IARPA Janus Benchmark – C: Face Dataset and Protocol *

Brianna Maze † Jocelyn Adams † James A. Duncan † Nathan Kalka † Tim Miller †
Charles Otto † Anil K. Jain ‡ W. Tyler Niggel † Janet Anderson † Jordan Cheney †
Patrick Grother §

Abstract

Although considerable work has been done in recent years to drive the state of the art in facial recognition towards operation on fully unconstrained imagery, research has always been restricted by a lack of datasets in the public domain. In addition, traditional biometrics experiments such as single image verification and closed set recognition do not adequately evaluate the ways in which unconstrained face recognition systems are used in practice. The IARPA Janus Benchmark–C (IJB-C) face dataset advances the goal of robust unconstrained face recognition, improving upon the previous public domain IJB-B dataset, by increasing dataset size and variability, and by introducing end-to-end protocols that more closely model operational face recognition use cases.

IJB-C adds 1,661 new subjects to the 1,870 subjects released in IJB-B, with increased emphasis on occlusion and diversity of subject occupation and geographic origin with the goal of improving representation of the global population. Annotations on IJB-C imagery have been expanded to allow for further covariate analysis, including a spatial occlusion grid to standardize analysis of occlusion. Due to these enhancements, the IJB-C dataset is significantly more challenging than other datasets in the public domain and will advance the state of the art in unconstrained face recognition.

*This research is based upon work supported by the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

†B. Maze, J. Adams, J. Duncan, N. Kalka, T. Miller, C. Otto, T. Niggel, J. Anderson, and J. Cheney are with Noblis, Reston, VA, U.S.A.

‡A. K. Jain is with Michigan State University, East Lansing, MI, U.S.A.

§P. Grother is with the National Institute of Standards and Technology (NIST), Gaithersburg, MD, U.S.A.

1. Introduction

Within the field of computer vision, one of the most prominent and well-documented research goals is the development of a system that can efficiently and accurately recognize faces in a variety of environments. Though research has shown vision systems are performing at near-human levels of face recognition accuracy on constrained face imagery, the performance of these systems still lags behind human performance on unconstrained imagery [3] [4] [15]. Because many practical applications of face recognition, e.g. surveillance, necessarily operate on unconstrained imagery, it is critical to improve unconstrained face recognition performance. A practical unconstrained face recognition system must successfully perform face detection, verification, and identification regardless of subject conditions (pose, expression, occlusion) or acquisition conditions (illumination, standoff, etc.).

In order to continue development and testing of these systems, researchers must have access to large amounts of relevant training and testing data as well as protocols to properly design and evaluate their algorithms at operationally relevant assessment points (e.g., FAR of 0.001%). While several large datasets have become available in the public domain [13], few unconstrained datasets of suitable size have been released alongside annotations and protocols that can accurately evaluate end-to-end (e.g. joint detection and recognition) systems. To remedy this deficiency, this paper introduces the IARPA Janus Benchmark–C (IJB-C) dataset, which contains a corpus of annotated, unconstrained face imagery and operationally relevant protocols to advance the state of the art in unconstrained face recognition.

1.1. Unconstrained Face Imagery

With the advent of Labeled Faces in the Wild (LFW) in 2007, research activity in unconstrained face recognition accelerated rapidly [7]. LFW was influential in the release of subsequent datasets such as PubFig, YouTube Faces (YTF), and MegaFace [8] [10] [18]. Table 1 shows a summary of current public domain face datasets available. Cur-

Dataset	# subjects	# images	avg. # img/subj	# videos	avg. # vid/subj	pose variation
IJB-C	3,531	31,334	6	11,779	3	full
IJB-B[17]	1,845	21,798	6	7,011	4	full
IJB-A [9]	500	5,712	11	2,085	4	full
LFW [7]	5,749	13,233	2	0	N/A	limited
YTF [18]	1,595	0	N/A	3,425	2	limited
PubFig [10]	200	58,797	294	0	N/A	limited
VGG [13]	2,622	982,803	375	0	N/A	limited
MegaFace [8]	N/A	1M	N/A	0	N/A	full
MF2[12]	672,057	4.7M	7	0	N/A	limited
WIDER FACE [19]	N/A	32,203	N/A	0	N/A	full
CASIA Webface [20]	10,575	494,414	47	0	N/A	limited
UMDFaces [2]	8,277	367,888	44	22,075	3[1]	full

