제품에 기여하는 머신러닝 - 연구에서 고객까지

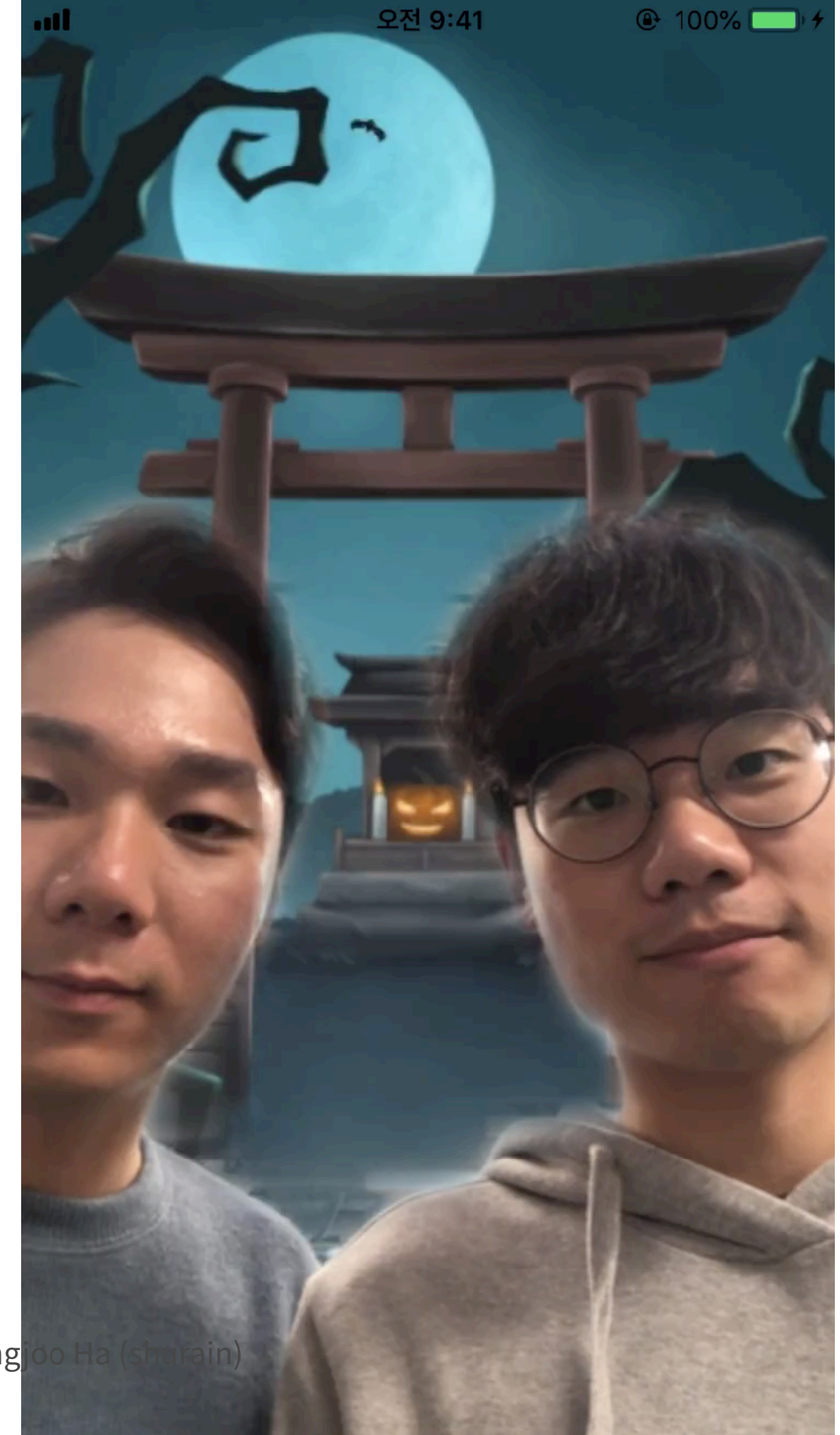
Hyperconnect AI Lab

Sungjoo Ha (shurain)

December 11th, 2019

오늘의 이야기

- 프로덕션서비스 중인/될 제품과 연구성공이 불확실한 기술 개발의 이야기
  - 팀에서 수행한 MarioNETte 연구 이야기를 토대로
- 하이퍼커넥트 AI Lab이 일하는 방식의 이야기
- 회사와 팀과 팀원 사이의 합을 맞추는 이야기



Hyperconnect AI Lab

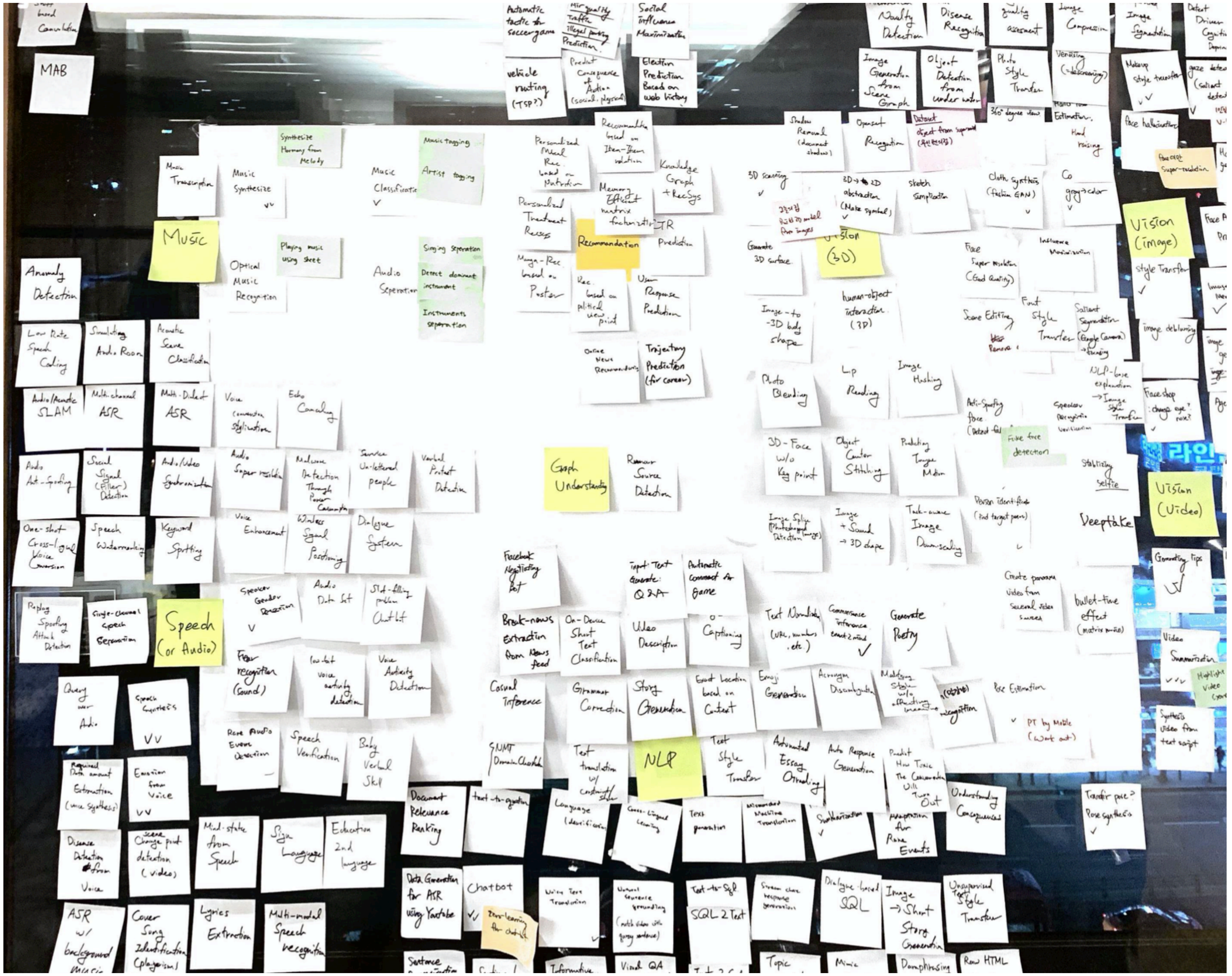
- 기계학습 관련된 업무 전반의 담당
  - 프로젝트 선정
  - 데이터 수집
  - 모델 개발 및 실험
  - 논문화
  - 기획 참여
  - 데이터 QA
  - 배포

