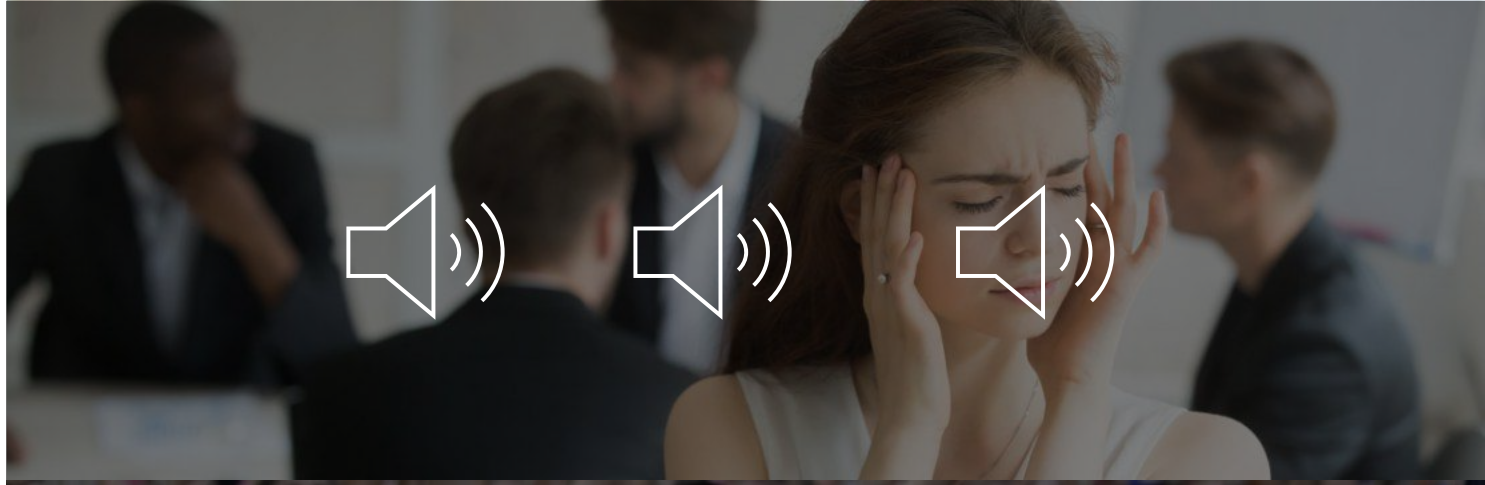
2021 데이터 청년 캠퍼스 [고려대]

