

2025.1.10

A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules

저자: Şefik İlkin Serengil, Alper Özpinar

저널: Bil. Teknol. Derg.

게재 년도: 2024년

01. Introduction

02. Dataset

03. Pipeline

04. Experiments

05. Results

06. Conclusion

Motivation

- 얼굴 탐지 파이프라인은 크게 네 가지로 구분할 수 있음
- LightFace에서 사용하는 각 단계가 결과값에 어떤 영향을 주는지 실험하고자 함

Methodology

- 정렬, 탐지, 특징 추출 등 다양한 조건들을 적절하게 조절하여 실험
- LightFace를 이용하여 실험을 계획

Results

- Detection module은 파이프라인 얼굴 탐지의 필수 구성 요소
- Alignment은 성능을 최대 17%까지 향상시킴

1. Introduction

얼굴 인식의 네 가지 기본 단계

Detection(탐지)

목표: 정확한 입력 제공

OpenCV Haar, SSD, MediaPipe, MTCNN,
Dlib HOG, Retina Face 등

Alignment(정렬)

Representation(표현)

제일 중요한 단계

FaceNet, VGG-Face, ArcFace, Dlib,
Sface, OpenFace, DeepFace,
DeepId 등

Verification(검증)

얼굴 이미지의 벡터 임베딩
사이의 거리를 계산

동일 인물인지 판단 및 거리
계산
거리 계산이 잘 될수록 좋은
모델

2. Dataset

LFW(Labeled Faces in the Wild) 데이터셋 사용

- 약 13,000개의 얼굴 이미지(250x250 픽셀)
- 약 5,749명의 인물로 동일 인물의 이미지가 여러 장 포함



Original Image

Original Image: 99.2%



Tight Crops

Tight Crops: 94.2%



Inverse Crops

Inverse Crops: 97.5%

얼굴 보다는 배경을 통해 많은 정보를
얻음

3. Pipeline

3.1. Detection and Alignment

- 오픈소스 Detection 모듈만 사용
(예시: OpenCv, Ssd, MtCnn, Dlib, RetinaFace 등)
- 모듈들의 역할은 얼굴 탐지 및 결과값 부스팅 방지
- 탐지 후 눈 위치로 직각 삼각형 생성 및 정면 이미지로 보정

3. Pipeline

3.1. Detection and Alignment



Figure 2. Alignment Procedure

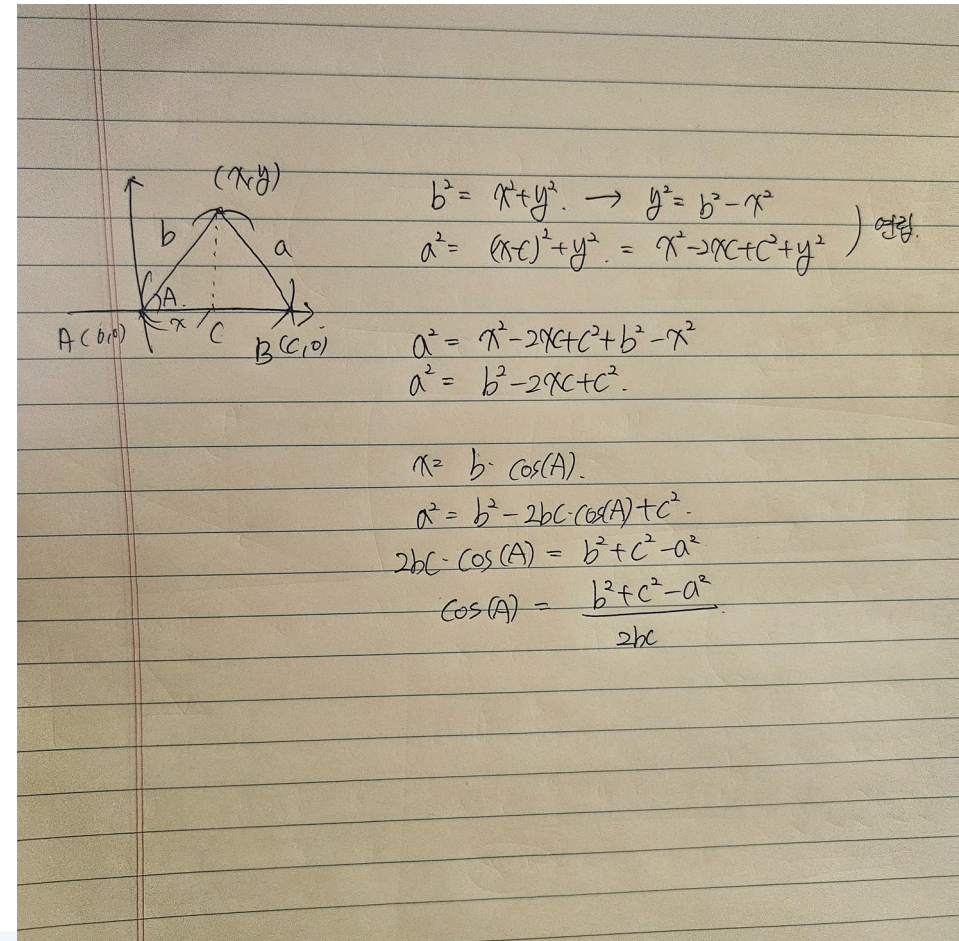
$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc}$$

