



Wrap UP Report (1)

CV 10조 조물주 김영민조의 랩업 리포트입니다.

Part 1. [팀] 프로젝트 Wrap Up

1-1. 프로젝트 개요

- 프로젝트 주제
 - 가상 피팅 서비스
- 프로젝트 개요
 - 국내 온라인 패션 시장이 코로나19의 영향으로 빠르게 성장했고, 앞으로도 그런 기조를 보일 예정입니다. 그러나 온라인에서 옷을 직접 입어보지 못해 고객들이 자신과 옷이 어울리는지 확인하기 어렵습니다. 저희는 Virtual Try-On 모델을 활용한 온라인 가상 피팅 서비스를 통해 많은 고객에게 편리한 서비스를 제공하려고 합니다.
- 활용 장비 및 재료 (개발 환경, 협업 tool 등)
 - GPU: V100
 - Visual Studio Code, Jupyter Lab
 - Git, Slack, Notion
- 데이터셋의 구조도(연관도)
 - 데이터셋
 - AI Hub의 패션상품 및 착용 영상 데이터
 - 이 중 목적에 맞는 데이터를 필터링하여 사용
 - 전체 이미지 개수 : 7068 (train) + 489 (test) pairs

```

data
|  -- my_test_pairs.txt
|  -- my_train_pairs.txt
|  -- test
|      |-- agnostic-v3.2
|      |-- cloth
|      |-- cloht-mask
|      |-- image
|      |-- image-densepose
|      |-- image-parse-agnostic-v3.2
|      |-- image-parse-v3
|      |-- openpose_img
|      |-- openpose_json
|
|  -- train
|      |-- agnostic-v3.2
|      |-- cloth
|      |-- cloht-mask
|      |-- image
|      |-- image-densepose
|      |-- image-parse-agnostic-v3.2
|      |-- image-parse-v3
|      |-- openpose_img
|      |-- openpose_json
|

```

- my_test_pairs.txt : test set에 대한 cloth / image pair list
- my_train_pairs.txt : train set에 대한 cloth / image pair list

1-2. 프로젝트 팀 구성 및 역할

전영주

- Paper review & experiment on VITON-HD model
- Generated cloth mask image by image background remove tool
- Implemented backend code for communication betw. main server and preprocessing server by FastAPI

김진섭

- Built and managed the Main Server(Back-end) using FastAPI
- Paper review & experiment on VITON-HD model
- Implement front-end using streamlit

함수민

- Paper review & experiment on VITON-HD model
- HR-VITON Image Generator finetuning
- Preprocessing raw data of AI-Hub

김의진

- Paper review for ACGPN and HR-VITON
- Model training using refocused image
- Data Preprocessing(data cleaning, image refocusing, Detectron Densepose)

김원회

- Paper review of HR-VITON for hyperparameter optimization
- Train condition generator and image generator
- Data Preprocessing(data cleaning, Human parse)

1-3. 프로젝트 수행 결과

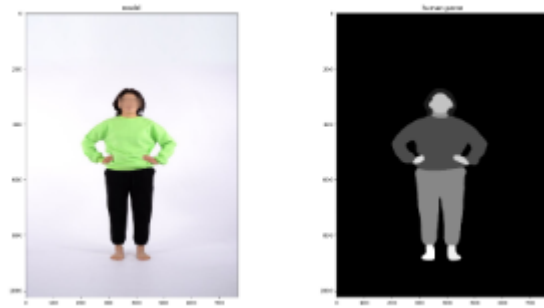
A. 데이터 전처리

1. Pose Estimation



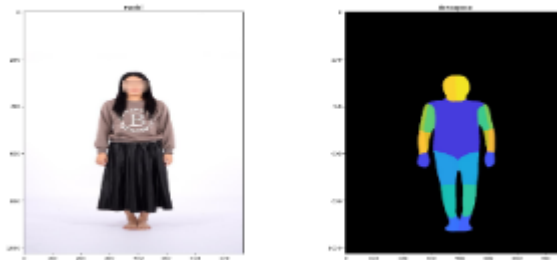
- 사람 포즈에 맞춰 옷을 warping 하기 위해 Openpose를 사용해 skeleton과 json 파일을 얻어내었음