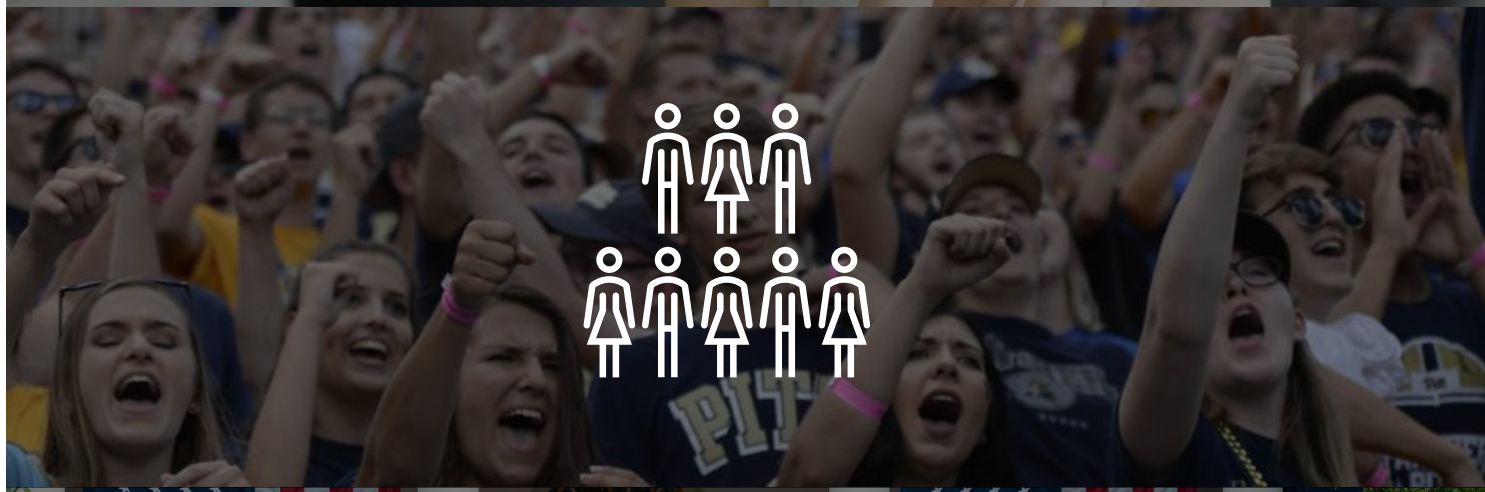
LIP 2 TEXT

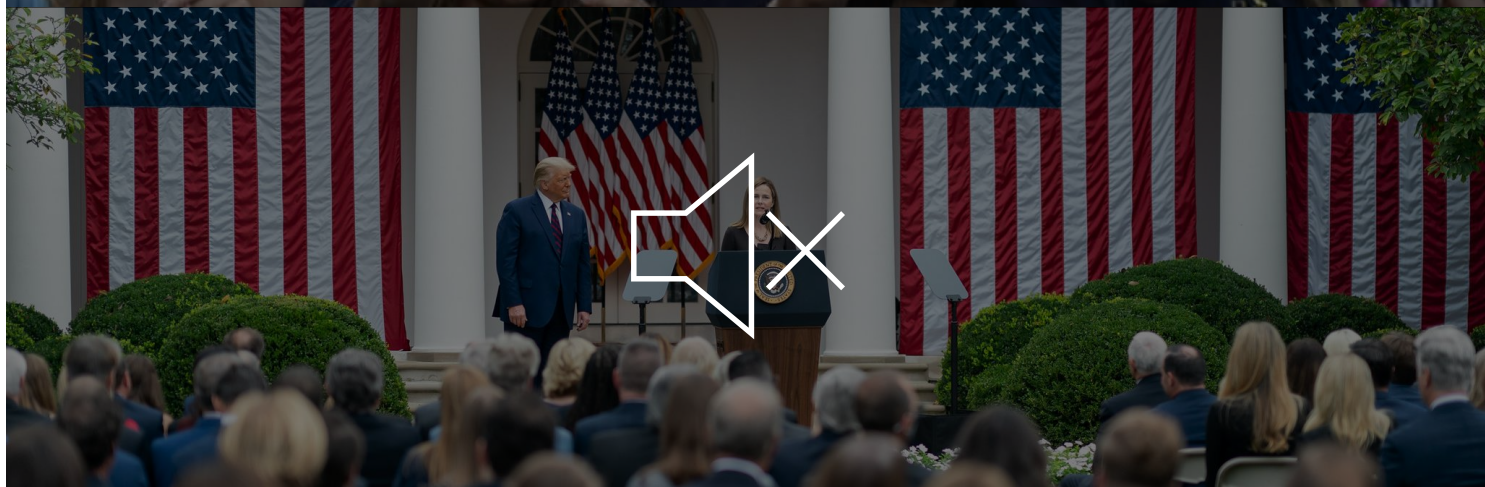
5조 이우준 안혜준 이태희 임준혁 송문영

Contents

주제 선정 배경







주제 선정 배경

주제 선정 배경

데이터 선정

