

DACON

SW중심대학 디지털 경진대회

SW와 생성AI의 만남



박웅 | 김홍권 | 박민규 | 박제현 | 유민주





데이터 분석

데이터 분석

데이터 전처리

사용 기법

모델 구조

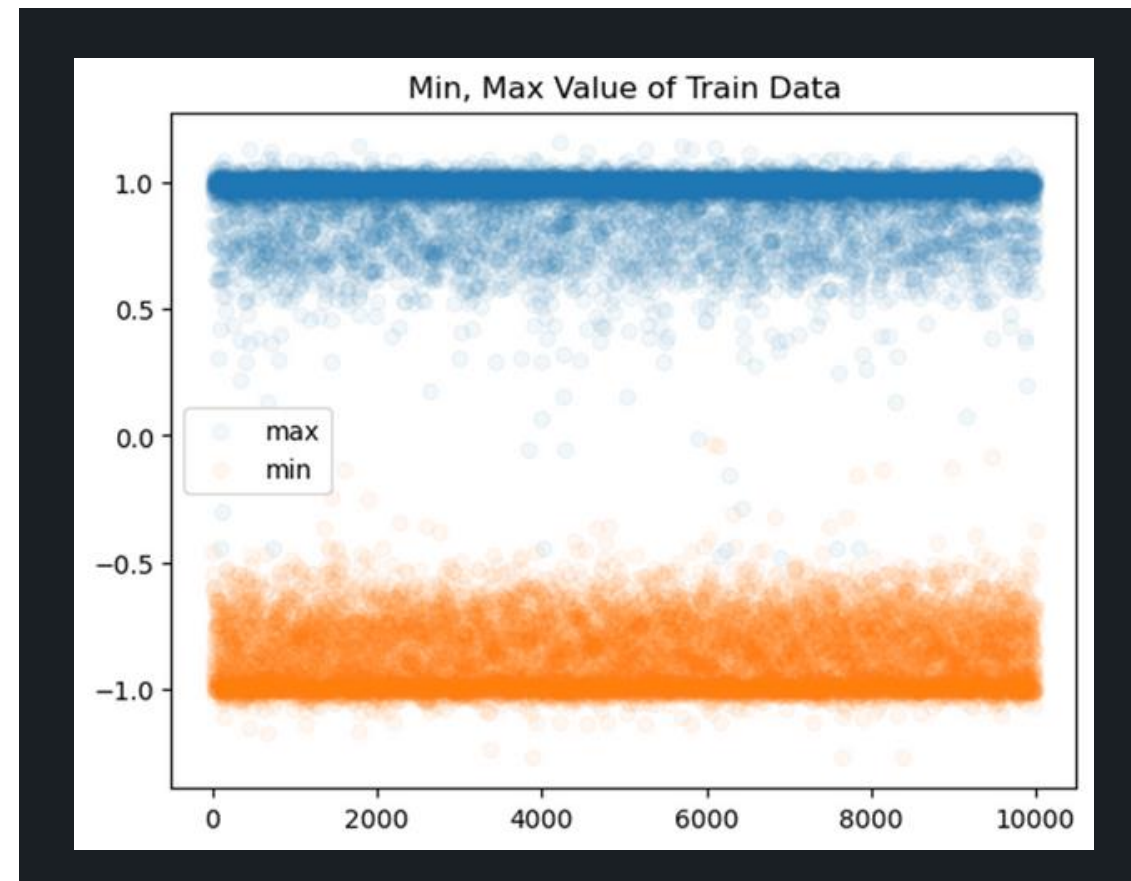
모델 학습

인퍼런스 전략

결과

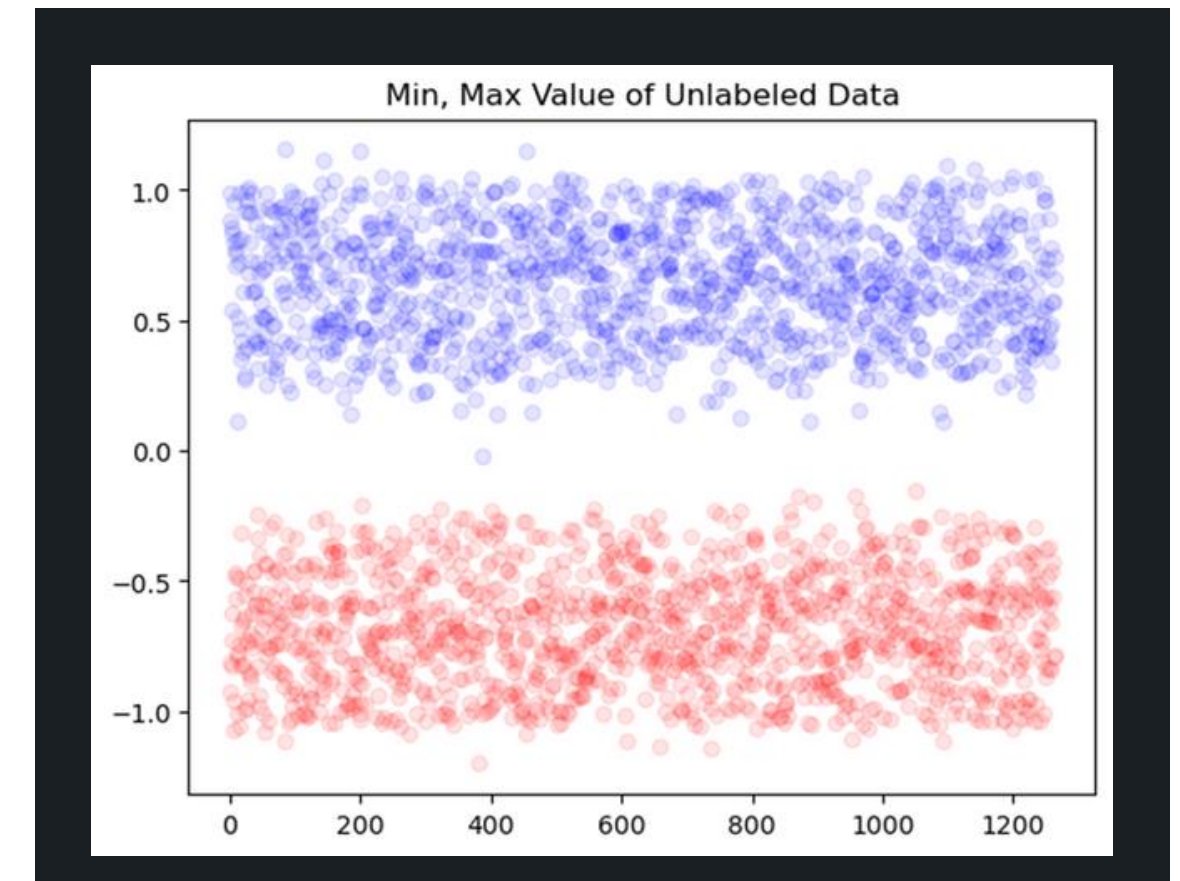
Train Data

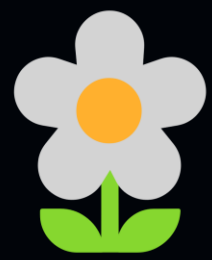
한 명의 화자
Fake or Real의 이진분류
대부분의 샘플이 $[-1, 1]$ 사이에 위치



Unlabeled Data

최소 0명 최대 두 명의 화자
Fake or Real 음성의 포함 확률
대부분의 샘플이 $[-1, 1]$ 사이에 위치





데이터 전처리 > 문제 해석

데이터 분석

데이터 전처리

문제 해석

Noise Generator

Data Generation

사용 기법

모델 구조

모델 학습

인퍼런스 전략

결과

본 문제에서 Test 데이터는 다음 6가지 경우의 수를 가짐

경우의 수	Fake Count	Real Count	Actual Label
0	0	0	[0, 0]
1	1	0	[1, 0]
2	0	1	[0, 1]
3	1	1	[1, 1]
4	2	0	[1, 0]
5	0	2	[0, 1]

6가지 경우 모두의 데이터를
생성해내는 Custom
Dataset 제작



데이터 전처리 > Noise Generator

데이터 분석

데이터 전처리

문제 해석

Noise Generator

Data Generation

사용 기법

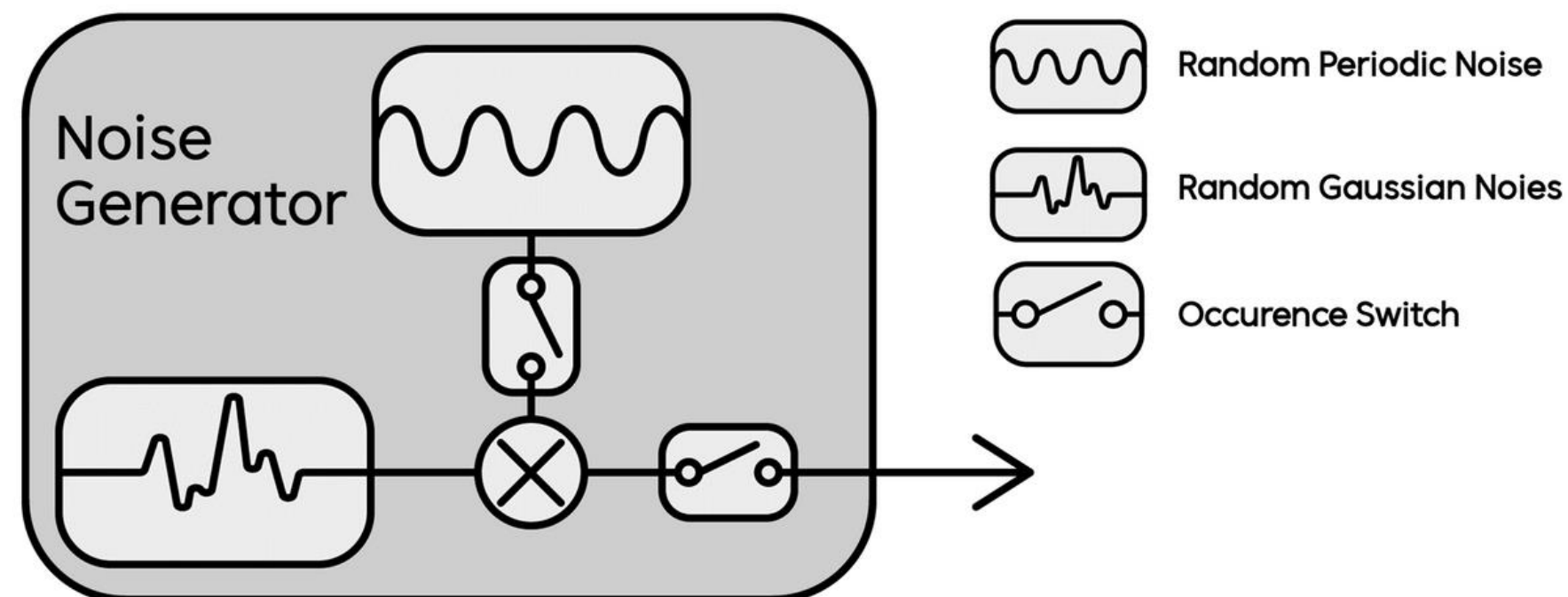
모델 구조

모델 학습

인퍼런스 전략

결과

다양한 데이터를 생성하기 위해
기존 데이터를 기반으로 데이터셋을 증강



Gaussian Noise와 Periodic Noise를
임의의 비율로 섞어 생성



데이터 전처리 > Data Generation

데이터 분석

데이터 전처리

문제 해석
Noise Generator
Data Generation

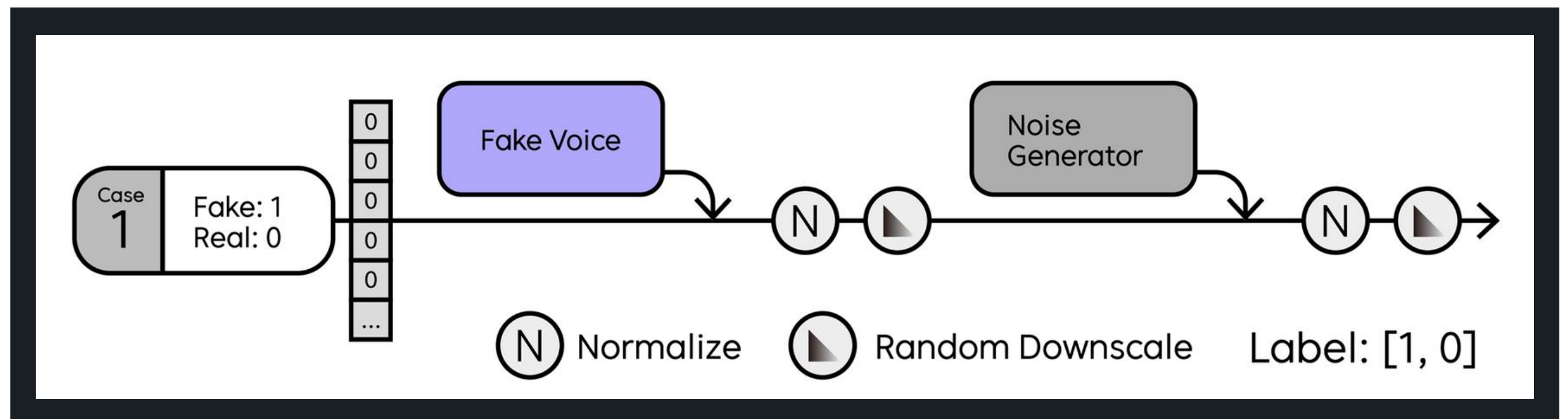
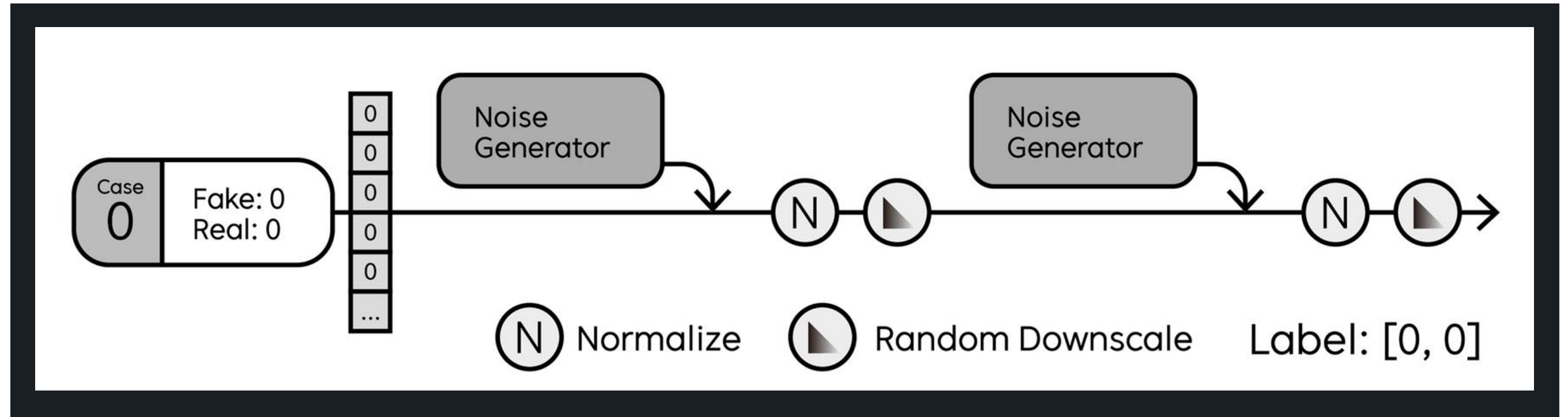
사용 기법

