Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет «Высшая школа экономики»

Факультет компьютерных наук
Прикладная математика и информатика

ДИПЛОМНАЯ РАБОТА

Исследовательский проект на тему
”Замена лица (FaceSwap) с помощью генеративных моделей”

Выполнил студент группы мИИАД22, 2 курс,
Рябыкин Алексей Сергеевич

Научный руководитель:
научный сотрудник Аланов Айбек

Москва 2024

РЕФЕРАТ

Дипломная работа содержит 15 страниц, N рисунок, N таблицу, N использованных источников.

Face swap, Deepfakes, ДИФФУЗИОННЫЕ МОДЕЛИ, ГЕНЕРАТИВНЫЕ МОДЕЛИ

Дипломная работа посвящена исследованию современных методов замены лиц на изображения с помощью диффузионных моделей. Исследована проблема

Теоретическая часть работы содержит исследовательский обзор статей о диффузионных моделях для задачи замены лиц на изображении (Face swap)...

В практической части содержатся результаты поставленных экспериментов, их анализ и постановка последующих исследований.

Contents

ВВЕДЕНИЕ

TBD

ОСНОВНАЯ ЧАСТЬ

1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1 Обзор литературы

1.1.1 Общие сведения о диффузионных моделях

Долго лидирующие на поприще генерации изображений генеративно-состязательные сети обладают набором проблем, с которыми боролось и продолжает бороться научное сообщество. Основными проблемами, сохранившимися до сих пор, являются нестабильность обучения [6], решаемая с помощью спектральной нормализации [7], контролирующей константу Липшица, вводя ограничение весов, для стабилизации обучения дискриминатора; Mode Collapse [8]. Последняя является одной из главных проблем моделей такого типа. Проблема заключается в том, что в процессе обучения генератор приходит к состоянию, при котором генерируется лишь ограченный (существенно меньший оригинального пространства изображений) набор выходов. Предлагаемые решения этой проблемы – WGAN [1] (авторы используют метрику Вассерштейна внутри лосс-функции, тем самым мотивируя дискриминатор выявлять повторяющие выходы, в которых стабилизировался генератор) и UGAN [9] (также адаптация функции потерь, однако теперь с оценкой выходов генератора на основе предсказаний будущих версий дискриминатора).

В отличие от генеративно-состязательных сетей эти же проблемы не свойственны новому классу моделей – диффузионным моделям. В оригинальной статье [2] авторы утилизируют идею из статистической термодинамики. Главной целью авторы видят определение двух процессов:

итеративный диффузный процесс, который преобразует любое комплекное распределение данных в более простое и контролируемое, постепенно уменьшая SNR (signal-to-noise ratio), а также параметризованный обратный диффузионный процесс, обучаемый итеративно моделировать целевое распределение. В статье [5] объемно исследуется использование диффузионных моделей для генерации изображений. Оригинально диффузионные модели базируются на методе моделирования динамики молекулярных систем – динамике Ланжевена.

Рассмотрим набор данных из сложно контролируемого целевого распределения: $X = (x_1, \ldots, x_n)$. Построим прямой процесс диффузии, постепенно зашумляющий данные из целевого распределения (уменьшая SNR).

$$x_i^0 \to x_i^1 \to \ldots \to x_i^T$$
$$x_i^{t+1} = \sqrt{1-\beta} \cdot x_i^t + \sqrt{\beta} \cdot \varepsilon, \varepsilon \sim \mathcal{N}(\varepsilon | \ 0, \ I)$$

где β – скорость диффузии, x_i – семпл из выборки. Это также можно записать в виде:

$$q_{t+1}(x^{t+1}|x^t) = \mathcal{N}(x^{t+1}|\sqrt{1-\beta}x^T, \beta)$$

Причем можно также получить явное выражение для $q_T(x^{t+1}|x^0)$:

$$q_T(x^{t+1}|x^0) = \mathcal{N}(x^{t+1}| \ \sqrt{\overline{\alpha}_{t+1}}x^0, \ (1-\overline{\alpha}_{t+1}I)),$$

где $\overline{\alpha}_{t+1} = (1-\beta)^{t+1}$. При бесконечной диффузии $(T >> 1):$

$$\overline{\alpha}_T \to 0, \qquad q_T(x^T \mid x^0) \sim \mathcal{N}(x^T | \ 0, I)$$

Пусть $x_0 \to x_1 \to \ldots \to x_T$ – однородная Марковская цепь, порожденная динамикой Ланжевена. Запишем динамику Ланжевена для произвольного распределения $p(x)$:

$$x^{t+1} = x^t + \varkappa \left(\frac{1}{2} \frac{\partial}{\partial x} \log p(x) + \frac{\varepsilon}{\sqrt{\varkappa}} \right), \qquad \varepsilon \sim \mathcal{N}(\varepsilon | 0,\ I) \tag{1.1}$$

Важнейшим свойством динамики Ланжевена является сохранение плотности, т.е. $x^t \sim p(x) \Rightarrow x^{t+1} \sim p(x)$. (см. ПРИЛОЖЕНИЕ 1). Процедура 1.1 представляет собой градиентный подъем, который направлен на нахождение моды распределения, однако, что является большим преимуществом перед генеративно-состязательными сетями: происходит зашумление градиента, что релаксирует высокие значения плотности распределения.