2019년 초

- 기존 팀의 포커스는 **모바일** 환경에서 **실시간**으로 **이미지** 다루기[1]

[1] https://github.com/hyperconnect/MMNet/

Workshop



- 10개 학회
  - 3700편 논문
- 유저 니즈/비즈니스 트렌드 기반 아이디어 브레인스토밍
- 300가지 정도의 잠재적인 활용처 고민
- 1년 간 로드맵 구성

Project Selection

- 다양한 요소를 고려하여 프로젝트 선정
  - 실현 가능성
  - 임팩트
  - 기술적 중요도
  - 트렌드
- 그 중 하나인 **이미지 합성**(image synthesis)

다양한 **합성** 기술이 **범람**할 것이다

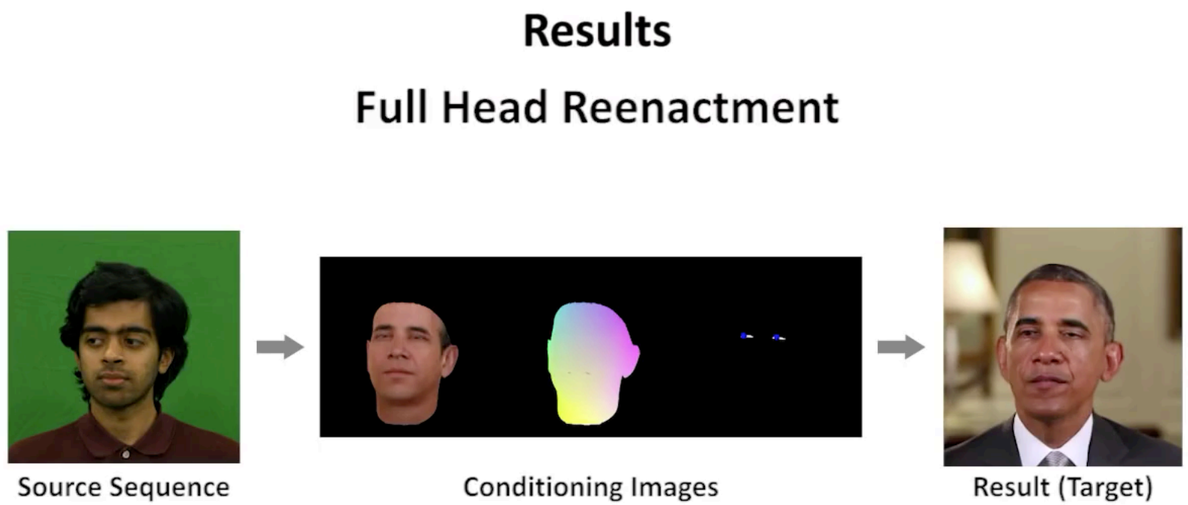
Why?

- 합성 기술이 **폭발적으로 발전** 중
- 관련된 새로운 문제를 빠르게 풀 수 있는 **기술 역량** 획득
- 회사가 집중하고 있는 **사람과 사람의 연결**이라는 관점에서 아주 큰 임팩트를 줄 수 있는 기술이라 판단

