — зашумленные данные на шаге t ; α_t, σ_t — функции от t , определяющие диффузионный процесс; c — условие; ϵ_θ — цель обучения модели диффузии.

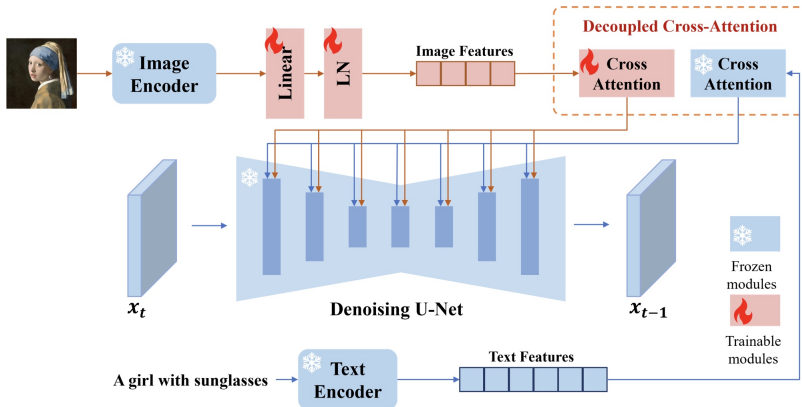
DreamBooth

DreamBooth принимает на вход несколько изображений одного объекта вместе с соответствующим названием класса и возвращает специальный токен, идентифицирующий объект, который затем встраивается в текстовую подсказку, по которой генерируется желаемое изображение.



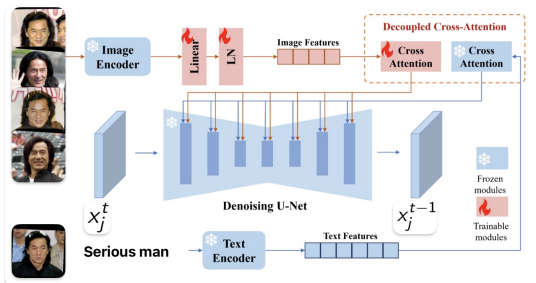
IP-Adapter

IP-Adapter состоит из двух частей: энкодера для извлечения признаков изображения, текста и адаптированных модулей с механизмом перекрестного внимания. Метод принимает на вход только одно изображение объекта.



Постановка задачи

$X = \{(x_i, \tau_i) : i = 1, \dots, n\}$, где x_i — входное изображение, τ_i — соответствующая ему текстовая подсказка. x_i^t — зашумленные данные изображения x_i на шаге t .



$$\mathcal{L} = \mathbb{E}_{\epsilon \sim N(0, I), c_j^T, G(c_i \setminus \{c^j\}), t, x_j^t} \|\epsilon - \epsilon_\theta(c_j^T, G(c_i \setminus \{c^j\}), t, x_j^t)\|^2$$

где G — агрегирующая функция, применяемая ко входным данным; c_j^T — текстовые признаки удаленного изображения; c_i — признаки изображений; c_j — признаки удаленного изображения; ϵ_θ — цель обучения модели диффузии.

- ▶ Разработать агрегирующую функцию
- ▶ Обучить модель
- ▶ Сравнить существующие модели с нашей используя метрики качества генерации изображений