$$A^\circ = \arccos\left(\frac{b^2 + c^2 - a^2}{2bc}\right) \frac{180}{\pi}$$

3. Pipeline

3.1. Detection and Alignment

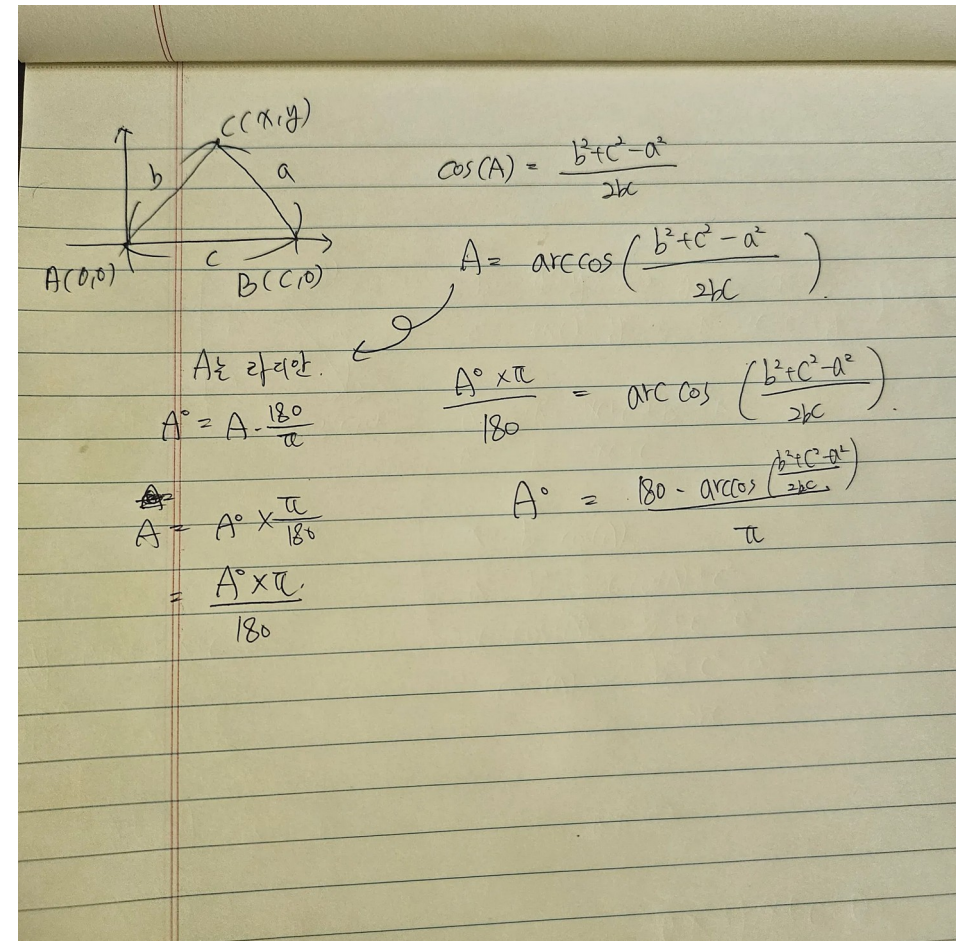
$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc}$$



3. Pipeline

3.1. Detection and Alignment

$$A^\circ = \arccos\left(\frac{b^2 + c^2 - a^2}{2bc}\right) \frac{180}{\pi}$$



3. Pipeline

3.2. Resizing

- CNN은 입력이 고정되어야 하지만 탐지된 얼굴 이미지는 그렇지 못함

Table 1. Facial Recognition Model's Architectures

Model	Input Shape	Embedding Dimensions	Total Params	Num of Layers
FaceNet-128d	160, 160, 3	128	22M	447
FaceNet-512d	160, 160, 3	512	23M	447
Vgg-Face	224, 224, 3	4096	134M	36
ArcFace	112, 112, 3	512	34M	162
Dlib	150, 150, 3	128	63M	34
SFace	112, 112, 3	128	9M	88
OpenFace	96, 96, 3	128	3M	166
DeepFace	152, 152, 3	4096	102M	9
DeepId	55, 47, 3	160	395K	17

