



Face Verification: Accuracy on LFW, Cross-Age and Cross-Pose Benchmarks

Engine version 3.1.1 · Measured 28 October 2026 · iApp Technology Co., Ltd., Bangkok, Thailand

Executive summary. The iApp Face Verification API decides whether two face images show the same person. We measured the production service on three public benchmarks of 6,000 image pairs each: LFW, the standard test of face verification on unconstrained photographs, and its cross-age (CALFW) and cross-pose (CPLFW) variants. At the API's default threshold, with each image centre-cropped to one face, the decision was correct for 98.50% of LFW pairs (95% CI [98.18, 98.82]). It accepted 97.00% of same-person pairs and 4 of 30,000 different-person pairs, a false acceptance rate of 1.3×10^{-4} . The default threshold therefore favours security. On pairs taken years apart (CALFW) accuracy falls to 94.83%, with 0.40% of different-person pairs accepted. Strongly turned heads (CPLFW) are the weakest case: the service declined 43.2% of those pairs and accepted 44.13% of same-person pairs. The service accepts one face per image; on LFW images as published it declined 16.3% of pairs because a second face was visible.

1 Introduction

Face verification (1:1 matching) compares a face in one image with a face in another, typically a selfie with the photograph on an identity document, and returns a similarity score and a decision. The iApp API applies a default threshold of 36 on a score that rises with similarity; a stricter threshold of 48 is documented for applications that need a lower false-acceptance rate, and a caller may set any threshold per request.

This paper measures the decision the API returns with its default settings on public benchmarks, reports the trade-off between false acceptance and false rejection over the whole score range, and states where accuracy drops.

2 Test Data

Table 1: Benchmarks. Each has 6,000 pairs in ten folds of 600: 3,000 same-person and 3,000 different-person pairs.

Set	Content
LFW [1]	13,233 news photographs of 5,749 people; the standard View 2 pair list
CALFW [2]	Cross-age LFW: same-person pairs chosen for a large age gap; different-person pairs of the same gender and ethnicity
CPLFW [3]	Cross-pose LFW: same-person pairs chosen for a large difference in head pose
LFW, extra impostors	27,000 further different-person LFW pairs, drawn with seed 1337 and excluding the View 2 pairs, to measure low false-acceptance rates

The photographs are of public figures, mostly adults of European descent photographed by news agencies; the sets do not represent Thai users or document photographs (Section 7). None of the images is shown in this paper.

3 Method

Service. Every pair was sent, with at most three requests in flight and no threshold parameter, to the production comparison engine that the public endpoint `/v3/store/ekyc/face-verification` calls, so the default threshold applied. A seeded sample of 300 LFW pairs was also sent through the public API to confirm that it returns the same answers (Section 6).

One face per image. The service compares exactly one face per image and declines an image in which it finds more than one, with the code E423 and score 0. LFW images are 250×250 crops of news photographs and often show part of a second person. For the main results, every LFW and CALFW image was centre-cropped to 150×150 pixels, which keeps the subject’s face and removes most background faces. The same crop was applied to every image of a set. CPLFW images, 224×224 crops aligned on the face, were sent as distributed. Results on LFW without the crop, and on CPLFW with a border added around each face, are reported separately (Table 5).

Metrics. A pair is *accepted* when the API reports *matched*. The true acceptance rate (TAR) is the share of same-person pairs accepted; the false acceptance rate (FAR) is the share of different-person pairs accepted. A pair the service declines counts as not accepted, which is what a caller receives. Accuracy at the default threshold counts every pair. From the returned scores we also report the area under the ROC curve, the TAR at fixed FAR, and the standard ten-fold LFW accuracy, in which the threshold is chosen on nine folds and tested on the tenth, for comparison with the literature. 95% confidence intervals are percentile bootstrap intervals (2,000 resamples, seed 1337).

4 Results

Table 2: Accuracy at the default threshold (36), and threshold-free measures. Declined: pairs the service did not compare because it found no face or more than one in an image.