Literature Survey

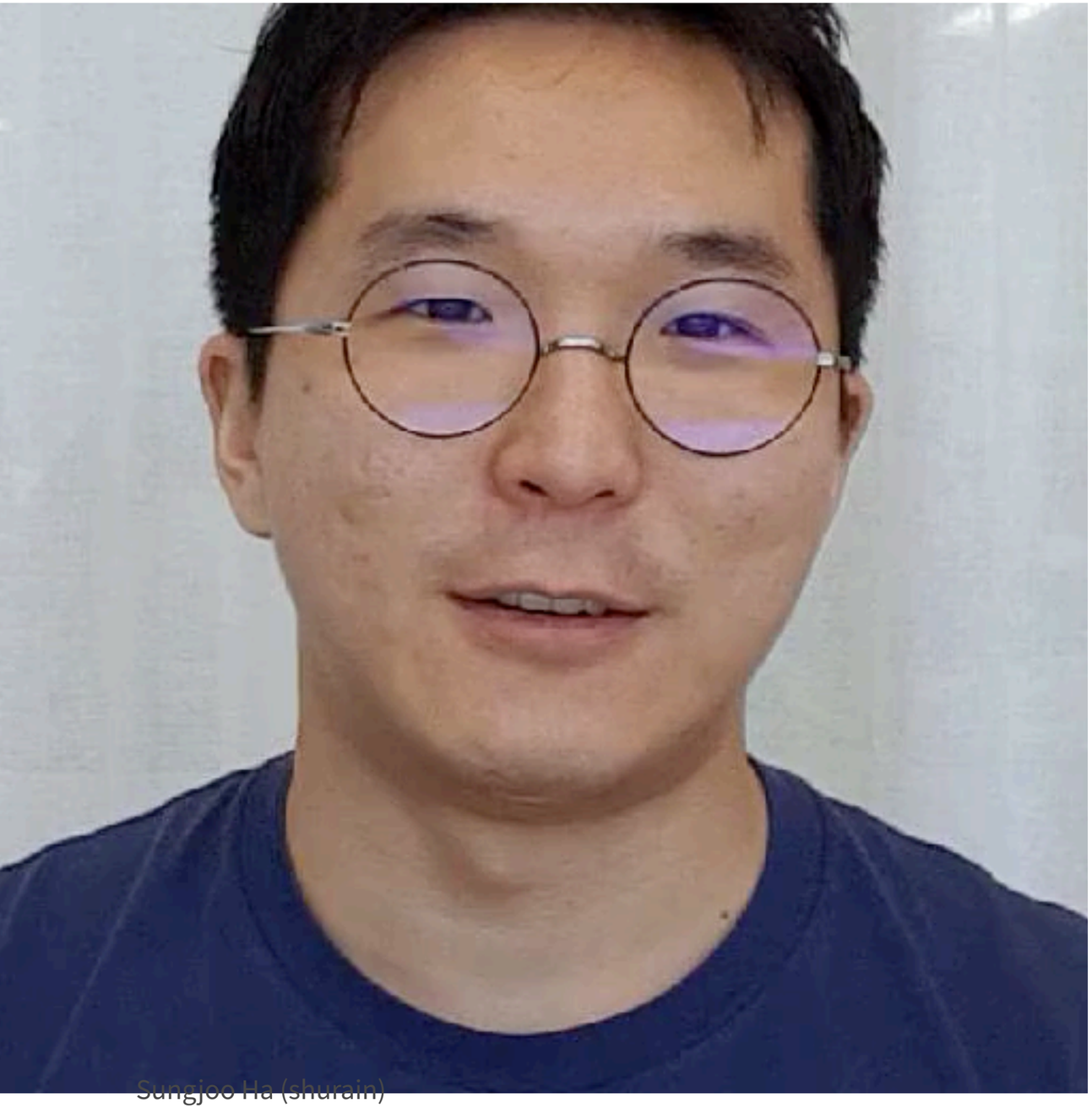
How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks)	https://arxiv.org/abs/1703.07332	facial landmarks	2017		yes	no	https://github.com/1adrianb/face-ali	
Joint Face Detection and Facial Motion Retargeting for Multiple Faces	https://arxiv.org/abs/1902.10744	face detection, facial expression	2019		no	yes		
Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network	https://arxiv.org/abs/1803.07835	3D face reconstruction	2018				https://github.com/YadiraF/PRNet	
CNN-based Real-time Dense Face Reconstruction with Inverse-rendered Photo-realistic Face Images	https://arxiv.org/abs/1708.00980		2018		yes			
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks	https://arxiv.org/abs/1703.10593	GAN	2017	ICCV			https://junyanz.github.io/CycleGAN/	
Recycle-GAN: Unsupervised Video Retargeting	https://arxiv.org/abs/1808.05174	GAN	2018				https://github.con	Use temporal-spa
A Style-Based Generator Architecture for Generative Adversarial Networks	https://arxiv.org/abs/1812.04948	GAN	2019?	CVPR				
Face Transfer with Generative Adversarial Network	https://arxiv.org/abs/1710.06090	GAN	2017					
A Face-to-Face Neural Conversation Model	http://www.cs.toronto.edu/face2face		2018	CVPR				neural conversatic
Face2Face: Real-time Face Capture and Reenactment of RGB Videos	http://niessnerlab.org/projects/thies2016face.html		2016	CVPR				
Deep Video Portraits	https://web.stanford.edu/~zollhoef/papers/SG2018_DeepVideo/page.h		2018	Siggraph				
HeadOn: Real-time Reenactment of Human Portrait Videos	https://niessnerlab.org/papers/2018/7headon/headon_preprint.pdf							
Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set	https://arxiv.org/abs/1903.08527						https://github.com/Microsoft/Deep3D	
paGAN: Real-time Avatars Using Dynamic Textures	http://vgl.ict.usc.edu/Research/pagan/		2018					They claim that "h
Avatar Digitization From a Single Image For Real-Time Rendering	http://www.hao-li.com/publication	hair	2017					
ReenactGAN: Learning to Reenact Faces via Boundary Transfer	https://arxiv.org/abs/1807.11079	GAN						Map 2D image to
X2Face: A network for controlling face generation by using images, audio, and pose codes	https://arxiv.org/abs/1807.10550							Map 2D image to
ICface: Interpretable and Controllable Face Reenactment Using GANs	https://tutvision.github.io/icface/	GAN	2019					
Video-to-Video Synthesis	https://arxiv.org/abs/1808.06601	GAN	2018	NeurIPS			https://github.com/NVIDIA/vid2vid	
Synthesizing Obama: Learning Lip Sync from Audio	https://grail.cs.washington.edu/projects/AudioToObama/		2017	Siggraph				
MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction	https://arxiv.org/abs/1703.10580		2017	ICCV			https://github.com/waxz/MoFA	
Image-to-Image Translation with Conditional Adversarial Nets	https://phillipi.github.io/pix2pix/	GAN	2017	CVPR			https://github.com/phillipi/pix2pix	
High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs	https://tcwang0509.github.io/pix2	GAN	2018	CVPR			https://github.com/NVIDIA/pix2pixHI	
Photographic Image Synthesis with Cascaded Refinement Networks	https://cqf.io/ImageSynthesis/	Image synthesis	2017	ICCV			https://github.con	Image generation
Everybody Dance Now	https://carolineec.github.io/everyt	Dance video synthesis	2018	Siggraph				2D pixel -> 2D stic
A Deep Learning Approach for Generalized Speech Animation	http://www.yisongyue.com/public	Speech animation generation	2017	Siggraph				
StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation	https://github.com/yunjey/StarGA	GAN	2018	CVPR			https://github.com/yunjey/StarGAN	
Progressive Growing of GANs for Improved Quality, Stability, and Variation	https://github.com/tkarras/progressive_growing_of_gans		2018	ICLR			https://github.com/tkarras/progressi	
FaceVR: Real-Time Facial Reenactment and Eye Gaze Control in Virtual Reality	http://gvv.mpi-inf.mpg.de/projects/FaceVR/		2018	Siggraph				
HoloGAN: Unsupervised learning of 3D representations from natural images	https://www.monkeyoverflow.com/#/hologan-unsupervised-learning-of-3d-representations-from-natural-images/							Map 2D to 3D ten
Deforming Autoencoders: UnsupervisedDisentangling of Shape and Appearance	http://www3.cs.stonybrook.edu/~cvl/content/papers/2018/Shu_ECCV18.pdf							
GANimation: Anatomically-aware Facial Animation from a Single Image	https://www.albertpumarola.com/research/GANimation/index.html		2018	ECCV			https://github.com/albertpumarola/G	

Baseline



- **비교** 대상이 되는 모델[2]이 있어야 개선이 의미가 있음
  - 점진적으로 비교할 수 있는 모델을 늘려가며 다양한 컴포넌트를 확보
- 프로덕션에서는 이미 연구된 모델을 구현하는 것이 충분할 수도 있음