뉴스 앵커 데이터

- 정확한 입모양
- 텍스트 형태의 대사 존재
- 온라인에서 크롤링을 통한 데이터 확보 용이성

웹캠 영상 데이터

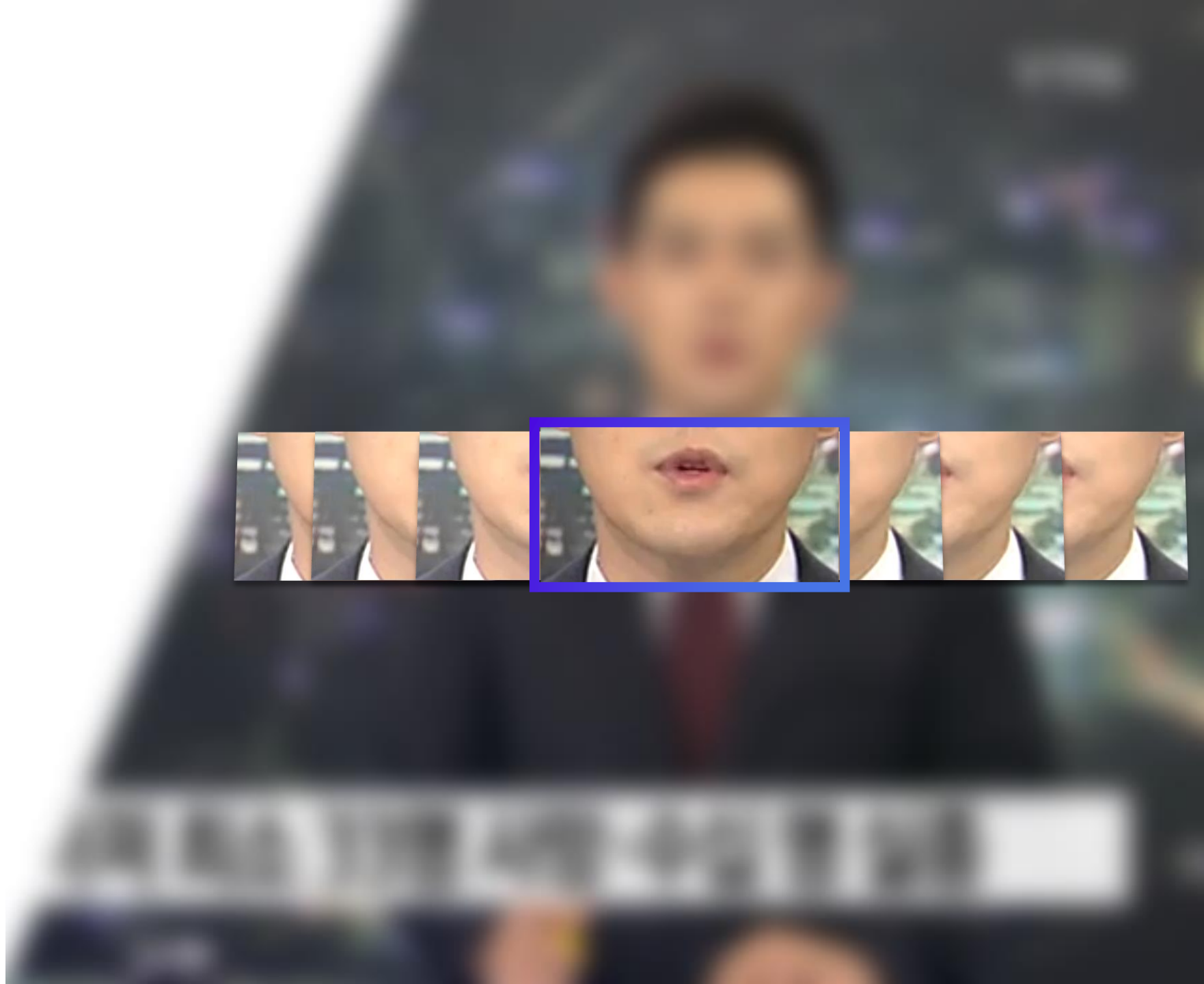
- 저해상도 영상 모델 학습 가능
- 직접 영상 촬영 방법 등 데이터 확보의 용이성