1.1.2 Обуславливаемые диффузионные модели (Conditioned diffusion models)

В статье [4] предлагается диффузионный процесс не в пространстве изображений, а в пространстве эмбеддингов, полученных с помощью автоэнкодера (VQ-VAE). Так же в предлагаемой U-Net подобной сети авторы предлагают условное обучение (возможность, обуславливать изображения, например, на текст). В статье [2] предлагается постепенное зашумление

В [3] предлагается использование концептов.

1.1.3 Обучение моделей

1.1.4 Задача замены лиц (Face swap)

1.2 Особенности поставленных экспериментов

1.2.1

2 ПРАКТИЧЕСКАЯ ЧАСТЬ

2.1 Используемые инструменты

- Python 3
- PyTorch
- TorchVision
- Matplotlib
- hugging-face
- Plotly
- transformers
- TorchIntegral
- TorchPruning
- diffusers
- wandb

2.2 Экспериментальные результаты

2.3 Выводы

2.4 Дальнейшая работа

„”

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Arjovsky M., Chintala S., Bottou L. Wasserstein GAN. — 2017. — arXiv: 1701.07875 [stat.ML].
2. Deep Unsupervised Learning using Nonequilibrium Thermodynamics / J. Sohl-Dickstein [et al.]. — 2015. — arXiv: 1503.03585 [cs.LG].
3. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation / N. Ruiz [et al.]. — 2023. — arXiv: 2208.12242 [cs.CV].
4. High-Resolution Image Synthesis with Latent Diffusion Models / R. Rombach [et al.]. — 2022. — arXiv: 2112.10752 [cs.CV].
5. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. — 2020. — arXiv: 2006.11239 [cs.LG].
6. On Convergence and Stability of GANs / N. Kodali [et al.]. — 2017. — arXiv: 1705.07215 [cs.AI].
7. Spectral Normalization for Generative Adversarial Networks / T. Miyato [et al.]. — 2018. — arXiv: 1802.05957 [cs.LG].
8. Thanh-Tung H., Tran T. On Catastrophic Forgetting and Mode Collapse in Generative Adversarial Networks. — 2020. — arXiv: 1807.04015 [cs.LG].
9. Unrolled Generative Adversarial Networks / L. Metz [et al.]. — 2017. — arXiv: 1611.02163 [cs.LG].

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ 1

Динамика Ланжевена как стохастическая реализация уравнения Фоккера-Планка

Рассмотрим стохастическое дифференциальное уравнение (СДУ) в общем виде:

$$dx = f(x,t)dt + g(t)dW,$$

где W – Винеровский процесс.

Теорема 0.1. Эволюция распределения $p(x|t)$ по времени t детерменированно описывается дифференциальным уравнением Фоккера-Планка:

$$\frac{\partial p(x|t)}{\partial t} = -\frac{\partial}{\partial x}\left(f\left(x,t\right)p(x|t)\right) + \frac{1}{2}g^2(t)\frac{\partial^2}{\partial x^2}\left(p(x|t)\right)$$

Proof. Докажем отдельно для детерминированной и стохастической частей. Детерменированная часть $(g(t) = 0)$. Пусть $\hat{x} = x + f(x,t)dt$. Тогда:

$$p(\hat{x}|t+dt) = p(\hat{x} - f(x,t)dt|t)\left|\frac{\partial}{\partial x}(x + f(x,t)dt)\right|^{-1}.$$

Разложем в ряд Тейлора в точке $\hat{x}$ функцию $f(x,t)$:

$$f(x,t) = f(\hat{x},t) + \frac{\partial f(\hat{x},t)}{\partial x}f(x,t)dt + o(dt).$$

Для упрощения выкладок рассмотрим одномерный случай.

$$p(\hat{x}|t+dt) = p\left(\hat{x} - f(\hat{x},t)dt - \underbrace{\frac{\partial f(\hat{x},t)}{\partial x} f(x,t)dt^2}_{\text{т.к. } dt\to 0, \text{ то } o(dt)} + o(dt)\right)\left(1 + \frac{\partial f(x,t)}{\partial x}dt\right)^{-1} =$$

$$= p(\hat{x} - f(\hat{x},t)dt + o(dt)|t) \cdot \left(1 + \frac{\partial f(x,t)}{\partial x}\right)^{-1} =$$

$$= p(\hat{x} - f(\hat{x},t)dt + o(dt)|t) \cdot \left(1 - \frac{\partial f(x,t)}{\partial x}dt + o(dt)\right) =$$