Table 1: A comparison of IJB-C to other unconstrained face benchmark datasets. Full pose variation is defined as -90 to +90 degrees of yaw; anything less is regarded as limited pose variation. MegaFace and WIDER FACE are distractor and face detection sets, respectively, and as such do not contain subject labels. Note that IJB-C is the only dataset listed in the table that includes end-to-end protocols.

rently, recognition performance on LFW has saturated, with the best performance exceeding 99% true accept rate at a 1.0% false accept rate. Due to the limitations of the standard LFW protocol, performance cannot be reliably estimated at lower, operationally relevant FAR values.

One of the key limitations of the above datasets is that commodity face detectors such as Viola-Jones (V-J) [16] were used to collect the faces in the dataset. The V-J face detector was not designed to detect faces with significant degrees of roll, pitch, or yaw, so using such a detector to construct a dataset excludes truly unconstrained imagery from it, reducing the dataset's relevance in solving the unconstrained face recognition problem. Fig. 1 shows examples of images with constrained, frontal faces and unconstrained, non-frontal faces with V-J detections superimposed over the ground truth bounding boxes included in the IJB-C release. Note that the V-J detector misses the vast majority of faces in the right-hand image.

The MegaFace and MF2 datasets were constructed using the HeadHunter algorithm to detect faces within potential media [8] [12]. MegaFace includes one million faces and is intended to be used only as a distractor set, whereas MF2 has 672K unique identities but is intended to be used only as a training set. The identity labels in MF2 are noisy, and no evaluation protocols are included.

WIDER FACE, a large scale face detection dataset released in 2016, made significant strides towards addressing the data quantity problem associated with evaluation of face detection algorithms [19]. However, utility of this dataset is limited to advancing face detection only, since subject identity labels are not provided. Similarly, the UMDFaces dataset, which includes images and frames for 8,277 subjects, only includes face verification protocols and could not be used for a full evaluation pipeline [2]. UMDFaces also has no clear authority for redistribution.

The release of the NIST Face Challenge and the IARPA Janus Benchmark–A (IJB-A) dataset in 2015 marked a milestone in unconstrained face recognition research [6][9]. When released, results from multiple submissions to the challenge showed significantly worse recognition performance compared to the previously mentioned datasets. As of 2017, performance on IJB-A is approaching saturation, with a top true accept rate of 96.1% at a 1.0% false accept rate [6].

The successive dataset, IARPA Janus Benchmark–B (IJB-B), released in 2017, continued to push the state of the art in unconstrained face recognition [17]. It included 1,845 subjects and protocols supporting face detection, verification, recognition, and clustering. While this dataset allowed for evaluation at more operationally relevant points at low ends of the ROC curve (e.g. FAR at 0.01% and 0.001%), the dataset did not support the evaluation of end-to-end systems. To allow evaluation of such an end-to-end system, a dataset is needed to support more operationally relevant protocols. We present the IJB-C dataset to address this need.

1.2. Paper Organization

The remainder of this paper is organized as follows: Section 2 descibes the IJB-C dataset and the collection methodology used to curate it, in detail. Section 2.2 describes the protocols released with IJB-C and the performance metrics to be reported on each. Section 3 reports benchmark results from multiple algorithms on the provided protocols. Finally, discussion and conclusions are provided in Section 4.

2. IJB-C Dataset

The IARPA Janus Benchmark–C (IJB-C) dataset contains Creative Commons[2] licensed face imagery and video

[1] UMDFaces does not guarantee a video for every subject, and the average listed above is the average number of videos across all subjects within the dataset.

[2] Creative Commons licenses allow for free distribution of content, provided all attributions and licensing conditions are met. Images within this paper are listed with the image author's name, in accordance with the Creative Commons licenses. Full attributions are listed in the supplementary material.