Set	Pairs	Declined	Accuracy (%) [95% CI]	TAR (%)	FAR (%)	ROC AUC	10-fold (%)
LFW	6,000	115	98.50 [98.18, 98.82]	97.00	0.00	0.9846	98.55 ± 0.55
CALFW (cross-age)	6,000	70	94.83 [94.30, 95.38]	90.07	0.40	0.9577	94.80 ± 1.14
CPLFW (cross-pose)	6,000	2,590	71.95 [70.83, 73.08]	44.13	0.23	0.6640	72.37 ± 2.18

Table 3: Operating points on LFW (centre-cropped): the 3,000 same-person pairs against the 3,000 View 2 different-person pairs, and against all 30,000 different-person pairs including the extra impostors.

Operating point	Threshold	TAR (%)	False accepts
Default threshold	36	97.00	4 of 30,000
Documented strict threshold	48	94.20	0 of 30,000
$\text{FAR} \leq 10^{-3}$ (View 2 impostors)	25	97.10	3 of 3,000
$\text{FAR} \leq 10^{-4}$ (all impostors)	38	96.80	3 of 30,000

Table 4: The default (36) and the documented strict (48) threshold on each set: TAR and accepted different-person pairs.

Set	Threshold 36		Threshold 48	
	TAR (%)	False accepts	TAR (%)	False accepts
LFW	97.00	0 of 3,000	94.20	0 of 3,000
CALFW (cross-age)	90.07	12 of 3,000	78.93	0 of 3,000
CPLFW (cross-pose)	44.13	7 of 3,000	33.00	1 of 3,000

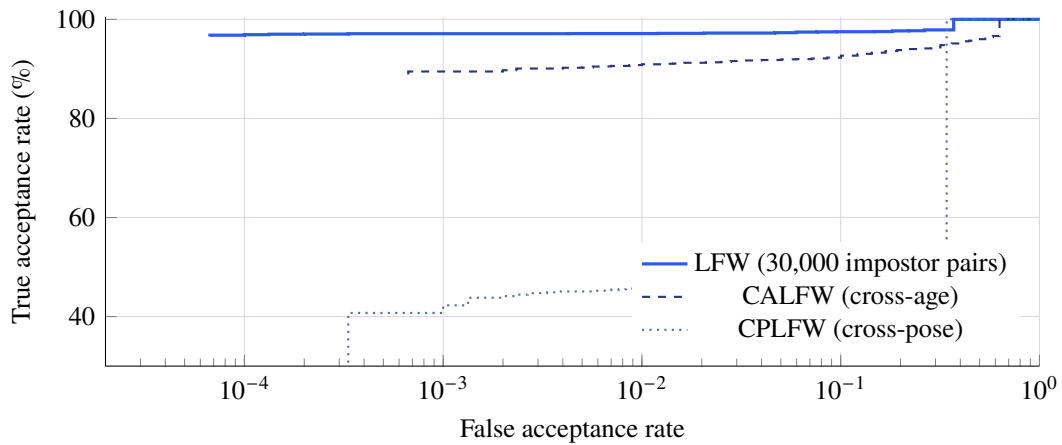


Figure 1: ROC curves from the returned scores. Each curve starts at its lowest measurable FAR, one false accept among its impostor pairs.

Table 5: Variants: LFW images as published (no crop), and CPLFW images with a uniform 64-pixel grey border. Declined pairs count as not accepted.

Set	Declined	Accuracy (%)	TAR (%)	FAR (%)
LFW, no crop	978 (16.3%)	90.87	81.73	0.00
CPLFW, with border	2,718 (45.3%)	68.05	36.90	0.80

5 Discussion

The default threshold rarely admits the wrong person. Of 30,000 different-person LFW pairs, 4 reached the default threshold of 36: a false acceptance rate of 1.3×10^{-4} (95% CI 3.6×10^{-5} to 3.4×10^{-4}), with scores from 37 to 40, close to the threshold. The documented strict threshold of 48 accepted 0 of them, at a TAR of 94.20% against 97.00% at the default; on the harder different-person pairs of CALFW and CPLFW it removes most false accepts (Table 4).