Step 1

Crop
Lip Area

Step 2

Listed by Time Frame



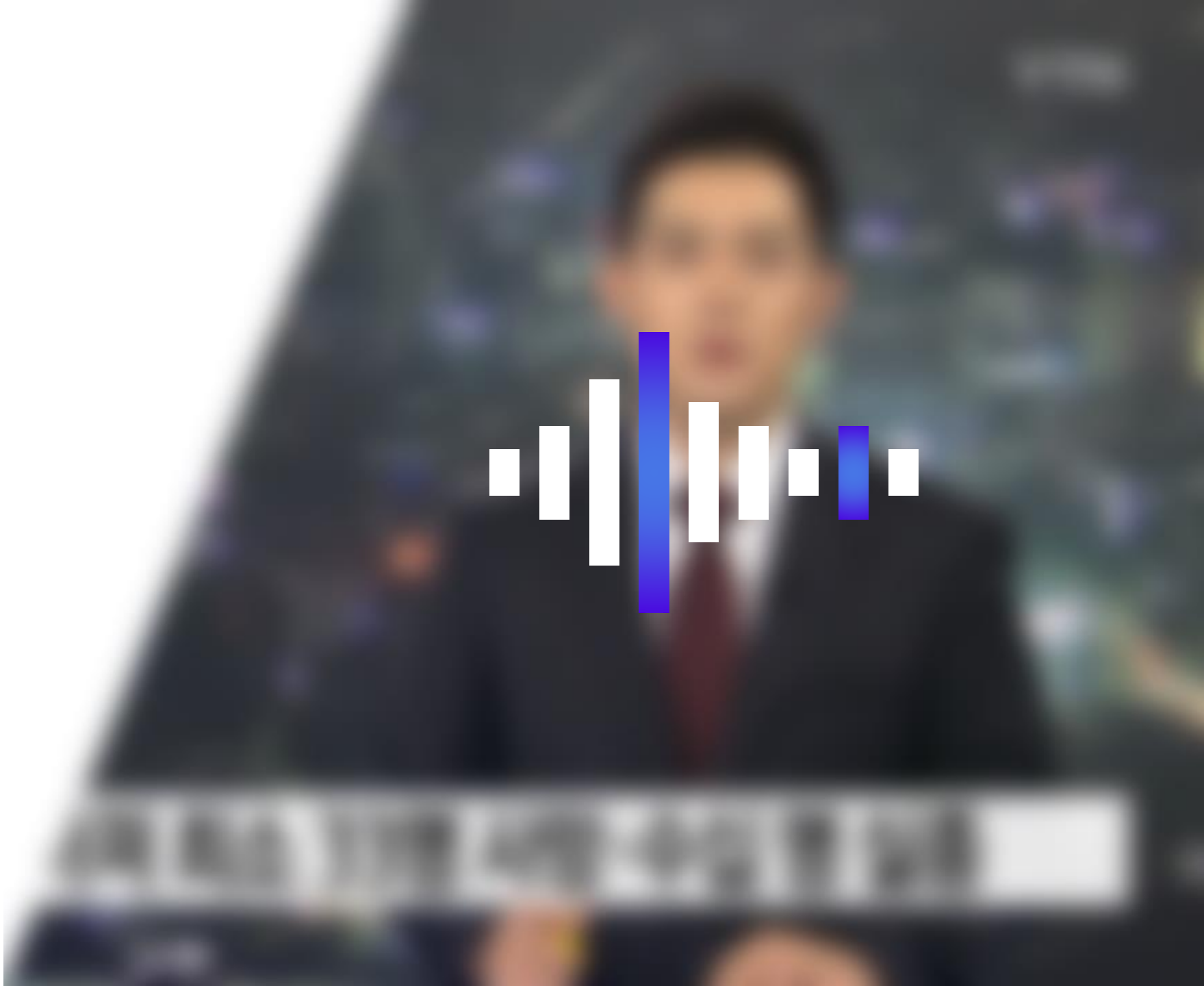
Step 3