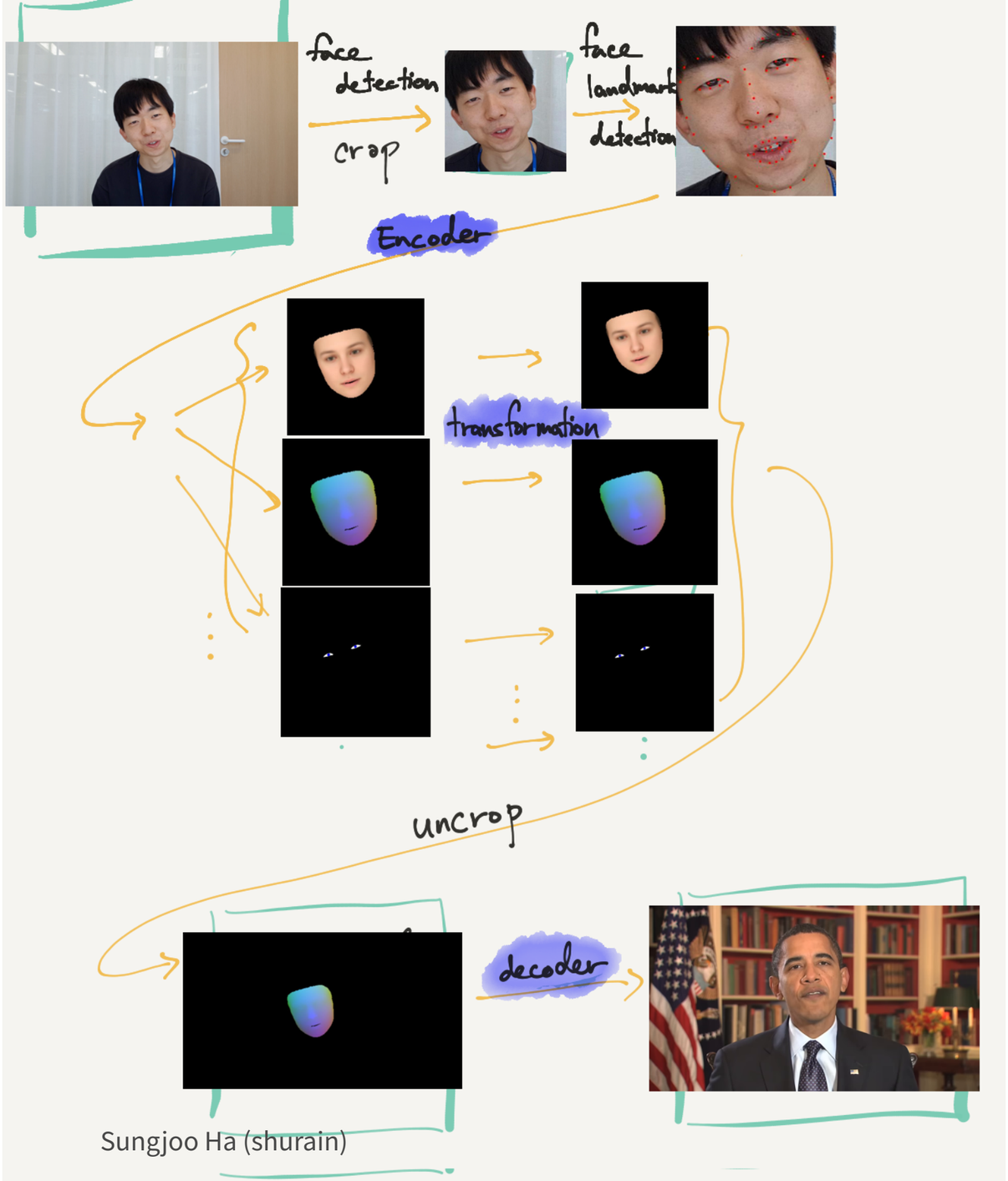
[2] Deep Video Portrait



Data

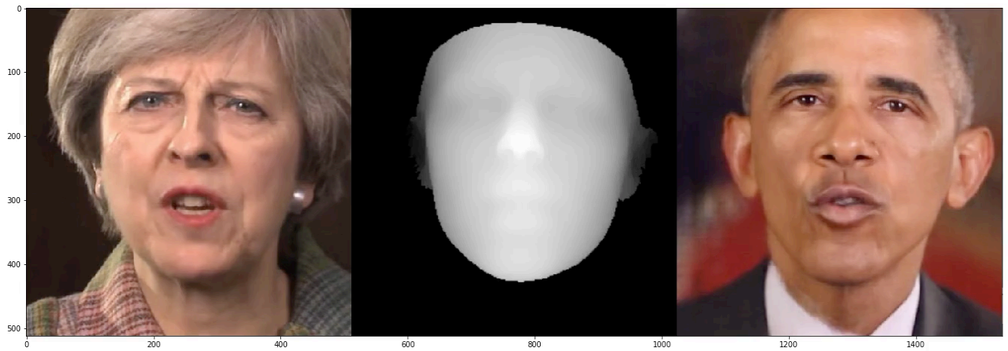
- 공개 데이터셋
  - 논문들이 공통적으로 사용하는 데이터셋을 최대한 확보
  - 공정한 비교를 해야함
  - 불필요하게 기존 모델보다 안 좋은 모델을 만드는데에 큰 노력을 기울이지 않도록
  - 아쉽게도 학외 소속에게는 데이터를 공개하지 않는 경우가 무척 많음[3]
- 비공개 데이터셋
  - 내가 관심있는 도메인에서의 모델 성능은 다를 수 있음
  - 데이터 수집에 대한 고민
    - 어노테이션
    - 정합성 확인
- 데이터 탐색이 필수적

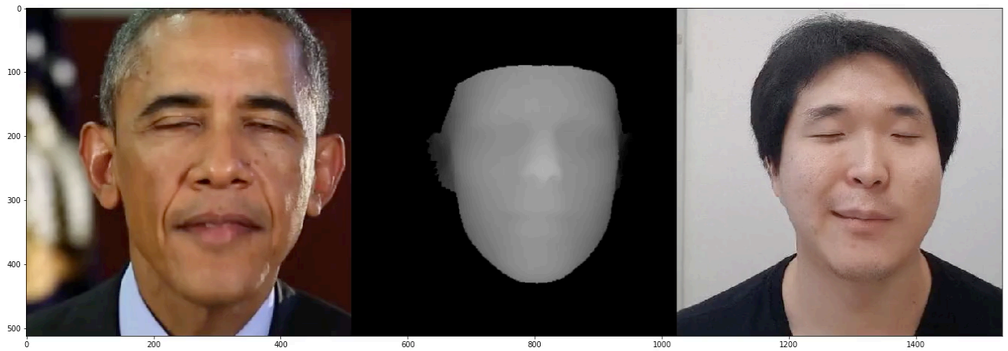
[3] 오픈된 데이터이고 논문 작성을 위한 용도라 하더라도 거절하는 경우가 많음