2. Human Parsing



- 이미지 생성 시 가이드를 제공하기 위해 Human Parsing을 통해 parse map을 얻어 내었음

3. Densepose



- Model의 모든 픽셀을 매핑하기 위해서 사용되며 detectron2 라이브러리를 사용

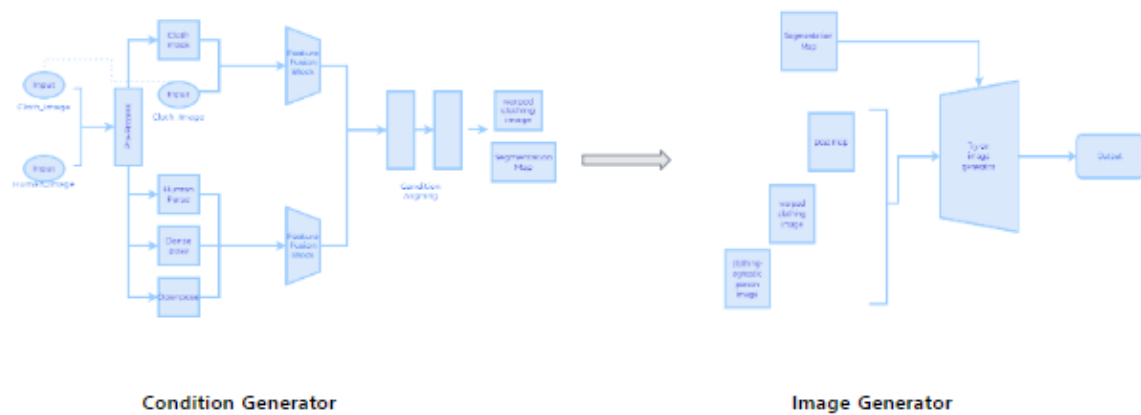
4. Cloth mask



- cloth image에서 필요한 부분을 사용하기 위해 배경을 제거하는 툴을 사용

B. 모델 개요

사용 모델인 HR-VITON은 크게 Condition Generator와 Image Generator로 나뉘어 있음



Condition Generator

- 기존 virtual try on task를 위해 나온 모델들은 cloth warping과 모델에 옷을 입힐 부위에 대한 segmentation map을 개별적으로 구하였음. 하지만 HR-VITON의 condition generator는 두 과정을 condition aligning이라는 과정을 통해 warped cloth와 segmentation map을 같이 구함.

Image Generator

- Image generator는 condition generator의 결과를 입력 받아 target cloth를 입은 결과물을 출력

C. 평가 및 개선 방향

Baseline을 통해 살펴보았을 때 아래와 같이 Logo가 잘 표현되지 못하거나 질감 표현에서 아쉬움이 존재하는 경우가 있었음

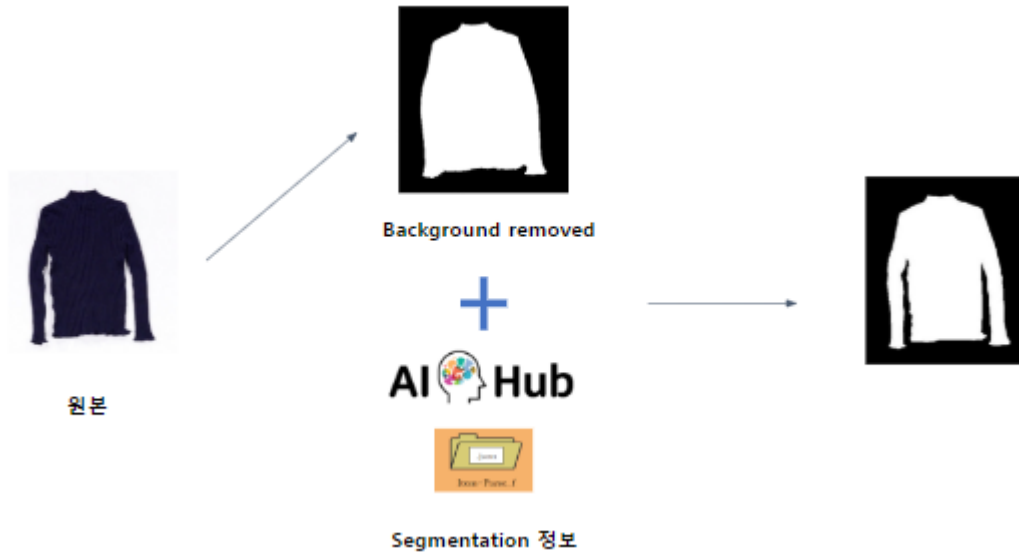


이를 해결하기 위해 아래와 같은 solution을 진행 하였음.