모델 구조

모델 학습

인퍼런스 전략

결과





데이터 전처리 > Data Generation

데이터 분석

데이터 전처리

문제 해석

Noise Generator

Data Generation

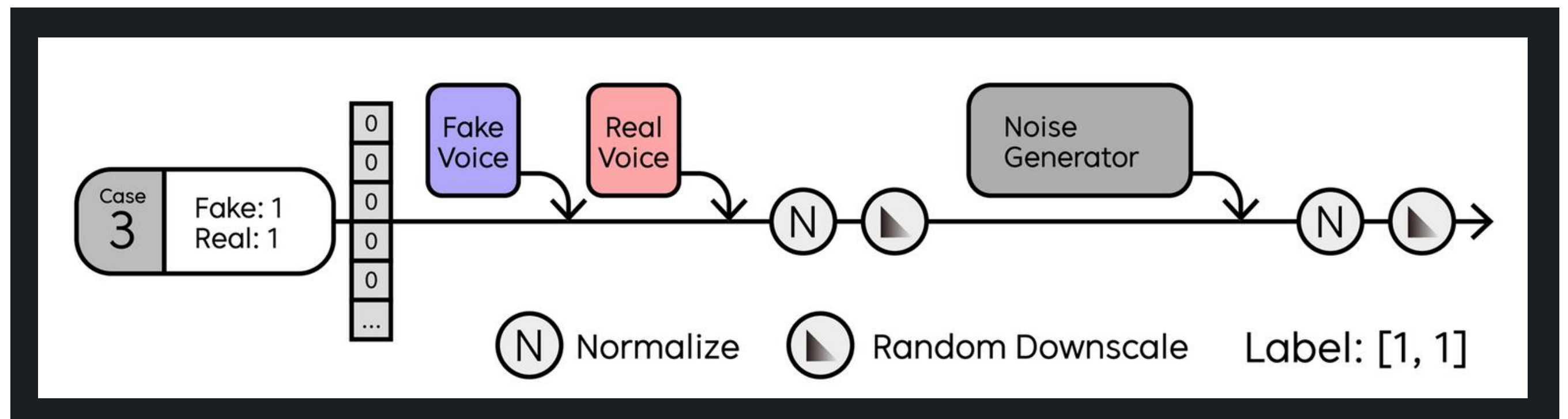
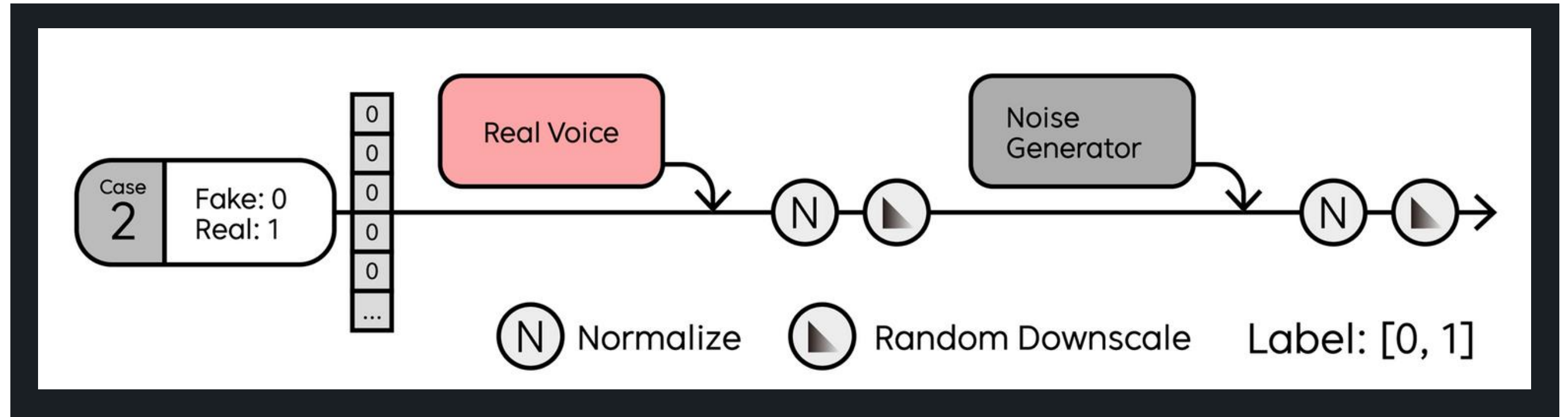
사용 기법

모델 구조

모델 학습

인퍼런스 전략

결과





데이터 전처리 > Data Generation

데이터 분석

데이터 전처리

문제 해석
Noise Generator
Data Generation

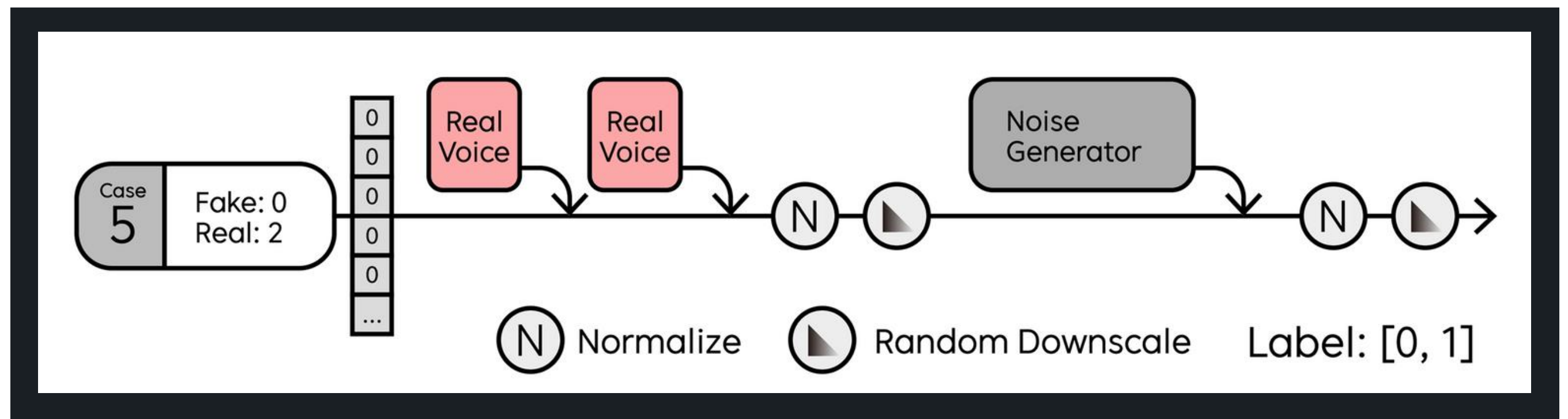
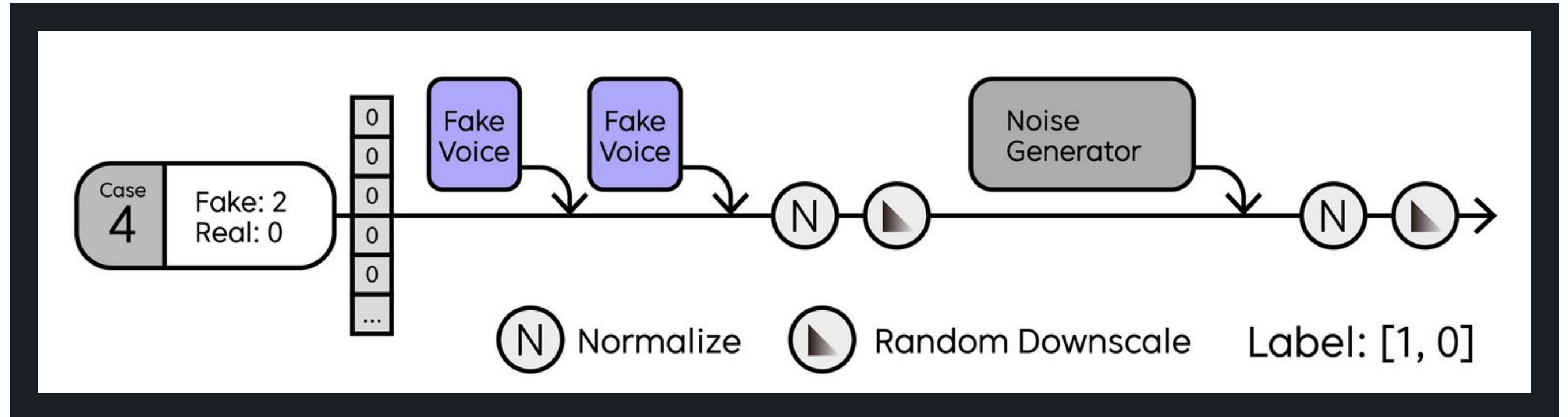
사용 기법

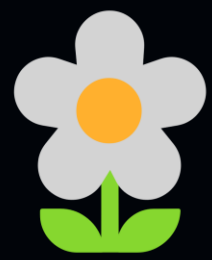
모델 구조

모델 학습

인퍼런스 전략

결과





사용 기법 > WeSpeaker Speaker Embedding Network

데이터 분석

데이터 전처리

사용 기법

WeSpeaker

Metric Learning

Siamese Net

모델 구조

모델 학습

인퍼런스 전략

결과

ResNet 기반 음성
임베딩 네트워크

Train 데이터, Augmented Train 데이터,
Test 데이터를 학습 이전에 모두 임베딩 벡터로 변환

Data preparation
UIO: raw / shard

On-the-fly Feature Preparation
Data slicing / augmentation / feature extraction

Model Export and deployment

Model training

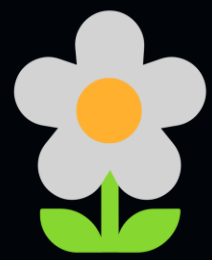
Speaker Embedding Extraction

Speaker Verification

Speaker Diarization

downstream

.....



사용 기법 > Metric Learning

데이터 분석

데이터 전처리

사용 기법

WeSpeaker

Metric Learning

Siamese Net

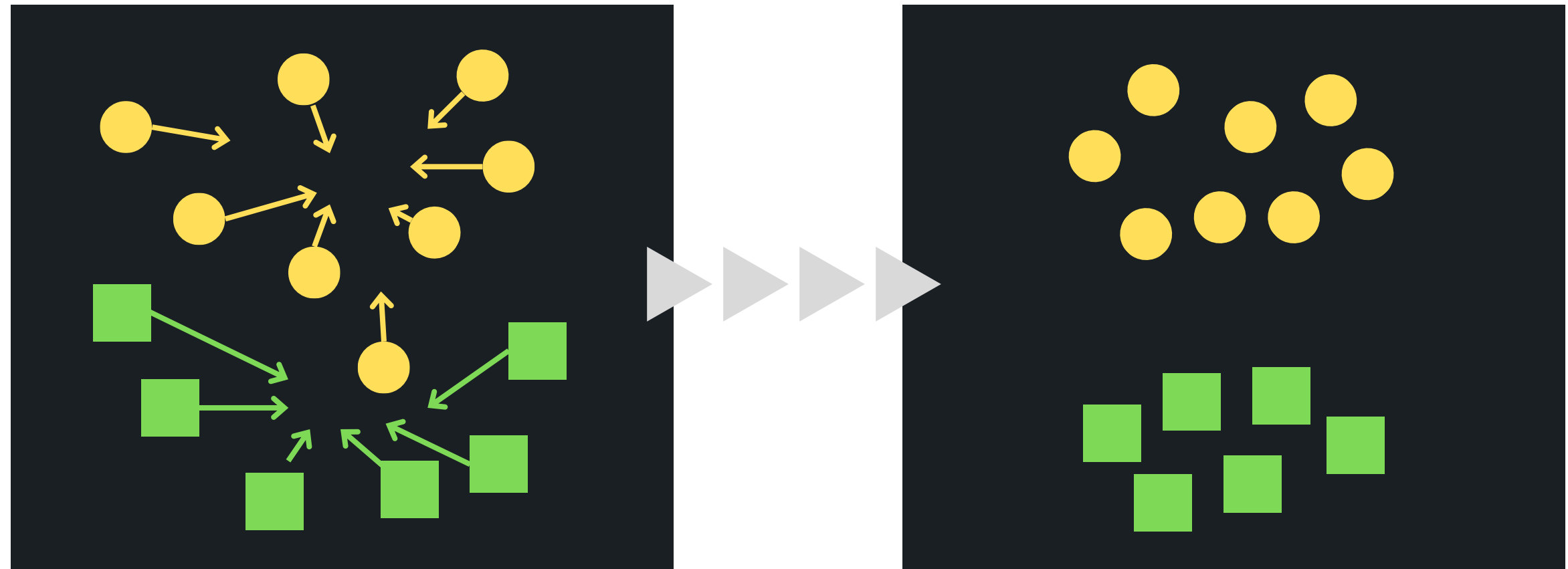
모델 구조

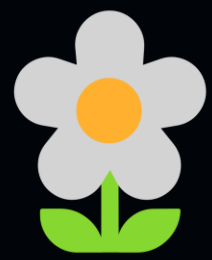
모델 학습

인퍼런스 전략

결과

데이터 샘플 간의 유사성을 측정하기 위해 거리 측정 방법을 학습하는 기계학습의 한 분야
비슷한 데이터 샘플들은 서로 가까이, 다른 데이터 샘플들은 서로 멀리 위치시키는 방법론