Figure 1: Example detections from Viola-Jones shown in purple, and ground truth detections from the IJB-C dataset shown in green on an image with frontal faces (left) and non-frontal faces (right).

for 3,531subjects, an addition of 1,661 subjects to IJB-B[3]. All subjects in the dataset are ensured to appear in at least two still images and one video. The bounding boxes and metadata labels were all labeled using the crowdsourcing platform Amazon Mechanical Turk (AMT). Subject names were deconflicted using fuzzy matching to ensure that subjects in IJB-C are disjoint from those included in the VGG-Face and CASIA datasets [13][20].

IJB-C includes a total of 31,334 (21,294 face and 10,040 non-face) still images, averaging to ∼6 images per subject, and 117,542 frames from 11,779 full-motion videos, averaging to ∼33 frames per subject and ∼3 videos per subject. The contributions of the IJB-C dataset to face recognition and biometrics communities are the following:

- Subjects with full variation in pose.
- Subjects with diverse occupations, avoiding one pitfall of "celebrity-only" media, as people in occupations strongly associated with physical appearance, such as actors and performers, may be less representative of the global population.
- Image- and frame-specific metadata annotations, including detailed information about occluded areas of the face.
- Protocols for face detection, 1:1 verification, 1:N identification (supporting open- and closed-set evaluation), clustering, and end-to-end system evaluation.
- Benchmark accuracy measures from a Government-Off-The-Shelf (GOTS) algorithm and state-of-the-art face recognition algorithms that utilize deep neural networks.
- Stable download of images which remains consistent over time, unlike datasets consisting of links which are subject to change.
- Clear authority for redistribution through several Creative Commons licensing variants.

Sample imagery from IJB-C and other datasets can be seen in Fig. 2, and example imagery for each geographic region represented in IJB-C is presented in Fig. 3.

[3]Similar to IJB-A and IJB-B, IJB-C will be made available in the public domain.

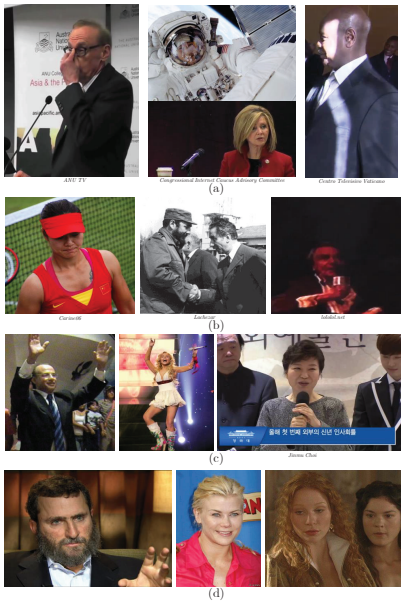

Figure 2: Sample imagery included within (a) IJB-C; (b) IJB-B; (c) IJB-A; and (d) LFW datasets. Note the greater variations in IJB-C imagery, especially occlusions.

2.1. Collection Methodology

Collection for the dataset began by identifying Creative Commons subject videos, which are often more scarce than Creative Commons subject images. Search terms that resulted in large quantities of person-centric videos (e.g. "interview") were generated and translated into numerous languages including Arabic, Korean, Swahili, and Hindi to increase diversity of the subject pool. Certain YouTube users who upload well-labeled, person-centric videos, such as the World Economic Forum and the International University Sports Federation were also identified. Titles of videos pertaining to these search terms and usernames were scraped using the YouTube Data API and translated into English using the Yandex Translate API[4]. Pattern matching was performed to extract potential names of subjects from the translated titles, and these names were searched using the Wikidata API to verify the subject's existence and status as a

[4]http://translate.yandex.com/

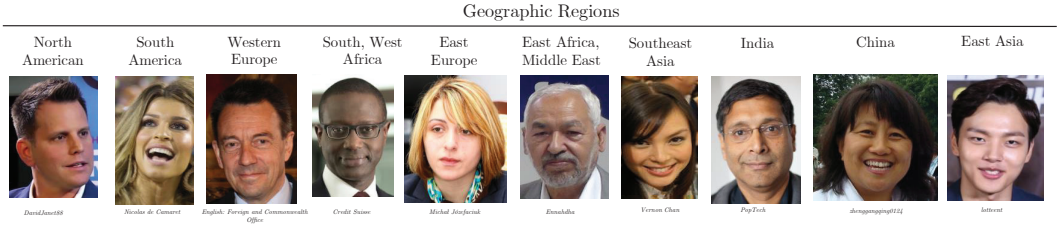

Figure 3: Examples of subjects included in IJB-C from various geographic regions.

public figure, and to check for Wikimedia Commons imagery. Age, gender, and geographic region were collected using the Wikipedia API.