C-1. Data Cleaning

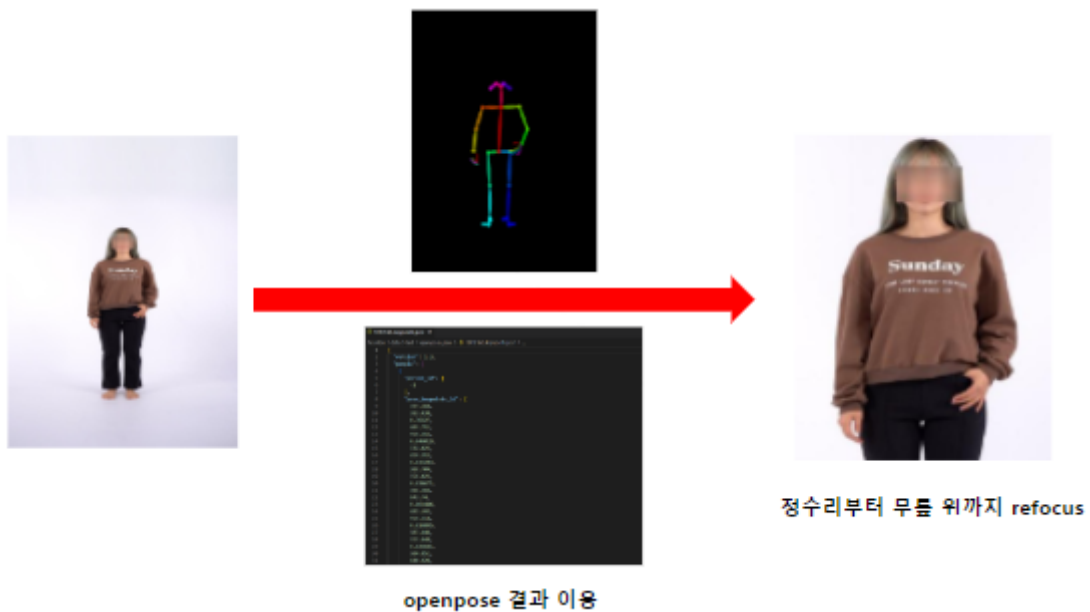


- 전처리를 위해 여러 모듈들을 사용하면서 여러 noise들이 발생하였고 제대로 생성되지 않는 결과들을 제거



- cloth image에 대한 segmentation 정보를 통해 cloth mask를 강화하였음.

C-2. Refocus Model



- 모델이 상체에 집중할 수 있도록 openpose 결과를 이용하여 model image crop / resize

위 두 솔루션을 활용한 결과 더 좋은 성능을 얻게 되었음.

로고 및 질감 표현 우수해짐



기본보다 나아진 색 표현



C-3. Increase Dataset

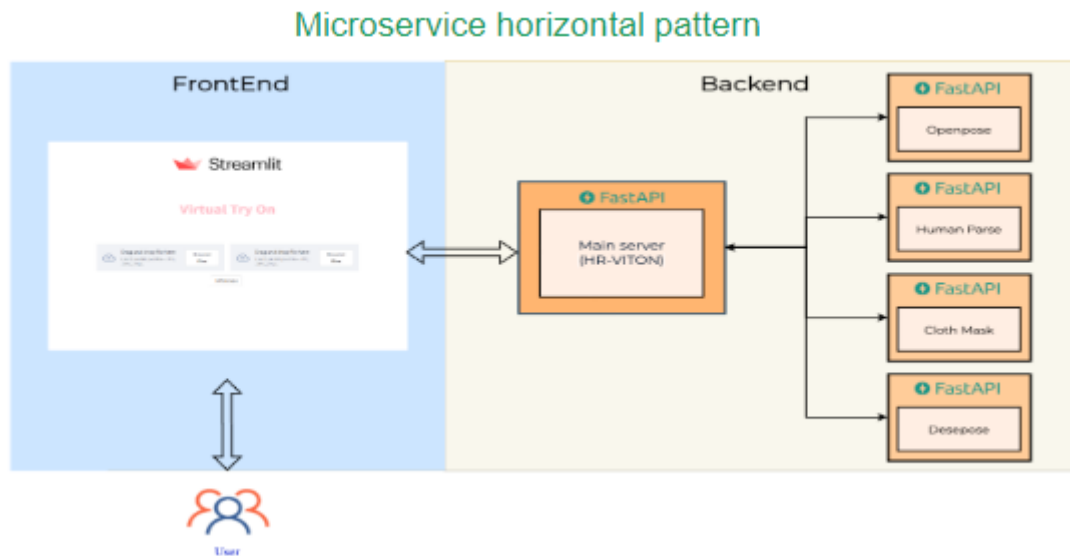


- 마지막으로 더 많은 데이터를 훈련에 사용



- 더 자연스러운 이미지를 생성

D. Product Serving



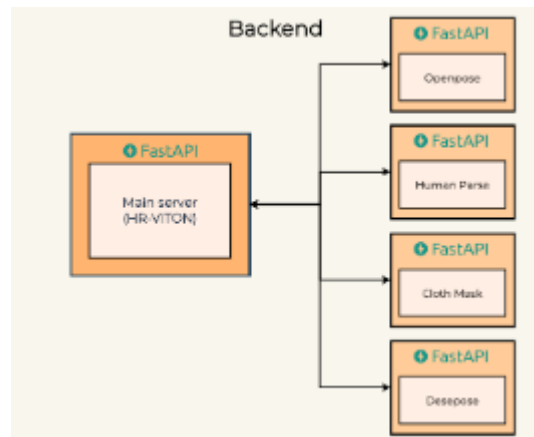
- FrontEnd: Streamlit
- Backend: FastAPI
 - microservice horizontal pattern으로 구성