사용 기법 > Siamese Network with Cosine Similarity

데이터 분석

데이터 전처리

사용 기법

WeSpeaker

Metric Learning

Siamese Net

모델 구조

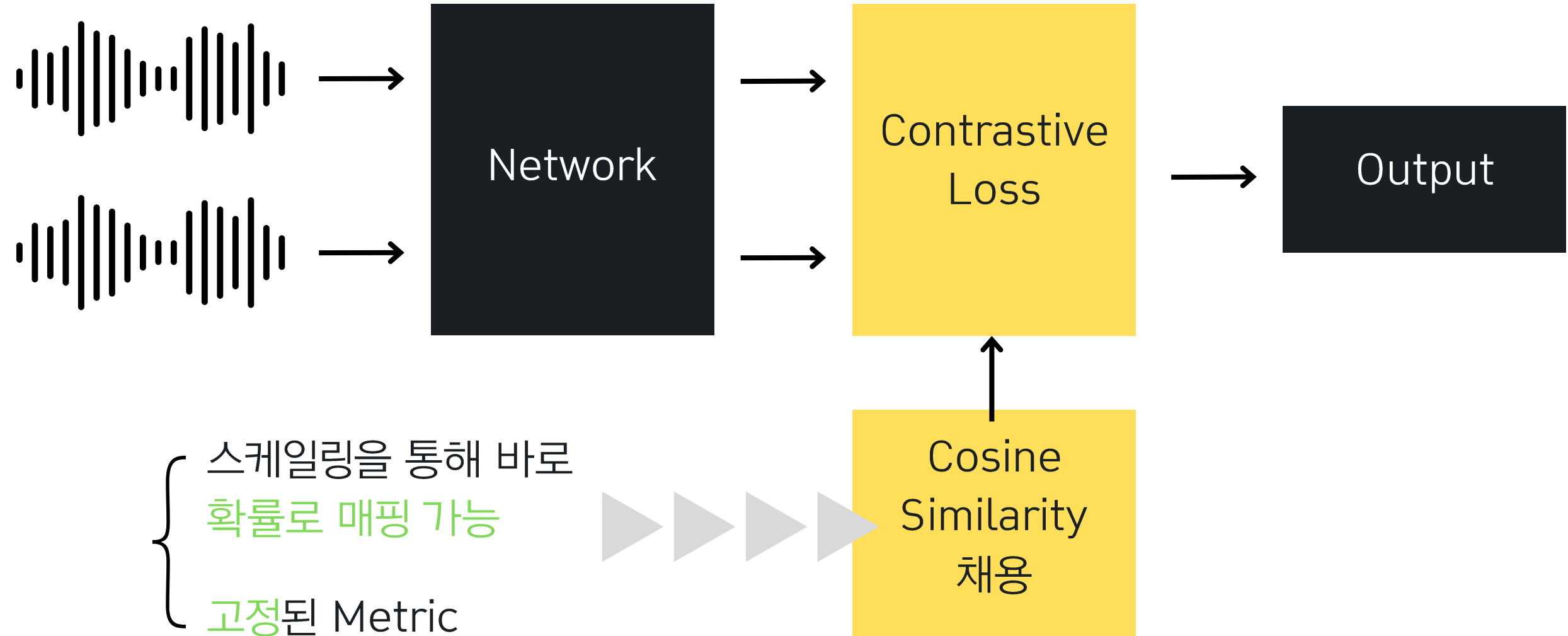
모델 학습

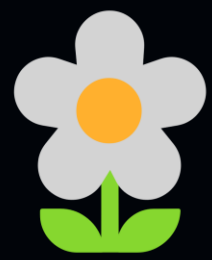
인퍼런스 전략

결과

Siamese
Network

두 데이터 간 유사도 혹은 거리를 예측하는 Metric Learning 방법론





모델 구조 > Embedding Transform Network

데이터 분석

데이터 전처리

사용 기법

모델 구조

Embedding Transform Network

Siamese Net

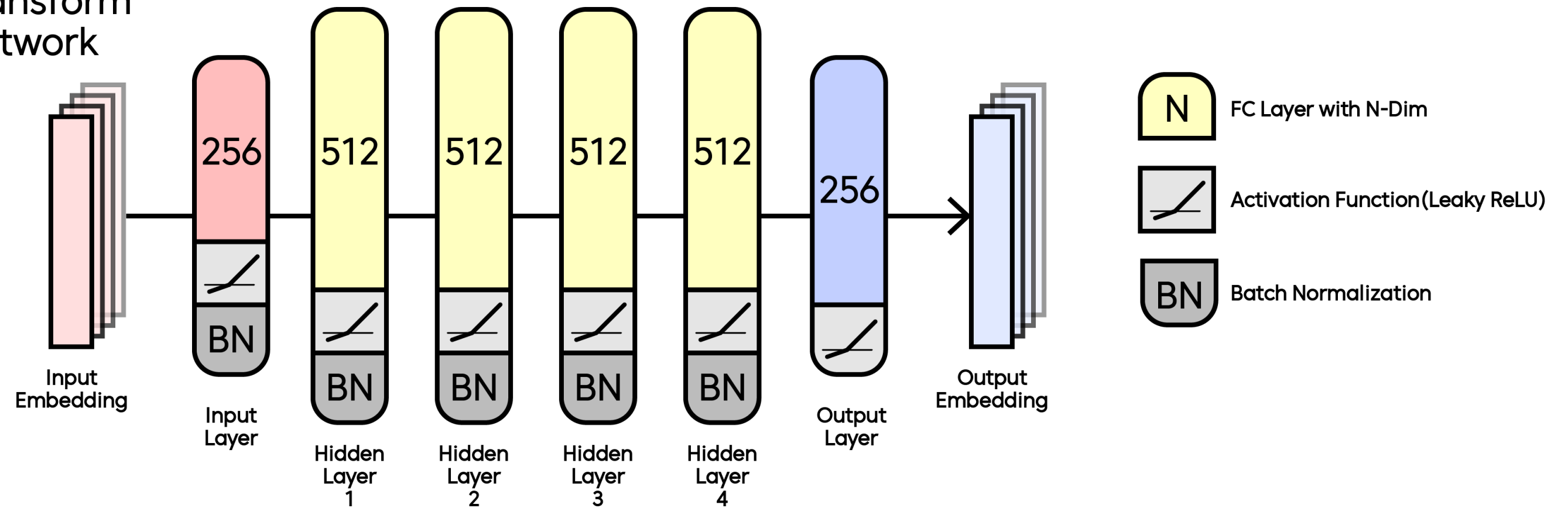
모델 학습

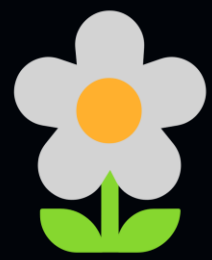
인퍼런스 전략

결과

입력 임베딩을 여러 FC Layer를 거쳐서 Transform시키는 Network

Embedding Transform Network

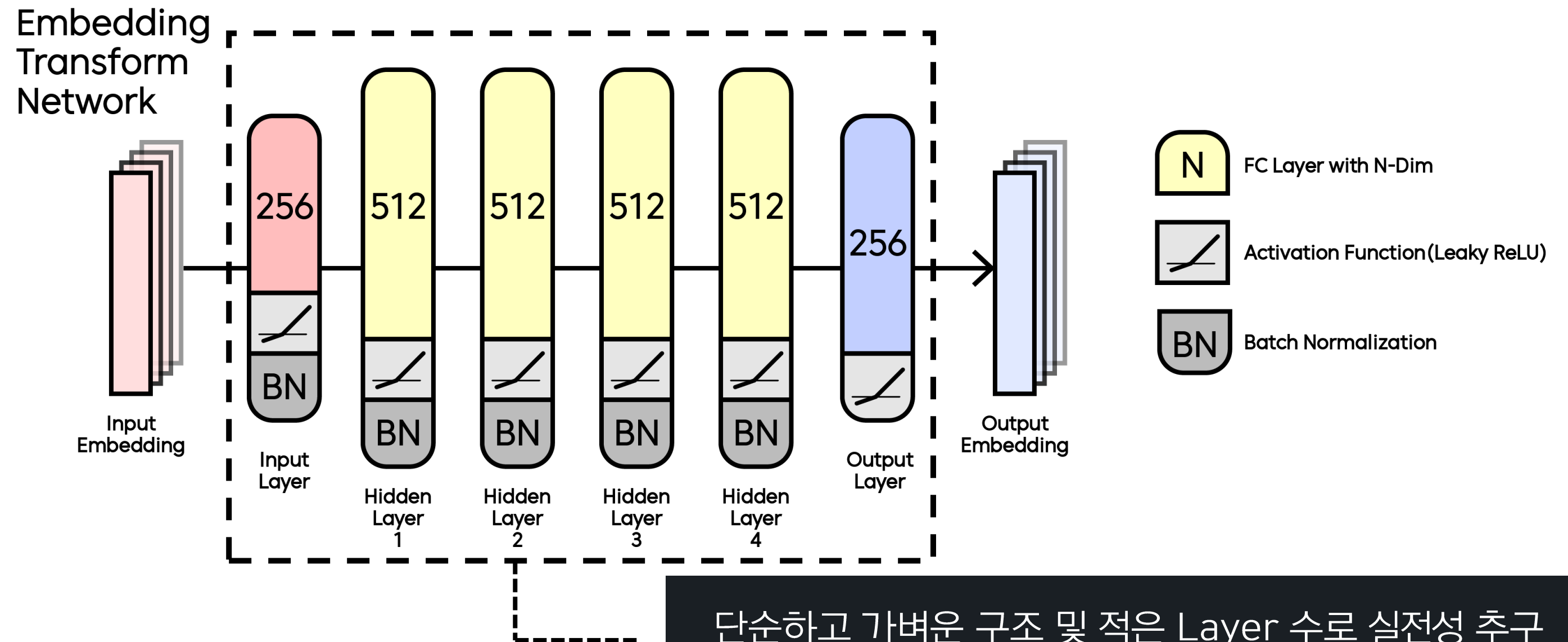


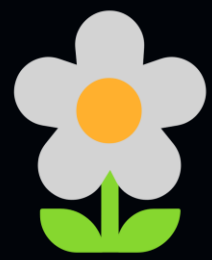


모델 구조 > Embedding Transform Network

- 데이터 분석
- 데이터 전처리
- 사용 기법
- 모델 구조
 - Embedding Transform Network
 - Siamese Net
- 모델 학습
- 인퍼런스 전략
- 결과

입력 임베딩을 여러 FC Layer를 거쳐서 Transform시키는 Network





모델 구조 > Siamese Network with Cosine Similarity

데이터 분석

데이터 전처리

사용 기법

모델 구조

Embedding Transform Network

Siamese Net

Example

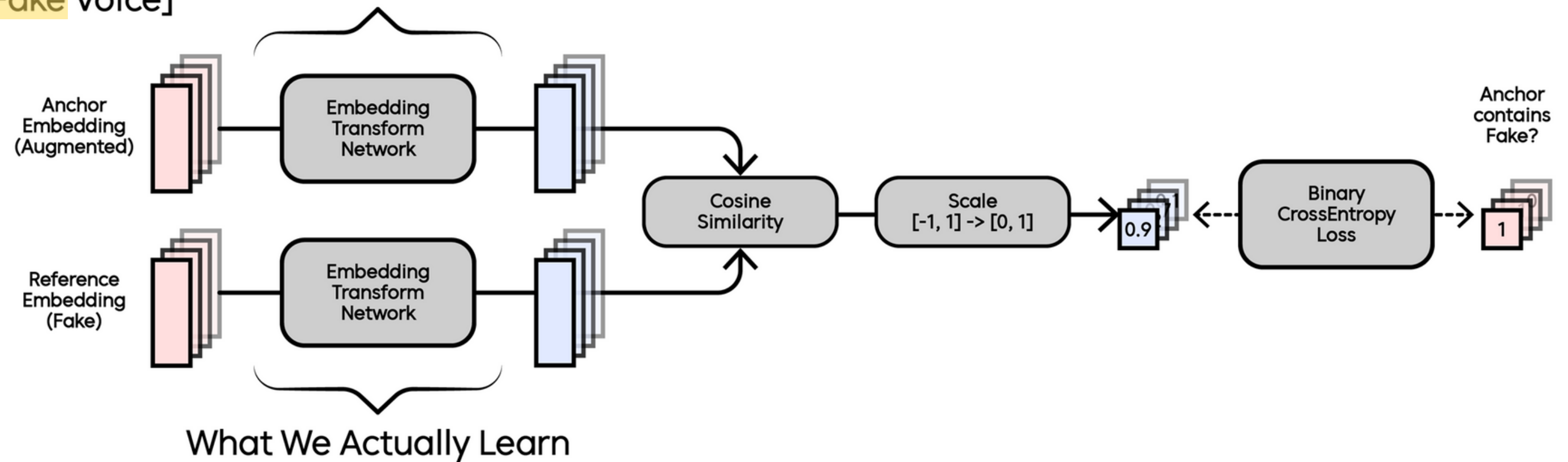
모델 학습

인퍼런스 전략

결과

Anchor와 Reference를 입력받아 두 벡터간의 Cosine Similarity를 계산
이를 확률로 해석하여 가짜 / 진짜 음성 탐지를 수행

Siamese Network
with Cosine Similarity
[Fake Voice]