3. Pipeline

3.2. Resizing

- 이미지 크기를 조절하는 대신 배경에 검은색 픽셀을 추가



Original Image



RetinaFace



MtCnn



Dlib



MediaPipe



Ssd



OpenCv

3. Pipeline

3.3. Representation

- 모든 출력 노드의 확률을 활용해 특징 벡터를 생성

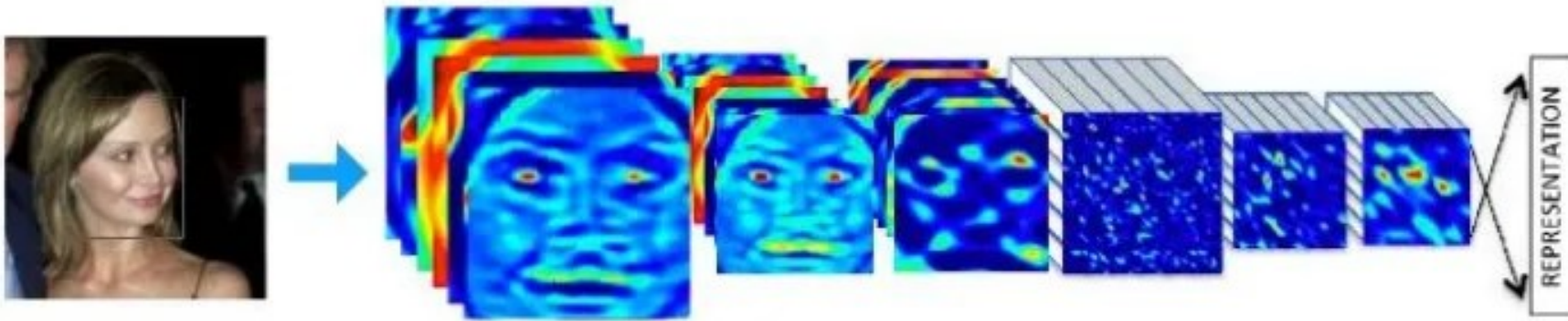


Figure 4. Vector Representation with DeepFace [1]

3. Pipeline

3.4. Verification

- Representation 모듈이 얼굴 쌍에 대한 벡터를 반환
- 반환 된 벡터를 Verification 모듈이 거리 계산

유클리드 공식

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

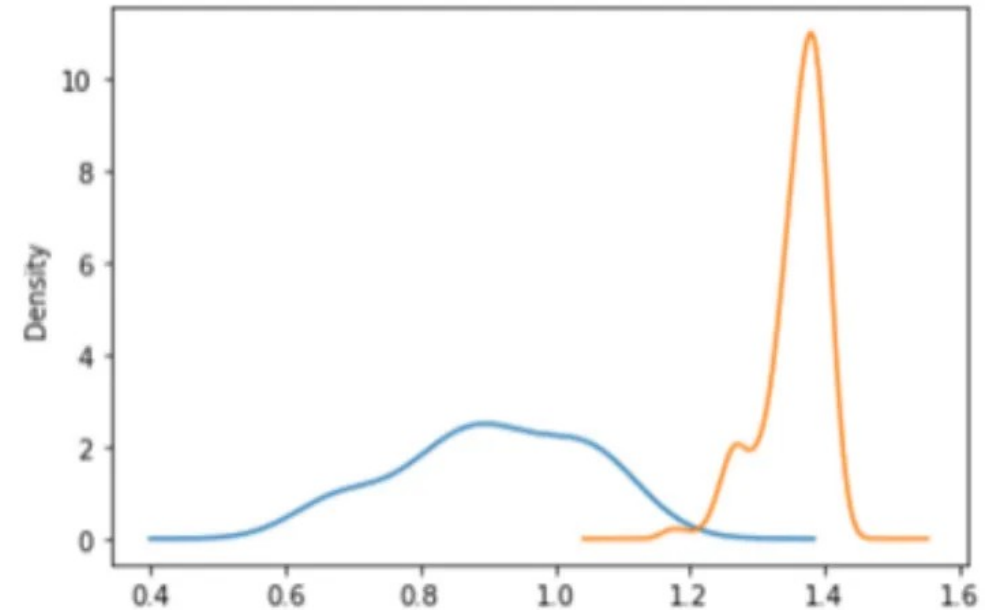
코사인 공식

$$d(p, q) = 1 - \frac{\sum_{i=1}^n p_i q_i}{\sqrt{\sum_{i=1}^n p_i^2} \sqrt{\sum_{i=1}^n q_i^2}}$$