Front_End



- 옷 이미지와 자신의 이미지를 올린 뒤 inference버튼을 누르게 되면 가상 피팅된 사진을 볼 수 있도록 설계

Back_End



- front에서 받은 이미지를 저장한 뒤 preprocessing을 위해 4개의 서버에 요청
- 각 서버에서 inference된 정보를 모두 response 받게 되면, 메인 모델인 HR-VITON의 inference가 시작, 모든 과정이 끝나면 최종 이미지를 front단에 전송

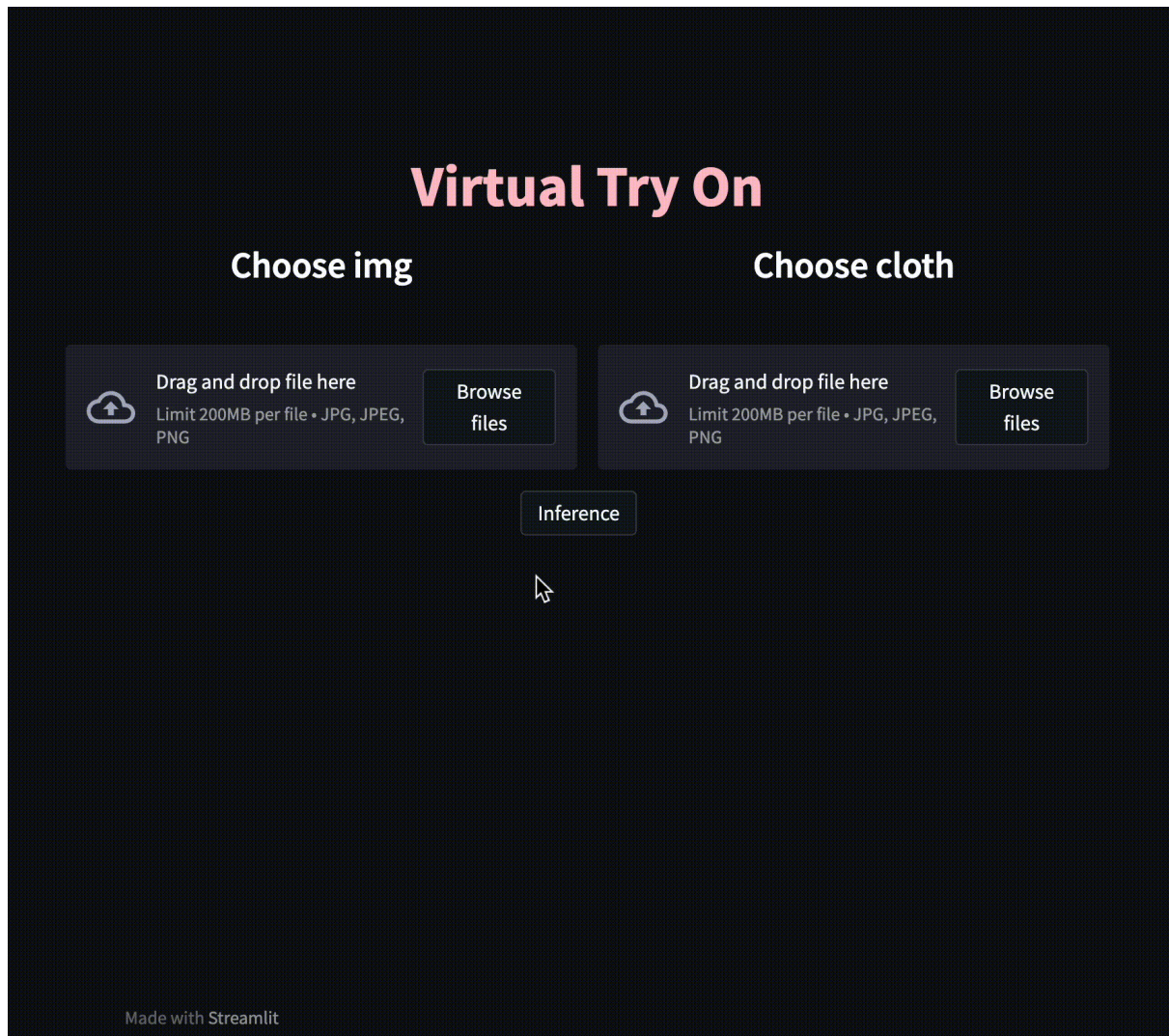
E. Results

E-1. Metrics

	V2		V3	
	Before Refocus	Refocused-image	Training Step 40000	Training Step 30000
LPIPS(paired)	0.1034	0.0963	0.111	0.1097
FID(unpaired)	58.5913	46.69	38.7989	36.1009

