AI면접에서의 실시간 거짓말 탐지를 위한

SSUL-NET:

SoongSil University LieNet

AI & Healthcare Lab.

CONTENTS

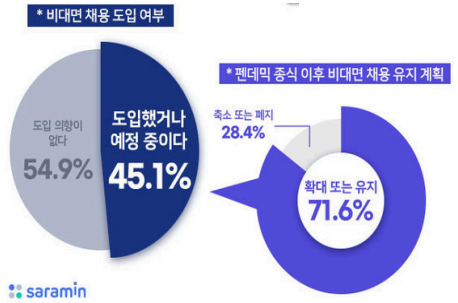
연구 배경

AI면접이란?

- 인공지능(AI)을 활용해 지원자의 역량을 평가하는 기술
- 지원자의 영상 및 음성 데이터를 기반으로 역량 평가
- 신뢰도, 자신감 외 37개 평가요소를 바탕으로 지원자의 역량 예측

- 코로나19의 영향으로 비대면 채용 방식인 AI면접 도입 증가

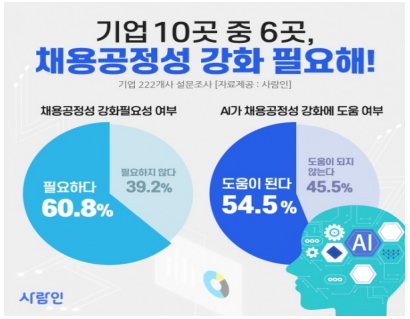
[AI면접 도입 상승 추세]



[기업의 비대면 채용 도입 계획]

AI면접의 문제점

- 비대면 AI면접의 특성상
  지원자의 신뢰성 검증에 취약



[AI면접의 채용공정성 강화 도움 여부]

- 기존 신뢰감 분류 모델의 낮은 정확도
  RandomForest 이용 → 머신러닝의 한계

AI면접 종합 모델의 분류 정확도

신뢰감 : 69%

자신감 : 76%

호감도 : 72%

긴장수준 : 67%

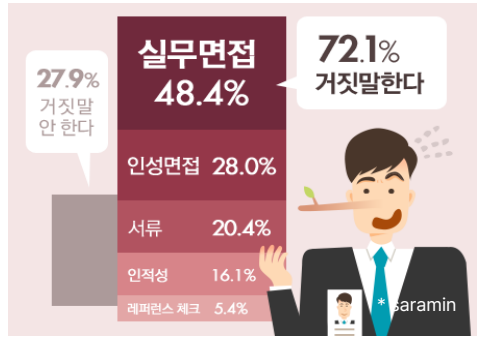
의사전달능력 : 72%

감정전달능력 : 71%

* Midas IT

신뢰성 검증 필요성 대두

- 면접자의 거짓 답변 증가로 인한
  채용 과정의 신뢰성 감소



[채용과정 중 지원자의 거짓 비율]

[면접 과정 중 지원자의 거짓 증가 비율]

선행 연구

논문 제목	Data set	사용 모델	요약
Bag-of-Lies :A Multimodal Dataset for Deception Detection	**Bag-Of-Lies** Video, EEG, Gaze로 구성	KNN, Random forest	• Video, Audio, EEG, Gaze의 각 feature 추출 • 각 feature를 concatenated 한 후 머신 러닝 모델에 적용 • Multi-Modal 한 속임수 탐지 시스템을 구축
Deception Detection in Videos Using Robust Facial Features	**Bag-Of-Lies,** **Real-life Trial Data**	KNN, SVM, MLP, AdaBoost	• FAU와 Gaze의 feature를 채널별로 연결하여 입력 값으로 사용 • Video feature 추출 후 TCN 모델 사용 • Video, FAU, Gaze를 기반의 미세 표현 감지 프레임워크 구축
LieNet : A Deep Convolution Neural Network Framework for Detecting Deception	**Bag-Of-Lies,** **Real-life Trial Data**	LieNet	• Video, audio, EEG feature를 추출한 이미지를 입력 값으로 사용 • CNN모델을 사용해 Multi-Modal한 속임수 탐지 시스템을 구축