L2 노름

$$l_2(q) = \frac{q}{\sqrt{\sum_{i=1}^n q_i^2}}$$

파란색: 동일 인물 주황색: 다른 인물



VGG-Face 테스트 결과 / 두 분포가 명확히 분리되어야 높은 모델

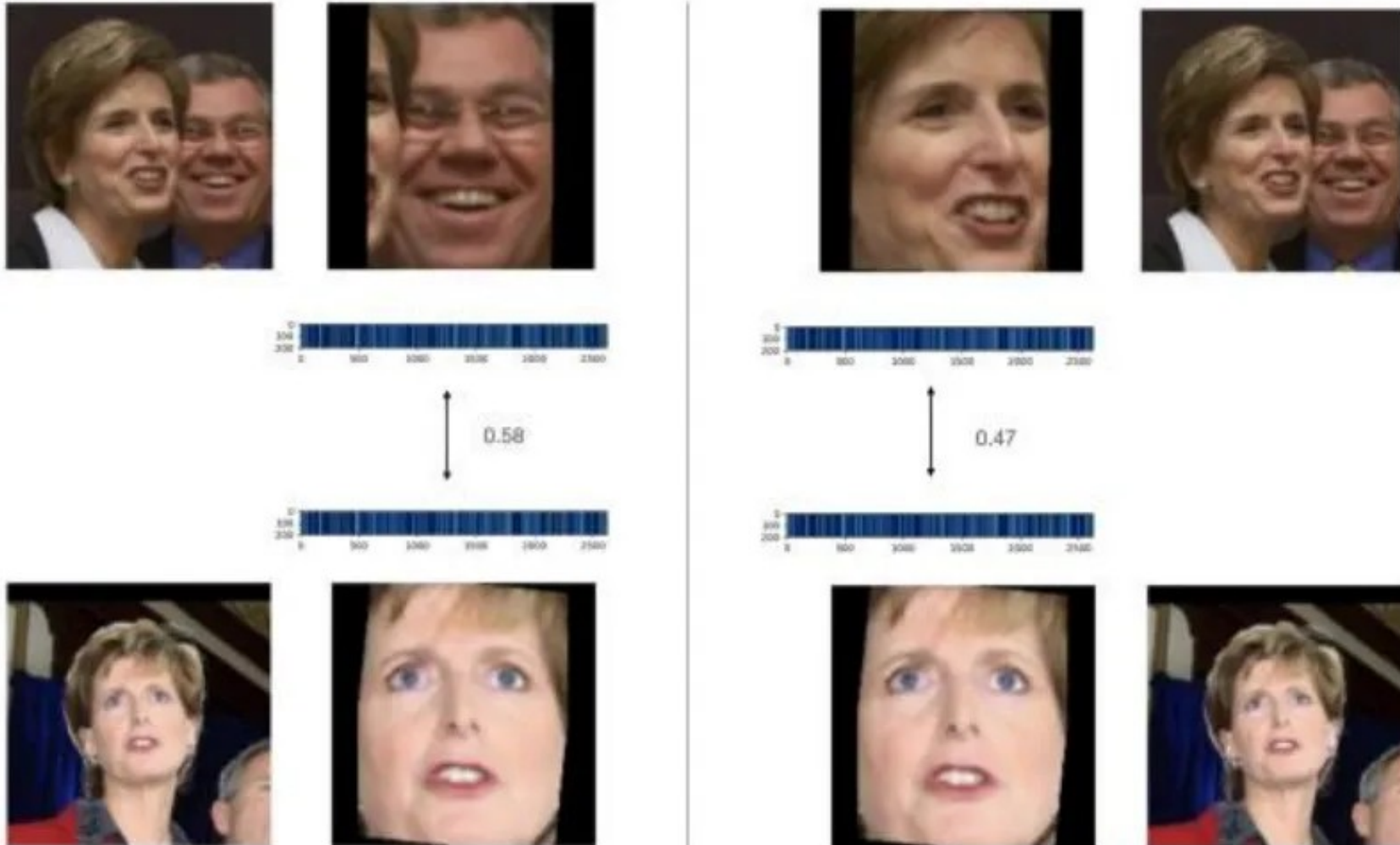
3. Pipeline

3.5. Handling Many Faces Issue

- LFW에는 다수의 얼굴이 포함되거나, 라벨이 맞지 않는 경우가 존재
- 모든 얼굴 쌍 추출을 위해 얼굴 인식 모델을 이용해 벡터로 표현
- 유사도 알고리즘(코사인, 유클리드)을 통해 모든 벡터 후보의 거리 측정

3. Pipeline

3.5. Handling Many Faces Issue



4. Experiments

- 네 가지 파트를 조합하여 총 378번의 실험을 실행
- 얼굴 인식을 위한 최적의 구성 세트 식별을 목표

Table 8. Different Configuration Sets for Experiments

Configuration	Possible Values
Facial Recognition Models	FaceNet128d, FaceNet512d, Vgg-Face, ArcFace, Dlib, SFace, OpenFace, DeepFace, DeepId
Face Detector Models	OpenCv, Ssd, MtCnn, Dlib, RetinaFace, MediaPipe, No Detection
Distance Metrics	Cosine, Euclidean, L2-Norm Euclidean
Alignment Mode	Enabled, Disabled

4. Experiments

Euclidean Distance about Alignment

Table 2. Accuracy Metric for Euclidean Distance and Disabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	92.8	96.1	95.7	84.1	88.3	78.6	70.8	67.4	64.3
MtCnn	92.5	95.9	95.5	81.8	89.3	76.3	70.9	65.9	63.2
Dlib	89.0	96.0	94.1	82.6	96.3	73.1	75.9	61.8	61.9
MediaPipe	87.1	94.9	93.1	71.1	91.9	73.2	77.6	61.7	62.4
Ssd	94.9	97.2	96.7	83.9	88.6	82.0	69.9	66.7	64.0
OpenCv	90.2	94.1	95.8	89.8	91.2	86.9	71.1	68.4	61.1
None	64.1	92.0	90.6	56.6	69.0	81.4	57.4	60.8	60.7