Refocus 하기 이전의 결과는 이 후의 결과들과 비교하는 것은 적절하지 않지만 비교를 위해 명시해 두었음

E-2. 시연 영상



1-5. 자체 평가 의견

1-5-1. HR-VITON



패턴, 로고 등은 정확히 표현하지 못하는 것이 아직 많음



옷의 목, 손 등의 부위가 제대로 생성되지 못함

위와 같이 실험이 지속되며 점차 자연스러운 이미지를 생성하였지만 팀 목표였던 FID 20 이하를 달성하지 못함(최종 FID:36.1), 여전히 특정 부분에 취약점이 존재하였음

본 프로젝트에서는 프로젝트 기간을 고려하여 약 7000장의 이미지 페어만을 사용하였는데, 기존 생성 모델이 요구하는 데이터의 수보다 많이 모자람. 더 많은 데이터를 사용했다면 좋은 결과가 있었을 것으로 생각됨.

1-5-2. Product Serving

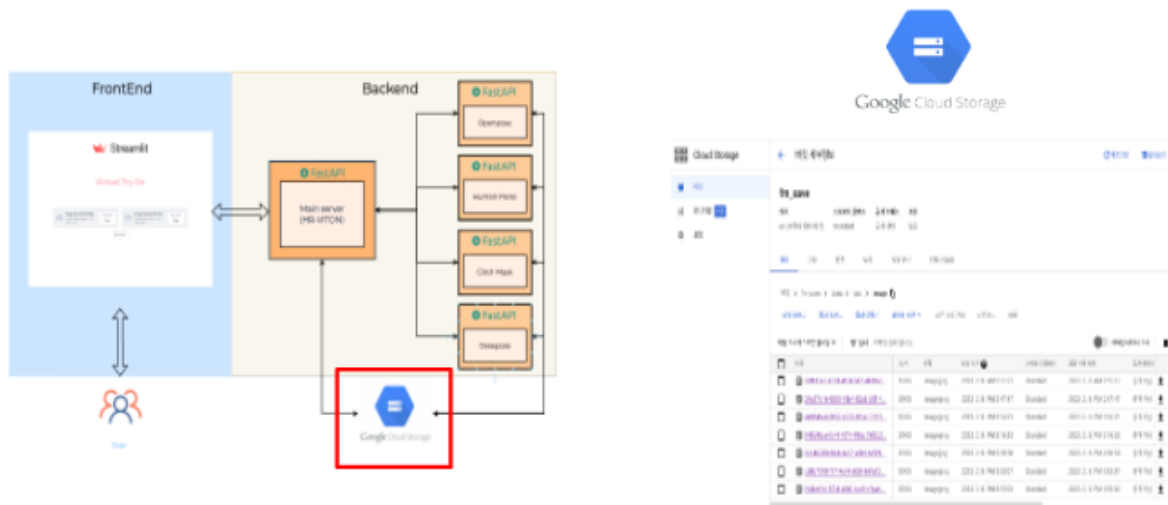
사용자 입장

1. 최종 Inference Time : 약2분
결과 도출까지 오래걸림
2. UI 콘텐츠 및 퀄리티 부족
고객 만족도 부분에서 감점요소

개발자 입장

1. DataBase의 부재
입력 이미지와 생성된 이미지의 관리 부실
2. 완성된 제품에 대한 실험(테스트)이 부족
다수의 사용자가 사용했을 때, 가이드대로 실행하지 않았을 때 등 예외/에러처리 미흡

Product Serving의 아쉬움을 개선하고자 후속 개발을 진행하였음. GCS를 통해 데이터를 관리하도록 하여 더 좋은 서비스를 제공할 수 있게 되었음



Part 2. [개인] 개인 회고

▼ 김원희_T4042

- 나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?
 - 우리 팀과 나의 학습 목표는 무엇이었나?
 - 개인 학습 측면
 - HR-VITON FID 20이하 달성

- 공동 학습 측면
 - 프로젝트의 완성

• 나는 어떤 방식으로 모델을 개선했는가?