실시간 거짓말 탐지 불가능

영상마다의 거짓말 여부만 판별 가능,답변의 진실정도를 확인 불가

최종 목표

- 실시간 특성을 고려해 짧은 시간 단위로 프레임을 뽑아 프레임 단위로 거짓말을 탐지
- 답변에 대한 신뢰성을 수치화(진실된 프레임 개수/총 프레임 개수)

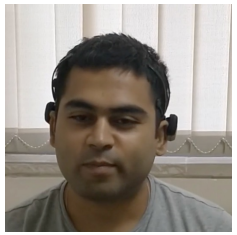
OVERVIEW

STEP 1 데이터 수집 및 전처리 → **STEP 2** 각 Video, Audio 모델 설계 → **STEP 3** Multi-modal 모델 설계 → **STEP 4** 성능 평가 및 프로토타입 구성

데이터 수집

Bag of Lies Data

- 총 325개 (162 lies+163 truths)
- 제시된 이미지에 대한 피실험자의 거짓/진실 답변 영상

데이터 전처리

1) 영상 및 음성 데이터 전처리

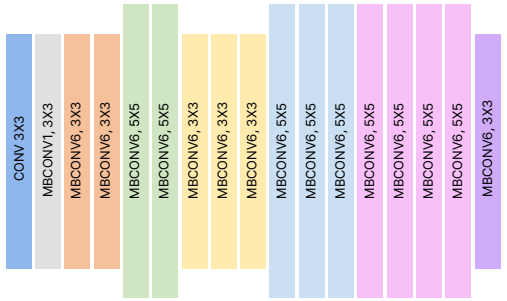
- LBP, Mel Spectrogram 기법 적용

2) 실시간 특성을 반영하기 위하여 특정 프레임 단위로 추출 및 가공

- Sliding Window 기법을 통한 프레임 추출
- 10개의 프레임을 Concatenation

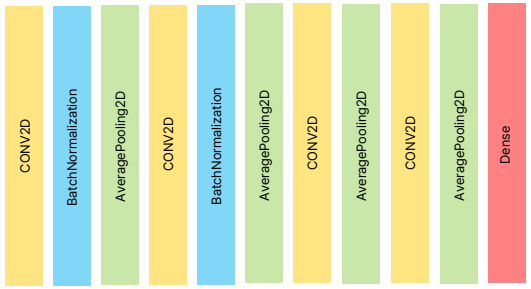
Video

- Efficient Net

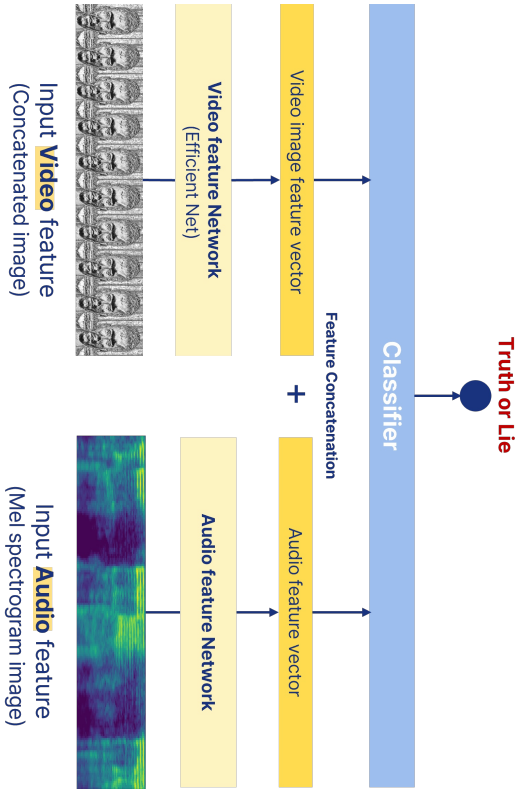


Audio

