

Watchlist Challenge: 3rd Open-set Face Detection and Identification

F. Kasim, T. E. Boulton, R. Mora, B. Biesseck, R. Ribeiro, J. Schlueter, T. Repák, R. Vareto, D. Menotti, W. R. Schwartz, M. Günther

INTERNATIONAL JOINT CONFERENCE ON BIOMETRICS, IJCB 2024

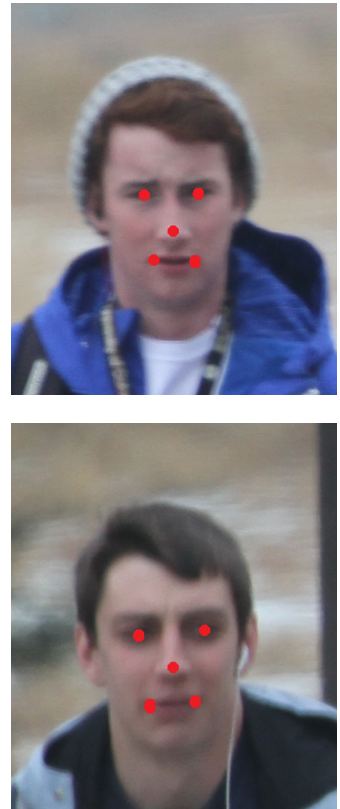
1. Introduction

- Face recognition technology faces challenges like blurriness, occlusion, poor lighting, non-frontal face poses and open-set settings in real-world surveillance environments.
- The UCCS Watchlist Challenge uses the UnConstrained College Students (UCCS) dataset, mirroring real-world conditions with non-cooperative behavior and environmental variability.
- The challenge comprises two parts: **(I)** face detection in all captured images and **(II)** open-set face recognition to only identify individuals on a watchlist.
- Past iterations [1, 2] underscored the challenges of open-set recognition, the need for better evaluation protocols to address temporal biases, and the importance of correcting label inaccuracies.

2. Dataset and Protocol

- The UCCS dataset, collected at the University of Colorado Colorado Springs, features diverse facial images with varied orientations, occlusions, and blurriness, alongside a thorough data cleaning process to correct mislabeling issues.
- This year's challenge introduces a separate watchlist (gallery) of cropped face regions, focusing on faces with high-quality and appearing in multiple sequence.
- The watchlist includes 1,000 identities, with each identity represented by 10 faces from the same sequence only, and annotations for five facial landmarks.

Gallery:



Probe Images:



- Validation set images feature bounding boxes with an integral identity label or marked as unknown, usually coming from the same sequence as the enrollment images.
- The test set mostly includes **unannotated** images from different sequences to avoid same-day bias. Unknown samples include unknown identities and false positive detections of the face detector:

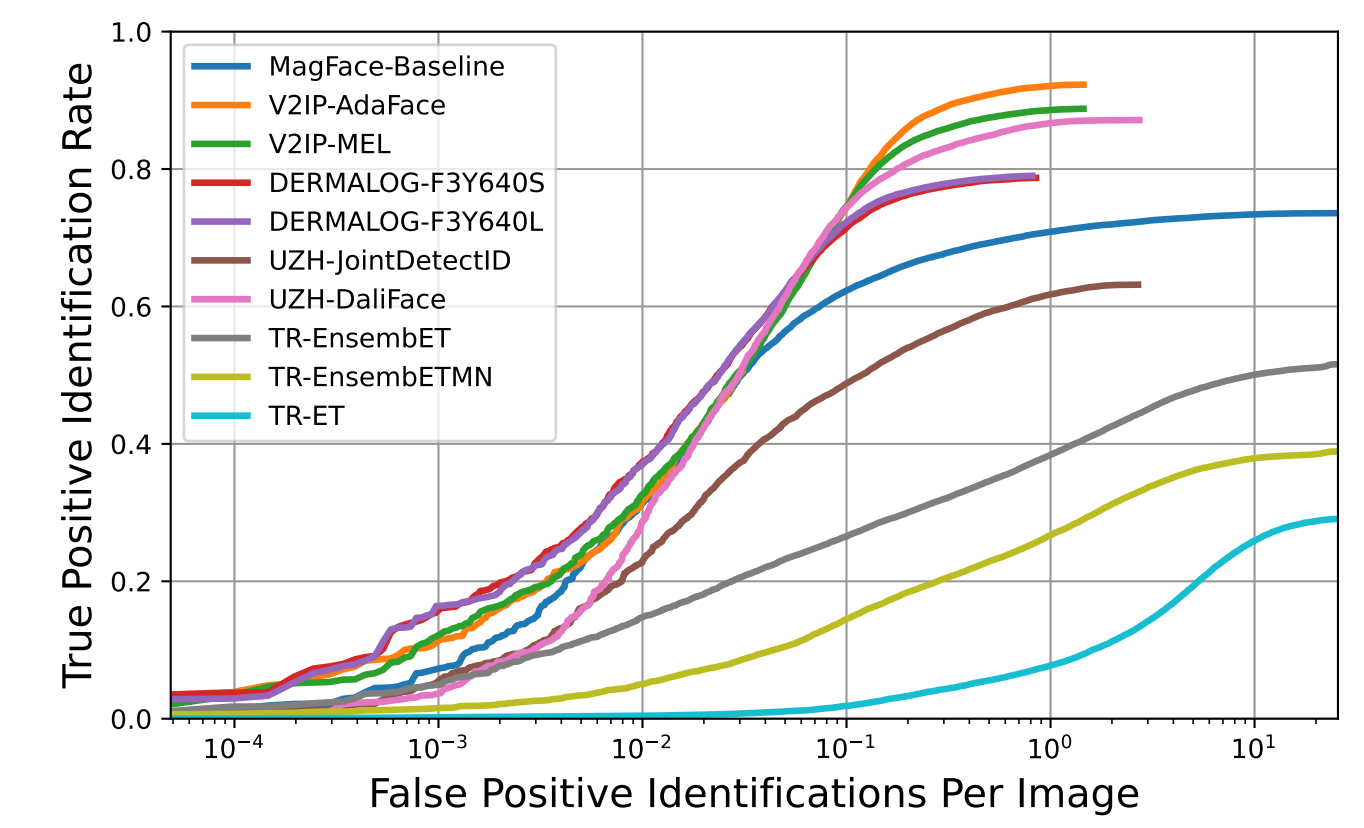
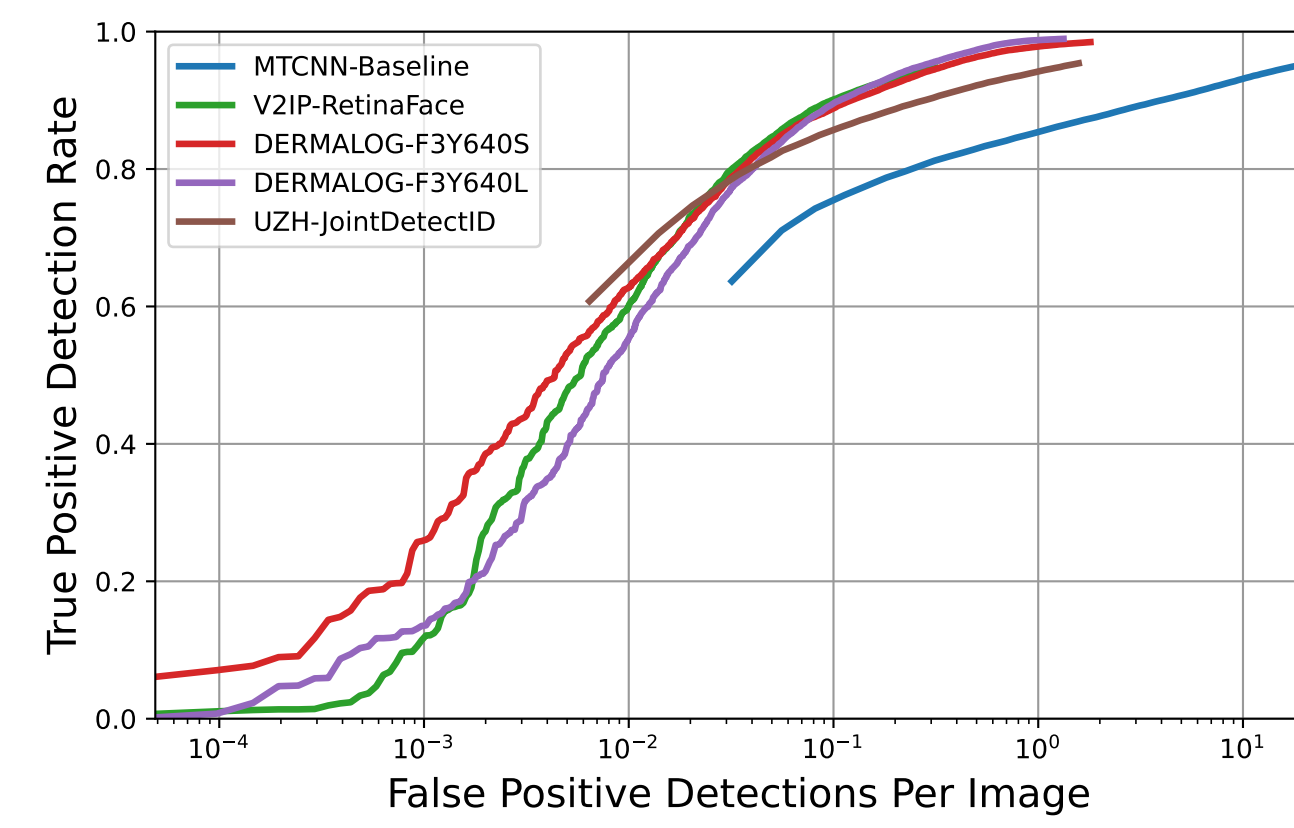
Sets	Identities	Images	Faces	Known	Unknown
Watchlist	1000	—	10'000	10'000	—
Validation	932	7'584	17'689	9'396	8'293
Test	996	20'534	57'368	31'512	25'856