Extract Speech Data

Step 4

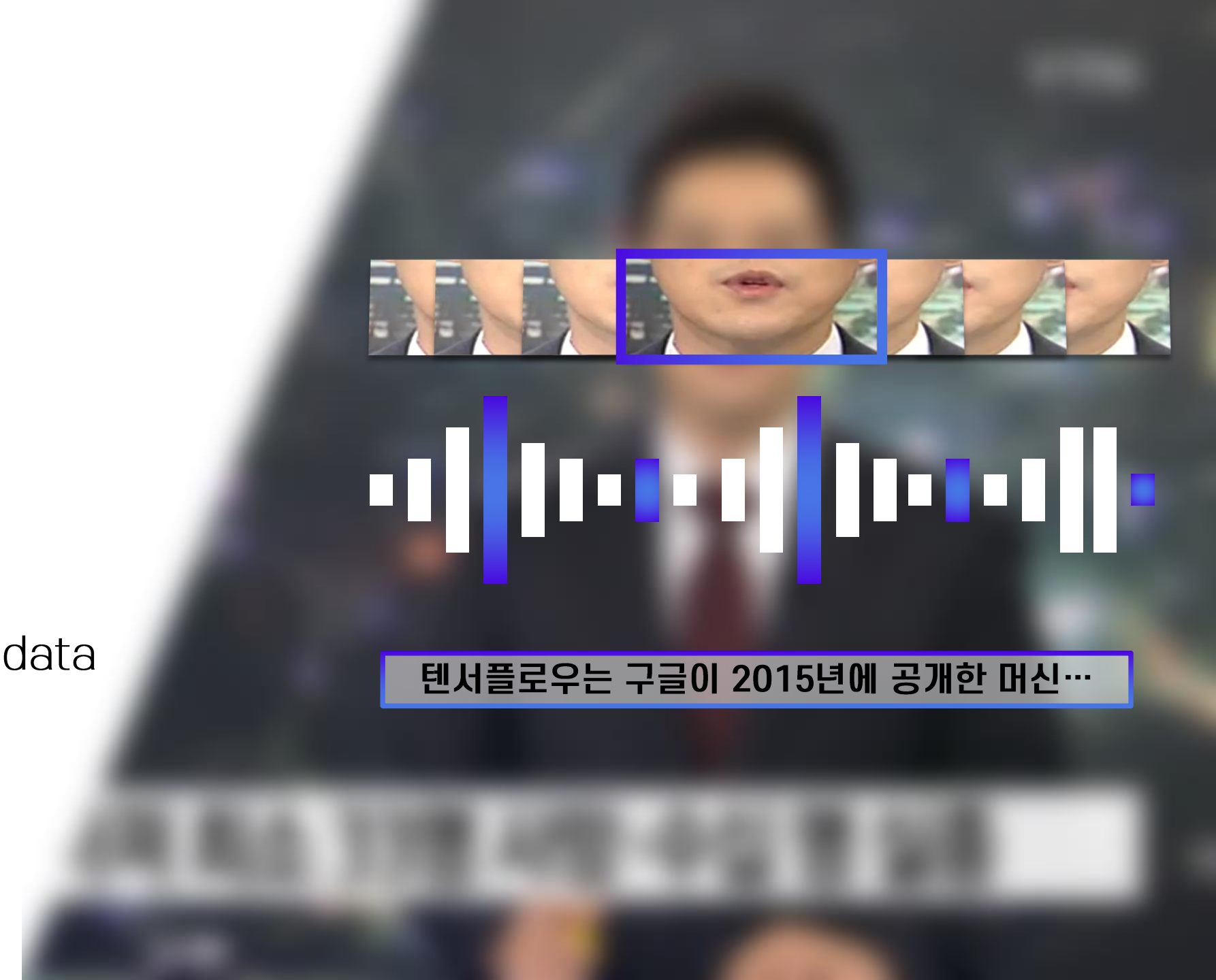
Feature Extraction

(from speech data)