- Network Structure of Audio



Model Architecture

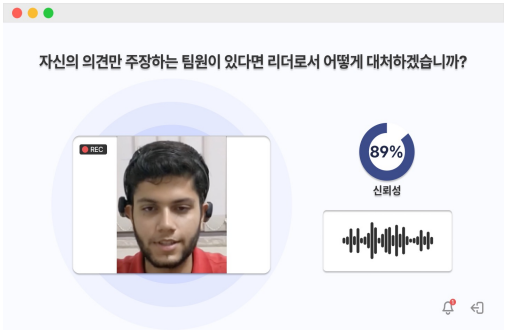


성능 평가

모델의 Accuracy 및 Recall

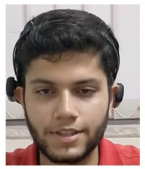
	Accuracy	Recall
Video	90.2%	-
Audio	71.1%	-
Video + Audio (Multi Modal)	96.3%	82.6%

제안 프로토타입



- 평가 결과를 AI 면접화면에 신뢰성 점수로서 실시간으로 제공

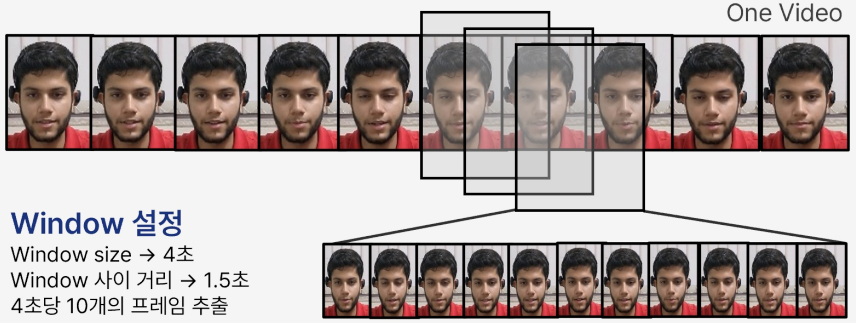
1) 데이터 전처리



Bag-of-Lies

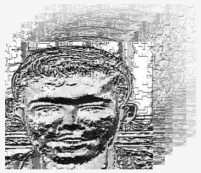
Video

Sliding Window 적용



Window 설정
Window size → 4초
Window 사이 거리 → 1.5초
4초당 10개의 프레임 추출

추출 프레임 LBP적용



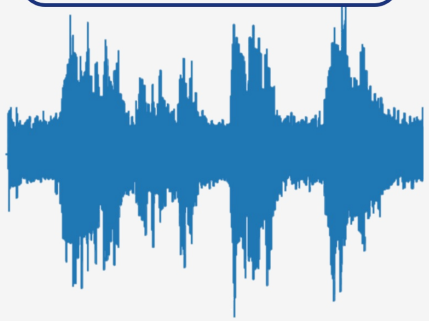
LBP
Local Binary Patterns(LBP)는 이미지의 질감표현 및 얼굴 인식 등에 활용되는 알고리즘

10개의 이미지 결합

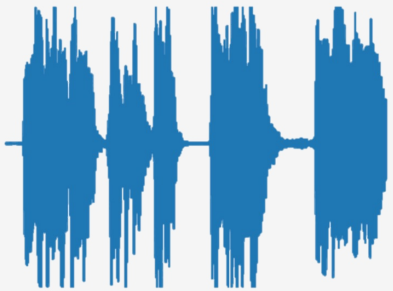
최종 Input Image 생성
LBP 알고리즘이 적용된 프레임 10개의 이미지를 결합
최종적으로 모델의 입력 값으로 사용