3. Challenge Participants

- Four teams – UZH, V2IP, DERMALOG, and TR (details in the paper) – including participants from various universities/institutions contributed to the challenge with distinct models.
- Besides the baseline face detector MTCNN [4], participants introduced three face detection algorithms, including RetinaFace, F3Y640S and F3Y640L.
- The baseline extractor MagFace [3] was accompanied by seven face recognition algorithms from participants, including AdaFace, MEL, F3Y640S/F3Y640L, DaliFace, and EnsembET, EnsembETMN, and ET.
- An algorithm, JointFaceDetectID, also tackled the integration of face detection and identification into a single model to streamline tasks.

4. Evaluation

- Face detection results are evaluated using an adapted Free-response Receiver Operating Characteristic (FROC) curve, which plots the True Positive Detection Rate (TPDR) over the number of False Positive Detections Per Image (FPDPI).
- To show open-set face identification results, the modified Open Receiver Operating Characteristic (O-ROC) curve is exploited, which plots True Positive Identification Rate (TPIR) over False Positive Identifications Per Image (FPIPI).

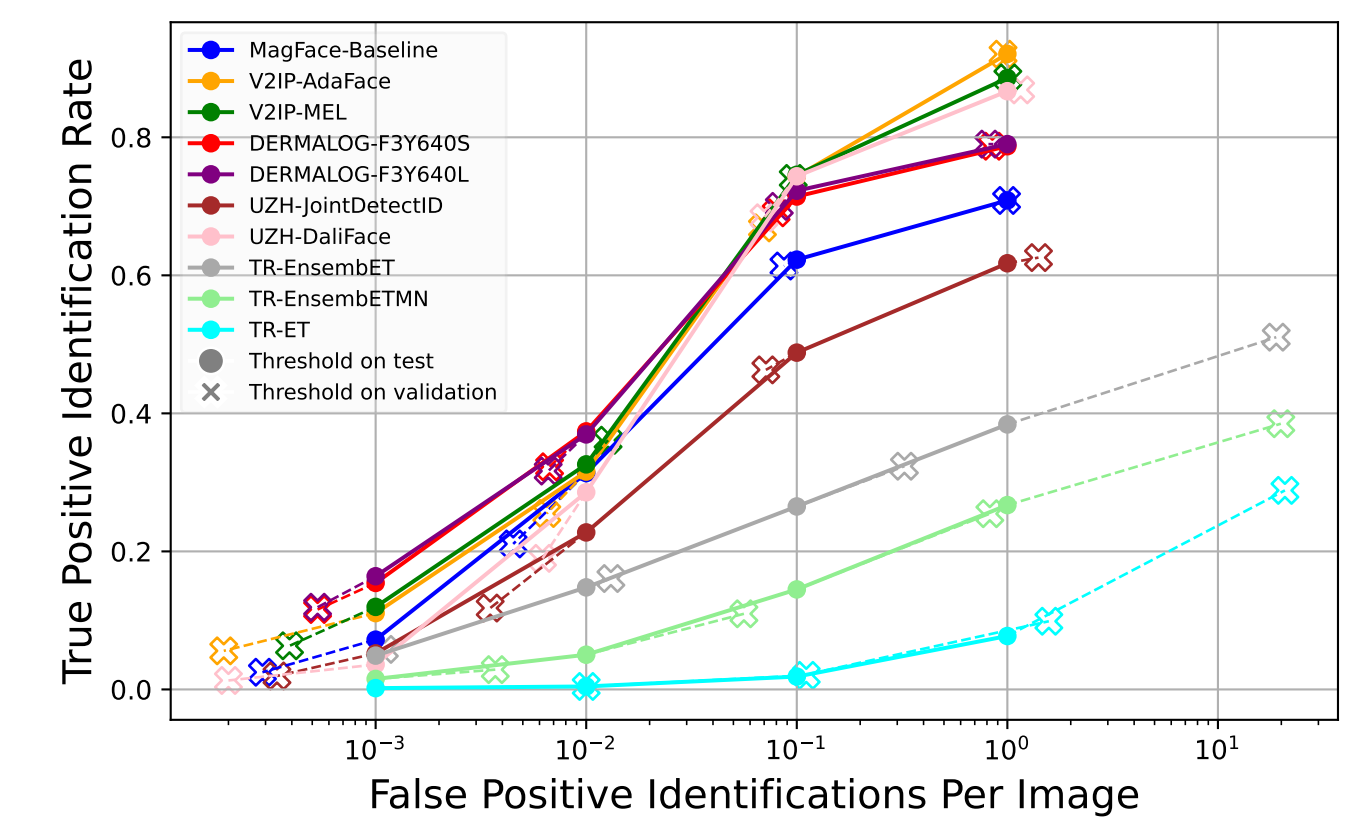
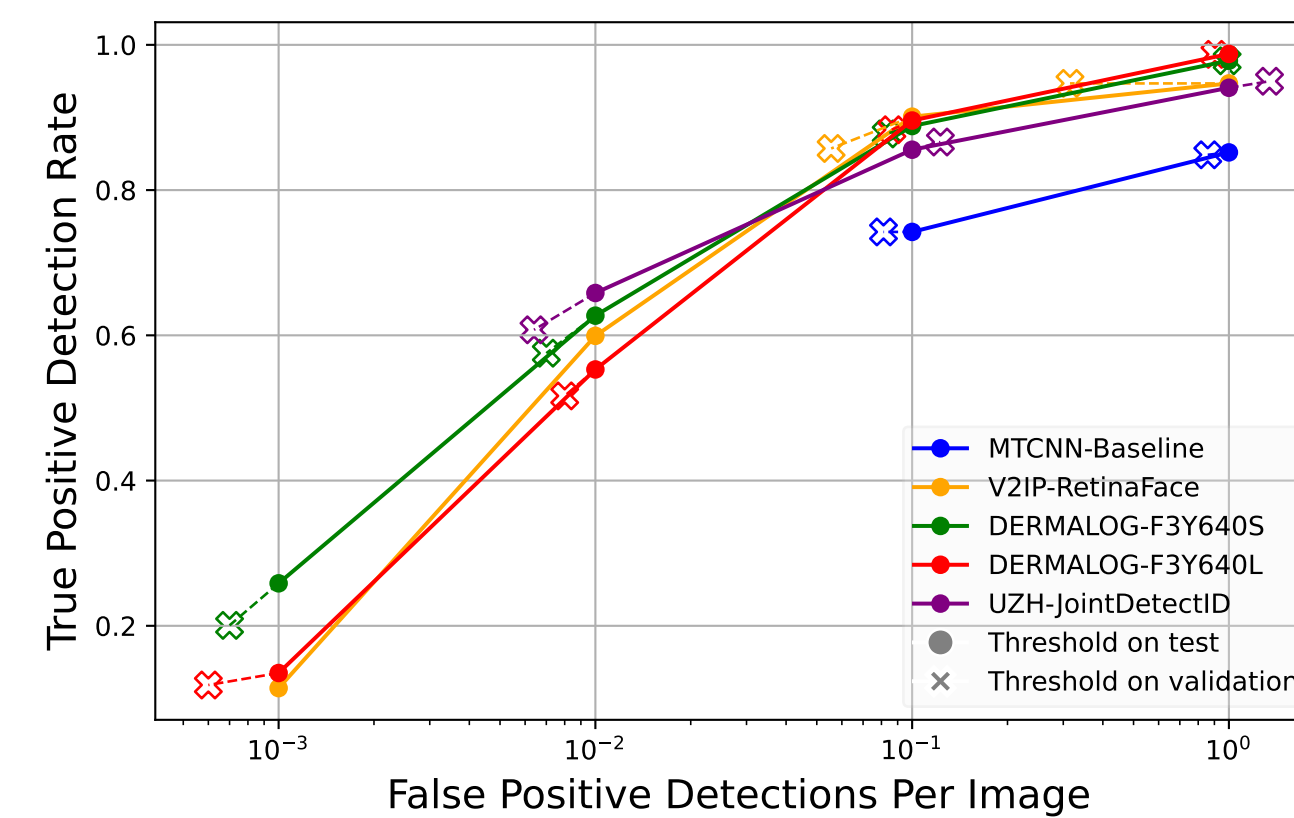


