

헤어스타일 시장 확장과 다양성을 위한 헤어스타일 변환 인공지능

연구 개요

1. 얼굴 데이터셋으로 **CelebA**와 **FFHQ** 데이터셋을 다운로드
2. 데이터셋 **전처리** (앞머리 여부 라벨링 등)
3. (2)에서 전처리한 데이터셋으로 변환 모델인, **StarGAN-v2 모델** 학습
4. 성능을 향상시키기 위해 자체 개발한 **결과 보정 알고리즘 적용**

연구 동기

<새로운 헤어스타일은 용기가 필요하다>

새로운 헤어스타일을 시도하는 것은 예상대로 되지 않았을 때 느껴야 하는 본인의 헤어스타일에 대한 부정적인 감정과 주변의 시선 등 여러 요인으로 인해 용기가 필요한 도전이다.

<인공지능으로 해결해보자>

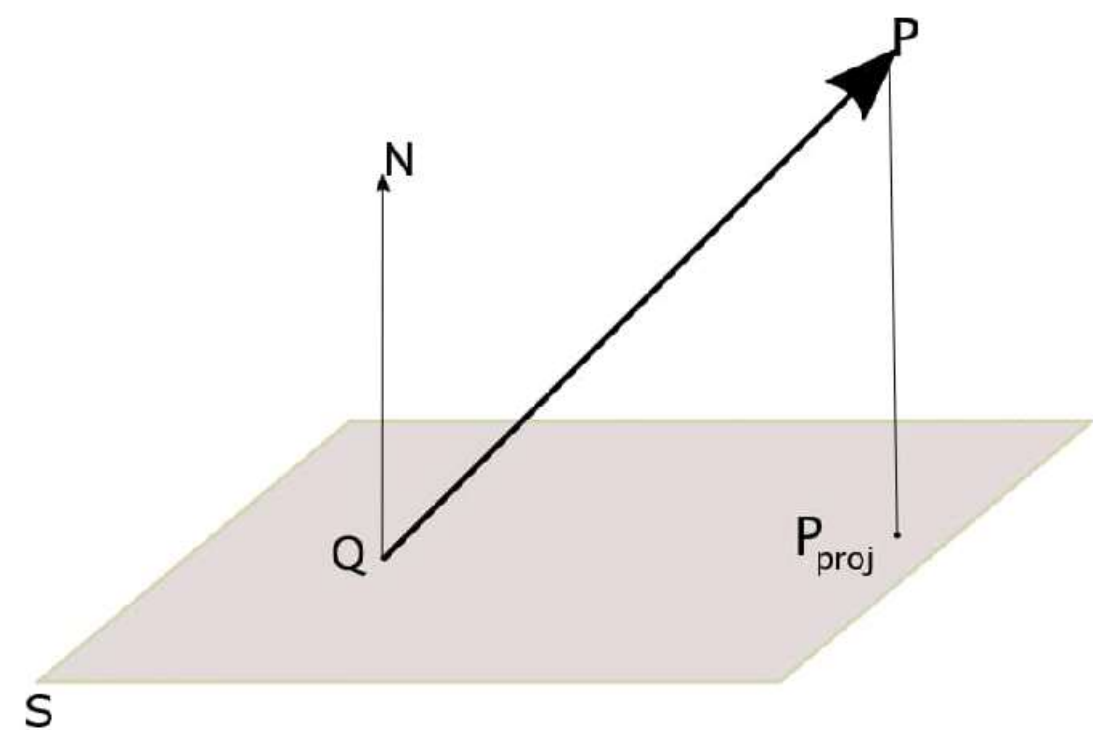
인공지능 기술이 발전하고 있는 시대에 더 이상 새로운 헤어스타일을 머릿속으로만 고민하며 시간을 낭비하고 헤어스타일 다양성이 제한되지 않도록 하는 것이 본 탐구의 목적이다.

이론적 배경

1. PCA (Principal Component Analysis)

Why? 탐구 과정 중 잠재적인 고차원 공간을 시각화하여 확인하기 위하여 차원축소를 해야 했다.
핵심 아이디어: 최대 분산을 만족시키는 벡터에 정사영하며 차원축소를 시킨다.

$$\text{projection} : \left(\frac{\vec{a}^T \vec{b}}{\vec{b}^T \vec{b}} \right) \vec{b}$$
$$\Rightarrow (X\vec{e})\vec{e} \quad (\vec{e}^T \vec{e} = 1)$$
$$(X : n \times m, \vec{e} : m \times 1)$$



[X: 데이터 행렬, \vec{e} : 정사영할 벡터]

정사영 결과의 분산을 최대화 하는 식은 (1), (2)과 같다. 라그랑지 승수법을 이용하여 보조함수를 작성하고(3), 분산을 최대로 하는 \vec{e} 를 구해보면 X의 공분산 행렬의 고유벡터와 동일하다.