Audio

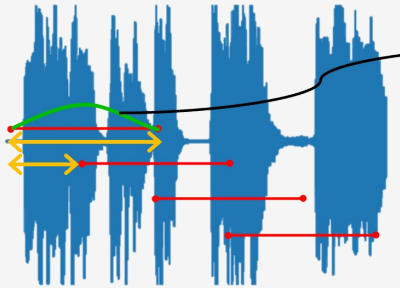
영상으로부터 음원 추출



노이즈 제거

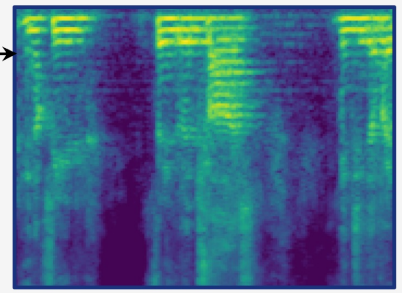


Sliding Window 적용



Window 설정
Window size → 4초 / hop length→ 1.5초

Mel Spectrogram 추출



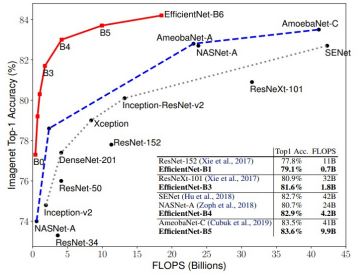
최종 Input Image 생성

3) 모델 설계

✓ Video

Efficient Net 선정 이유

- 적은 파라미터로 효율적인 성능



- 타 모델과 비교 시 우수한 성능

Model	Accuracy(%)
VGG-16	87%
VGG-Face	83%
ResNet	89%
Efficient Net	90.2%

Efficient Net architecture

- NAS(Neural Architecture Search) 기술을 사용하여 설계된 CNN 아키텍처
- Compound Scaling이라는 기법을 사용하여 아키텍처의 크기를 조정
- 네트워크의 depth, width, resolution을 동시에 스케일링 후 모델 성능을 향상

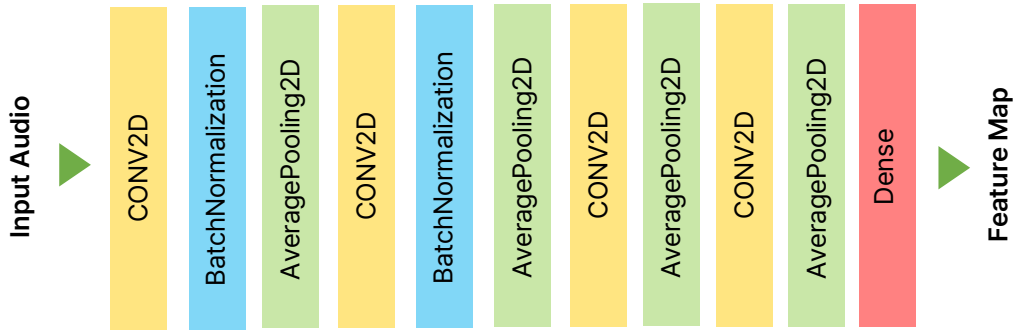
✓Audio

Network Structure of Audio

- 기존 Inhalation and Exhalation Detection을 위한 모델 아키텍처
- 모델 도입 시 고려사항

  1) 오디오 데이터로부터 추출한 feature를 합성곱 신경망을 사용

  2) Binary Classification의 특성



- 타 모델과 비교 시 우수한 성능

Model	Accuracy(%)
ResNet18	53.2%
LieNet	60.4%
Network of Audio	71.1%

3) Multi-modal 모델 설계

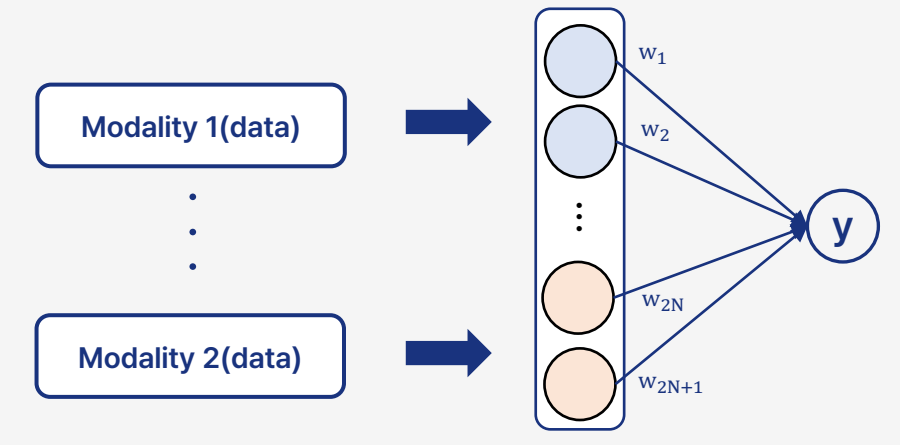
Feature Concatenation

Multi-modal이란?