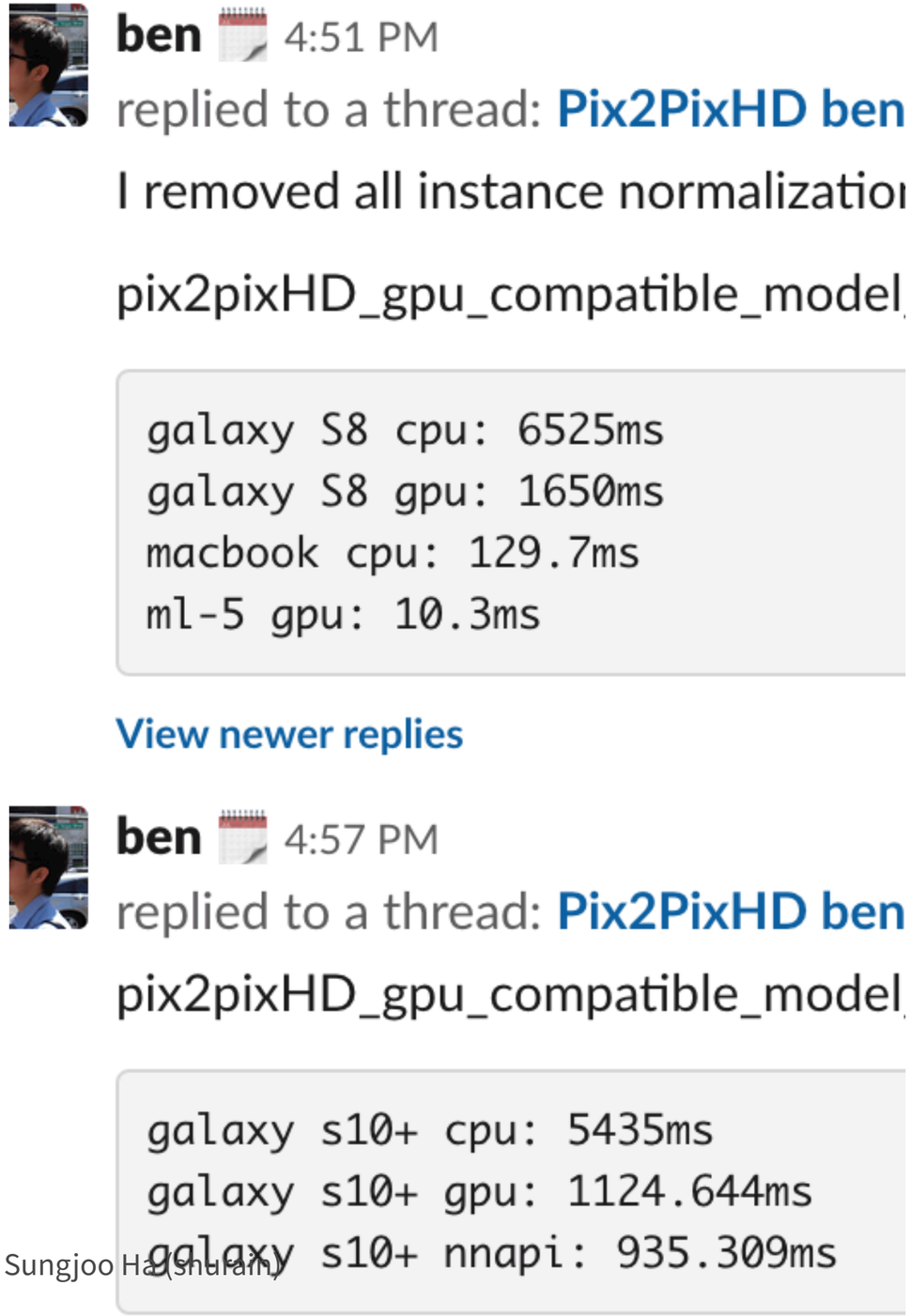


PoC

- 새로운 아이디어 테스트
  - Baseline 모델 개선
- 달성하고 싶은 **목표**를 설정하고
  - 충분한 품질 + 모바일 실시간 + few-shot
- **단계적**으로 도착할 수 있는 방법을 설정
  - 충분한 품질 먼저
  - 모바일 속도는 그 다음
  - Few-shot도 그 다음
- **중간 산출물**
  - 프로덕션에 활용될 수 있는지 치열하게 고민해야 함
  - 이를 고려한 마일스톤을 잡아야 함


0
100
200
300
400
500
0
200
400
600
800
1000
1200
1400


0
100
200
300
400
500
0
200
400
600
800
1000
1200
1400



Process

- 모델이 제품에 적용되기 위해 필요한 부분을 모두 만들어 한 바퀴 사이클 돌리기
  - 도메인에 적합한 전처리 기법 구현
  - 학습 및 검증 파이프라인 만들기
  - 디플로이를 위한 고려
    - 단순 연구와 달리 실제 서비스 환경에서의 모델 행동 양식에 대한 고민이 필요
    - 분산 처리 가능성 및 TF-Lite 활용 가능성 등

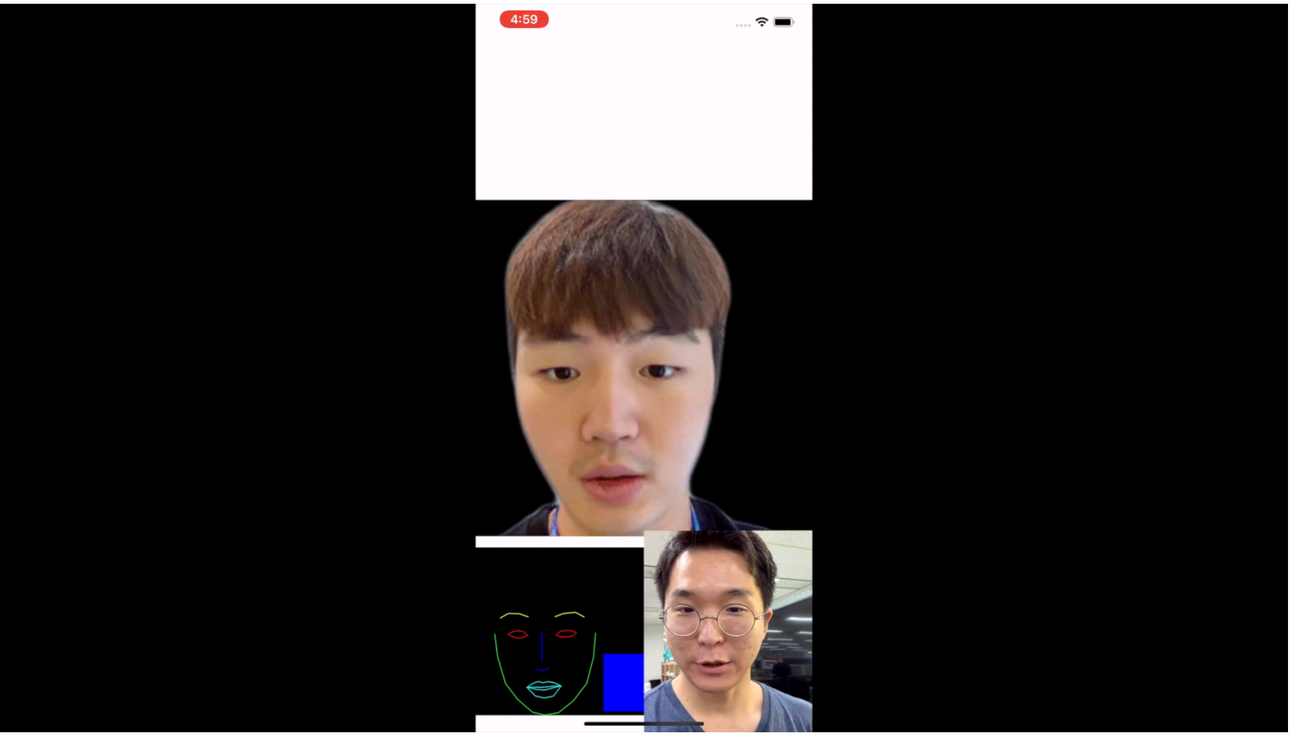
Evaluation

- 여러 모델을 구현/비교할 때에는 공정하게
- 같은 **데이터**, 같은 종류의 **어그멘테이션** 등을 활용해야
- 모델 최적화하는 사람의 역량에 따라 다를 수 있음
  - 논문 재현 시 리포팅된 결과보다 좋은 결과가 종종 나옴
- 논문에서는 측정하기 용이한 메트릭을 주로 보지만 프로덕션 환경에서는 **다양한 메트릭**을 보아야 할 수도 있음
- 콘텐츠를 생성하는 모델은 평가도 어려움