Table 5. Accuracy Metric for Euclidean Distance and Enabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	93.5	95.9	95.8	85.2	88.9	80.2	69.4	67.0	65.6
MtCnn	93.8	95.2	95.9	83.7	89.4	77.4	70.2	66.5	63.3
Dlib	90.8	96.0	94.5	88.6	96.8	66.3	75.8	63.4	60.4
MediaPipe	88.6	95.1	92.9	73.2	93.1	72.5	78.7	61.8	62.2
Ssd	85.6	88.9	87.0	75.8	83.1	76.9	66.8	63.4	62.5
OpenCv	84.2	88.2	87.3	73.0	84.4	81.1	66.4	65.5	59.6
None	64.5	92.0	90.6	56.6	69.0	81.4	57.4	60.8	60.7

4. Experiments

L2 Normalized Euclidean Distance about Alignment

Table 3. Accuracy Metric for L2 Normalized Euclidean Distance and Disabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	95.9	98.0	95.7	95.7	88.4	90.6	70.8	67.7	64.6
MtCnn	96.2	97.8	95.5	95.9	89.2	91.1	70.9	67.0	64.0
Dlib	89.9	96.5	94.1	93.8	95.6	75.0	75.9	62.6	61.8
MediaPipe	90.0	96.3	93.1	89.3	91.8	74.6	77.6	64.9	61.6
Ssd	97.0	97.9	96.7	96.6	89.4	93.0	69.9	68.7	64.9
OpenCv	92.9	96.2	95.8	93.2	91.5	91.7	71.1	68.3	61.6
None	67.6	91.4	90.6	57.2	69.3	83.4	57.4	62.6	61.6

Table 6. Accuracy Metric for L2 Normalized Euclidean Distance and Enabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	96.4	98.4	95.8	96.6	89.1	92.4	69.4	67.7	64.4
MtCnn	96.8	97.6	95.9	96.0	90.0	90.5	70.2	66.4	64.0
Dlib	92.6	97.0	94.5	95.1	96.4	69.8	75.8	66.5	59.5
MediaPipe	90.6	96.1	92.9	90.3	92.6	75.4	78.7	64.7	63.0
Ssd	87.5	88.7	87.0	86.2	83.3	84.6	66.8	64.1	62.6
OpenCv	84.8	87.6	87.3	84.6	84.0	83.6	66.4	63.8	60.9
None	67.6	91.4	90.6	57.2	69.3	83.4	57.4	62.6	61.6

4. Experiments

Cosine Distance about Alignmnet

Table 4. Accuracy Metric for Cosine Distance and Disabled Alignment

	<i>FaceNet-128d</i>	<i>FaceNet512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	95.9	98.0	95.7	95.7	88.4	90.6	70.8	67.7	63.7
MtCnn	96.2	97.8	95.5	95.9	89.2	91.1	70.9	67.0	64.0
Dlib	89.9	96.5	94.1	93.8	95.6	75.0	75.9	62.6	61.7
MediaPipe	90.0	96.3	93.1	89.3	91.8	74.6	77.6	64.9	61.6
Ssd	97.0	97.9	96.7	96.6	89.4	93.0	69.9	68.7	63.8
OpenCv	92.9	96.2	95.8	93.2	91.5	91.7	71.1	68.1	61.1
None	67.6	91.4	90.6	54.8	69.3	83.4	57.4	62.6	61.1

Table 7. Accuracy Metric for Cosine Distance and Enabled Alignment

	<i>FaceNet128</i>	<i>FaceNet-512d</i>	<i>VGG-Face</i>	<i>ArcFace</i>	<i>Dlib</i>	<i>SFace</i>	<i>OpenFace</i>	<i>DeepFace</i>	<i>DeepId</i>
RetinaFace	96.4	98.4	95.8	96.6	89.1	92.4	69.4	67.7	64.4
MtCnn	96.8	97.6	95.9	96.0	90.0	90.5	70.2	66.3	63.0
Dlib	92.6	97.0	94.5	95.1	96.4	69.8	75.8	66.5	58.7
MediaPipe	90.6	96.1	92.9	90.3	92.6	75.4	78.7	64.8	63.0
Ssd	87.5	88.7	87.0	86.2	83.3	84.5	66.8	63.8	62.6
OpenCv	84.9	87.6	87.2	84.6	84.0	83.6	66.2	63.7	60.1
None	67.6	91.4	90.6	54.8	69.3	83.4	57.4	62.6	61.1