Using the candidate subject names, Creative Commons images were scraped from Google and Wikimedia Commons, and Creative Commons videos were scraped from YouTube. After images and videos of the candidate subject were identified, AMT Workers were tasked with validating the subject's presence throughout the video. The AMT Workers marked segments of the video in which the subject was present, and key frames were then extracted with this information using FFmpeg[5].

AMT Workers were tasked with annotating bounding box locations for all faces in the imagery, after which the person of interest was identified. Metadata attributes were labeled by the AMT Workers on a per-media or per-subject basis. For more information about this process, see [17].

A new grid-based annotation approach to occlusion was introduced in IJB-C to allow for more fine-grained occlusion analysis. AMT Workers were tasked with labeling specific regions of the face that were fully occluded by objects such as glasses or microphones. Fig. 4 shows an example image and grid the AMT Workers were responsible for labeling based on the subject of interest's face. For a full list of metadata labels included within IJB-C, see Fig. 5.

To address the fact that crowd-sourced bounding boxes are often somewhat noisy, IJB-C also introduces the concept of ignore flags. After the initial annotation collection, at least 3 AMT Workers examined each image and flagged any they believed had errors, such as oversized, missing, or extra boxes. If any of the reviewers observed an error in the image's bounding boxes, that was denoted by an "ignore" flag on the image.

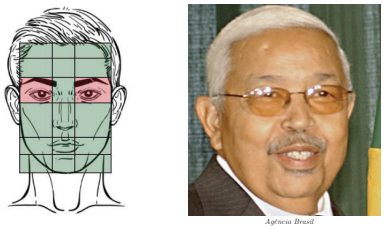

Figure 4: Example of labeled occlusion grid. In this example, the subject of interest (right) has full eye occlusion, so the areas of the grid covering the eyes are marked. Glasses (tinted or clear) are considered occlusion.

[5] www.ffmpeg.org

In total, over 9.7 million manual annotations were performed by AMT Workers determined to be reliable through qualification tests. This annotation process produced (i) an accurate ground truth corpus of imagery containing bounding boxes for face detection evaluations, (ii) subject labels for face recognition and clustering evaluations, and (iii) attribute metadata for understanding the effects of different covariates (occlusion, facial hair, gender, capture environment, skin tone, age, and face yaw) on an algorithm's facial recognition performance.

2.2. Protocol Description

IJB-C is released along with 8 protocols testing face detection, verification, recognition, and clustering at various scales. Throughout these protocols, IJB-C utilizes the concept of subject-specific modeling, in which a single template is generated for a subject based upon the available pieces of media – a paradigm shift from the traditional process of creating a template for every available piece of media (e.g., still images and frames). This is a more operationally relevant approach to the problem of face recognition. Also, by utilizing multi-image templates, the inherent difficulty of IJB-C is obfuscated, since algorithms have the ability to pool information from multiple pieces of media through subject-specific modeling.

Table 2 outlines key statistics of the protocols released with IJB-C. Due to space constraints, the focus of this paper will be on the novel protocols, but the following serves as a brief overview of the remaining protocols:

- The face detection protocol tests an algorithm's ability to detect faces, and is augmented with an additional 10,040 non-face images to test operationally relevant use cases. All media includes an ignore flag that identifies media with potential bounding box issues.
- The baseline 1:1 mixed verification protocol tests an algorithm's abilities in subject verification scenarios.
- The covariate 1:1 verification protocol utilizes single-image templates to allow further analysis of an algorithm's performance on individual covariates.
- The baseline 1:N mixed recognition protocol tests an algorithm's abilities in identification scenarios.
- The clustering protocol, which includes 4 subprotocols ranging from 32 to 3,531 subjects, tests an algorithm's ability to cluster faces at different scales.