Most rejected same-person pairs were not compared. Of the 90 same-person LFW pairs that were not accepted, 56 were declined because the service found more than one face or none, and 34 scored below the threshold. Lowering the threshold to 25, which keeps the FAR at or below 10^{-3} , would raise the TAR only from 97.00% to 97.10%.

Age and pose cost accuracy. On cross-age pairs (CALFW) the default decision was correct for 94.83% of pairs and accepted 90.07% of same-person pairs. 12 of its 3,000 different-person pairs, which are chosen to share gender and ethnicity, were accepted, all with scores from 36 to 39, just above the threshold; the strict threshold of 48 rejected all of them, at a TAR of 78.93%. Cross-pose pairs (CPLFW), which pair a frontal face with a strongly turned one, are the weakest case: the service declined 43.2% of them, 2,352 because it found no face in one or both images, and accepted 44.13% of same-person pairs. Adding a border around each face did not help (Table 5). The service is built for frontal faces, such as a selfie and an identity-document photograph; a photograph with the head strongly turned should be retaken.

One face per image. On LFW images as published, 16.3% of pairs were declined, 946 of them because a second face was visible. Of the 4,986 pairs compared both with and without the crop, the crop changed the decision for 43. An application should send a photograph that shows the user alone, or crop to the user's face.

The claim the evidence supports: on LFW, with one face per image, the API's default decision is correct for 98.50% of pairs; it accepts 97.00% of same-person pairs and accepted 4 of 30,000 different-person pairs.

6 Production Performance

The seeded sample of 300 centre-cropped LFW pairs sent through `api.iapp.co.th` returned the same decision as the engine for 300 pairs and the identical score for 300. Every call returned HTTP 200, including the 5 pairs the service declined, which carry their reason in the response body. Measured from a client in Bangkok with two requests in flight, the median end-to-end time was 0.64 s (90th percentile 1.16 s), of which the service reported a median of 0.46 s of processing. Face verification costs 0.3 IC per call.

7 Limitations

- **Population.** LFW and its variants are news photographs of public figures, mostly adults of European descent. They do not measure accuracy on Thai faces, on identity-document photographs, or on selfies taken with phone cameras, which is the main use of this API.
- **Public benchmarks.** LFW has been public since 2007. We cannot rule out that its images were seen during the development of the comparison engine; CALFW and CPLFW reuse LFW identities.
- **Low false-acceptance rates.** With 30,000 different-person pairs, false acceptance rates well below 10^{-4} cannot be resolved; the rate at the default threshold is known only to within its confidence interval.
- **Pre-processing.** The main results use a fixed centre crop. Callers who send photographs with more than one face receive a decline, not a decision.
- **Label noise.** A few LFW pairs are known to be mislabelled; we used the official lists unchanged.

8 Reproducibility

The benchmarks are public. The scripts `get_data.sh`, `crop.py`, `bench.py` and `gen_results.py`, the per-pair results (pair index, label, score, decision; no images) and the seeds are available from iApp Technology on request: sale@iapp.co.th.

References

- [1] G. B. Huang, M. Ramesh, T. Berg, E. Learned-Miller. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. University of Massachusetts, Amherst, Technical Report 07-49, 2007.
- [2] T. Zheng, W. Deng, J. Hu. *Cross-Age LFW: A Database for Studying Cross-Age Face Recognition in Unconstrained Environments*. arXiv:1708.08197, 2017.
- [3] T. Zheng, W. Deng. *Cross-Pose LFW: A Database for Studying Cross-Pose Face Recognition in Unconstrained Environments*. Beijing University of Posts and Telecommunications, Technical Report 18-01, 2018.

About iApp Technology. iApp Technology Co., Ltd. is a Thai artificial-intelligence company providing OCR, document understanding, speech, identity verification and large-language-model services for the Thai language. Web: iapp.co.th · E-mail: sale@iapp.co.th · Tel: +66 2 124 4041