Single-modal 데이터의 한계를 극복하고자 Multi-modal의 데이터를 사용해 주어진 문제를 해결하는 모델을 구축하는 방법론

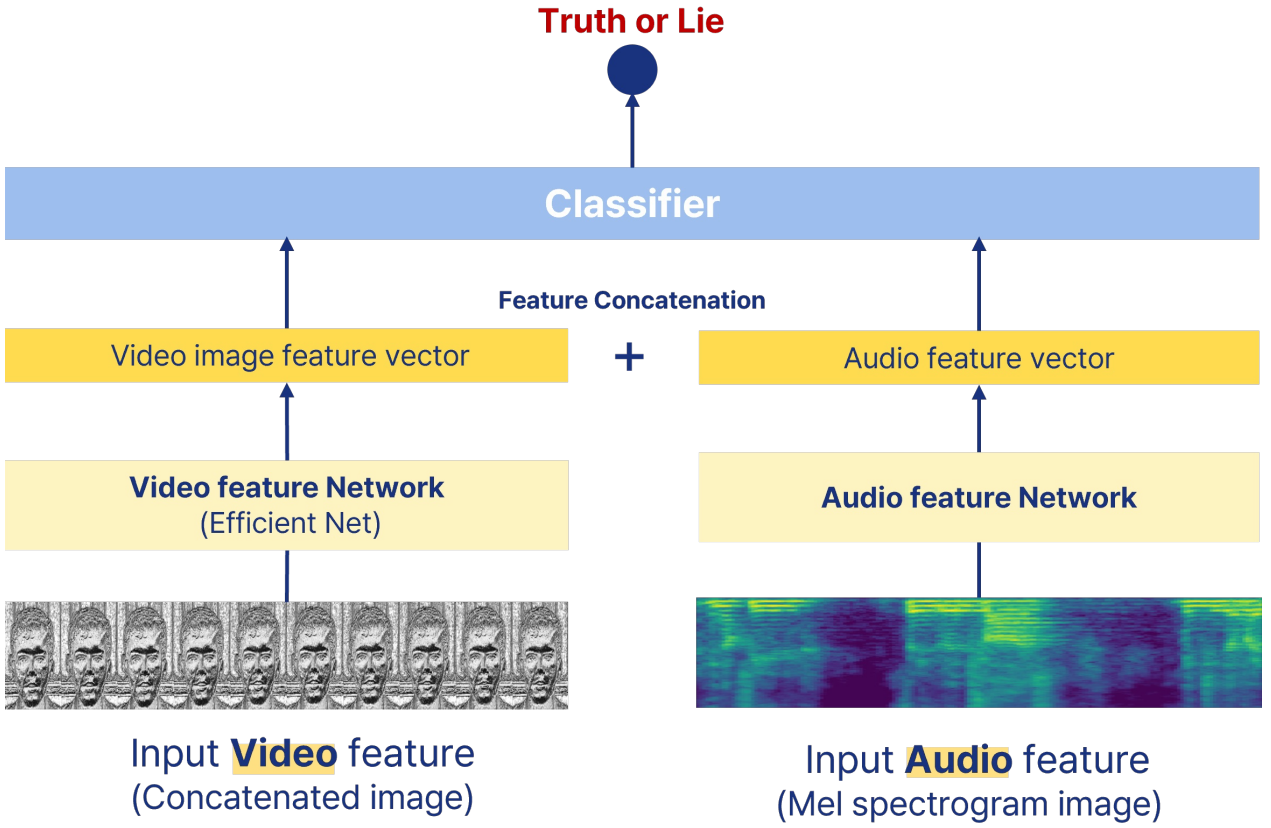
Feature Concatenation

- 각 모달의 feature들을 추출한 후 이어 붙이는 방식
- 하나의 이어진 벡터가 input으로 사용
- 이는 각각의 모달이 가진 정보를 보존하는 것