Figure 5: Annotation Labels included within IJB-C

		Num. of Subjects	Num. of Templates	Still Images and Frames
1:N Identification	Gallery-G1	1,772	1,772	5,588
	Gallery-G2	1,759	1,759	6,011
	Mixed	3,531	19,593	127,152
1:1 Verification	Verify	3,531	23,124	138,836
	Verify-Cov	3,531	140,739	138,836
Face Detection	Face Partition	>3,531	N/A	138,836
	Non-face Partition	0	N/A	10,040
Clustering	Clustering-32	32	955	955
	Clustering-1024	1,021	41,074	40,709
	Clustering-1845	1,839	71,392	70,056
	Clustering-3545	3,531	140,623	138,136
End-to-End	End-to-End Probe Still Images/Frames	–	–	137,275
	End-to-End Probe Video	–	N/A	11,739
	End-to-End Probe Mixed	–	N/A	31,415

Table 2: Overview of the different protocols developed for IJB-B. Verify-Cov refers to Covariate Verification; Clustering-X denotes clustering with "X" number of subjects. 1:N identification supports both closed- and open-set evaluations.

Included with the protocols are two disjoint galleries, gallery 1 (G1) and gallery 2 (G2). Each gallery contains one template per subject, created by randomly selecting half of the subject's still imagery. The remaining media instances are reserved for the probe set. G1 includes 1,772 subjects with 5,588 still images, and G2 includes 1,759 subjects with 6,011 still images. These galleries are disjoint from each other so that open-set identification scenarios (i.e. where probe templates do not have a mate in the gallery) can be tested.

Receiver Operating Characteristic (ROC) curves are reported for the 1:1 verification protocols, while Cumulative Match Characteristic (CMC) and Identification Error Tradeoff (IET) curves are reported for the 1:N mixed recognition protocol [11]. BCubed precision, recall, and F-measure values are reported for the clustering protocol [1]. For more detail about the above protocols and their evaluation metrics used, please see [17].

2.3. End-to-End Protocols

An important feature that differentiates IJB-C from previous releases of this series is the inclusion of 3 end-to-end protocols, designed to test an algorithm's ability to perform end-to-end face recognition. The end-to-end protocols utilize galleries G1 and G2 (discussed in Section 2.2) for evaluation.

2.3.1 1:N End-to-End Still

For the 1:N end-to-end still protocol, the algorithm is responsible for detecting faces in the image and searching each detected face against galleries G1 and G2. This resembles the operational work performed today by law enforcement. The 1:N end-to-end still protocol includes 136,734 still images and frames.

2.3.2 1:N End-to-End Video

A 1:N end-to-end video protocol is also provided. This protocol follows the same outline as the above 1:N still protocol, but instead uses full-motion videos as input. The protocol includes 11,779 videos. For evaluating this protocol, frames are weighted such that all of a subject's frames from a single video have the same combined weight as a single still image.

2.3.3 1:N End-to-End Mixed

The 1:N end-to-end mixed protocol is designed to test an algorithm's ability to perform end-to-end recognition tasks. The protocol contains both still images and full-motion videos, for a total of 31,415 pieces of media. The algorithms are responsible for detecting faces in still images and videos, clustering the detected faces, creating a multi-image template from each cluster and then searching against the galleries. Frames in this protocol are weighted as described in Section 2.3.2.

2.3.4 Performance Metrics for End-to-End Protocols

Performance on the end-to-end protocols is evaluated according to two metrics: (i) End-to-End Retrieval Rate (EERR) and (ii) a variant on the Identification Error Tradeoff (IET) [17].

The EERR evaluates accuracy in a closed-set identification scenario relative to rank. Similar to the CMC used in earlier protocols, the EERR expresses the proportion of mated searches returning a match at or above a particular rank, where a mated search is defined as having a corresponding mate in the gallery. For end-to-end probes, how-

ever, there are two scenarios in which a mated search may result in a miss: (1) the face of the subject of interest was not detected, or (2) the face of the subject of interest was detected but the resulting candidate list did not contain the mate. A correct detection is defined as a (normalized) predicted bounding box which has an intersection over union score of at least 50% with the ground truth bounding box. Predicted boxes are rigidly normalized by increasing or decreasing the area of the predicted box until it matches the area of the ground truth bounding box. If the percent difference of the normalized box area and original predicted box area is greater than 150%, then the predicted boxes are considered to be a false alarms.