모델 구조 > Siamese Network with Cosine Similarity

데이터 분석

데이터 전처리

사용 기법

모델 구조

Embedding Transform Network

Siamese Net

Example

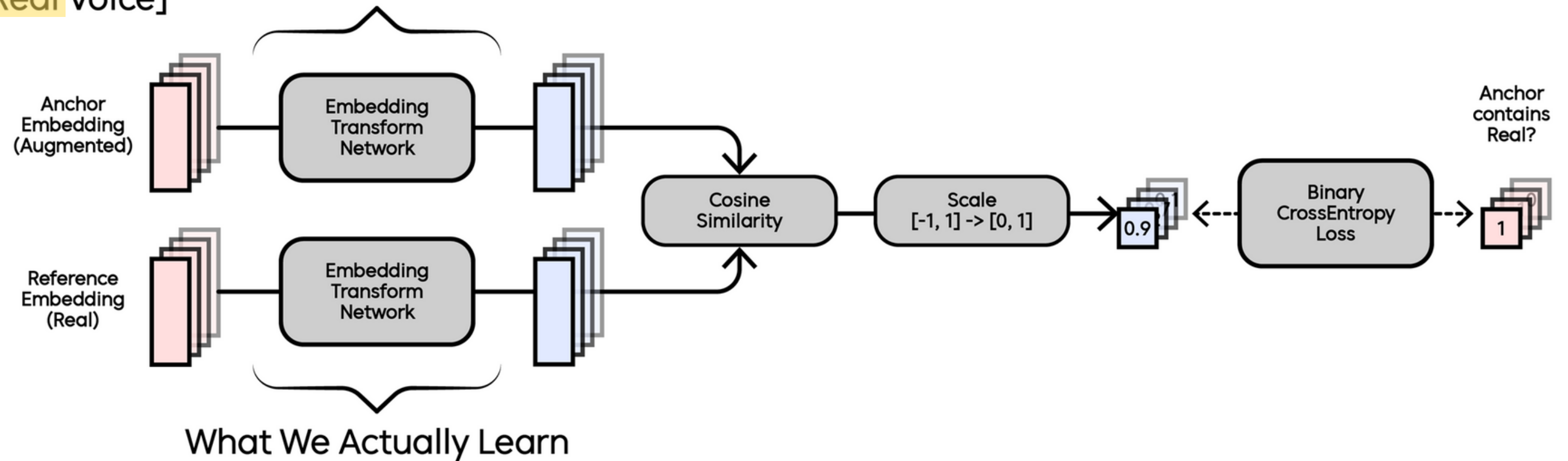
모델 학습

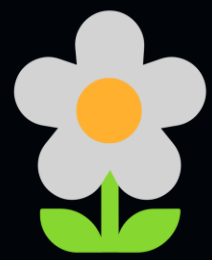
인퍼런스 전략

결과

Anchor와 Reference를 입력받아 두 벡터간의 Cosine Similarity를 계산
이를 확률로 해석하여 가짜 / 진짜 음성 탐지를 수행

Siamese Network
with Cosine Similarity
[Real Voice]





모델 구조 > Example

데이터 분석

데이터 전처리

사용 기법

모델 구조

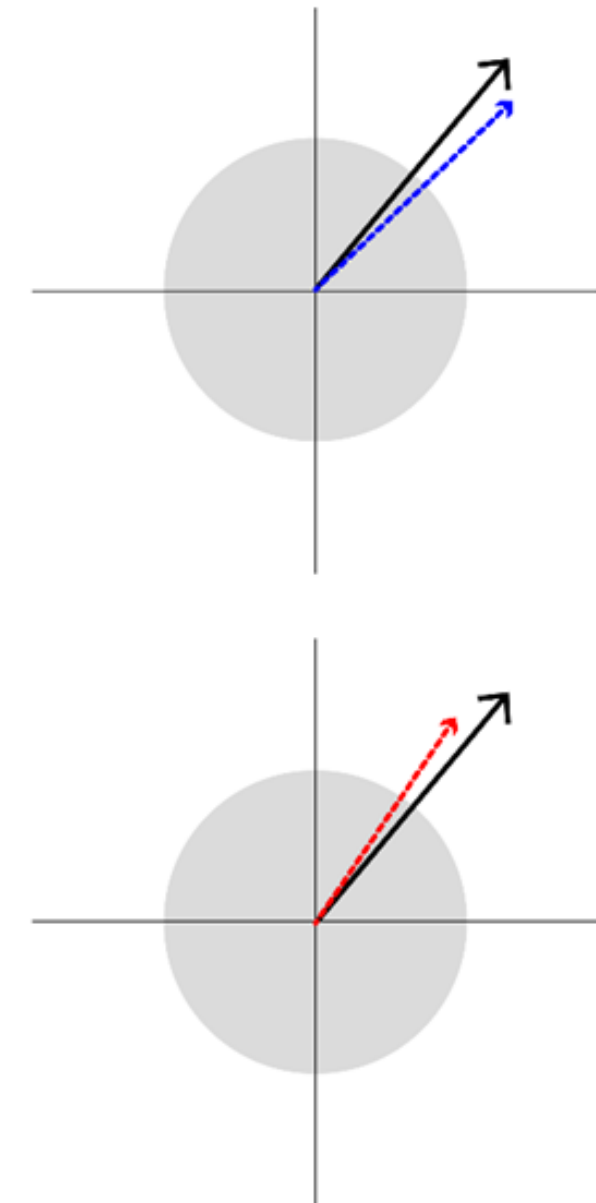
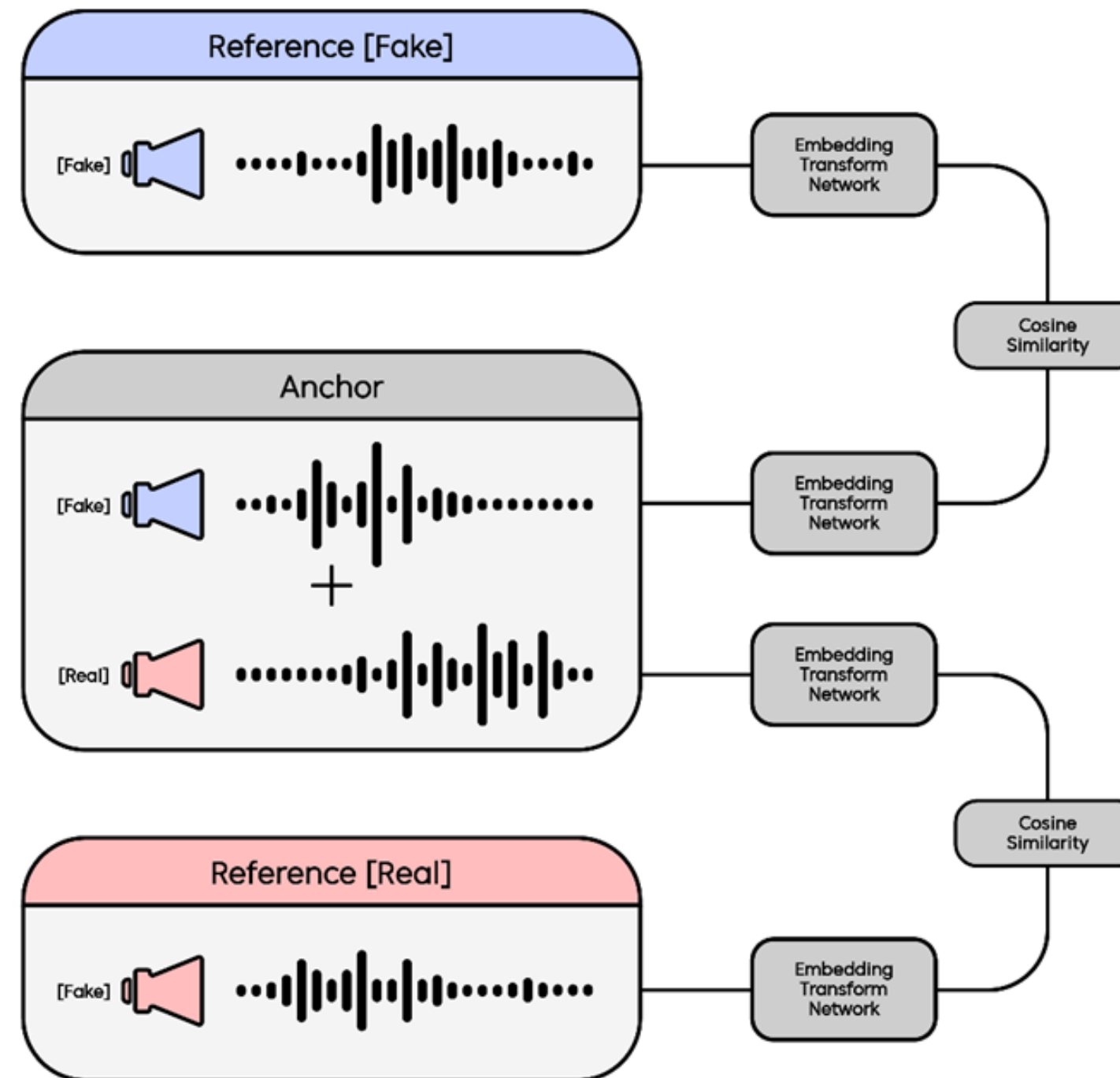
Embedding Transform Network
Siamese Net
Example

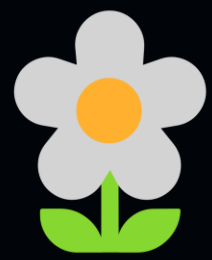
모델 학습

인퍼런스 전략

결과

[Fake, Real]





모델 구조 > Example

데이터 분석

데이터 전처리

사용 기법

모델 구조

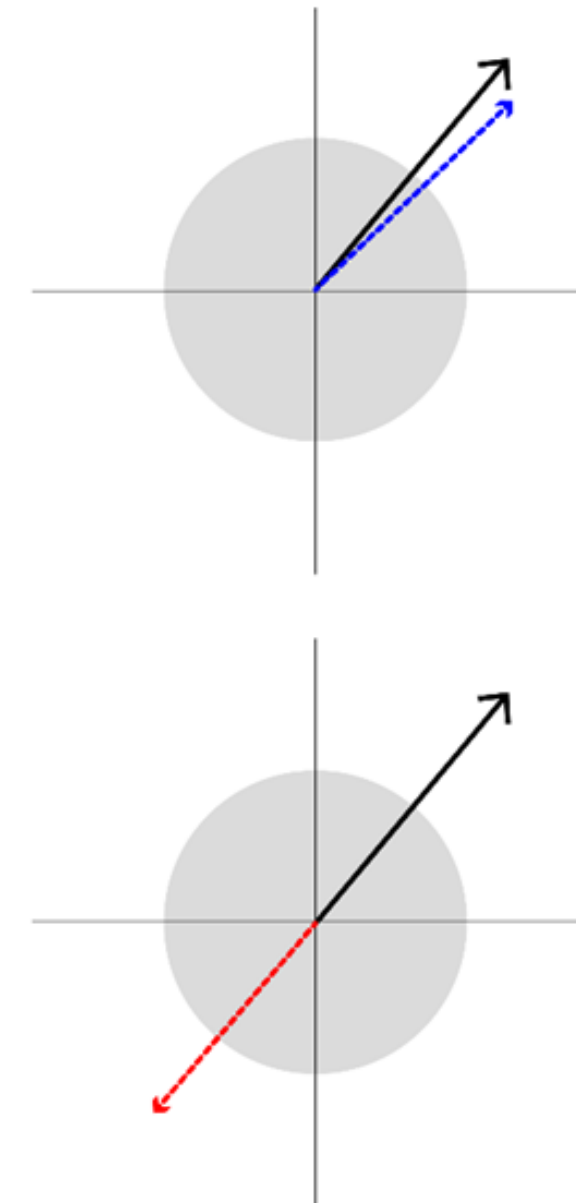
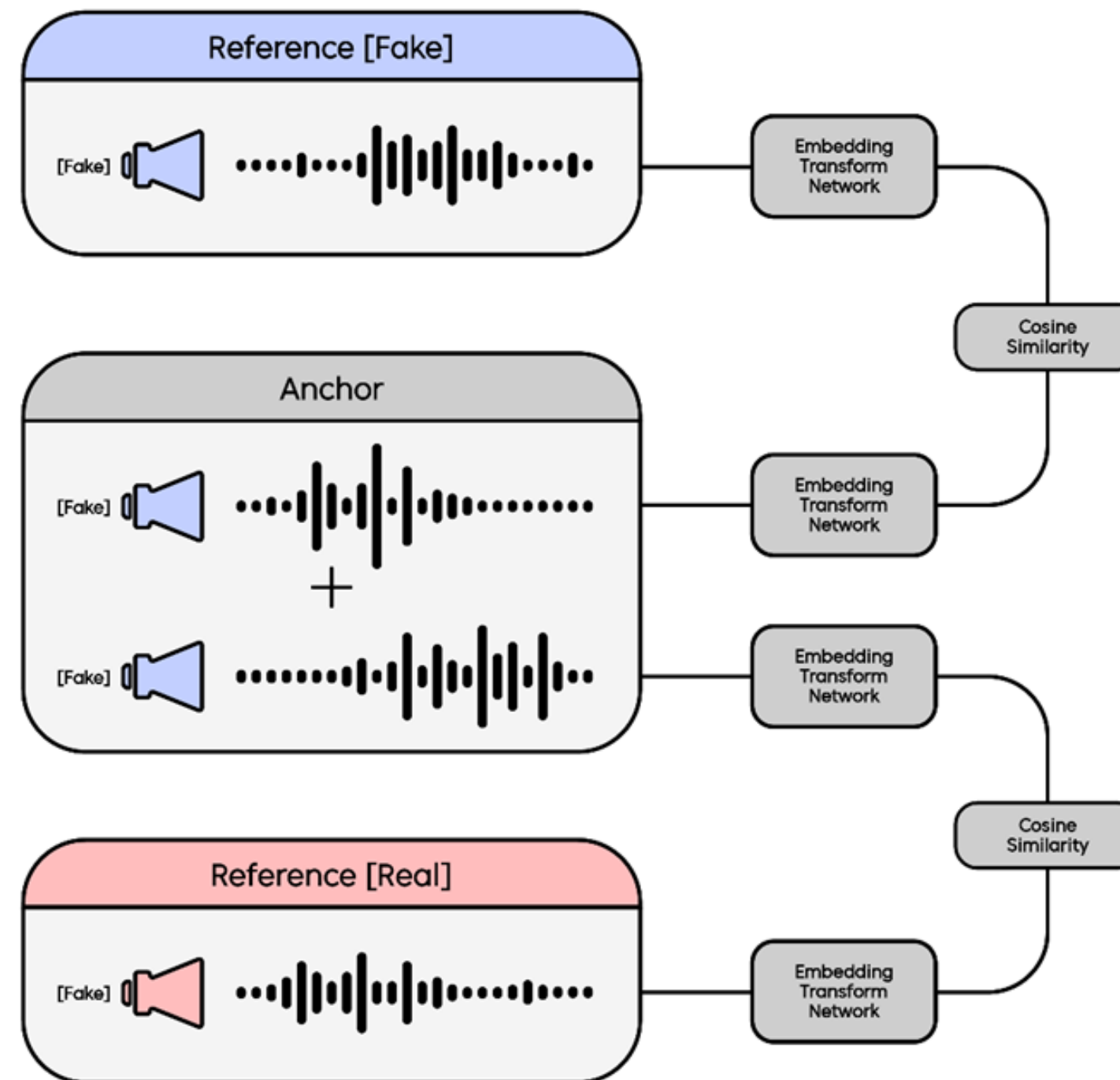
Embedding Transform Network
Siamese Net
Example