[Feature Concatenation Structure]

Multi-modal Model Structure



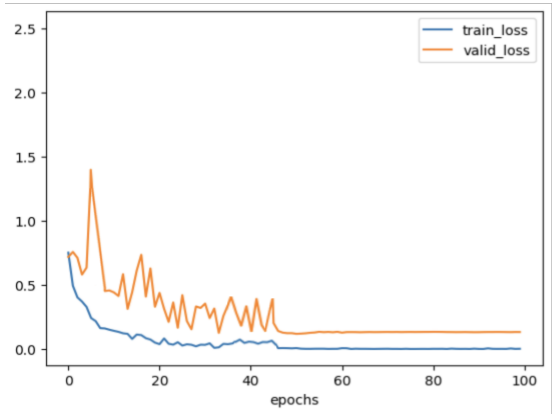
성능 평가 및 결과

분류 성능 평가 지표

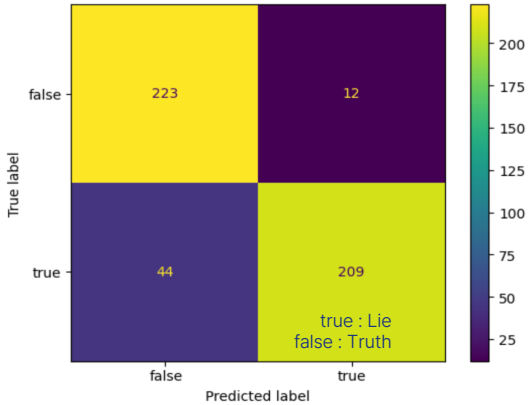
Accuracy & Recall

	Accuracy	Recall
Video	**90.2%**	**-**
Audio	**71.1%**	**-**
Video + Audio (Multi-modal)	**96.3%**	**82.6%**

Loss Graph & Confusion Matrix

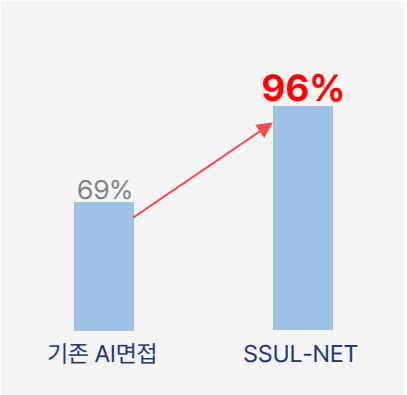


[Loss Graph]



[Confusion Matrix]

신뢰성 측정 및 성능 향상



- 머신 러닝 기반의 기존 AI면접 분류 모델은 **69%** 정확도
- Multi-modal 딥러닝 기반의 SSUL-NET은 **96%** 로 보다 정확한 성능

결과 해석

- **Multi-modal 모델 도입으로 인한 성능 향상**
  단일 기능 모델보다는 Video와 Audio를 모두 고려하는 Multimodal 모델을 사용했을 때 모델이 더욱 향상됨을 확인
- **높은 재현율(Recall 값)**
  거짓말 탐지의 도메인 특성상, 실제 거짓말 데이터를 정확하게 예측하는 것이 중요함 → Recall = 82.6%

최종 시스템 및 GUI구성

AS-IS

[면접 종합 평가 화면]

[면접 역량 평가 지표]

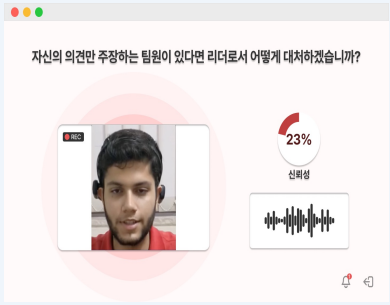
다른 AI면접의 평가 분류 모델 보다 비교적 **낮은 분류 정확도**를 보이는 신뢰도

사용자가 면접을 보는 도중 **실시간**으로 **신뢰도 판별 불가**

모든 답변에 대해 **종합한 신뢰도만**을 확인 가능
(개별 답변에 대한 신뢰도는 확인불가)