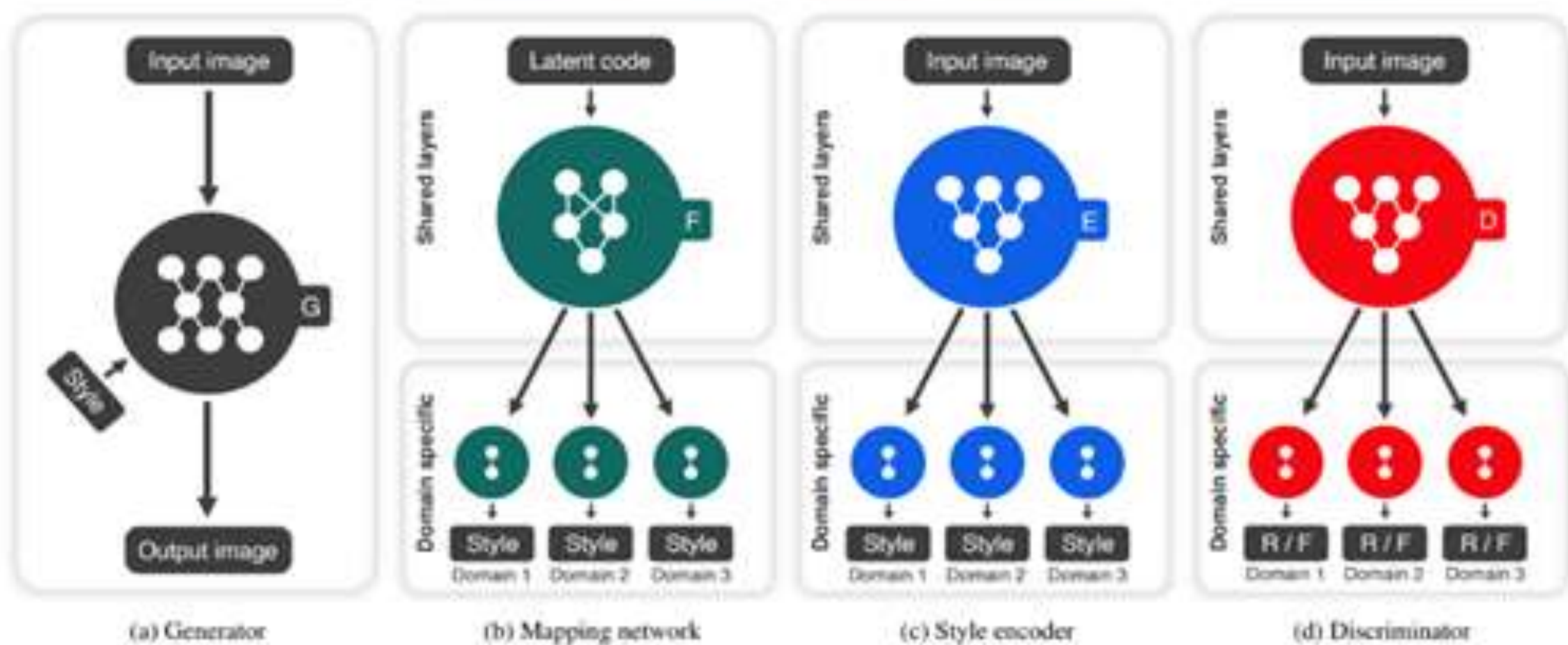
$$\arg \max_{\vec{e}} Var(X\vec{e}) = E[(X\vec{e})^2] - (E[X\vec{e}])^2$$
$$= E[(X\vec{e})^2] \quad (\because E[X\vec{e}] = 0)$$

$$\frac{1}{m}(X\vec{e})^T(X\vec{e}) = \frac{1}{m}\vec{e}^T X^T X \vec{e}$$
$$= \vec{e}^T \frac{X^T X}{m} \vec{e} = \vec{e}^T \sum \vec{e}$$

$$\mathcal{L}(\vec{e}, \lambda) = \vec{e}^T \sum \vec{e} - \lambda(\vec{e}^T \vec{e} - 1)$$
$$\frac{d\mathcal{L}(\vec{e}, \lambda)}{d\vec{e}} = 2 \sum \vec{e} - 2\lambda \vec{e} = 0$$
$$\Rightarrow \sum \vec{e} = \lambda \vec{e}$$