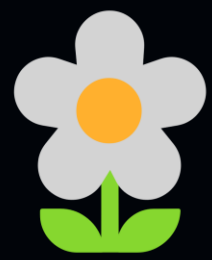
모델 학습

인퍼런스 전략

결과

[Fake, Fake]





모델 구조 > Example

데이터 분석

데이터 전처리

사용 기법

모델 구조

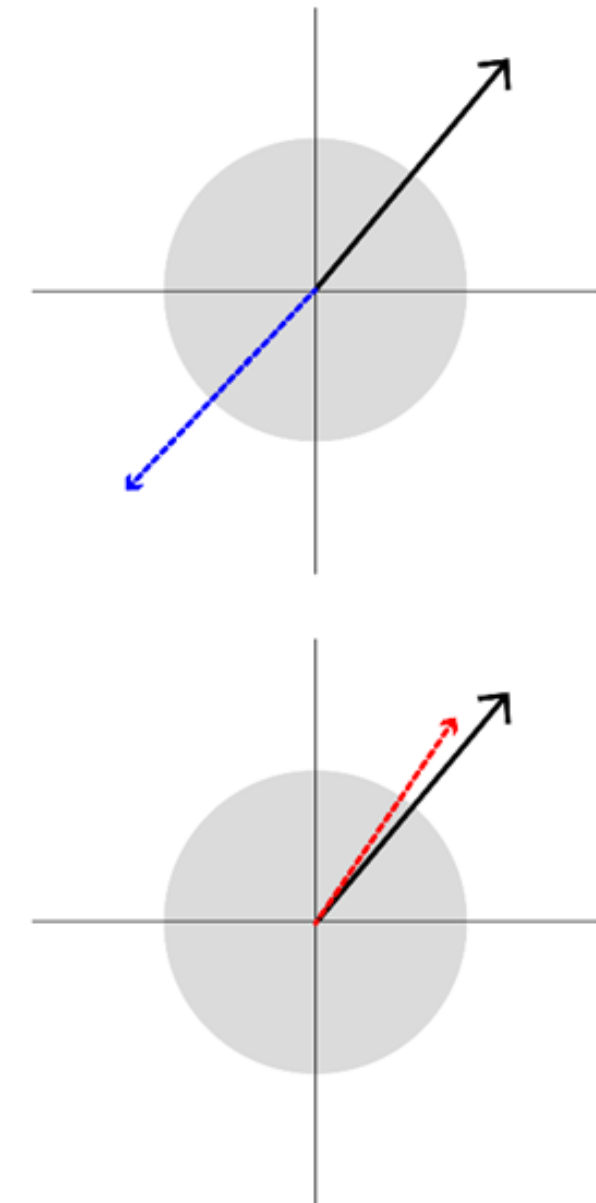
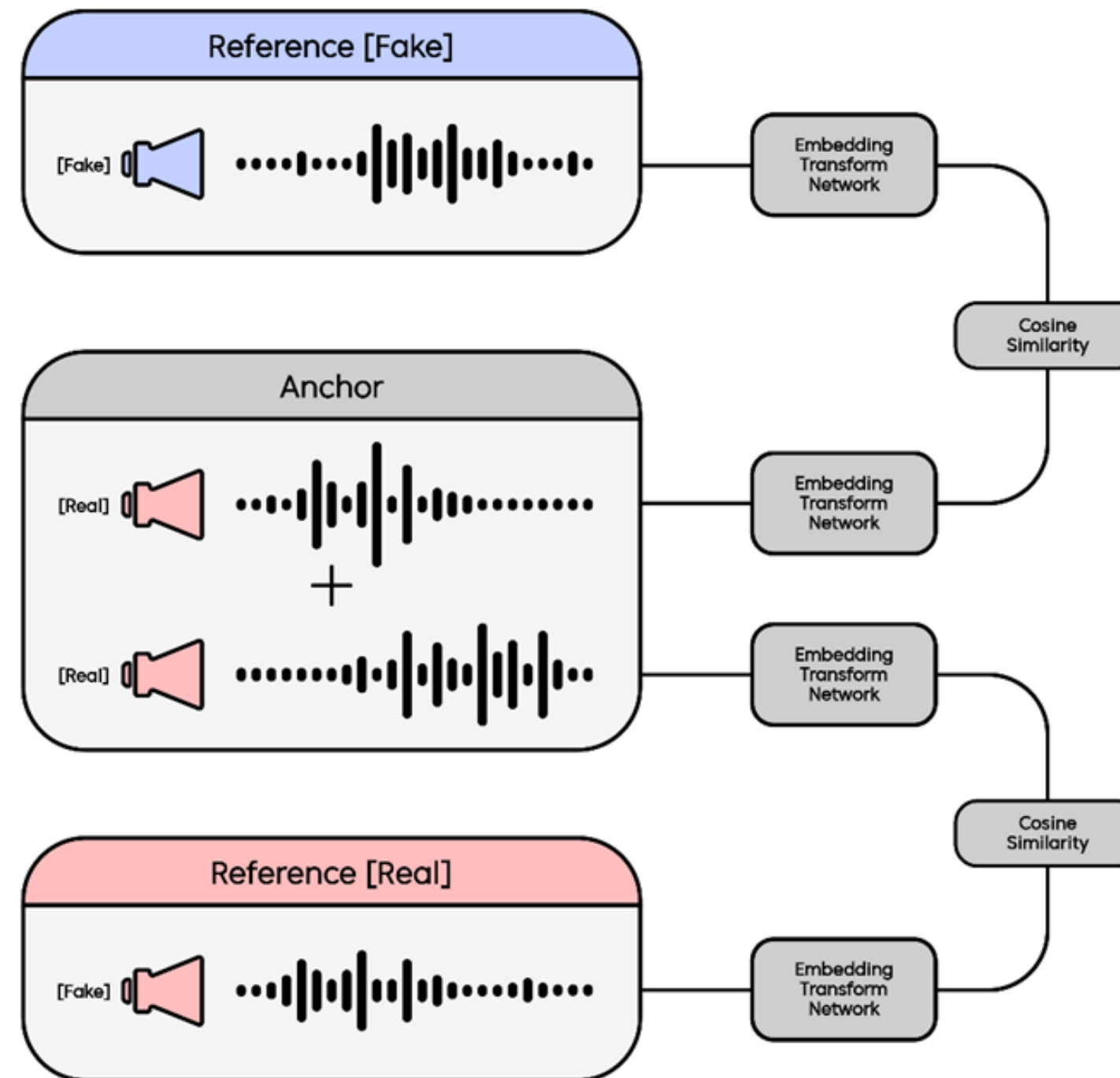
Embedding Transform Network
Siamese Net
Example

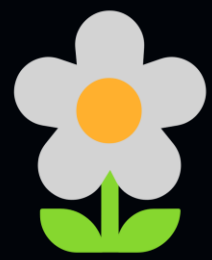
모델 학습

인퍼런스 전략

결과

[Real, Real]





모델 구조 > Example

데이터 분석

데이터 전처리

사용 기법

모델 구조

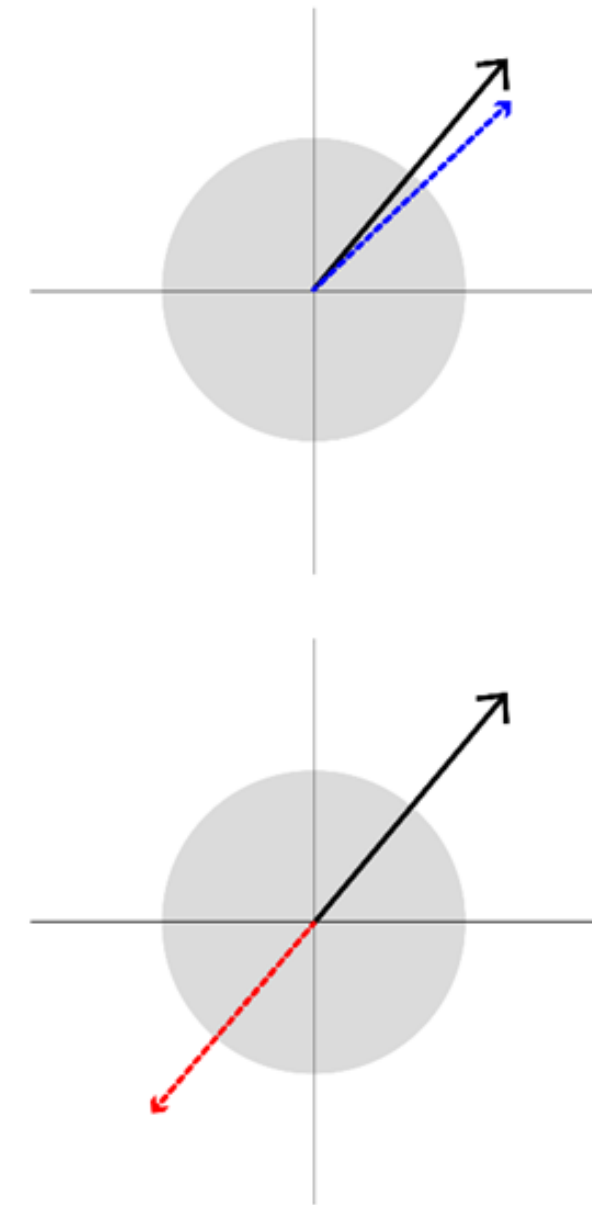
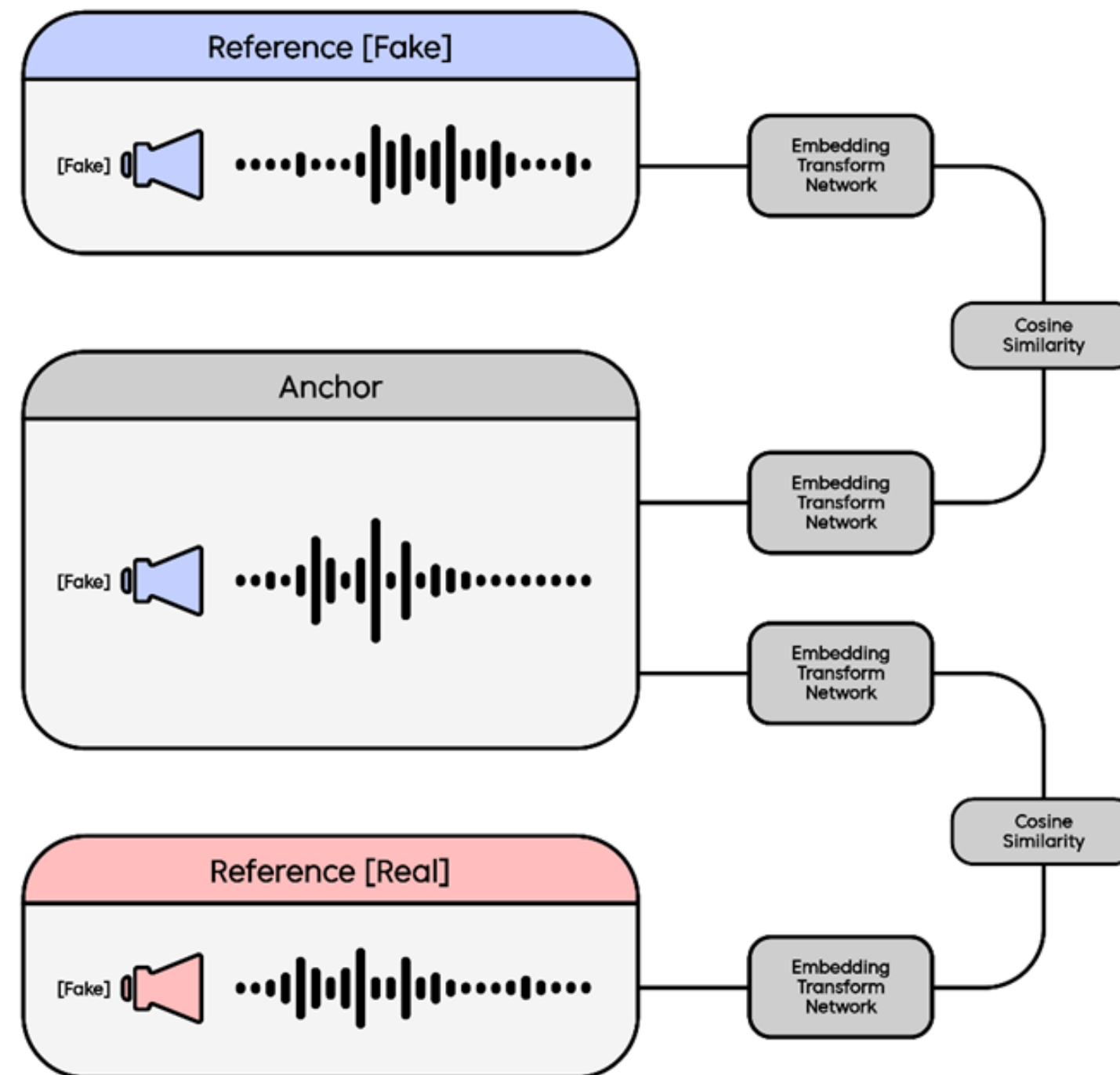
Embedding Transform Network
Siamese Net
Example