The IET variant (henceforth referred to as the weighted IET) expresses how the False Positive Identification count (FPI) varies with respect to the FNIR. FPI is the number of non-mated probe searches that return a candidate at rank one with a score greater than a threshold, *t*. FNIR is the proportion of mated searches that do not return the mated gallery template at or above the same threshold *t*. Unlike in earlier IJB protocols, the false positives are not normalized by the proportion of non-mated searches. Otherwise, algorithms could configure their detector to have many more false detections, thus lowering their FPIR.

2.4. Synthetic Degradation Experiments

To further analyze performance in the presence of various media artifacts, experiments were conducted on synthetically degraded media. Performance on these protocols was evaluated before and after rotation, blur, and resolution changes were applied. An example image is shown below in Fig. 6 before and after each degradation type was applied. To test the impact of rotation on face detection, images had an equal chance of being rotated 90° in either direction. Blur experiments were run using the OpenCV Gaussian blur function with mean 0 and standard deviation parameters of 2, 5, and 10. In order to maintain a minimum bounding box side size of 20 pixels for resolution experiments, resizing levels were selected on a per-image basis as fractions (0, 0.33, and 0.66) of the range between the reduction level that would reach this minimum size (0), and the original size of the image (1). For blur and resize testing, degradation was only applied to images in the probe set in order to enable comparisons between original and degraded media. The scripts used to synthetically degrade the media will be released along with IJB-C to facilitate repeatability and reproducibility.

3. Experimental Results

Baseline results for select protocols are shown below. Due to space limitations, only benchmark results for face detection, 1:1 mixed, 1:N mixed, and 1:N end-to-end still

Figure 6: Examples of original and synthetically degraded media. From left to right: the original, blurred at a standard deviation parameter of 10, rotated 90°, and resized by the minimum factor 0.

are discussed. Results from all protocols released with IJB-C are available from the authors.

We do not provide a detailed meta-analysis of confounding factors that impact face recognition performance due to space constraints. However, it should be noted, as illustrated in [6], extreme pose continues to be a primary confounding factor of performance in the IJB datasets. Specifically, the difference in yaw between subject's pose in compared templates is a limiting factor in face recognition performance. In other words, face recognition algorithms perform best when the two compared templates closely resemble one other. This also holds true across the dataset for confounding factors such as occlusion and "environment". The authors are preparing a full manuscript outlining the IJB-C confounding factors and their impact on face recognition performance.

3.1. Face Detection

Three baseline algorithms were used to test the IJB-C face detection protocol. First, two government-off-the-shelf (GOTS) algorithms were utilized. The GOTS algorithms were both designed specifically to detect faces in unconstrained imagery, and are shown to be the top performing face detectors in a recent face detection benchmark [5]. Secondly, we report performance from a TensorFlow implementation of a multi-task cascaded convolutional neural network (MTCNN) [21][6]. Results are provided in Table 3 for all media not containing an ignore flag, as well as for all media inclusive of flagged images and frames. Note that the GOTS-1 has the best performance of 49.0% true detect rate (TDR) at a false detect per image (FDPI) of 10^{-2} on all media, while the MTCNN algorithm has the best performance of 66.7% TDR at a FDPI of 10^{-2} when only media containing no bounding box errors was used.

3.2. 1:1 Mixed Verification

To test the 1:1 mixed identification protocol, we report performance from three baseline algorithms. We utilize a GOTS algorithm previously mentioned in Section 3.1.

[6]The implementation of the face detector can be found at `github.com/davidsandberg/facenet/tree/master/src/align`

	With Flag TDR (%)		Without Flag TDR (%)	
	FDPI 10^{-1}	FDPI 10^{-2}	FDPI 10^{-1}	FDPI 10^{-2}
GOTS-1	68.0	49.0	71.7	63.2
GOTS-2	75.0	41.7	78.7	66.4
MTCNN	81.0	41.1	86.0	66.7

Table 3: True detect rates (TDR) at operating points of 10^{-1} and 10^{-2} false detects per image (FDPI) for the benchmark algorithms. "Without flag" refers to the performance of the algorithm on only media containing no bounding box errors, while "With flag" refers to the performance of the algorithm on all media within IJB-C.