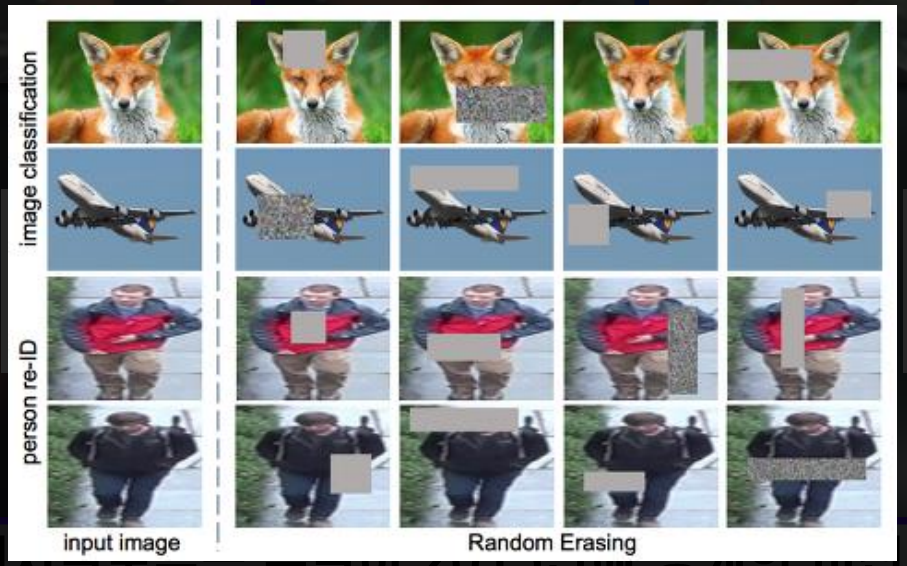
Step 5

Match
Lip image
with voice and text data



데이터가 부족하다면

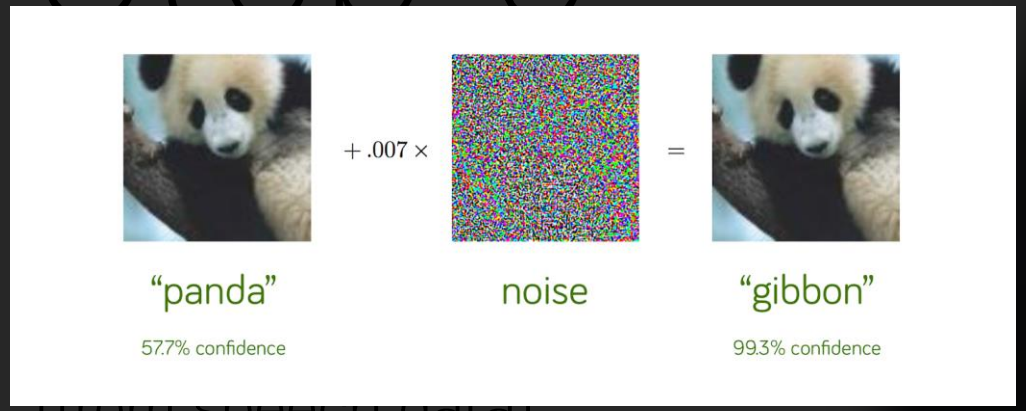
Horizontal Flipping

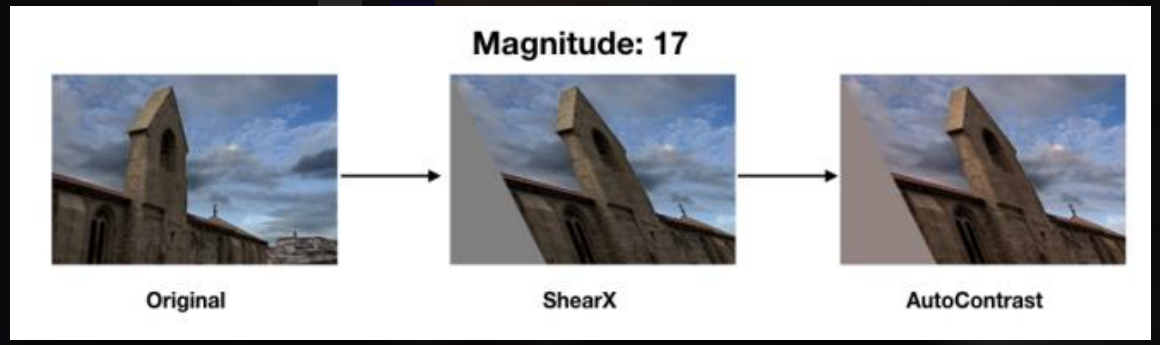


Removal of
Random frame

데이터가 부족하다면

Step 5





Adversarial Attack

Rand Augmentation

구현 방법

데이터 전처리

입 부분 추출

사진 생성
(1초당 30장)

사진, 텍스트,
음성 데이터
라벨링

I

AI 알고리즘 구현

Vision-
Transformer
변형 모델
바탕 구현

II

AI 모델 학습

데이터 활용하여
인공지능 모델
학습

III

AI 모델 테스트

Test 데이터
활용 생성된
인공지능 모델의
성능을 검증

IV

ViT (Vison Transformer)

ViT (Vison Transformer)