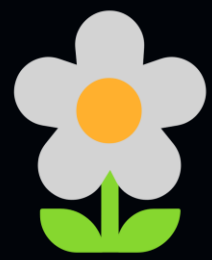
모델 학습

인퍼런스 전략

결과

[Fake]





모델 구조 > Example

데이터 분석

데이터 전처리

사용 기법

모델 구조

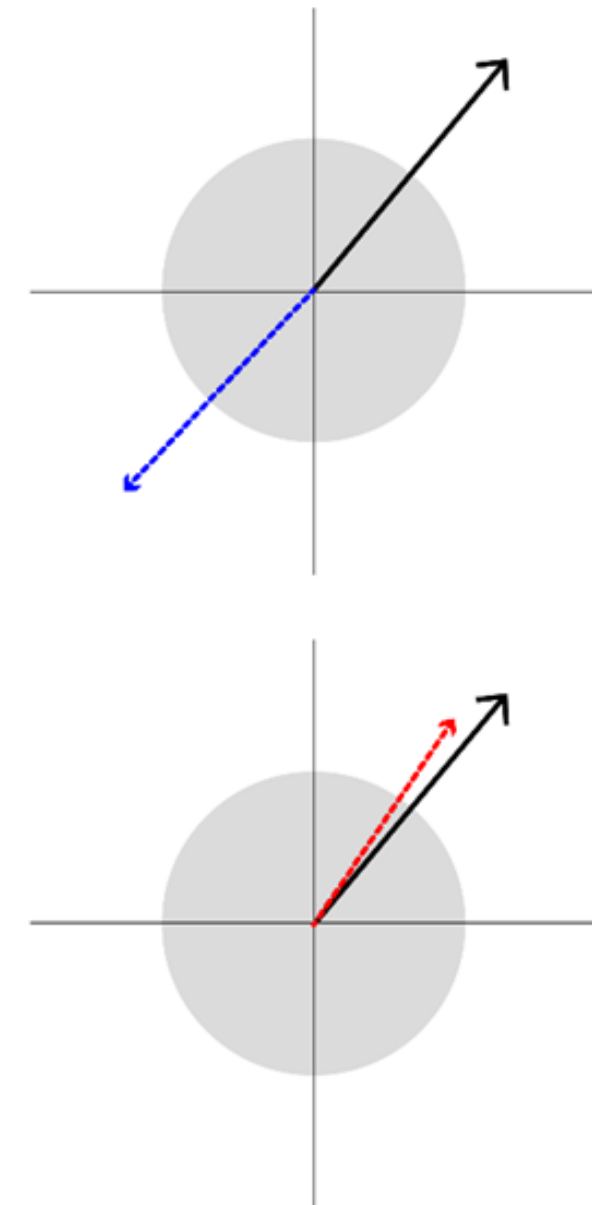
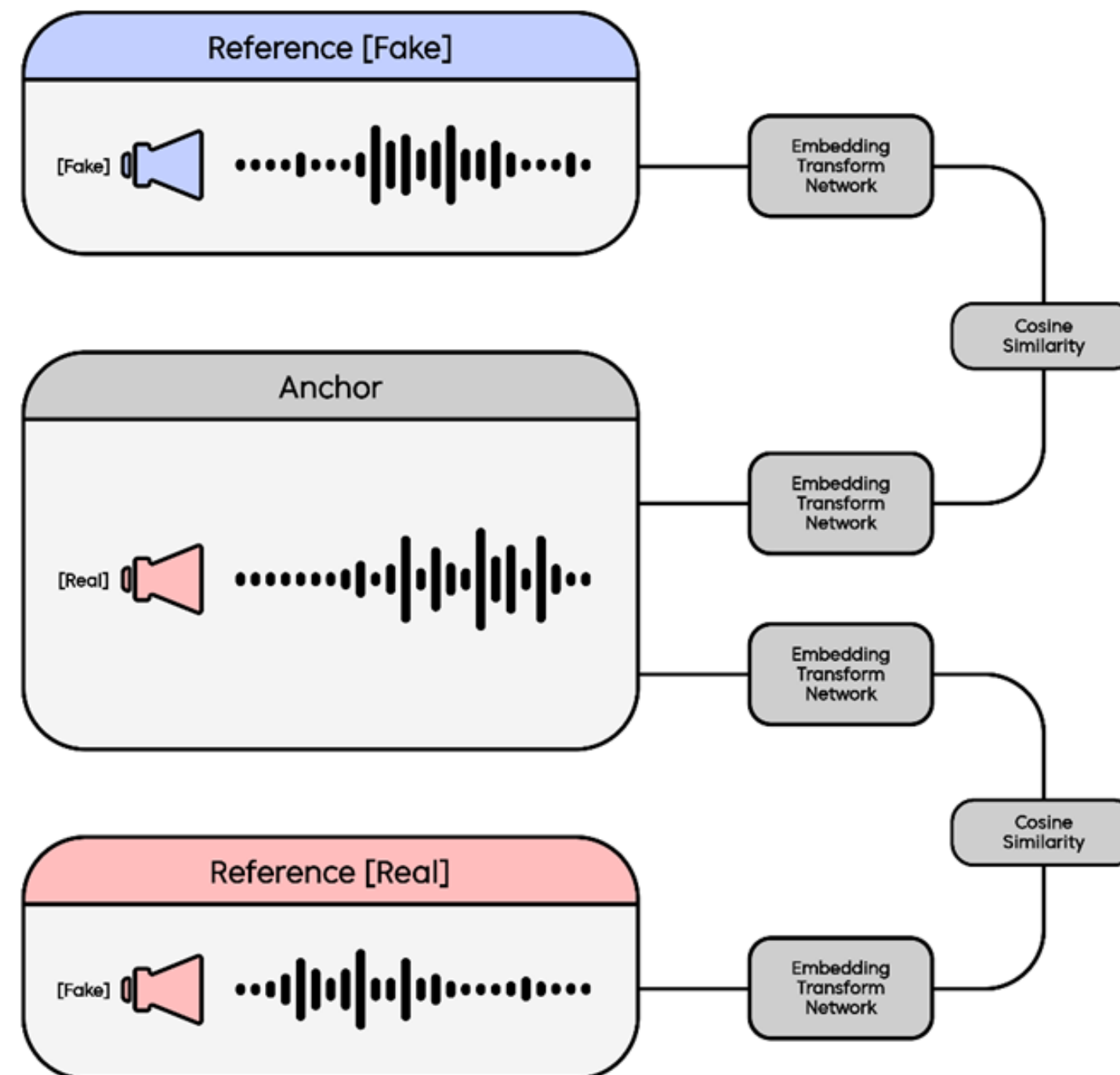
Embedding Transform Network
Siamese Net
Example

모델 학습

인퍼런스 전략

결과

[Real]





모델 구조 > Example

데이터 분석

데이터 전처리

사용 기법

모델 구조

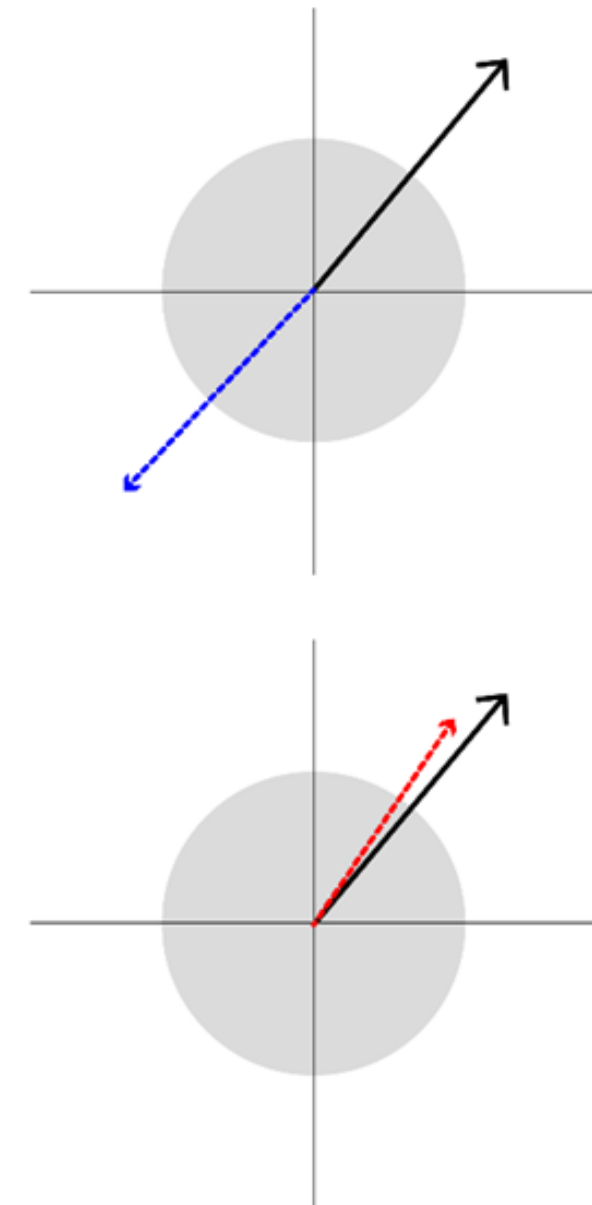
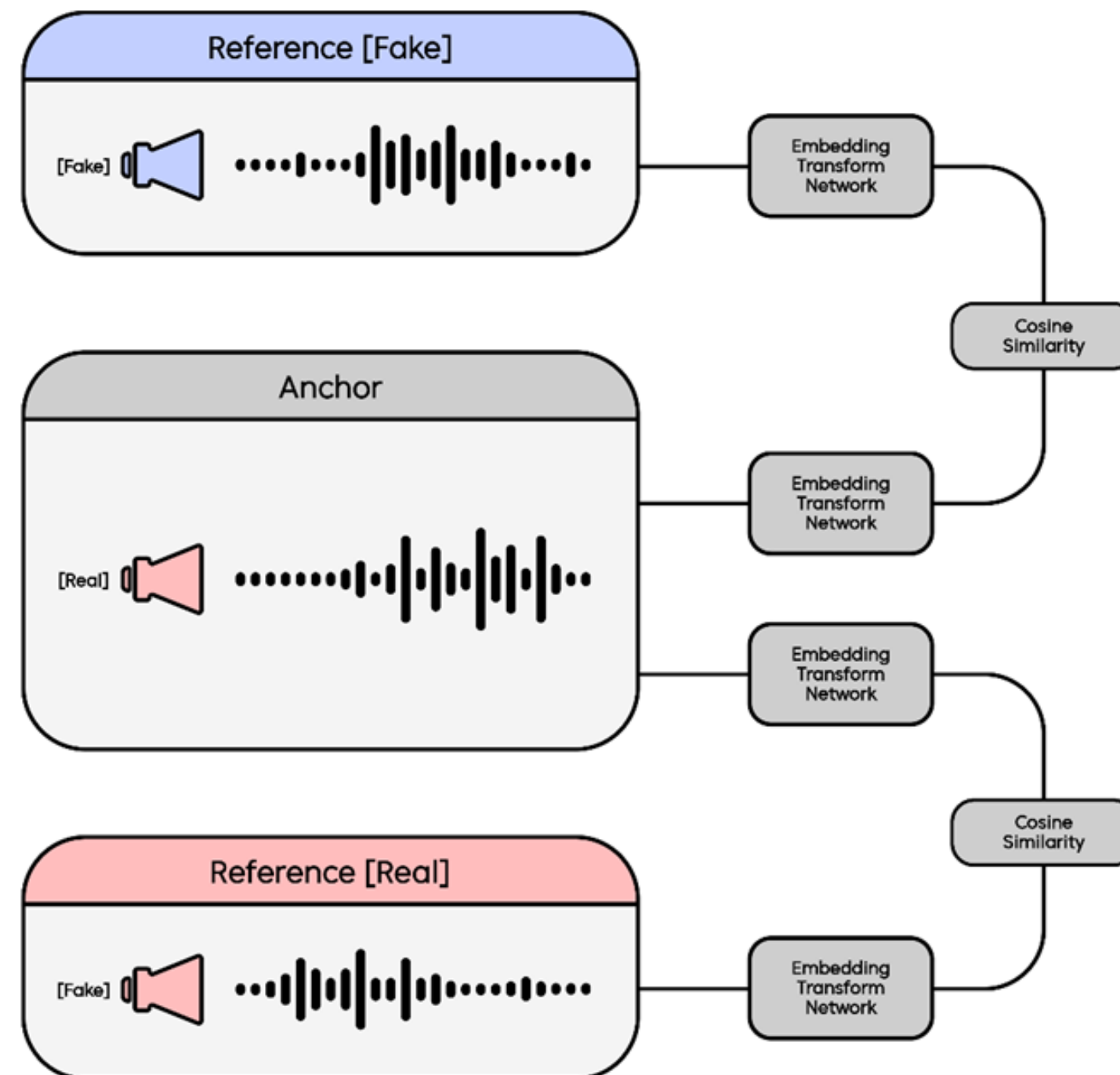
Embedding Transform Network
Siamese Net
Example

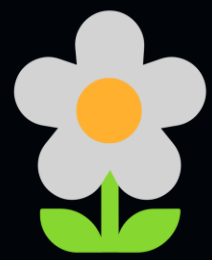
모델 학습

인퍼런스 전략

결과

[Real]