We also report performance from an implementation of Google's FaceNet[7], which was shown to achieve a 98.7% accuracy on LFW [14]. Finally, we report performance of a VGG CNN model described in [13]. To handle the multi-image IJB-C templates, a single feature vector was composed using a weighted average of the images in the template, such that all frames belonging to the same subject within a video have a combined weight equal to a single still image. Results are illustrated in Figure 7. It is illustrated that the VGG-CNN algorithm provides the best performance at all operating points. Note that this protocol provides 19,557 genuine matches and 15,638,932 impostor matches to allow for evaluations of performance at low FAR values.

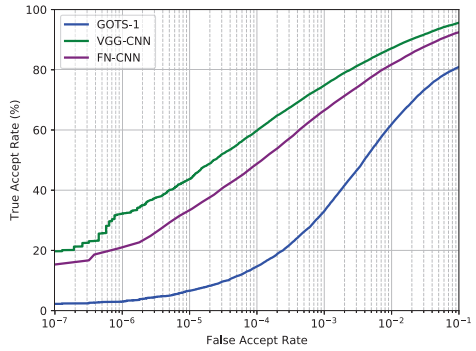

Figure 7: Average ROC performance across gallery sets G1 and G2 for the 1:1 Mixed Verification protocol.

3.3. 1:N Mixed Identification

To evaluate the 1:N mixed identification protocol, the benchmark algorithms described in Section 3.2 are utilized. CMC results for this protocol can be see in Fig. 8, and the IET results can be seen in Fig. 9. Note that the VGG-CNN algorithm provides the best CMC and IET performance in comparison to the FN-CNN and GOTS-1 algorithms.

[7]The implementation of the face recognizer can be found at `https://github.com/davidsandberg/facenet`.

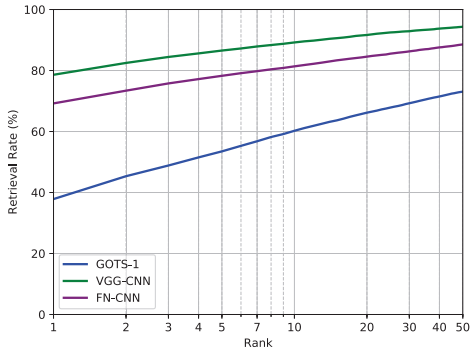

Figure 8: Average CMC performance across gallery sets G1 and G2 for the 1:N Mixed Media Identification protocol.

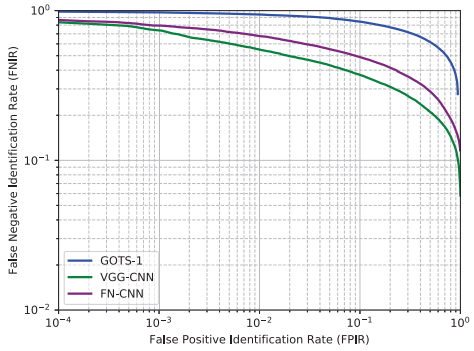

Figure 9: Average IET (Identification Error Tradeoff) performance across gallery sets G1 and G2 for the 1:N Mixed Media Identification protocol.

3.4. 1:N End-to-End Still

For the 1:N end-to-end still protocol, we utilize the same three baseline algorithms as described in Section 3.3. The MTCNN detector is utilized for both VGG and FaceNet CNN algorithms. The EERR for this protocol can be seen in Fig. 10, and the weighted IET can be see in Fig. 11. As show in the figures, overall performance results are lower across the board. This is a challenging protocol and performance is lower without subject specific modeling.