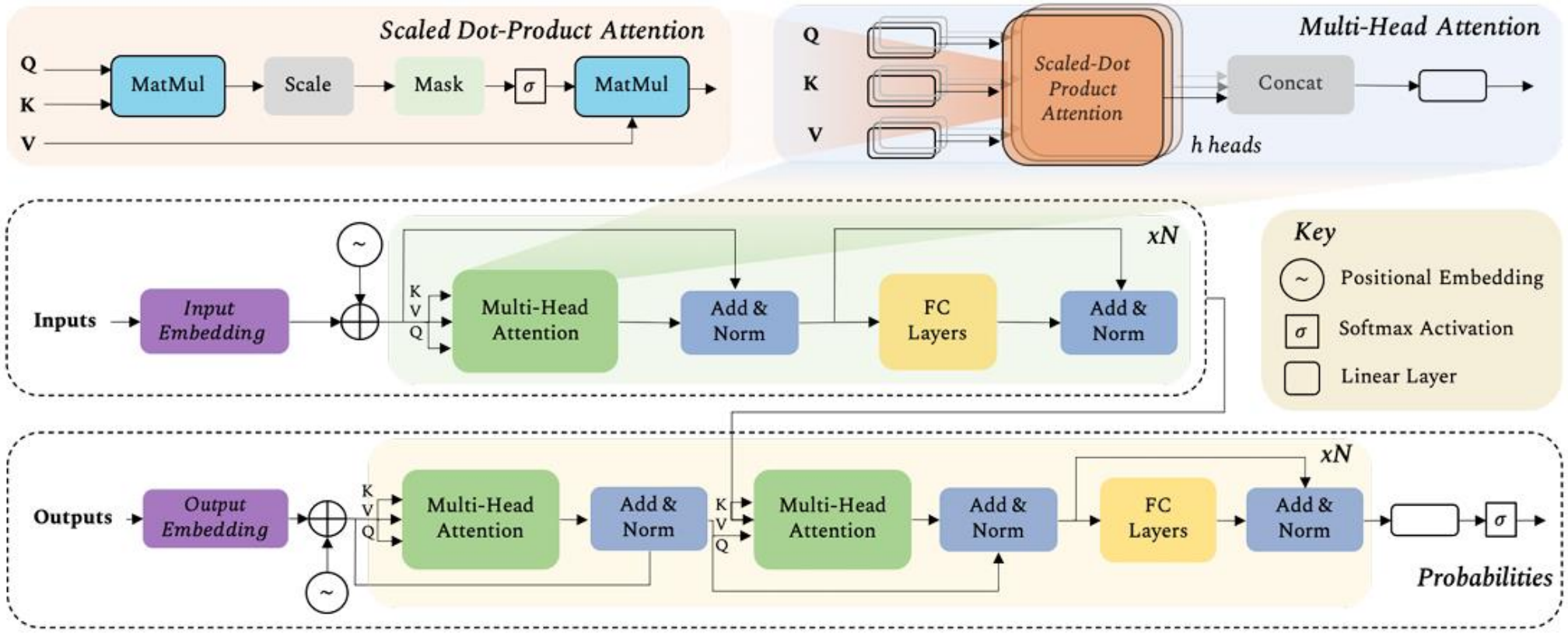
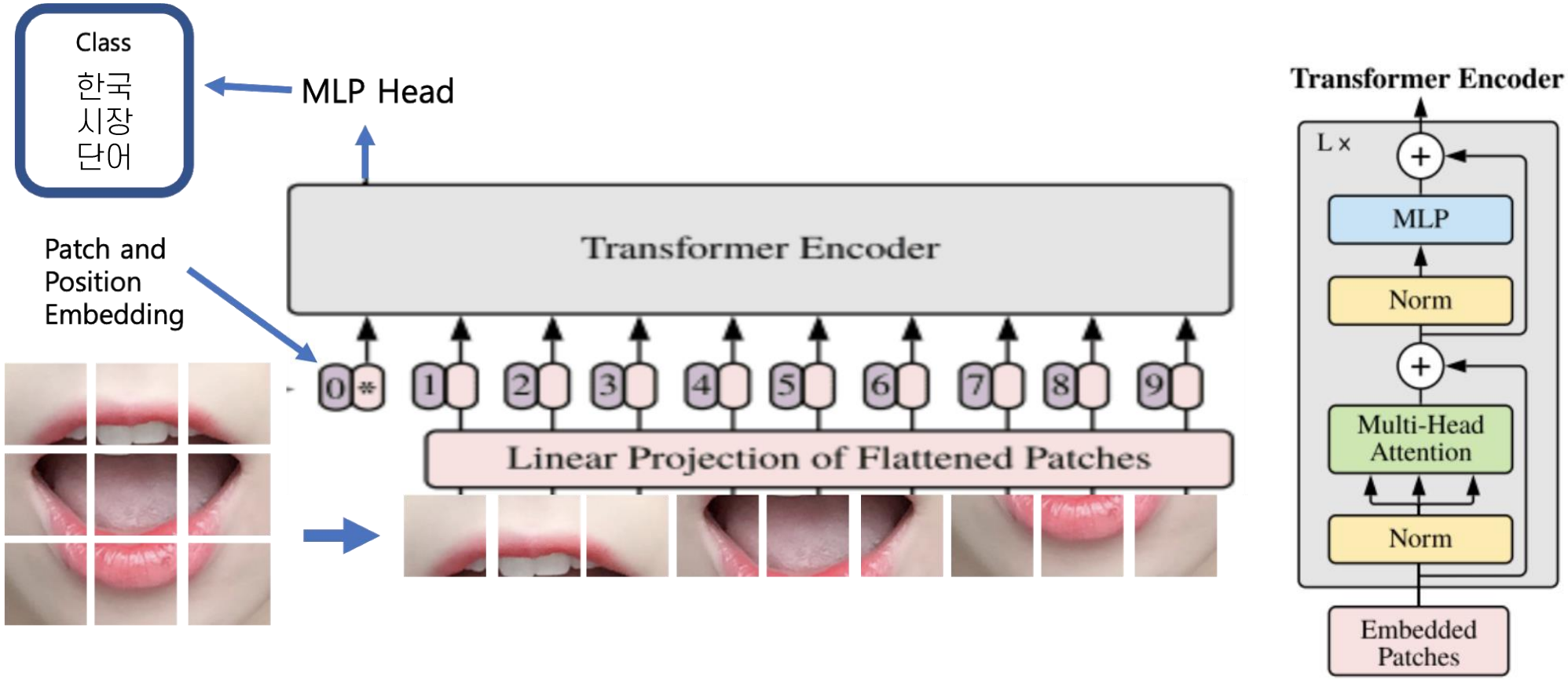