모델 구조 > Example

데이터 분석

데이터 전처리

사용 기법

모델 구조

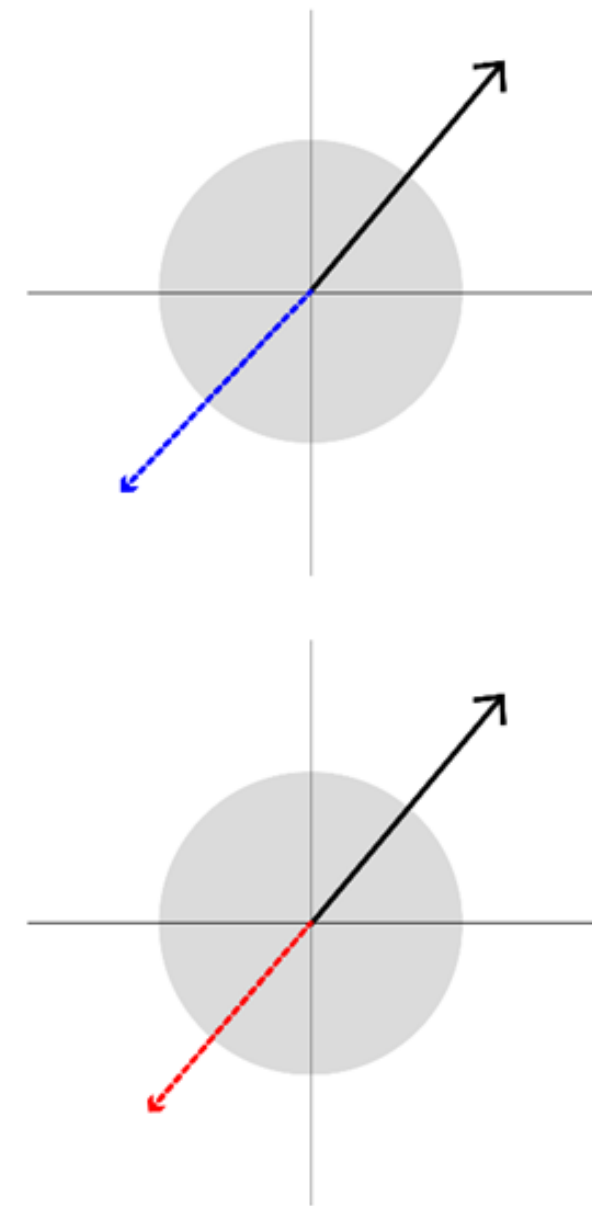
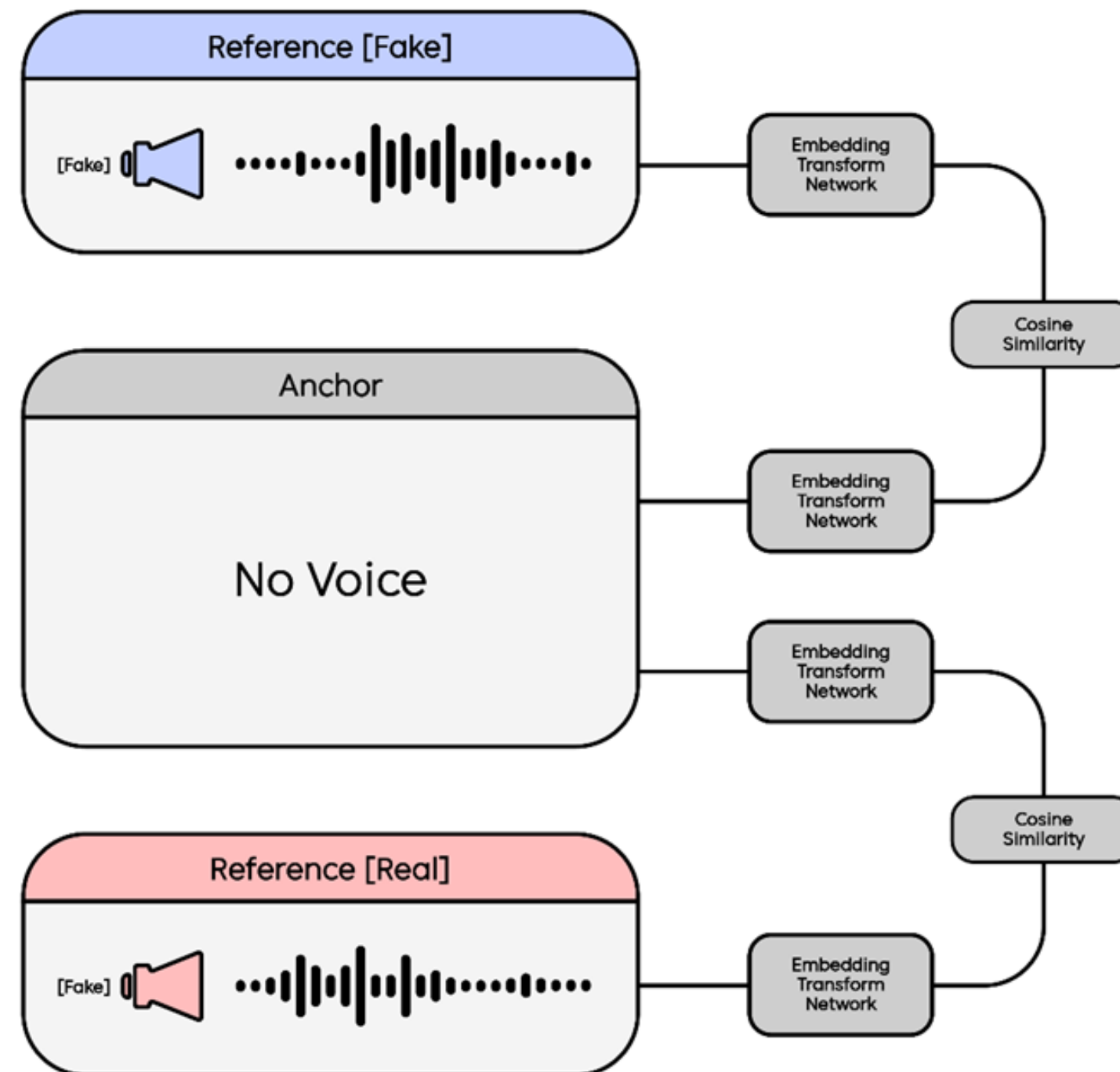
Embedding Transform Network
Siamese Net
Example

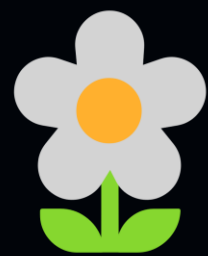
모델 학습

인퍼런스 전략

결과

[No Voice]





모델 학습 > 환경 및 하이퍼파라미터

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

환경 및
하이퍼파라미터

데이터 Split

학습 그래프,
예측 성능
모델 검증

인퍼런스 전략

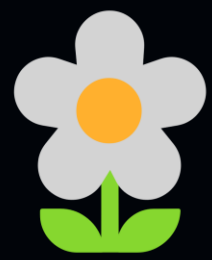
결과

학습 환경

Component	Version
Python	3.9.13
Torch	2.0.1+cu117
CUDA	11.7
cuDNN	8.5.0
GPU	NVIDIA RTX 4060 Mobile

학습 하이퍼파라미터

Component	Version
Total Epochs	50
Learning Rate	3e-4
Batch Size	64
Criterion	Binary CrossEntropy Loss
Optimizer	Adam



모델 학습 > 데이터 Split

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

환경 및
하이퍼파라미터

데이터 Split

학습 그래프,

예측 성능

모델 검증

인퍼런스 전략

결과

Anchor(Augmented)

Train Data(60%)

Valid Data(20%)

Test Data(20%)

Reference(Fake)

Train Data(60%)

Valid Data(20%)

Test Data(20%)

Reference(Real)

Train Data(60%)

Valid Data(20%)

Test Data(20%)