- 베이스라인 구축 이후, VITON Dataset으로 pretrain 된 weight로 확인한 결과 배경과 잘 구분되지 않을 정도로 흰 옷이거나, 외투 등 단추가 존재하는 경우, 로고를 가지고 있거나 니트 같은 질감이 있는 경우 등 여러 부분에서 취약점이 존재했습니다.

로고와 질감은 훈련을 통해 개선해 나가야겠다고 판단했지만 나머지 부분은 train set에서 제거하는 것이 맞다고 판단하여 data cleaning을 1차적으로 진행했습니다.

이후 cloth와 image를 refocus하는 아이디어가 도출되어 이를 비교하였습니다. 실험 결과 refocus 해주는 것이 훨씬 좋은 결과를 생성하였습니다.

• 내가 시도한 것은 무엇이고, 어떤 효과가 있었는가?

- 메인 모델인 HR-VITON은 GAN의 구조를 사용하기 때문에 generator와 discriminator를 적절히 tuning하는 것을 주 과제로 삼았습니다. 실험 중반, discriminator의 loss가 saturated 되는 것을 확인하여 이를 개선하고자 spectral normalization을 사용하여 효과를 보았습니다. 하지만 이 후 model refocus를 진행하고 나서는 큰 효과를 보지 못했습니다.
- cloth mask를 개선하기 이전, parse map과 cloth mask를 교차하여 보강하는 occlusion handling을 시도하였습니다. 좀 더 깔끔한 mask가 생성되는 것을 확인하였고, 이에 따른 cloth warping도 개선되는 것을 확인할 수 있었습니다. 이 후 segmentation map을 통해 cloth mask를 강화하여 더욱 큰 성능 향상을 얻을 수 있었습니다.

• 아쉬웠던 점은 무엇인가?

- 앞서 실험의 솔루션 중 하나로, segmentation map을 이용한 cloth mask 강화가 있었습니다. Human parse map도 마찬가지로 segmentation 정보를 이용하여 개선하려 하였지만 aligning에 실패하여 사용하지 못했습니다. Parse map을 강화하였다면 cloth warping이 더 깔끔해졌을 것입니다. 이를 개선하지 못한 것이 가장 큰 아쉬움으로 남습니다.
- 모델의 학습 시간이 대략 70시간 정도 되었기에, wandb의 free trial로는 한계가 있어 tensorboard로 실험 logging을 진행했습니다. 때문에 실험 내용 공유에 있어 아쉬움이 남습니다. mlflow 등을 이용하여 실험 관리를 진행해보지 못한 것이 아쉬움으로 남습니다.

• 한계/교훈을 바탕으로 다음 프로젝트에서 스스로 새롭게 시도해 볼 것은 무엇일까?

- 이번 프로젝트는 전체적인 프로젝트 관리에서 미흡했다고 생각합니다. Notion을 통한 의견 교환 및 피드백을 진행하였는데, Github issue / project 등을 이용하여 프로젝트 일지를 작성하였다면 더 원활한 커뮤니케이션이 가능했을 것으로 생각합니다.

▼ 김의진_T4046

● 이번 프로젝트에서 나의 목표는 무엇이었는가?

- : 1. FID 20이하를 목표로하는 Virtual Try On 모델을 만드는 것
- : 2. 그럴듯한 Output을 내는 모델을 만드는 것

● 나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?

: 저는 모델에 사람이미지와 옷 이미지와 더불어 추가적인 입력으로 필요한 5개 (densepose, openpose, humanparse, cloth mask, agnostic image)의 preprocessing과정 중 densepose mask와 openpose mask를 생성하는 역할을 맡았습니다. 그 후 해당 입력들을 모아 HR-VITON Paper review를 근거로 실험을 진행하였습니다.

● 나는 어떤 방식으로 모델을 개선했는가?

: 저는 주로 입력 데이터를 개선하는 방향으로 모델의 성능을 개선하였습니다. 처음에는 옷의 특징(로고, 질감)이 제대로 반영되지 않는 것을 발견하여 주어진 옷의 이미지를 확대하여 입력으로 넣는 방식으로 개선하여 성능 향상의 효과를 보았습니다. 또 그 이후로도 모델의 착장 이미지가 흐릿하게 보이는 것을 개선하기 위해 모델의 이미지도 확대하여 실험을 진행하였습니다. 다만, 해당 실험들에도 불구하고 저희가 Preprocessing으로 사용하는 모델의 개수가 많았기 때문에, Preprocessing의 결과에 HR-VITON모델의 결과가 많이 의존하는 경향을 발견하였고, 이 문제를 해결하기 위해 Preprocessing후에 data들을 일일이 확인하여 정제하고, aihub에서 데이터셋과 함께 주어진 meta data를 활용하여 Preprocessing결과를 조절하도록 해 주었습니다. 그 결과 이전보다 나아진 성능을 확인할 수 있었습니다.