Fig. 2. Architecture of the Transformer Model [1].

ViT (Vison Transformer)



ViT (Vison Transformer)

ViViT

spatio-temporal tokens 추출,
연속된 transformer layer 인코딩

효율적 model regularize
작은 Dataset 좋은 성능

DeiT

적은 GPU로 빠른 학습이 가능

distillation token을 추가 따로 학습
teacher model의 output과 비교

VATT

순수히 Attention만을 사용한 모델
multimodal 영상에 대해 학습

큰 규모의 자기 지도 학습 기반 사전학습
데이터 확보에 대한 부담을 덜어

Drop Token 방식 사용 트랜스포머가 가진
quadratic training complexity
문제 완화

ViT (Vison Transformer)

ViViT

spatio-temporal tokens 추출,
연속된 transformer layer 인코딩

효율적 model regularize
작은 Dataset 좋은 성능

DeiT

적은 GPU로 빠른 학습이 가능

distillation token을 추가 따로 학습
teacher model의 output과 비교

VATT

순수히 Attention만을 사용한 모델
multimodal 영상에 대해 학습

큰 규모의 자기 지도 학습 기반 사전학습
데이터 확보에 대한 부담을 덜어

Drop Token 방식 사용 트랜스포머가 가진
quadratic training complexity
문제 완화

기존 특허와의 차이점

현재 Lip Reading 분야에서
Transformer 기법을 적용한 사례는 매우 **희귀함.**

(가장 최신 분야이기 때문에 적용 사례가 적음),
ViT 기법을 통해 전처리과정이 단축되었음

음절 단위로써의 학습이 아니라 **단어**, **문장** 단위로써의 학습 진행

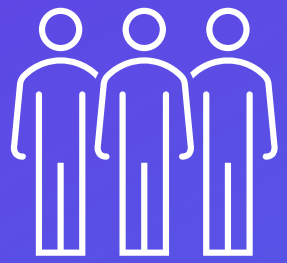
입술의 포인트를 추적해 이동거리 분석을 통한 학습이 아닌
이미지 자체로써의 모델 학습

기존 RNN 모델보다 **높은 성능**이 예상됨
(모델 자체 성능의 우수성)

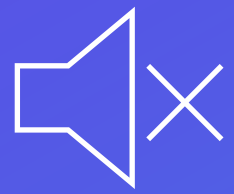
기대 효과

청각장애인을 위한
효과적인 소통 방법을 제공

주변 소음으로부터의 자유



여러 화자들 사이에서의
의사전달 용이



소리가 제한되는 곳에서
입모양으로만 소통이 가능

Q & A

Thank You