TO-BE

[실시간으로 신뢰성 점수 표시]



[실시간으로 신뢰성 점수 표시]

[답변의 신뢰도 추이 그래프]

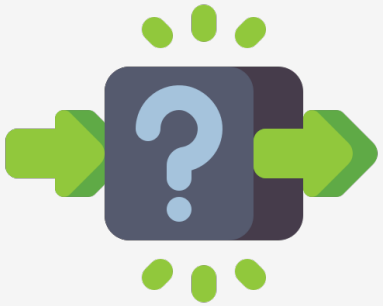
면접자의 영상과 음성 특징을 기반으로 **더 높은 정확도**를 가진 모델

평가 결과를 화면에 **신뢰성 점수**로서 **실시간**으로 제공

각 답변에 있어서 시간에 따른 **신뢰도 추이를 그래프**로 제공

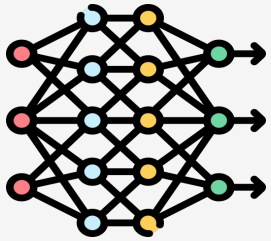
실시간 신뢰도에 대한 **종합 평가**를 제공

기대효과



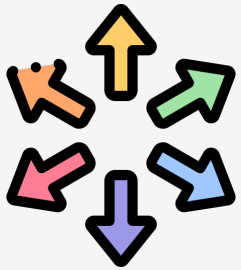
AI 면접의 블랙박스 문제 해결

기업과 지원자 모두에게 신뢰성 요소
평가 결과에 대한 근거 제공



딥러닝 모델 유연한 적용 가능

기존 AI면접의 머신러닝 모델에서 벗어나
다양한 요소를 종합적으로 고려



SSUL-NET의 적용 도메인 확장

AI면접에 더해 거짓 탐지를 위한
다양한 영역에서 활용 가능

향후 개선 과제

1 모델 고도화를 위한 데이터 수집

- SSUL-NET의 정확도 향상을 위해 실제 AI면접 영상 등 **연구 방향에 적합한 양질의 데이터** 필요
- SSUL-NET 활용 목적에 따른 **다양한 학습 데이터** 수집으로 적용 가능 도메인 확장

2 추가 특징 도입으로 세밀한 거짓 탐지 실현

- SSUL-NET에서 제안한 Multimodal 딥러닝 모델은 Video, Audio 두가지 특징을 고려
- 향후 EEG, Gaze, Text 등 **다른 종류의 데이터를 추가**함으로써 보다 세밀한 거짓 탐지가 가능한 모델로 발전

Reference

1) V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh, and
M. Vatsa, "Bag-of-lies: A multimodal dataset for deception detection," in *Proc. IEEE Conf. Comput. Vis. Patter n Recognit. Workshops*, Long Beach, CA, USA, 2019, pp. 83–90.

2) Stathopoulos, A.; Han, L.; Dunbar, N.; Burgoon, J.K.; Metaxas, D. Deception Detection in Videos Using Robust Facial Features. In Proceedings of the Future Technologies Conference, Online, 5–6 November 2020; pp. 668–682.

3) M.Karnati , A.Seal , *S.Member, IEEE*, Anis Yazidi , *Senior Member, IEEE*, and O.Krejcar,"LieNet: A Deep Conv olution Neural Network Framework for Detecting Deception" in *Proc.* IEEE TRANSACTIONS ON COGNITIVE AN D DEVELOPMENTAL SYSTEMS, VOL. 14, NO. 3, SEPTEMBER 2022

4) M. Dörfler, R. Bammer, and T. Grill, "Inside the spectrogram: Convolutional neural networks in audio processing," in Proc. Sampling Theory Appl. Conf., 2017, pp. 152–155.

5) 『Local binary pattern(LBP)의 원리 및 활용』, bskyvision, https://bskyvision.com/entry/Local-binary-pattern-LBP%EC%9D%98-%EC%9B%90%EB%A6%AC-%EB%B0%8F-%ED%99%9C%EC%9A%A9, 2018-01-22