모델 학습 > 학습 그래프, 예측 성능

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

환경 및
하이퍼파라미터

데이터 Split

학습 그래프,

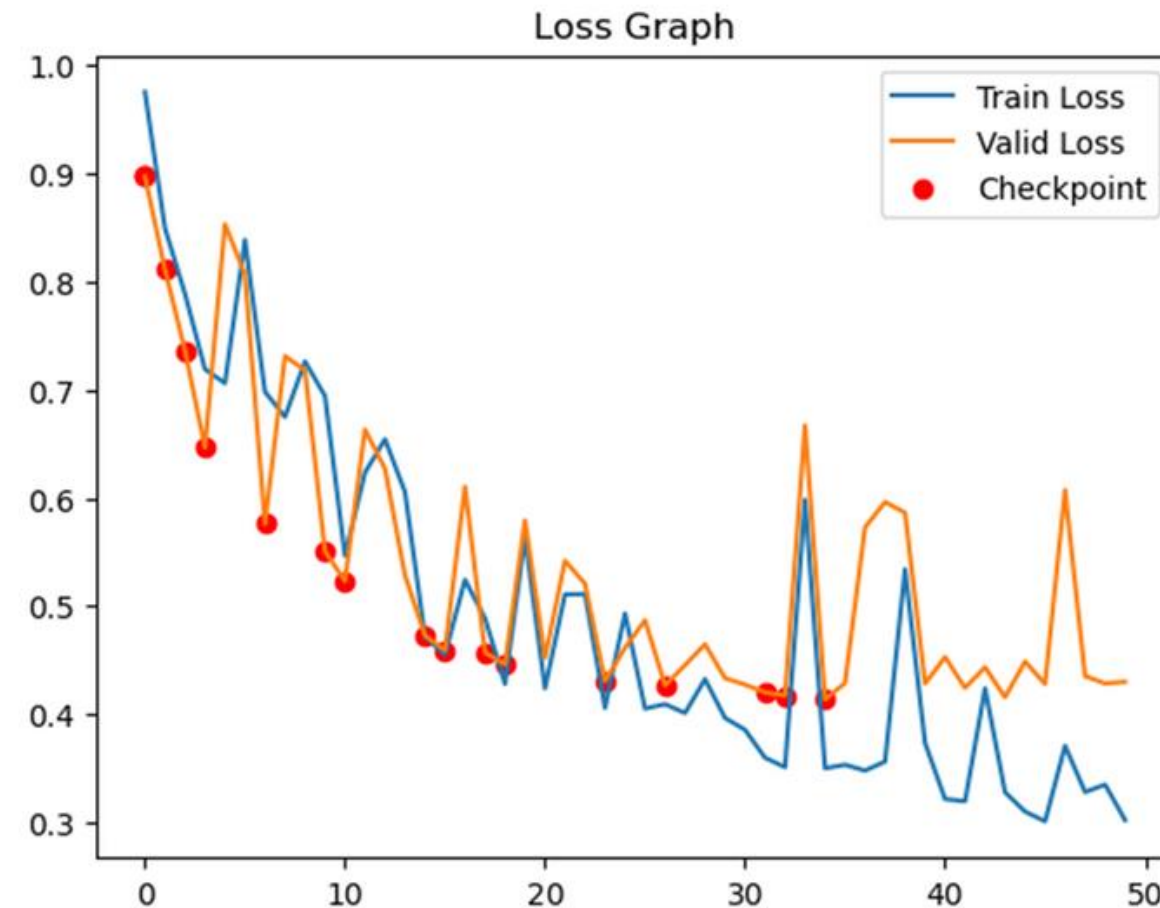
예측 성능

모델 검증

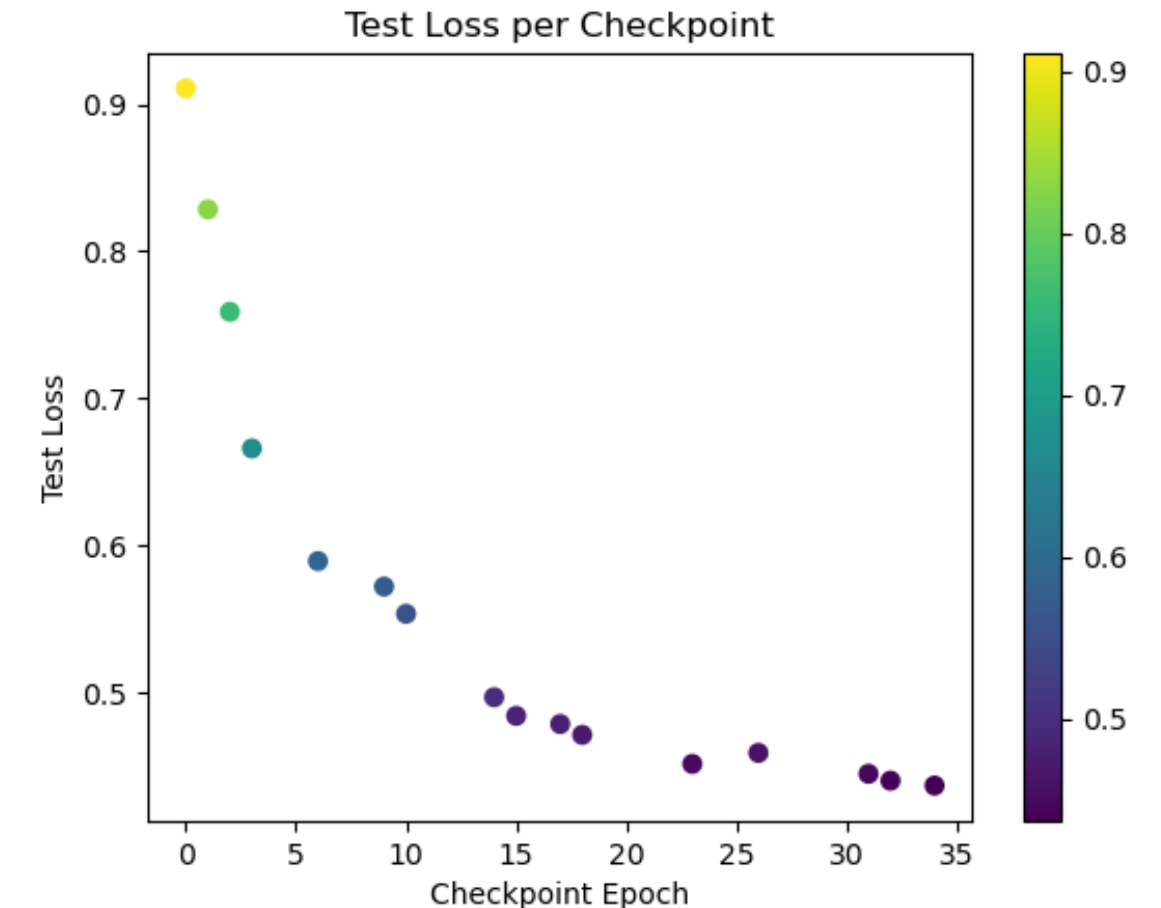
인퍼런스 전략

결과

학습 그래프



Checkpoint별
Test Split 예측 성능





모델 학습 > 모델 검증

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

환경 및
하이퍼파라미터

데이터 Split

학습 그래프,

예측 성능

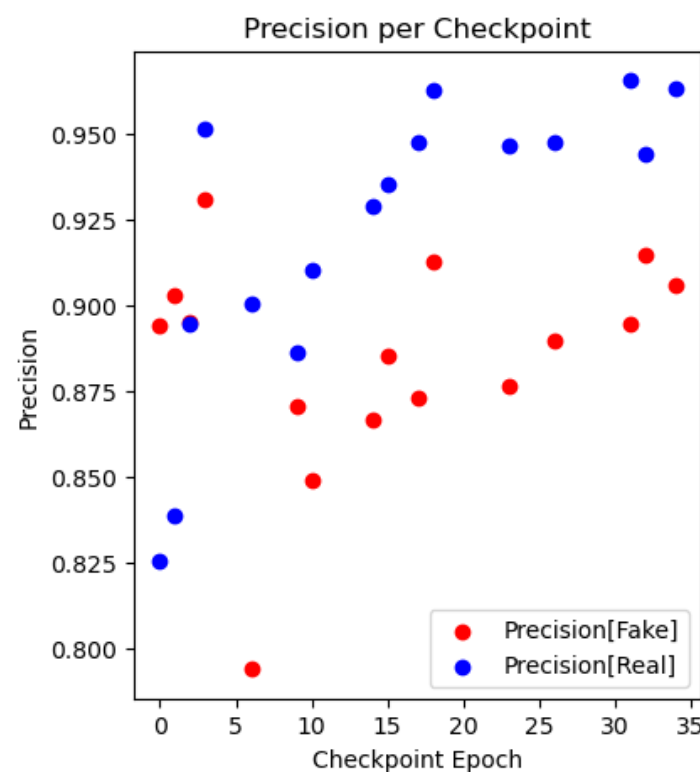
모델 검증

인퍼런스 전략

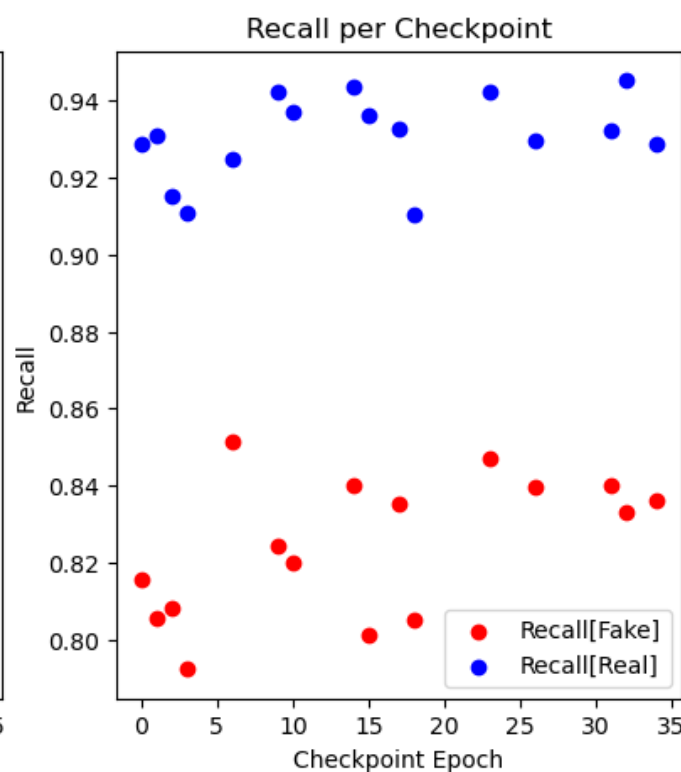
결과

체크포인트별 성능 지표

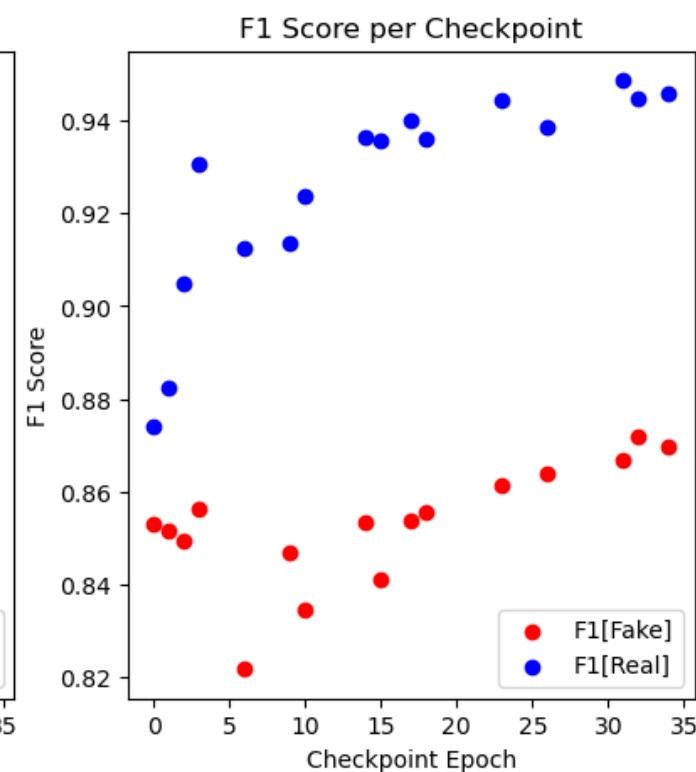
Precision



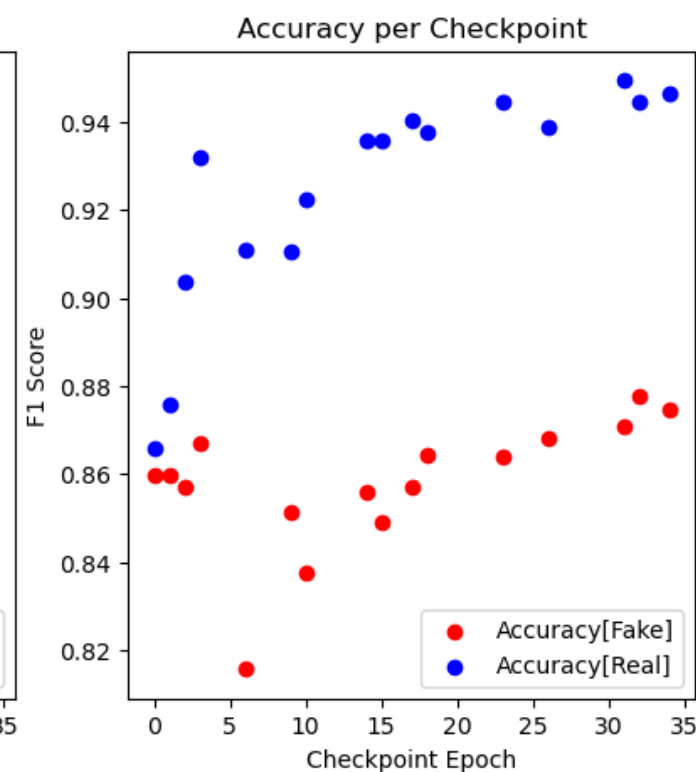
Recall

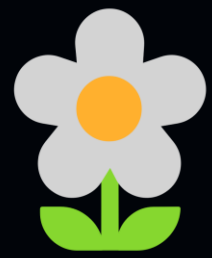


F1-Score



Accuracy





모델 학습 > 모델 검증

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

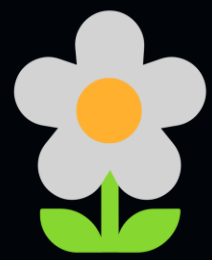
환경 및
하이퍼파라미터
데이터 Split
학습 그래프,
예측 성능
모델 검증

인퍼런스 전략

결과

Epoch 34

Metric	Fake	Real
Precision	0.9057	0.9634
Recall	0.8360	0.9286
F1 Score	0.8695	0.9457
Accuracy	0.8748	0.9465



인퍼런스 전략 > Multi-Reference Similarity

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

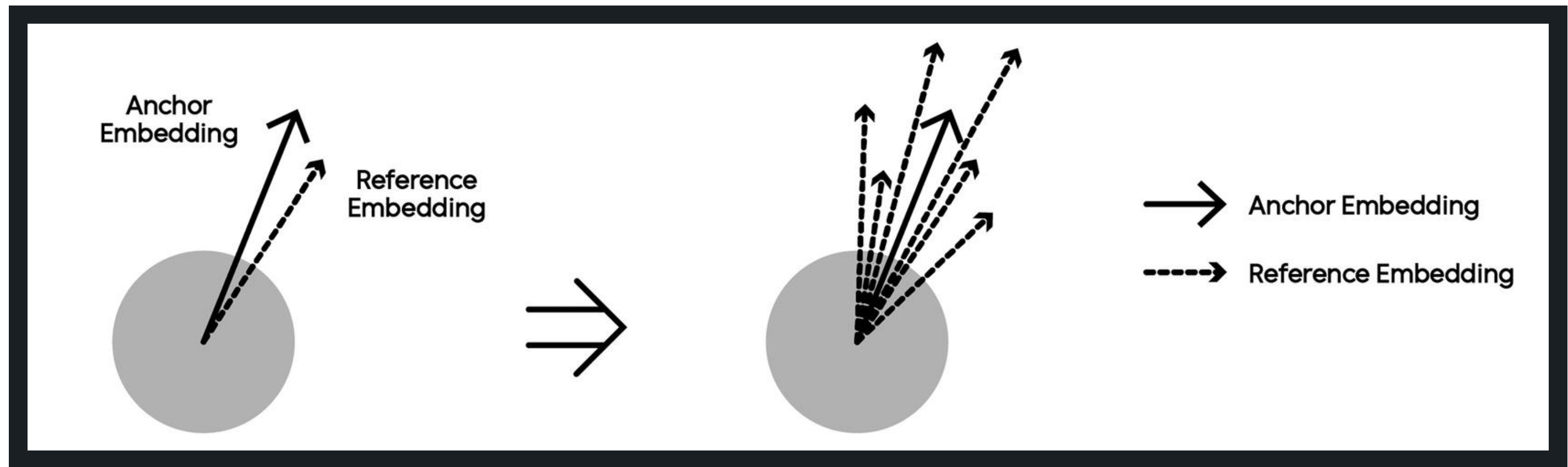
인퍼런스 전략

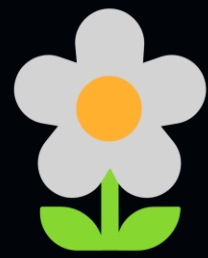
Multi-
Reference
Similarity
Ensemble
Prediction



결과

하나의 Test 샘플에 대한 예측을 수행할 때 여러 개의 Fake, Real Reference와의 유사도를 계산 후 이를 평균 내어 예측





인퍼런스 전략 > Ensemble Prediction

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

인퍼런스 전략

Multi-
Reference
Similarity
**Ensemble
Prediction**

결과

독립된 4개의 모델을 각각 학습
4개 모델의 예측값을 평균내어 예측

0.20X → 0.17X 로의 성능 향상

Model 0	Epoch 23
Model 1	Epoch 23
Model 2	Epoch 35
Model 3	Epoch 34

- 데이터 분석
- 데이터 전처리
- 사용 기법
- 모델 구조
- 모델 학습
- 인퍼런스 전략
- 결과**

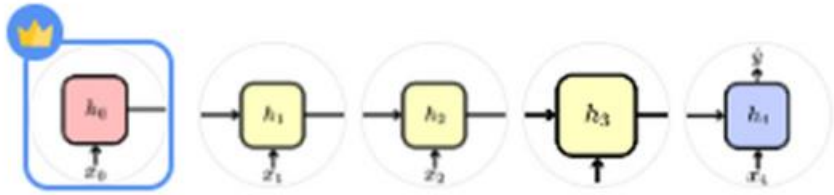
최종 제출 모델

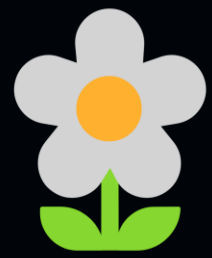
1031068 Epoch_22_prediction_id_17_reproduce_base_17_ensemble.csv 2024-07-18 15:28:27 0.1712580795

Head: 6, Ensemble [0, 3] edit

→ 4개의 모델 Ensemble
Test 샘플 당 6개의 Reference와 유사도 계산

Public Score	0.17125
Private Score	0.17077

#	팀	팀 멤버	최종점수	제출수	등록일
6	코스모스		0.17077	48	4일 전



적용 가능성

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

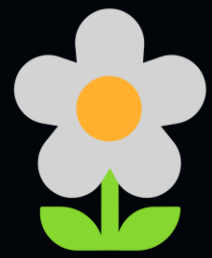
인퍼런스 전략

결과

더 적은 파라미터 수를 가지는 WeSpeaker ResNet 34에 대한 추가 실험 진행
(Test Split 기준)

Embedding Network	Parameter	Label	Precision	Recall	F1 Score	Accuracy
WeSpeaker ResNet 221 (Epoch 34)	23,792,224	Fake	0.9057	0.8360	0.8695	0.8748
		Real	0.9634	0.9286	0.9457	0.9465
WeSpeaker ResNet 34 (Epoch 34)	6,634,336	Fake	0.8833	0.8753	0.8793	0.8801
		Real	0.9581	0.9413	0.9496	0.9500

파라미터 수에 비해서 크게 차이 나지 않는 분류 성능
심지어 단일 모델 기준 ResNet221보다 미세하게 높은 성능



적용 가능성

데이터 분석

데이터 전처리

사용 기법

모델 구조

모델 학습

인퍼런스 전략

결과

On-device

간단하고 가벼운 네트워크를 설계함으로써
휴대폰이나 저사양 PC와 같은 환경에서도 로컬 환경에서 실행 가능

최근에 생성형 AI를 악용한 보이스 피싱
사례가
지속적으로 보고되고 있는 상황에
네트워크가 되지 않는 상황에서도
온 디바이스로 가짜 음성을
감지해낼 수 있다.

군대와 같은 보안이 중요한
시설에서도 가짜 음성의
탐지는 중요하다.
온 디바이스 AI를 사용하면 외부망과의
연결 없이도 가짜 음성을 탐지해낼
수 있다.