2. StarGAN v2

얼굴 데이터셋인 CelebA-HQ에서 Image-to-Image Translation 분야로 SOTA (State-of-the-Art)를 차지하고 있는 StarGAN-v2를 사용하였다. 다음은 대략적인 모델 구조와 Loss Function이다.



$$\mathcal{L}_{adv} = \mathbb{E}_{\mathbf{x}, y} [\log D_y(\mathbf{x})] + \mathbb{E}_{\mathbf{x}, \tilde{y}, z} [\log (1 - D_{\tilde{y}}(G(\mathbf{x}, \tilde{s})))]$$

$$\mathcal{L}_{ds} = \mathbb{E}_{\mathbf{x}, \tilde{y}, z_1, z_2} [\|G(\mathbf{x}, \tilde{s}_1) - G(\mathbf{x}, \tilde{s}_2)\|_1]$$

$$\mathcal{L}_{sty} = \mathbb{E}_{\mathbf{x}, \tilde{y}, z} [\|\tilde{s} - E_{\tilde{y}}(G(\mathbf{x}, \tilde{s}))\|_1]$$

$$\mathcal{L}_{cyc} = \mathbb{E}_{\mathbf{x}, \tilde{y}, z} [\|\mathbf{x} - G(G(\mathbf{x}, \tilde{s}), \hat{s})\|_1]$$

$$\min_{G, F, E} \max_D \mathcal{L}_{adv} + \lambda_{sty} \mathcal{L}_{sty} - \lambda_{ds} \mathcal{L}_{ds} + \lambda_{cyc} \mathcal{L}_{cyc}$$

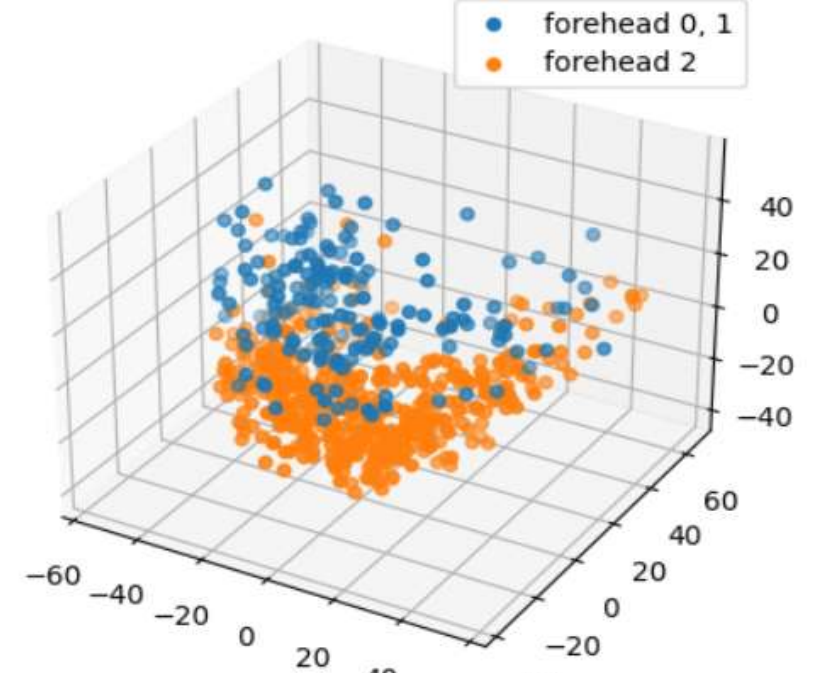
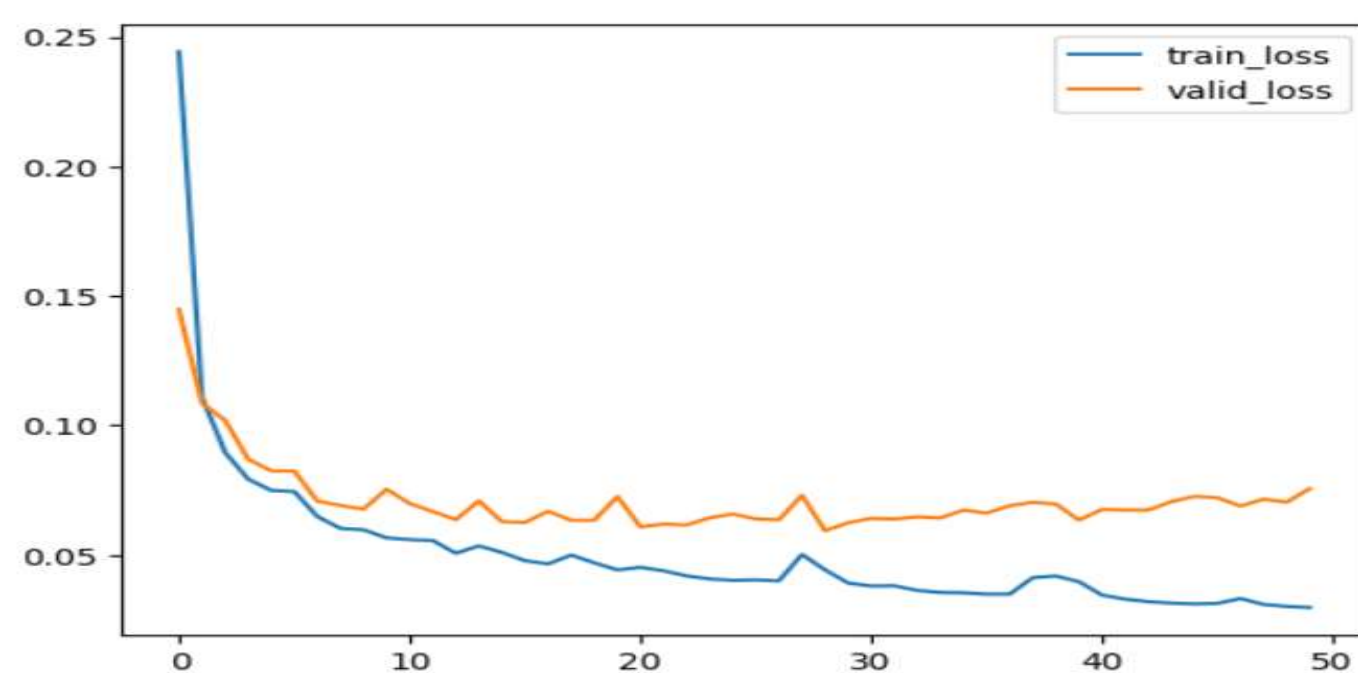
연구 방법