Research

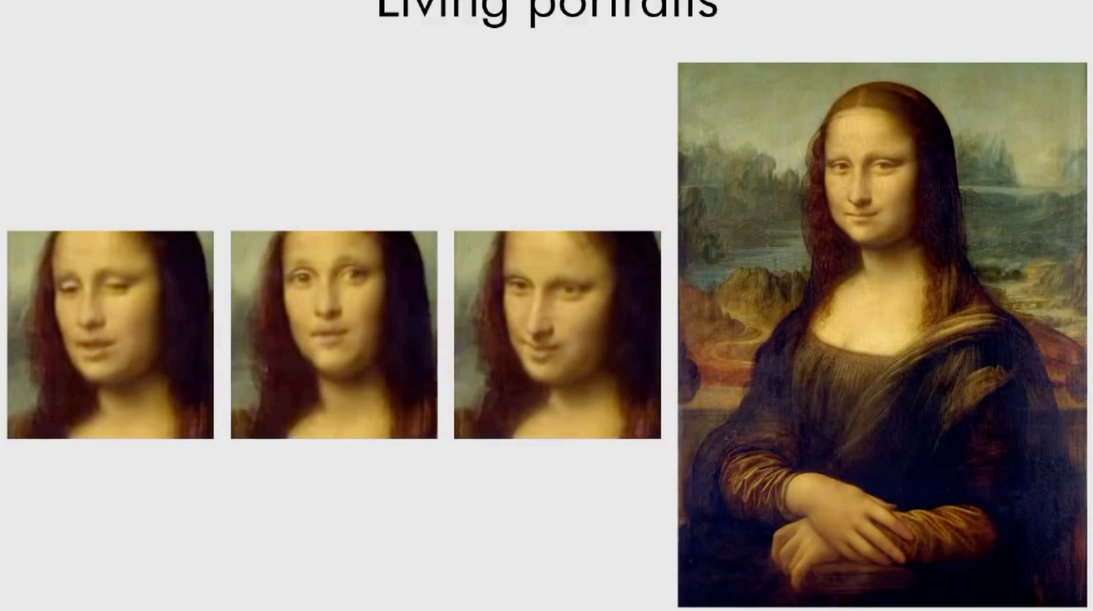
- 원하는 목표에 도달하기까지 계속해서 다양한 시도
- 보통 생각대로 잘 되지 않고 시행착오를 많이 겪게 됨
  - 모두의 인내와 이해가 필요한 시간
- 몇 가지 접근 방법
  - 리터러쳐 서베이에서 유망해보였던 **모델 재현**하면서 아이디어 얻기
  - 다른 도메인의 아이디어 훔쳐오기
  - 팀원들과의 토론

Stretch Goal



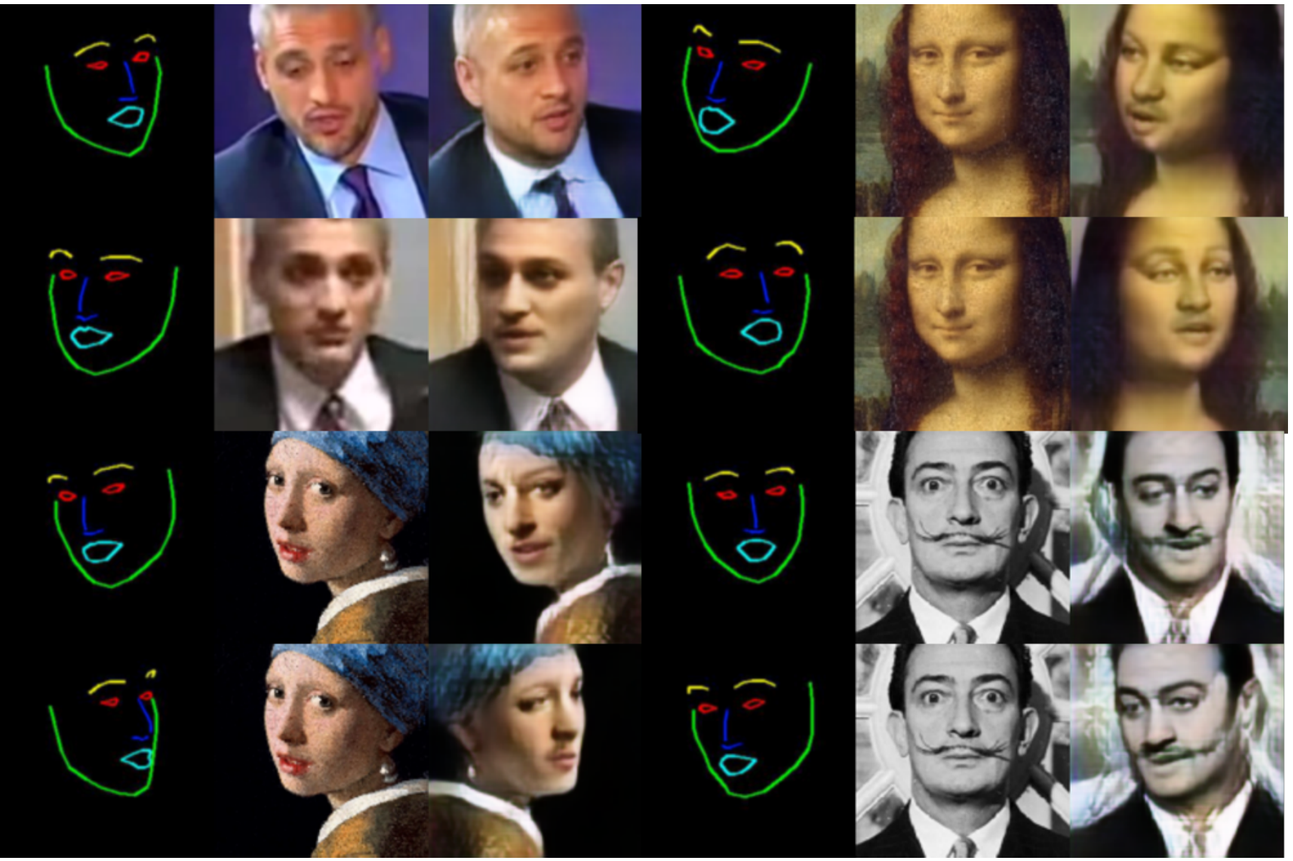
- 목표를 달성하면 그 다음의 목표로
  - DVP 재현하는 것 한 달 정도 걸림
  - 결국 폰에서 실시간으로 돌아가는 모델 만드는데에 성공
    - 한 달 정도 더 작업
  - 그 다음은 few-shot 모델!