$$= \left[p(\hat{x}|t) - \frac{\partial p(\hat{x}|t)}{\partial x}\frac{\partial f(x,t)}{\partial x} f(\hat{x},t)dt^2 - \frac{\partial p(\hat{x}|t)}{\partial x} f(\hat{x},t)dt - p(\hat{x}|t)\frac{\partial f(x,t)}{\partial x}dt + o(t)\right] =$$

$$= p(\hat{x}|t) - \frac{\partial p(\hat{x}|t)}{\partial x} f(\hat{x},t)dt - \frac{\partial f(x,t)}{\partial x} p(\hat{x}|t)dt + o(t) =$$

$$p(\hat{x}|t) - dt\left(\frac{\partial p(\hat{x},t)}{\partial x} f(\hat{x},t) + \frac{\partial f(x,t)}{\partial x} p(\hat{x}|t)\right) + o(dt) =$$

$$= p(\hat{x}|t) - dt\left(\frac{\partial p(\hat{x}|t)}{\partial x} f(\hat{x},t) + \frac{\partial f(\hat{x},t)}{\partial x} p(\hat{x},t)\right) + o(dt)$$

Перенесем налево $p(\hat{x}|t)$, разделим обе части на dt и рассмотрим $\lim\limits_{dt\to 0}$:

$$\underline{\frac{\partial p(\hat{x}|t)}{\partial t} = -\frac{\partial}{\partial x}\left(f(\hat{x},t)p(\hat{x}\ t)\right).}$$

Рассмотрим теперь стохастическую часть. Пусть $\hat{x} = x + \varepsilon$, где $\varepsilon \sim \mathcal{N}(\varepsilon|0, g^2(t)dt)$:

$$p(\hat{x}|t+dt) = \int p(\hat{x} - \varepsilon|t)\mathcal{N}(\varepsilon|0, g^2(t)dt)d\varepsilon =$$

$$\int \left(p(\hat{x},t) - \varepsilon\frac{\partial p(\hat{x}|t)}{\partial x} + \frac{1}{2}\varepsilon^2\frac{\partial^2 p(\hat{x}|t)}{\partial x^2} + o(\varepsilon^2)\right) \cdot \mathcal{N}(\varepsilon|0, g^2(t)dt)d\varepsilon =$$

$$p(\hat{x}|t)\underbrace{\int \mathcal{N}(\varepsilon|0, g^2(t)dt)d\varepsilon}_{1} - \frac{\partial p(\hat{x}|t)}{\partial x}\underbrace{\varepsilon\mathcal{N}(\varepsilon|0, g^2(t)dt)d\varepsilon}_{\mu=0} +$$

$$+\frac{1}{2}\frac{\partial^2 p(\hat{x}|t)}{\partial x^2}\underbrace{\int \varepsilon^2\mathcal{N}(\varepsilon|0, g^2(t)dt)d\varepsilon}_{\sigma=g^2(t)dt} + \int\underbrace{\varepsilon^2 \cdot o(1) \cdot \mathcal{N}(\varepsilon|0, g^2(t)dt)d\varepsilon}_{o(1)\cdot g^2(t)dt=o(dt)} =$$

Упростив последнее выражение, получим:

$$= p(\hat{x}|t) + \frac{1}{2}\frac{\partial^2 p(\hat{x}|t)}{\partial x^2} g^2(t)dt + o(dt).$$

Перенесем первое слагаемое налево и поделим обе части на dt:

$$\frac{p(\hat{x}|t+dt) - p(\hat{x}|t)}{dt} = \frac{1}{2}\frac{\partial^2 p(\hat{x}|t)}{\partial x^2} g^2(t) + o(1)$$

При $\lim\limits_{dt \to 0}$:

$$\frac{\partial p(\hat{x}|t)}{\partial t} = \frac{1}{2}\frac{\partial^2 p(\hat{x}|t)}{\partial x^2} g^2(t)$$

По принципу суперпозиции, получаем доказываемое выражение:

$$\frac{\partial p(x|t)}{\partial t} = -\frac{\partial}{\partial x}\left(f\left(x,t\right) p(x|t)\right) + \frac{1}{2} g^2(t) \frac{\partial^2}{\partial x^2}\left(p(x|t)\right)$$

$\square$

Рассмотрим теперь стохастическое дифференциальное уравнение с $g(t) = 1,\ f(x,t) = \frac{1}{2}\frac{\partial}{\partial x}\log p(x|t)$. Запишем эволюцию распределения для этого СДУ:

$$\frac{\partial p(x|t)}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{1}{2}\frac{\partial}{\partial x}\log p(x|t) p(x|t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2} p(x|t) =$$
$$= -\frac{\partial}{\partial x}\left(\frac{1}{2}\frac{1}{p(x|t)}\frac{\partial}{\partial x} p(x|t) p(x|t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2} p(x|t) = 0.$$

Таким образом, мы показали, что динамика Ланжевена как стохастическая реализация уравнения Фоккера-Планка сохраняет распределение.

ПРИЛОЖЕНИЕ 2