● 내가 시도한 방법을 통해 어떠한 깨달음을 얻었는가?

: 과거 진행했던 Classification모델이나 detection모델과 같은 경우 일부 데이터는 예외로 두고 모델 training을 하는 것이 오히려 좋았습니다. 하지만 GAN모델의 경우 데이터의 질과 양이 매우 중요하다는 것을 깨달았습니다. 또한 다음번 실험때 이번 aihub데이터셋과 같이 meta data에 대한 정보가 같이 주어진 경우 이를 최대한 활용하는 것을 고려해야 겠다고 생각했습니다.

● 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

: densepose와 openpose는 비교적 최신 모델을 사용하여 괜찮은 성능을 가졌지만 humanparse 모델을 과거 모델로 사용했던 것이 아쉬웠습니다. 그 결과 비교적

치마와 드레스를 구분하지 못했던 것 같고, 이를 해결하기 위해 meta data를 사용하고 싶었지만, 제공된 meta data와 저희가 사용해야 할 humanparsing 부위가 달라 직접적으로 사용해 보지 못하였습니다. 또한 초반에 주제를 바꾸어 2~3주라는 비교적 짧은 시간동안 실험을 했던 부분도 아쉬웠습니다.

- **한계/교훈을 바탕으로 다음 프로젝트에서 스스로 새롭게 시도해볼 것은 무엇일까?**
: 조금 더 다양한 실험을 해보기 위해 다음 프로젝트에서는 최대한 빠르게 baseline을 잡고 실험을 진행할 것이고, metadata와 같이 활용할 수 있는 모든 정보들을 활용해 볼 것입니다.

▼ 김진섭_T4053

- **학습 목표**

- 공동 목표
 - 팀원과 함께 프로젝트를 완성
- 개인 목표
 - 고객에서 만족할 수 있는 서비스를 제공

- **학습 목표를 달성하기 위해 무엇을 어떻게 했는가?**

- Main Server 구축 및 관리 - Front_end, 다른서버와의 연결
- 고객입장에서 만족할 수 있는 UI 디자인 및 속도개선
- 입력 이미지 데이터와 Inference를 통해 생성된 이미지 데이터의 관리

- **내가 시도한 것은 무엇이고, 어떤 효과가 있었는가?**

- preprocessing단계에서 4개의 모델을 inference 해야하기 때문에 'microservice horizontal pattern' 으로 시스템을 구성
 - 한정된 GPU메모리로 인해 하나의 서버에서 4개의 모델을 돌릴 수 없었던 문제를 해결할 수 있었습니다.
- Streamlit을 이용해 간단한 prototype 제작 및 Frontend 단 구성
 - python만으로 빠르게 prototype을 제작 후 서비스화 할 수 있었습니다.
- 이미지 데이터 처리를 위해 Google Cloud Platform (GCP)의 Google Cloud Storage 사용
 - 서버에 계속 쌓이던 이미지를 효율적으로 관리할 수 있었습니다

- 이번 프로젝트에서의 한계 및 아쉬운 점

- 비동기 동작을 통해 빠른 서비스를 제공하려고 했지만, 구현하지 못한 것이 아쉽습니다
- 이번 프로젝트에선 Product Serving과 관련된 다양한 시도들을 했지만, 더 많은 도구들을 사용해 보지 못한게 많이 아쉽습니다
- 사용자입장에서 많은 테스트를 하지못해 실제 서비스를 제공한 뒤 생기는 문제점에 대해 충분히 파악하지 못한점이 아쉽습니다.

- 한계/교훈을 바탕으로 다음 프로젝트에서 스스로 새롭게 시도해볼 것

- 시도하지 못한 Product Serving기술을 도입해보고 다양한 경험을 해보는 것이 중요하다고 생각합니다
- 해당 프로젝트뿐만 아니라 다른 이전에 했던 프로젝트도 서비스화해보면서 실력을 쌓을 것 입니다
- 다양한 테스트코드 및 실험을 통해 꼼꼼하게 오류나 부정적인 요소들을 발견하고 수정할 것 입니다.

▼ 전영주_T4189

- 학습목표

- 프로젝트를 팀원과 함께 잘 완성시키는 것
- 백엔드를 잘 완성시키는 것

- 나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?

모델을 선정하는 과정에서 VITON-HD 모델을 가지고 성능이 어떤지 확인하는 실험을 하였고, 데이터 전처리를 하는 과정에서 image background remove tool 로 cloth mask image 를 만들었다. 또한, microservice horizontal pattern으로 백엔드를 만들기 위해서 main server 와 preprocessing server 사이에서 이미지를 주고 받는 코드를 fastapi로 구현하였다.