Few-shot Face Reenactment



- 프로덕션에서 타겟의 영상을 구하는 것이 가능할까?
  - 어림없음
  - One-shot[4]
    - 이미지 한 장으로도 되야함

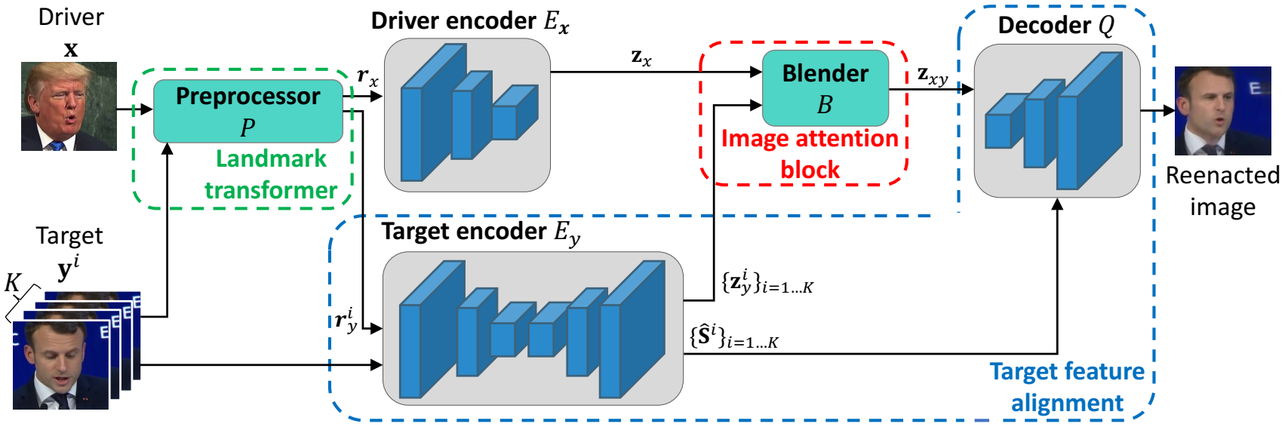
[4] Few-Shot Adversarial Learning of Realistic Neural Talking Head Models



Identity Preservation Problem

- 기존 연구는 타겟과 드라이버가 비슷하고
- 모델을 내가 원하는 타겟마다 추가 학습 (fine-tuning)을 해야 함
- 아니면 드라이버의 얼굴이 타겟에 녹아듦

Novelty



- 타겟의 스타일의 공간적 정보를 유지하도록 정보 추출
- 해당 정보를 잘 활용할 수 있는 구조 제안
  - Image attention block, target feature alignment
- 드라이버의 랜드마크로부터 표정만 추출하여 타겟에게 입힐 수 있는 기법 개발
  - Landmark transformer
  - 추가 레이블 데이터 없이

Result

- 추가 학습 없이
- 단 한장의 이미지만 받아서
- 타겟 얼굴을 유지하는 모델 개발

Publishing

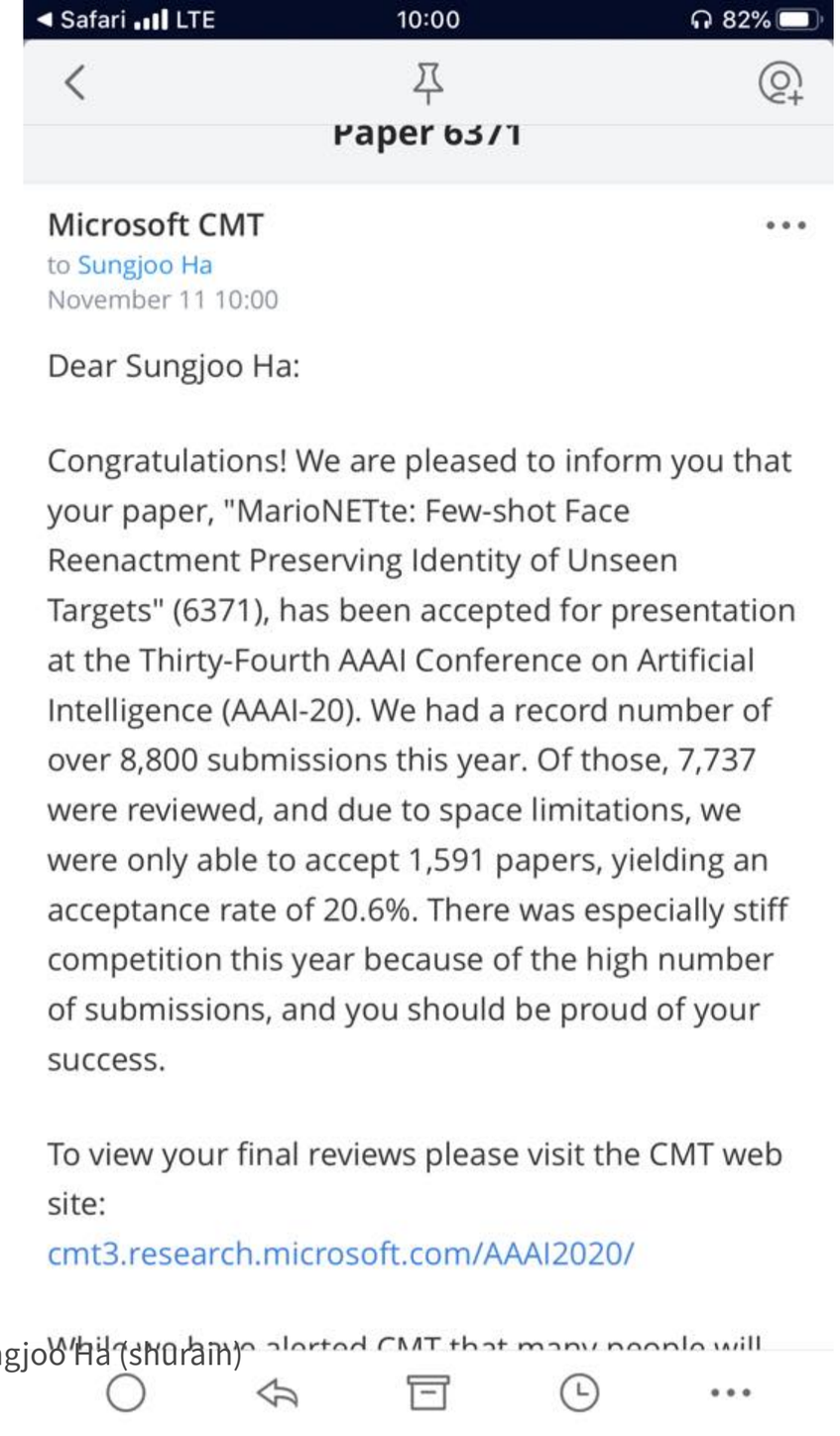
- 의미있는 결과는 최대한 논문화
- 논문을 쓰면서 얻을 수 있는 것
  - 우리가 **풀고자 하는 문제**가 무엇인지 명확하게 정의하는 것
  - **앞으로 진행**해야 하는 실험이 무엇인지 알게 되는 것
  - **Ablation 테스트** 등을 통해 불필요한 컴포넌트를 이해하는 것
- 어차피 숨기고 있어봐야 몇 달 내로 더 좋은 기술이 나옴[5]