4. Experiments

Recognition 모델 별 ROC 곡선

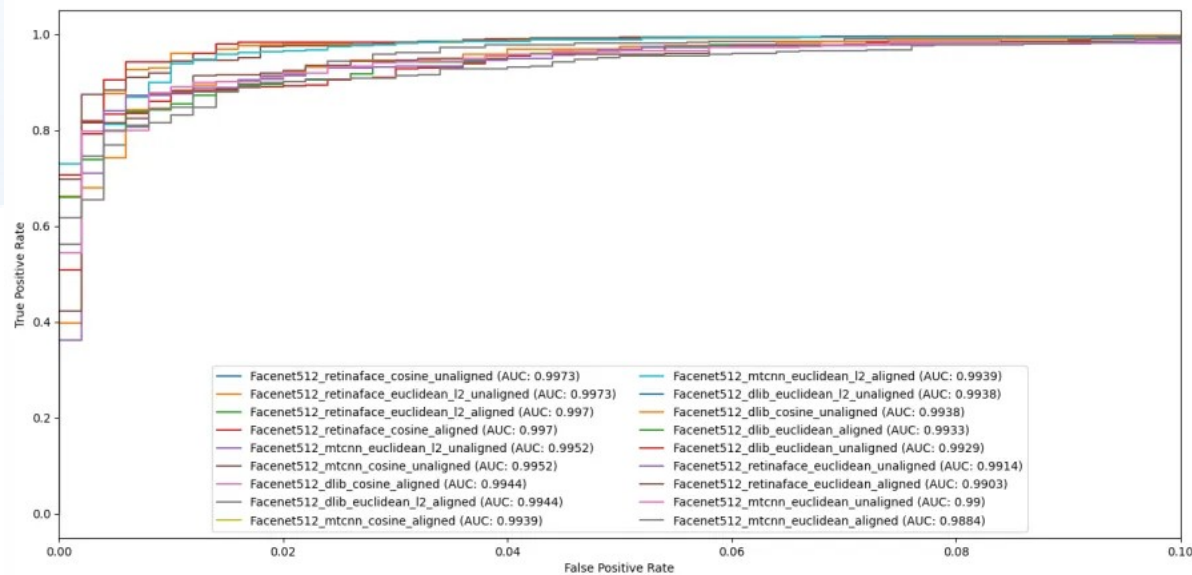


Figure 7. ROC Curves and AUC Scores for FaceNet512d

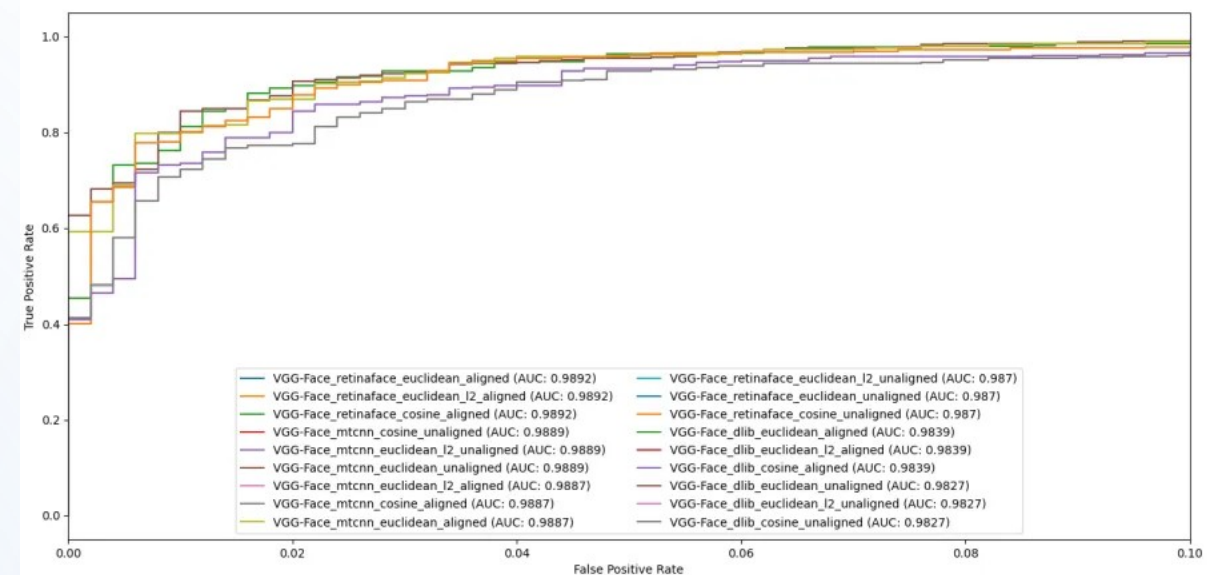


Figure 8. ROC Curves and AUC Scores for VGG-Face

4. Experiments

Recognition 모델 별 ROC 곡선

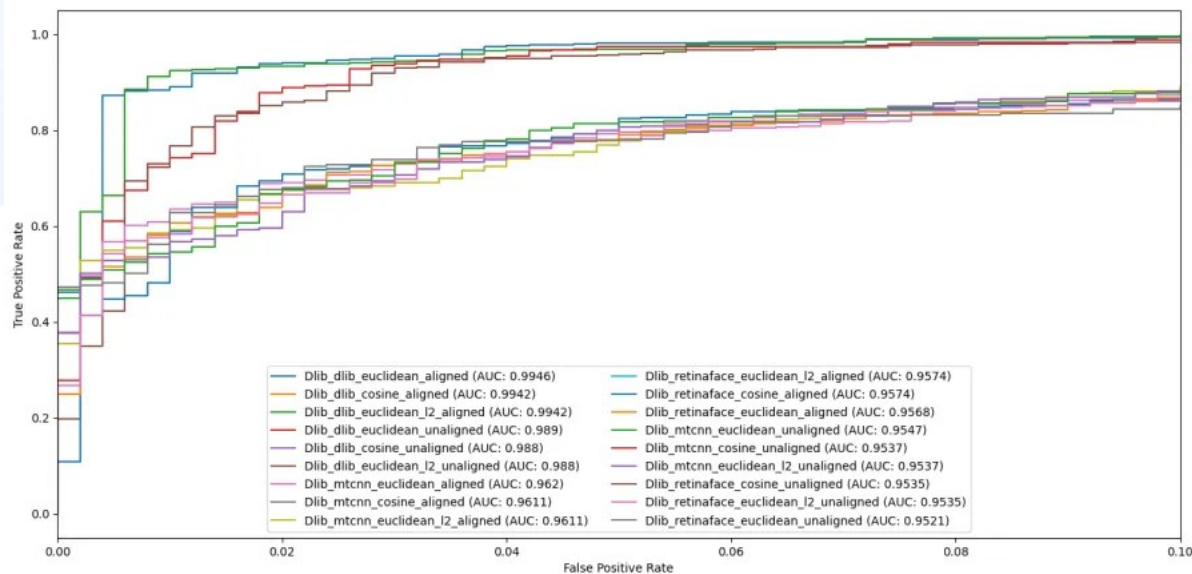


Figure 9. ROC Curves and AUC Scores for Dlib

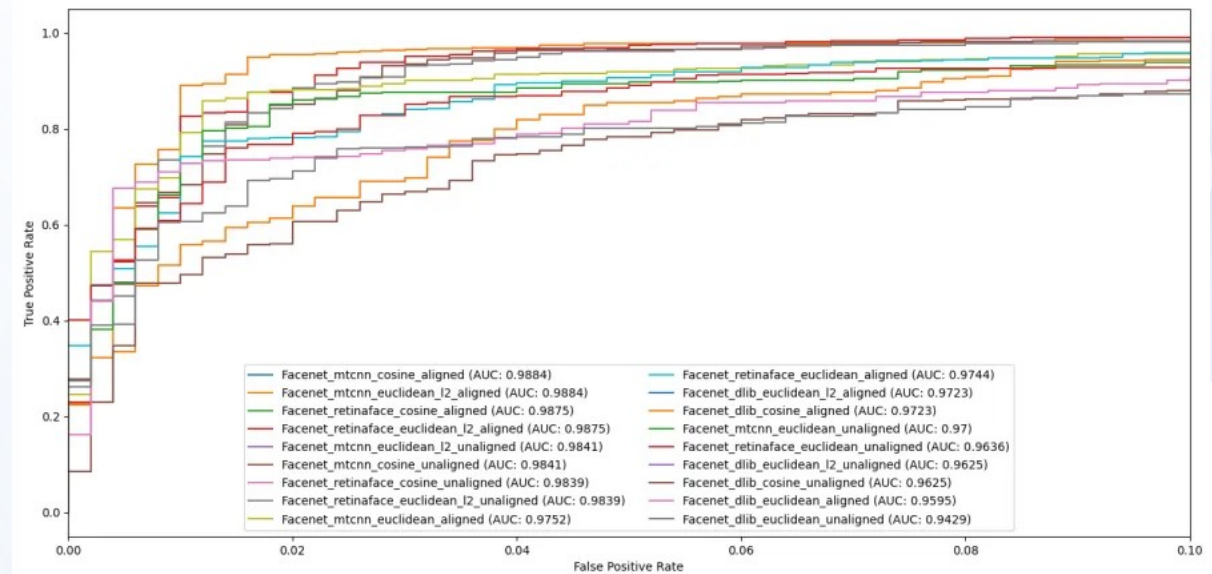


Figure 10. ROC Curves and AUC Scores for FaceNet128d

4. Experiments

Recognition 모델 별 ROC 곡선

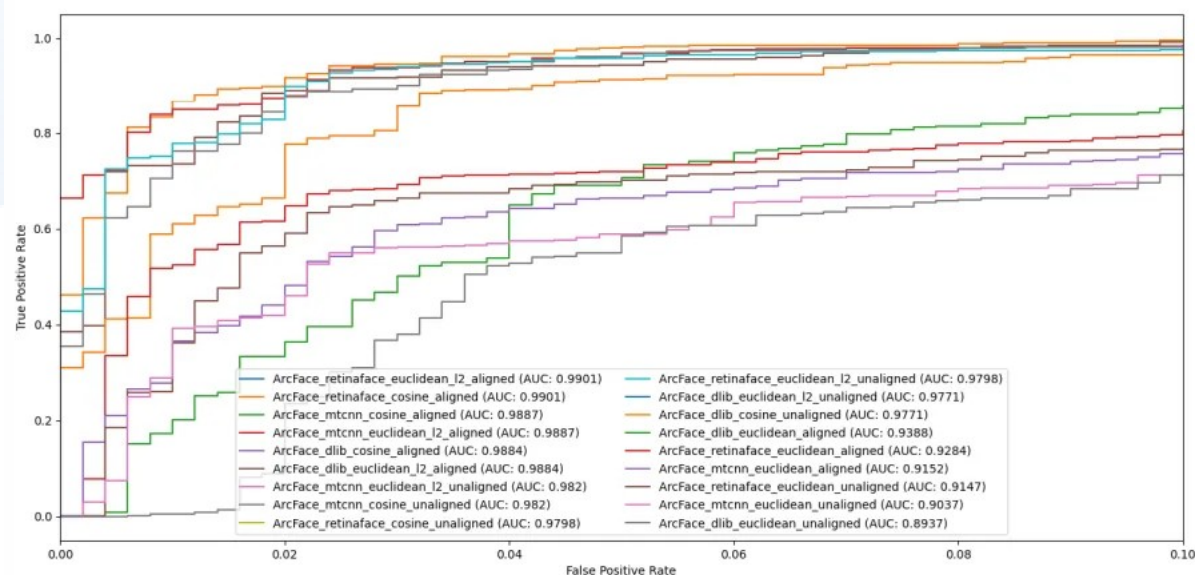