- To define rankings for both tasks, a single score to compare across all algorithms is calculated by summing TPDR and TPIR on the test set at four different thresholds:

Method	@FPDPI				Σ TPDR
	10 ⁻³	10 ⁻²	10 ⁻¹	10 ⁰	
DERMALOG-F3Y640S	0.2585	0.6271	0.8882	0.9782	2.752
DERMALOG-F3Y640L	0.1350	0.5530	0.8958	0.9875	2.5713
V2IP-RetinaFace	0.1142	0.5993	0.9011	0.9469	2.5615
UZH-JointDetectID	-	0.6583	0.8556	0.9409	2.4548
MTCNN-Baseline	-	-	0.7424	0.8519	1.5943

Method	@FPIPI				Σ TPIR
	10 ⁻³	10 ⁻²	10 ⁻¹	10 ⁰	
V2IP-AdaFace	0.1106	0.3157	0.7434	0.9208	2.0905
V2IP-MEL	0.1196	0.3261	0.7457	0.8859	2.0773
DERMALOG-F3Y640L	0.1640	0.3695	0.7226	0.7901	2.0462
DERMALOG-F3Y640S	0.1543	0.3739	0.7142	0.7871	2.0295
UZH-DaliFace	0.0361	0.2857	0.7435	0.8668	1.9321
MagFace-Baseline	0.0721	0.3135	0.6227	0.7086	1.7169
UZH-JointDetectID	0.0522	0.2275	0.4880	0.6175	1.3852
TR-EnsembET	0.0490	0.1479	0.2651	0.3840	0.8460
TR-EnsembETMN	0.0152	0.0503	0.1449	0.2671	0.4775
TR-ET	0.0019	0.0043	0.0186	0.0774	0.1022

- Thresholds for detection and recognition models, when determined on the validation set, do not always translate to test set performance, highlighting the need for more realistic evaluation metrics:



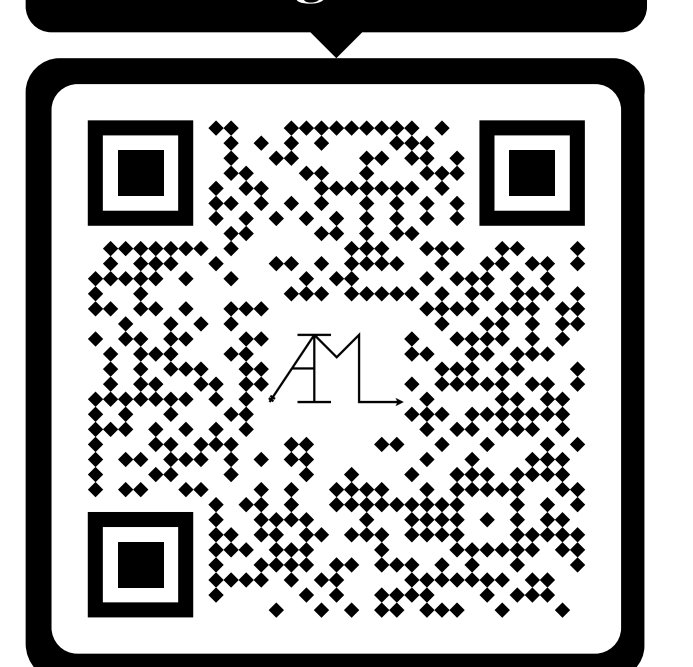
5. Conclusions

- The watchlist challenge evaluates real-world open-set face detection and recognition algorithms, providing a benchmark for assessing surveillance algorithms.
- Detection results are generally good, but some faces under extreme conditions remain undetected, highlighting areas for improvement.
- Identification models vary in performance, with some excelling at strict thresholds to reduce false positives and others achieving high TPIR at softer thresholds.
- Models pre-trained on large-scale datasets generally outperform those fine-tuned on the validation set, which show potential but need more training on diverse identities.
- Rejecting false positive detections as unknowns greatly varies across algorithms.
- Open-set performance needs improvement across all models, especially for rejecting faces of unknown identities at lower FPIPI thresholds.

References

- [1] Manuel Günther, Peiyun Hu, Christian Herrmann, Chi-Ho Chan, Min Jiang, Shufan Yang, Akshay Raj Dhamija, Deva Ramanan, Jürgen Beyrer, Josef Kittler, Mohamad Al Jazaery, Mohammad Iqbal Nouyed, Guodong Guo, Cezary Stankiewicz, and Terrance E. Boulton. Unconstrained face detection and open-set face recognition challenge. In *IJCB*, 2017.
- [2] Manuel Günther, Walter Scheirer, and Terrance E. Boulton. Open-set recognition challenge. Poster @ IAL Workshop, ECCV, 2018.
- [3] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. MagFace: A universal representation for face recognition and quality assessment. In *CVPR*, 2021.
- [4] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *Signal Processing Letters*, 2016.

Challenge Website



www.ifi.uzh.ch/en/aiml/challenge.html