

Supplementary Material for "Face Verification with Challenging Imposters and Diversified Demographics"

Adrian Popescu¹, Liviu-Daniel Ștefan², Jérôme Deshayes-Chossart¹, Bogdan Ionescu²

¹Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

²University Politehnica of Bucharest, Romania

{adrian.popescu, jerome.deshayes-chossart}@cea.fr, {liviu_daniel.stefan, bogdan.ionescu}@upb.ro

1. Introduction

In this supplementary material, we provide:

- An illustration of random and challenging imposters.
- Details about the annotation task and the interface used to validate genuine pairs in *FaVCI2D*.
- The distribution of genuine identities across countries.
- The detailed results obtained after the ablation of a percentage of genuine identities.

2. Illustration of imposters

Name	LFW			FaVCI2D		
	Genuine 1	Genuine 2	Random imposter	Genuine 1	Genuine 2	Similar imposter
Mireya Moscoso						
Akbar Hashemi Rafsanjani						
Angela Merkel						

Figure 1. Illustration of genuine and imposter pairs from LFW and *FaVCI2D*. LFW pairs are created randomly. Images in *FaVCI2D* genuine pairs are selected as different from one another among available faces. *FaVCI2D* imposters are samples of an identity which is visually similar to the genuine identity.

Random and challenging imposters are presented in Figure 1 with examples from LFW and from our dataset proposed in this article, *FaVCI2D* (Face Verification with Challenging Imposters and Diversified Demographics), respectively. The provided examples illustrate the fact that random imposters make the verification task too easy because the two faces from the random pairs are visually different from one another. The selection of similar imposters provides a more realistic and challenging testbed for face

verification. Note also that the annotators are asked to choose challenging imposters whenever there are more than two faces available for an identity.

3. Face annotation task

The annotation of genuine pairs was first performed by one participant. Two versions of the interface were created depending on the availability of age related information for each identity (person). We present the full instructions and interface, which include age-related information, in Figure 2. A similar annotation process was implemented for genuine pairs which did not have age-related information associated to them. In this case, the age-related instructions and associated parts of the interface were naturally omitted.

The participant who started the validation of genuine pairs received the following instructions: "Your participation is needed to create a dataset for face verification, a task whose objective is to automatically detect whether two faces represent the same identity or not. Each page of the interface contained the following elements:

- Up to twelve reference faces which were automatically ranked as representative for the tested identity. These faces are presented in order to facilitate the evaluation of candidate faces. Note that, in some cases, a minority of reference faces might be irrelevant and they should be ignored.
- The birth year of the person which is useful to evaluate if the age associated to candidate faces is realistic.
- At least two candidate faces with associated date taken information. The sources of the date taken information is indicated: EXIF, URL or ALT description. You should select two representative faces which, in addition of being relevant for the identity should: (1) be as different as possible from each other (face orientation, hairstyle, presence of glasses, beard etc.) and (2) have an age difference which is underrepresented in the bins

Task: Select two representative faces for the current identity. Give priority to images which are visually different from each other (face orientation, age, hairstyle, glasses etc.) and to images which have an age difference in one of the underrepresented bins

Top ranked images for 1009205

Birth year : 1981



Age difference stats



Candidate faces	Date info
	URL 2017
	URL 2012
	EXIF 2009
	EXIF 2014
	EXIF 2010

Save Ignore

Count : 4457/4457

Figure 2. Illustration of the interface used to collect visually challenging pairs for $FaVCI2D$.

to the right. Note that face selection is validated when its background turns to green.

- Age-related statistics capture the number of identities in each age difference bin. The objective is, to the extent possible, to have a balanced representation of these difference.
- The "Save" button allows you to validate a selected pair of faces and move to the next identity.
- The "Ignore" button allows you to skip the current identity if the candidate examples are not representative of the identity. This can be due to: (1) uncertainty about their relevance for the current identity or (2) wrong assignment of the date taken information."