1. 데이터셋 수집 및 전처리

- CelebA와 FFHQ 데이터셋을 다운로드 받고, 앞머리 여부 라벨링 등 전처리를 한다.

2. UNet 모델 학습 및 PCA 시각화

- (1)에서 라벨링한 일부 데이터셋을 이용하여 나머지 데이터셋들도 라벨링하기 위한.
- 머리카락 segmentation 모델로써 학습을 시킨다. 학습 결과는 다음의 좌측과 같다.
- UNet 모델로 얻은 잠재벡터를 PCA 시각화한 결과는 다음의 우측과 같다.



3. StarGAN v2 학습

- (1)과 (2)에서 라벨링한 데이터셋들을 이용하여 StarGAN v2 모델을 학습한다.
- 데이터셋 부족 문제를 해결하기 위해 StyleGAN2-ada 모델 사용함.
- StarGAN v2 모델의 구조를 수정하여 스타일 제공 정보를 제약함.
- 다음은 expr5, expr6, expr7 실험 결과이다. (좌측부터 우측 순)



4. 결과 보정 알고리즘 개발 및 적용

탐구 결과

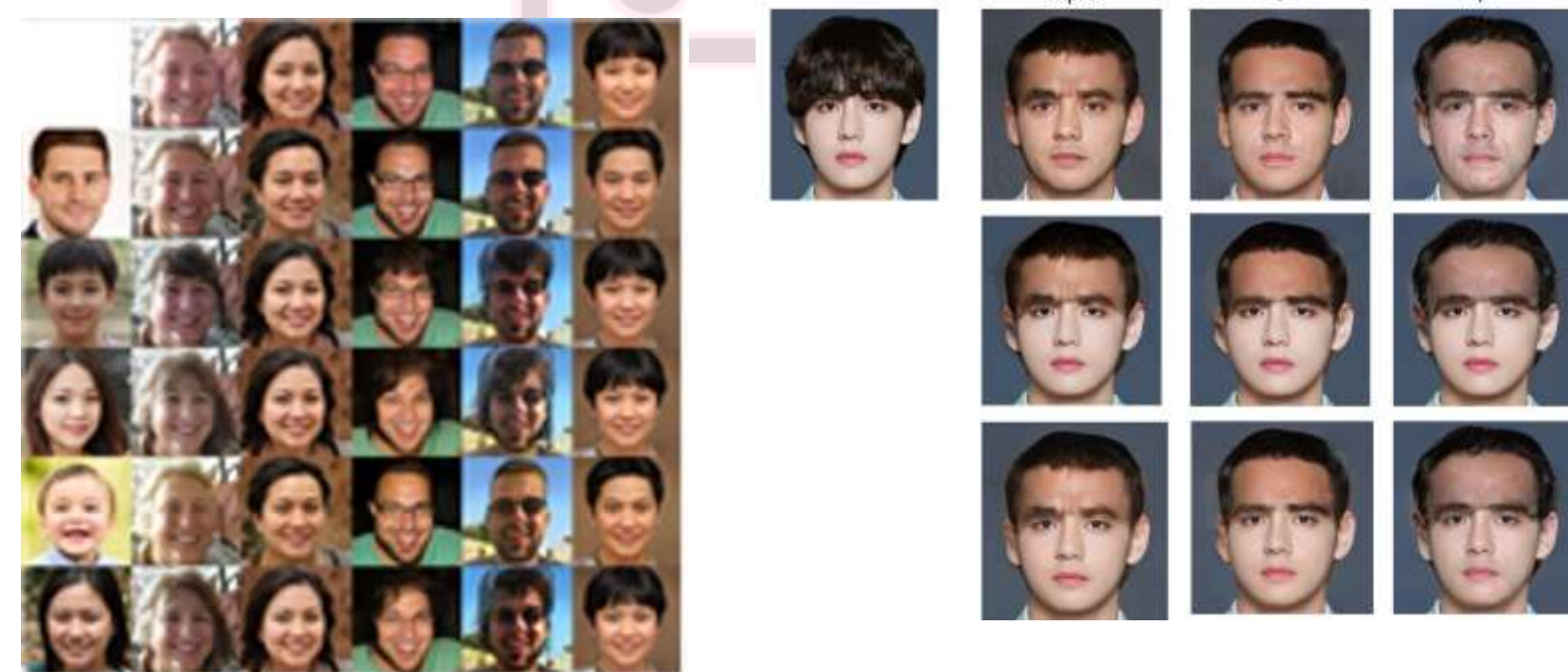
1. 데이터셋 부족 문제

- 데이터셋의 크기가 약 600~1200장으로 굉장히 작아 본 탐구에서 기대했던 것만큼 뛰어난 성능을 보이지 못했다.
- 다음은 KAIST에서 ICIP 2021에 발표한 K-hairstyle의 논문 결과이다. 해당 연구에서는 50만장의 초고해상도 이미지를 사용하여 학습을 시켜 생성 퀄리티가 높다. 하지만 빨간색 화살표로 표시된 부분처럼 바뀌면 안되는 부분이 바뀌는 경우가 많았으며, 이러한 문제는 해결하기 어려운 문제이다.



2. 모델 구조 수정 및 결과 보정 알고리즘