---

[5] Few-shot Video-to-Video Synthesis

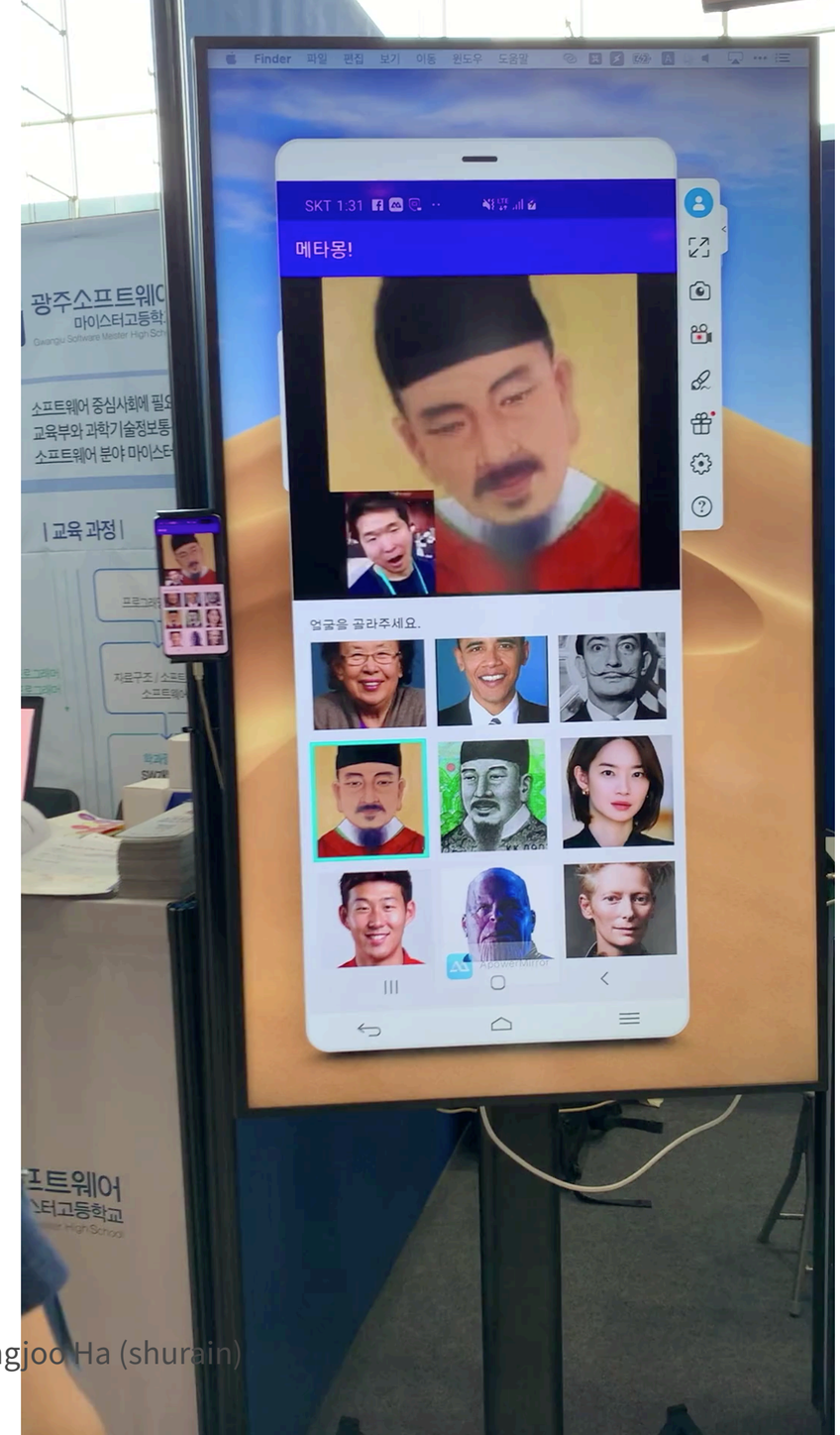
Ablation Test

- 프로덕션 연구 개발 과정에서는 많은 것들이 **점진적**으로 이루어짐
- 최종적인 모델에서 예전에는 의미가 있었으나 더 이상 의미 없는 부분이 있을 수 있음
- Ablation 테스트로 제거
  - 열심히 만들었던 컴포넌트가 사실 별로 쓸모 없었다는 결과는 무척 흔하게 나옴
  - 결과적으로는 프로덕션 환경에서 불필요한 부분을 제거하므로 **이득**



Retrospection

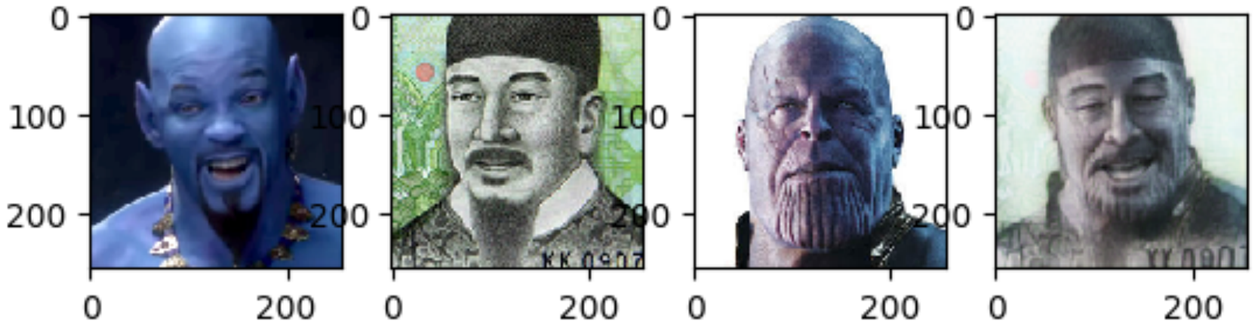
- 프로젝트 시작 4개월 후 SotA 기술 개발
  - **기계학습 전문성 > 도메인 전문성**의 예
  - AAAI 2020 개제
- **회사 홍보** 방법 고민
- **제품에 활용**할 방법 고민



Production + Research

- 제작을 중심으로 하는 회사/팀에서 성공적인 기계학습 조직 운영
  - 서로의 **기대를 맞춰야** 하고
  - 서로 윈-윈할 수 있는 **positive-sum 게임**을 만들어야 함

Expectation Management



- ML ≠ Magic
  - 시도하고 **실패할 수 있음**을 회사에서는 인지해야 함
  - 팀에서는 **리스크를 스스로 판단**하고 움직일 수 있어야 함
  - 하지만 전문성이 잘 맞는 분야에서는 놀라운 결과를 단시간에 낼 수도 있음
- 팀은 **제품에 기여** 해야함
  - 팀에서 해당 고민을 꼭 해줘야 함
    - 기계학습 기술은 약간의 변형을 통해 다방면으로 사용될 가능성이 있음
    - 능동적으로 다른 팀과 **기술의 활용**에 대한 이야기를 해야함
  - **소프트웨어 개발력 + 기계학습 연구력**

Positive-Sum Game

- 제로섬 게임이나 네거티브섬 게임보다는 **포지티브섬 게임**이 낫다
  - 회사와 팀 리더가 특히 고민해줘야 함
- 회사도 팀도 팀원도 연구의 성공 실패와 무관하게 득을 볼 수 있는 방법을 고민해야 함
  - 제품에 들어갈 수 있는 연구
  - 팀원의 성장과 커리어 디벨롭먼트에 대한 고민

Ownership

- 프로젝트의 결정 및 방향 설정에 팀원들이 함께 정함
  - 이 기술이 회사에 **쓸모** 있을까?
    - 연구를 위한 연구는 대부분의 회사에서는 빛을 발하기 힘듦
  - 내가 이 연구를 하면 **재미**있을까?
    - 연구에서 막히는 경우 인내심을 발휘할 수 있는 이유
- 기술을 가장 잘 **이해**하고 있는 것은 연구자 본인
  - 사내 다른 팀들과 지속적으로 이야기 해야함

Thank You