

EigenGAN: Layer-Wise Eigen-Learning for GANs

Zhenliang He, Meina Kan, Shiguang Shan

Николаев Михаил, Деб Натх Максим

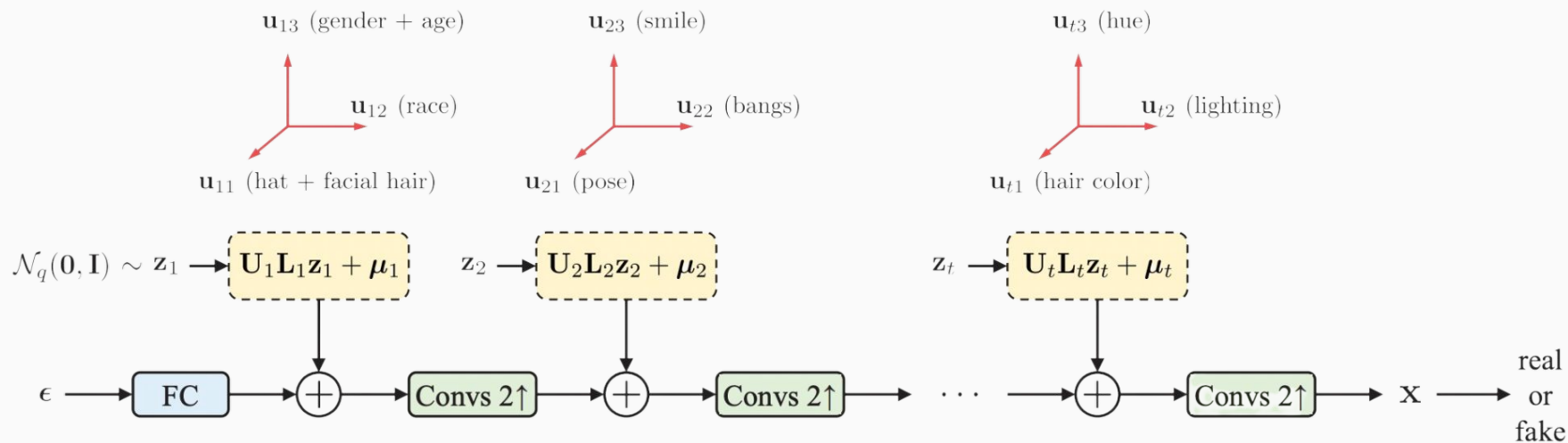
Проблема: интерпретируемость слоёв

- Дискриминативные и генеративные модели в разных слоях отвечают за различные свойства изображения
- Более глубокие слои управляют атрибутами более высокого уровня, такими как поза и цвет кожи, а более мелкие слои управляют функциями более низкого уровня, такими как наклон головы и положение глаз.
- Можно использовать алгоритмы постобработки для идентификации семантических атрибутов из разных слоев.

Предложенное решение

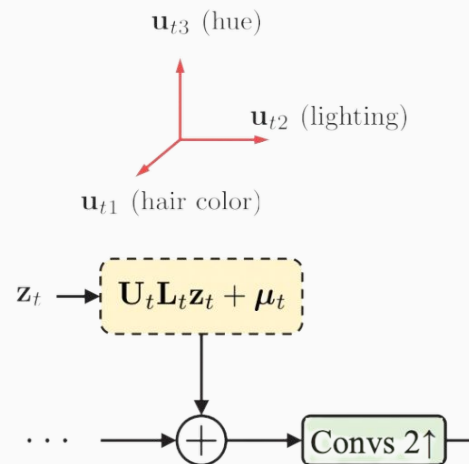
- Предлагаемая авторами статьи модель архитектура EigenGAN
- В каждый upsampling слой генератора добавляется точка из линейного подпространства меньшей размерности с ортогональным базисом
- Комбинацией векторов-базисов после обучения будет задавать признаки изображения

Архитектура



Подробнее про слой

1. Набор ортонормальных векторов $U_i = [u_{i1}, \dots, u_{iq}]$ отвечающий за семантические признаки;
2. Вектор весов L_i , который определяет какой вклад будет давать тот или иной вектор u_i ;
3. Смещение μ_i .



Функции потерь

- В генераторе и дискриминаторе был использован hinge loss
- Была использована R1-регуляризация генератора
- Для того, чтобы заставить матрицы U иметь ортогональные столбцы, использовалась следующая функция потерь:

$$\frac{\|U^T U - I\|^2}{q^2}$$

Обучение модели

- Картинки 64x64
- Латентное пространство 128 + 4 слоя по 6 линейно независимых переменных
- batch size 64
- Училось на CelebA, на A100 2.5 часа

Результаты генеративной модели ($\sigma=0.5$)



Исследование латентного пространства

L1D1 (24%): Яркость фона (оттенки серого)

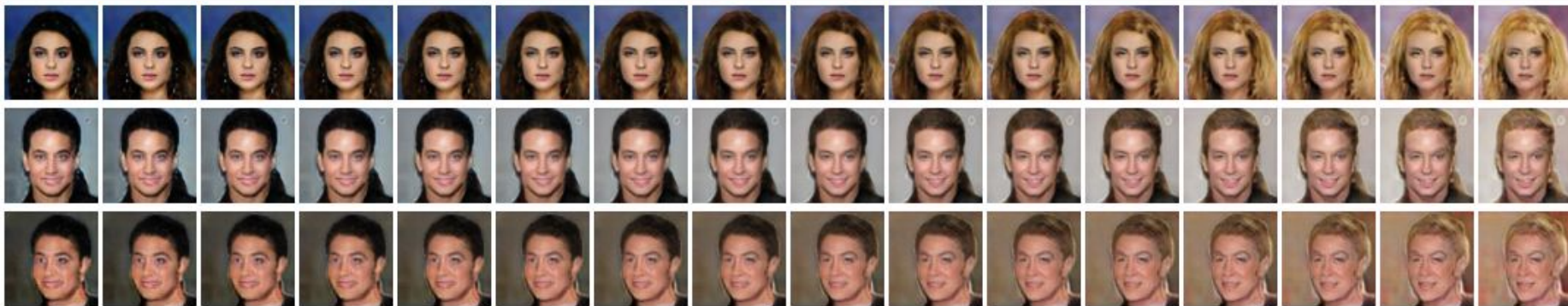


L1D2 (21%): Цвет фона (холодные-тёплые цвета)



Исследование латентного пространства

L1D3 (17%): Цвет волос и фона



L2D1 (23%): Поворот головы + гендер



Исследование латентного пространства

L2D2 (24%): Гендер + длина волос



L2D3 (16%): Улыбка (губы)



Исследование латентного пространства

L2D4 (13%): Наклон головы вверх-вниз



L2D5 (11%): Улыбка (зубы)



L3D2 (23%): Длина волос



L3D3 (18%): Пышность причёски



Исследование латентного пространства

L3D4 (15%): Видимый лоб, закрытый чёлкой



L3D5 (9%): Пробор (его сторона)



L4D3 (19%): Длина чёлки



Результаты

- Получилось воспроизвести результаты авторов в другом фреймворке
- Полученные латентные переменные действительно в некоторой степени интерпретируемы и ранжируемы по своей важности и вкладу в конечное изображение
- Некоторые из латентных переменных объединяют в себе несколько свойств (гендер + улыбка, цвет волос + фона)