Figure 11. ROC Curves and AUC Scores for ArcFace

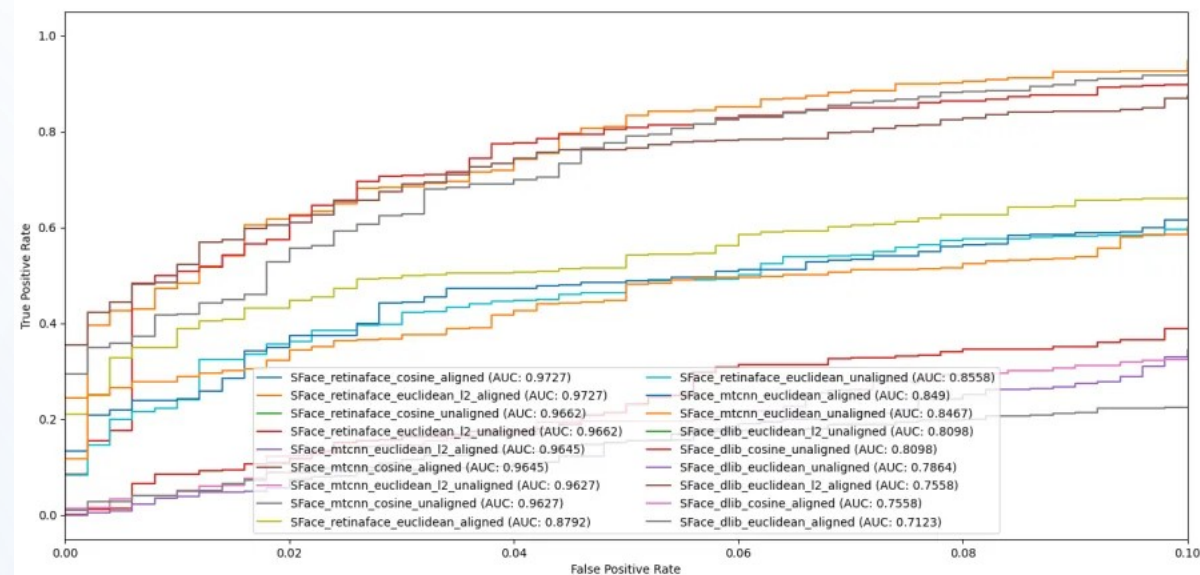


Figure 12. ROC Curves and AUC Scores for SFace

5. Results

Context 의존성과 비의존성의 비교

Context-independent(비의존적)

- 배경 정보 없이 실행

- 정확도: 97.5%

Context-dependent(의존적)

- 배경 정보 포함 실행

- 정확도: 99.2%

컨텍스트 정보를 포함하는 것은 얼굴 인식 정확도를 높이는데

탐지기와 정렬 모듈 비교

Face Detector와 Alignment Module이 얼굴 인식 파이프라인에 미치는 영향 평가

각 모듈이 파이프라인 성능에 중요한 역할을 함

탐지기와 정렬 모듈 비교

인간의 얼굴 인식 성능과 비교하여 모델 성능 평가

5. Results

Table 9. Comparison of Measured Accuracy Scores in LightFace and Declared Accuracy Scores in Their Original Researches

Model	Measured Accuracy	Declared Accuracy	Original Paper
FaceNet-512d	98.4	99.6	[2]
FaceNet-128d	97.0	99.6	[2]
Dlib	96.8	99.3	[5]
VGG-Face	96.7	98.9	[3]
ArcFace	96.6	99.5	[4]
SFace	93.0	99.5	[6]
OpenFace	78.7	92.9	[7]
DeepFace	68.7	97.3	[1]
DeepId	65.6	97.4	[8]

6. Conclusion

- 얼굴에 초점을 맞추기 위해 얼굴 이외의 영역은 버림
- AUC와 ROC곡선으로 얼굴 인식의 성능을 보여줌
- Detection module은 파이프라인 얼굴 탐지의 필수 구성 요소
 - 성능을 최대 40%까지 향상시킴
- Alignment은 성능을 최대 17%까지 향상시킴
 - OpenCv, SSD, MediaPipe는 정렬 시 성능이 저하됨
 - RetinaFace, MtCnn, Dlib은 정렬이 되어도 항상 성능이 높음