4. Summary

In this paper, the authors have introduced a new publicly available face dataset, IARPA Janus Benchmark–C (IJB-C), as an extension to the previously released IJB-B dataset. IJB-C focuses on unconstrained media, and includes 3,531 subjects, with a total of 31,334 images (21,294 face and 10,040 non-face), and 117,542 frames pulled from 11,779 full-motion videos. All media has manually annotated facial bounding boxes and covariate labels, including spatial occlusion results. This dataset includes subjects from diverse occupations, which increases the inherent variability of subject appearance and environment when compared to easily accessible celebrity-only media. All images within

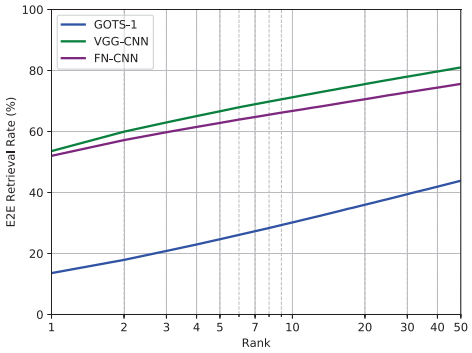

Figure 10: Average EERR performance across gallery sets G1 and G2 for the 1:N End-to-End Still Identification protocol. The end-to-end (E2E) retrieval rate on the y-axis indicates the proportion of mated searches returned at or above a rank, incorporating misses from failed bounding box association.

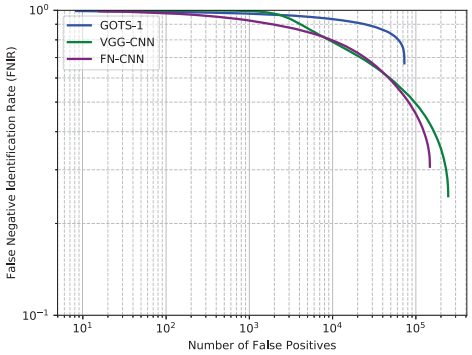

Figure 11: Average weighted IET performance across gallery sets G1 and G2 for the 1:N End-to-End Still Identification protocol.

the dataset can be publicly redistributed through Creative Commons licensing. IJB-C is the first dataset that provides end-to-end protocols based on operationally relevant use-cases to evaluate an algorithm's combined performance on detection, clustering and recognition. Along with the dataset, benchmark results from two GOTS algorithms and two academic algorithms are released to be used for comparative research. The IJB-C dataset is available through the NIST Face Projects website[8].

References

[1] E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval*, 12(4):461–486, 2009.

[2] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa. Umdfaces: An annotated face dataset for training deep networks. *arXiv preprint arXiv:1611.01484v2*, 2016.

[3] L. Best-Rowden, S. Bisht, J. C. Klontz, and A. K. Jain. Unconstrained face recognition: Establishing baseline human performance via crowdsourcing. In *IEEE IJCB*, pages 1–8, 2014.

[4] A. Blanton, K. C. Allen, T. Miller, N. D. Kalka, and A. K. Jain. A comparison of human and automated face verification accuracy on unconstrained image sets. In *IEEE CVPR Workshop on Biometrics*, 2016.

[5] J. Cheney, B. Klein, A. K. Jain, and B. F. Klare. Unconstrained face detection: State of the art baseline and challenges. In *IEEE ICB*, pages 229–236, 2015.

[6] P. Grother and M. Ngan. The IJB-A face identification challenge performance report. Technical report, National Institute of Standards and Technology, 2017.

[7] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, 07-49, University of Massachusetts, Amherst, 2007.

[8] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard. The MegaFace Benchmark: 1 million faces for recognition at scale. In *IEEE CVPR*, pages 4873–4882, 2016.

[9] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In *IEEE CVPR*, pages 1931–1939, 2015.

[10] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In *IEEE ICCV*, Oct 2009.

[11] S. Z. Li and A. K. Jain (Eds). *Handbook of Face Recognition, Second Edition*. Springer, 2011.

[12] A. Nech and I. Kemelmacher-Shlizerman. Level playing field for million scale face recognition. In *IEEE CVPR*, 2017.

[13] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.

[14] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.

[15] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE CVPR*, pages 1701–1708, 2014.

[16] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[17] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, J. Cheney, and P. Grother. IARPA Janus Benchmark-B face dataset. In *IEEE CVPR Workshop on Biometrics*, July 2017.

[18] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *IEEE CVPR*, pages 529–534, 2011.

[19] S. Yang, P. Luo, C. C. Loy, and X. Tang. Wider face: A face detection benchmark. In *IEEE CVPR*, 2016.

[20] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014.

[21] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, Oct 2016.

[8] https://www.nist.gov/programs-projects/face-challenges