- 위의 사진에서 볼 수 있듯이 본 탐구에서도 변하지 말아야 할 영역이 변하는 문제점이 발견되었다. 본 탐구에서는 모델 구조를 수정하고, 결과 보정 알고리즘을 개발 및 적용하여 문제를 완화시켰다. 결과는 다음과 같다.



3. 결론

- 모델 구조 수정 방법과 결과 보정 알고리즘은 생성 퀄리티 향상의 많은 기여를 할 것이다. 또한 데이터셋 크기 문제가 해결된다면 본 탐구에서 궁극적으로 목표하였던 '다양한 헤어스타일 변환을 통한 헤어스타일 시장 확장과 다양성'이라는 목적을 달성할 수 있을 것으로 기대한다.

참고문헌

- YouTube[1]: <https://www.perfectcorp.com/consumer/apps/ymk> K-hairstyle[2]: <https://arxiv.org/abs/2102.06288>
CelebA[3]: <https://github.com/switchablenorms/CelebAMask-HQ>
FFHQ[4]: <https://github.com/royorel/FFHQ-Aging-Dataset>
gender classifier[5]: <https://github.com/ndb796/Face-Gender-Classification-PyTorch>
U-Net[6]: <https://arxiv.org/abs/1505.04597> PCA[7]: <https://arxiv.org/abs/1404.1100>
StarGAN-v2[8]: <https://arxiv.org/abs/1912.01865> AdaIN Style Transfer[9]: <https://arxiv.org/abs/1703.06868>
StyleGAN2-ada[10]: <https://arxiv.org/pdf/2006.06676> WGAN-GP[11]: <https://arxiv.org/abs/1704.00028>