- 내가 시도한 것은 무엇이고, 어떤 효과가 있었는가?

메인서버와 전처리서버 사이에서 이미지를 주고 받을 때에 처음에는 예전에 네트워크 수업에서 socket을 활용해서 데이터를 서버사이에 주고받았던 것이 기억나서 socket을 활용해서 이미지를 주고받는 시도를 했으나, 단순한 string은 주고받는 것이 가능하지만, 이미지를 주고받는 것은 불가능하였다. 그래서, fastapi로 데이터를 주고받는 것으로 방법을 바꾸었더니 잘 되었다.

- **아쉬웠던 점은 무엇인가?**

product serving 하는 과정에서 옷 이미지와 모델 이미지를 전처리해주는 과정에서 동기식으로 구현하여 시간이 오래걸려서 전처리해주는 서버를 동기에서 비동기로 바꿔서 동시에 여러 서버로 전처리를 요청하는 것을 시도했지만, 잘 되지 않아서 전처리에 걸리는 시간을 줄이지 못한 것이 아쉽다.

또한, 데이터셋을 선정하는 과정에서 다른 대체제에 대하여 생각하지 못하여 데이터셋을 어떤 것으로 선정해야할지 꼼꼼히 따지지 못해서 모델이 제한적인 성능을 보인 것이 아쉽다.

- **한계/교훈을 바탕으로 다음 프로젝트에서 스스로 새롭게 시도해 볼 것은 무엇일까?**

모델을 선정할 때에는 여러 선택지들의 장단점을 잘 따졌지만, 데이터셋을 선정해야 할 때에는 여러 대안들을 조사해서 그것들의 장단점을 따져보고 데이터셋을 선정하지 않았던 점이 아쉽다. 그래서 다음에는 여러 안들을 조사해서 장단점을 잘 따지고 어떤 것을 선택했을 때에 확실한 이유를 세우려고 한다.

또한, 제한적인 시간 때문에 동기식에서 비동기식으로 바꾸는 것에 있어서 문제상황을 해결하지 못했다. 다음에는 시간을 더 내서라도 프로젝트의 완성도에 더 기여하고자 한다.

▼ 함수민_T4227

- **나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?**

- 시도는 여러가지 했다. Backend 구현도 시도해봤고, data preprocessing의 여러 과정중 openpose도 시도해봤으나 가상환경의 충돌문제 등의 이유로 실패만을 겪었다.

Image Generator 학습은 성공적으로 마쳤다.

- **내가 한 행동의 결과로 어떤 지점을 달성하고, 어떠한 깨달음을 얻었는가?**

- 최종 프로젝트를 하면서 실패의 연속만을 겪은 느낌이 있는데, 실패 속에서 배우는 것이 많다는 것을 느꼈다.
- 이전에는 문제(에러)가 생기면 그것을 바로 해결하려고 무작정 들이 박았다면, 이제는 멀리서 관조하며 차분히 문제의 발생 원인을 먼저 생각하고 들어갈 줄 알게 되었다. 이런 깨달음이 없었다면 openpose를 시도 할 때 마냥 이것저것 하다가 가상환경이 꼬여서 image generator 학습도 팀원들의 도움 없이는 아예 불가능했을 것이다.
- 문제상황 발생 시 바로 소통을 하는 것도 중요하다는 것을 느꼈다. 모두가 바빠 도움을 요청하면 방해될 것 같다는 생각에 혼자 끙끙대는 일이 많았는데, 그게

오히려 진행상황에 차질을 만들어 팀 전체에 폐를 끼친것 같다. 그리고 누군가의 문제는 누군가에겐 이미 겪어왔기에 아무것도 아닐수도 있기에 나 또한 추후에 도움을 준다는 마인드를 가져야겠다는 생각을 하게 되었다.

- **마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?**

- 가장 큰 문제는 역시 가상환경 문제였다. 정말 아쉬웠던 점은 당시에는 그 문제를 해결하는 것 자체가 까다롭다는 생각밖에 안들었는데, 돌이켜보니 차분히 공식문서를 잘 숙지하고 내 환경에 맞는 솔루션을 택했다면 간단히 해결했을 것이라는 점이다.

- **한계/교훈을 바탕으로 다음 프로젝트에서 스스로 새롭게 시도해볼 것은 무엇일까?**

- 문제상황 정의, 머리로만 알고 있었으나 제대로 실천하지 못한 소통 이 두 가지를 제대로 해보고싶다. 다른 대회때는 그나마 덜? 했던거같은데, 각자의 역할이 더더욱 중요해지는 최종 프로젝트를 하다보니 내 부족함이 더더욱 부각되었던 것 같다.