The essential elements of these instructions were then reminded for each identity. The annotation interface deployed for the annotation of genuine pairs is illustrated in Figure 2. In the example, the two selected faces differ in terms of face orientation and expression. They are also distant in time since the first candidate photo was taken in 2017, while the second was taken in 2009. This age difference also con-

Africa		America	
South Africa	239	United States	2100
Nigeria	213	Canada	396
Egypt	97	Brazil	322
Ghana	83	Argentina	267
Tunisia	68	Mexico	242
Asia		Europe	
India	917	United Kingdom	576
Japan	673	France	353
South Korea	540	Germany	341
Philippines	216	Italy	290
China	209	Ireland	249

Table 1. Distribution of the number of genuine pairs for the most represented five countries in each region of the world included in $FaVCI2D$. The total number of genuine pairs is 12468

tributes to the creation of a challenging genuine pair of images because faces tend to change over time.

Each genuine pair selected using the interface from Figure 2 is then verified by two more annotators using a simpler interface which displays only the preselected candidate faces and their age when this information is available. The

two participants were instructed to validate pairs only if they were certain that both faces represented the same identity. The pair is retained only if all three annotators agree that the two selected candidates are relevant.

The selection of most challenging imposters (Similar = 1 in Table 2 of the main paper) should be verified since the two identities can in some rare case be the same. This situation can be an effect of the selection of: (1) fictional characters illustrated by images of the real-life person which is also included in the dataset or (2) from an erroneous selection of reference images for the identity. To make sure that imposter pairs actually contained faces of different identities, we again displayed the pair of faces. Three participants were asked to validate that the two candidate faces belonged to different identities. An imposter pair was retained only if the candidate faces were attributed to different identities by all three participants.

Model	Ablation		
	50%	25%	0%
<i>insightface</i>	95.69 ± 0.05	95.72 ± 0.04	95.75 ± 0.0
<i>ir152</i>	89.43 ± 0.04	89.45 ± 0.04	89.48 ± 0.0
<i>seqface</i>	85.67 ± 0.02	85.59 ± 0.02	85.61 ± 0.0
<i>vgg</i>	85.03 ± 0.09	85.19 ± 0.07	85.28 ± 0.0
<i>facenet</i>	82.54 ± 0.05	82.66 ± 0.02	82.61 ± 0.0

Table 2. Verification accuracy for the ablation of 50% and 25% of genuine IDs (pairs) from the dataset. Results with the full dataset (0%) are also provided for reference. Results are reported with challenging imposters (Similar=1) selected among 52,410 IDs. The ablation was performed by randomly selecting five subsets of IDs and the reported results are averaged.

4. Distribution of genuine pairs across countries

We discuss the distribution of genuine pairs a selection of top countries *FaVCI2D*. The selected countries correspond those for which results were presented in Table 5 of the main paper. The distribution of identities per country is presented in Table 1. We note that while it was possible to ensure a balanced representation of each large region of the world, imbalance subsists for individual countries. This is mainly an effect of the highly imbalanced character of the frequency of country-related identities in Wikipedia. For instance, nearly a half of initial number of identities with associated photos were from the United States, a percentage which is reduced to less than 17% of genuine pairs kept in *FaVCI2D*. While surprising, the relatively low number of pairs from China is explained by the fact that this coun-

try has a highly skewed gender distribution in Wikipedia. Nearly 90% of Wikipedia identities with images describe Chinese men. Inversely, it is somewhat surprising to note the relatively large number of Irish identities in the dataset. This situation is probably explained by the fact that Ireland is an English-speaking country and has an active community of Wikipedia contributors.

5. Results for ablated dataset

The results from Table 2 complete the ablation study mentioned in Subsection 4.2 of the main paper. We ablate 50% and 25% of the total number of genuine identities from the datasets to verify that the number of unique IDs is sufficient. The results confirm this hypothesis since the variation between the ablated datasets and its full version are small. The maximum difference is obtained for *vgg* when 50% of IDs are removed and amounts to 0.25%. The differences are under 0.1% for all other models tested in Table 2. Note also that, intuitively, the difference is smaller when the number of removed identities is 25% compared to 50%.