Supplementary Material:
DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman Ming Yang Marc'Aurelio Ranzato Lior Wolf

Facebook AI Research
Menlo Park, CA, USA
{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University
Tel Aviv, Israel
wolf@cs.tau.ac.il

Illustration of the *SFC* dataset

Examples of images assigned to the same identity are shown in Fig. 1, demonstrating the *SFC* dataset. Note that typically there is about 3% error in labeling.

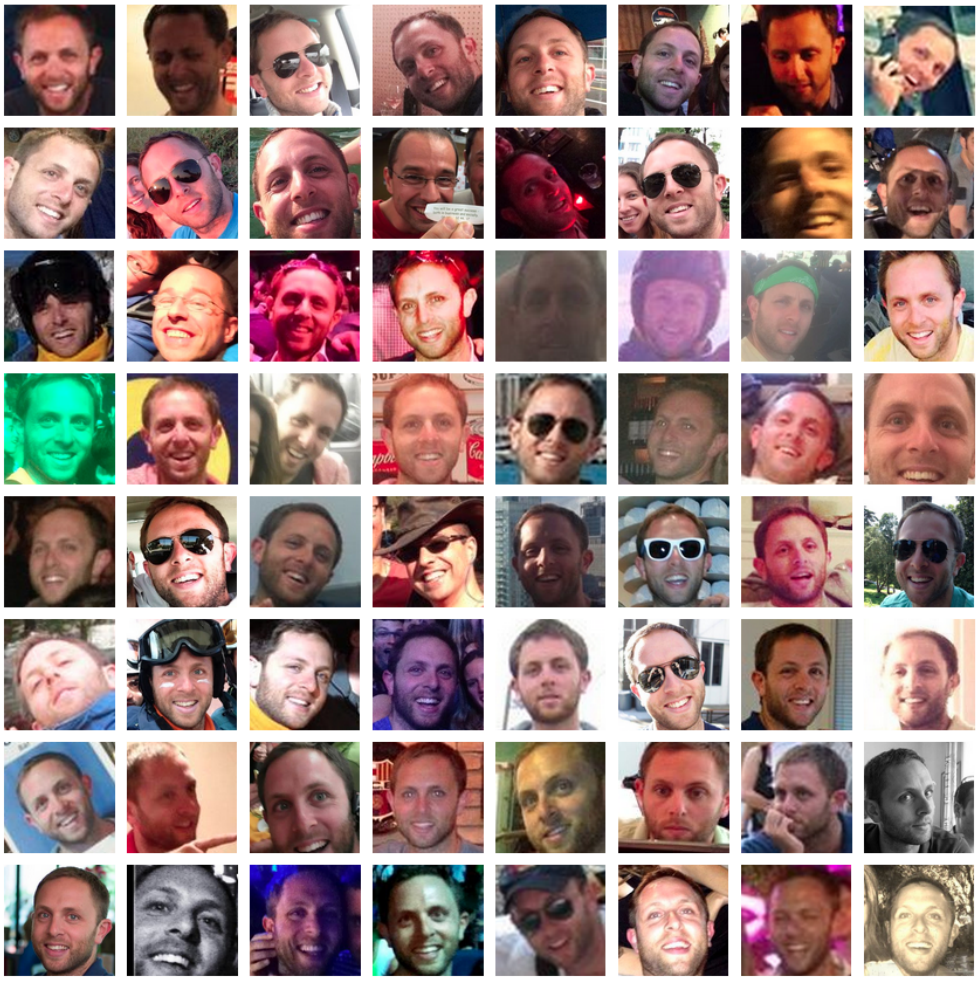

Figure 1. Samples from *SFC* (permission granted).

Error visualization on the *LFW*

For our best performing result (**97.35**%) on *LFW*, given by *DeepFace-ensemble*, there are 99 false negatives and 60 false positives, all shown in Fig. 2 and Fig. 3.

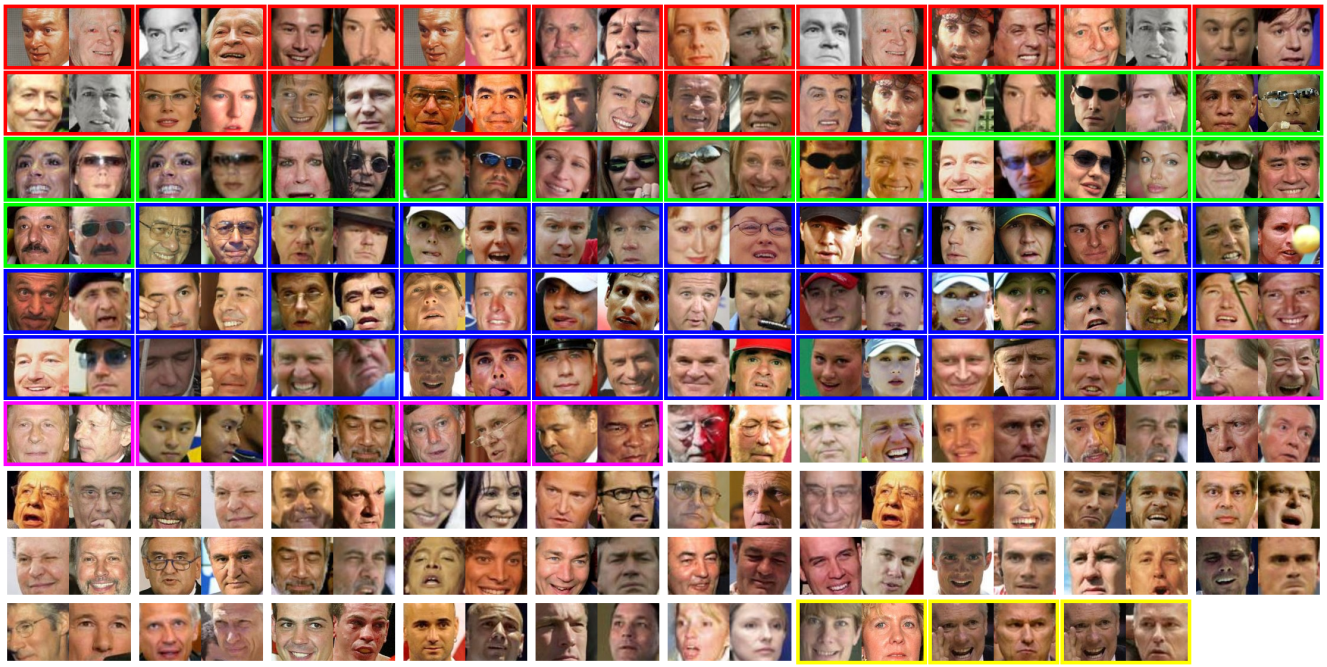

Figure 2. The **99 False-Negative** pairs (1.65%) on *LFW* by *DeepFace-ensemble*. Borders were colored by the authors according to their *belief* with respect to what contributed to the misses: Aging (red), Sunglasses (green), Occlusions/hats (blue), Profile faces (purple), Dataset errata (yellow), and assorted (white). Note however, that since *LFW* is a collection of celebrities, then most likely that human were well "trained" to recognize them, in contrast to algorithms that report accuracy on *LFW* protocols, which are not provided with training samples of the same identities they are tested upon.

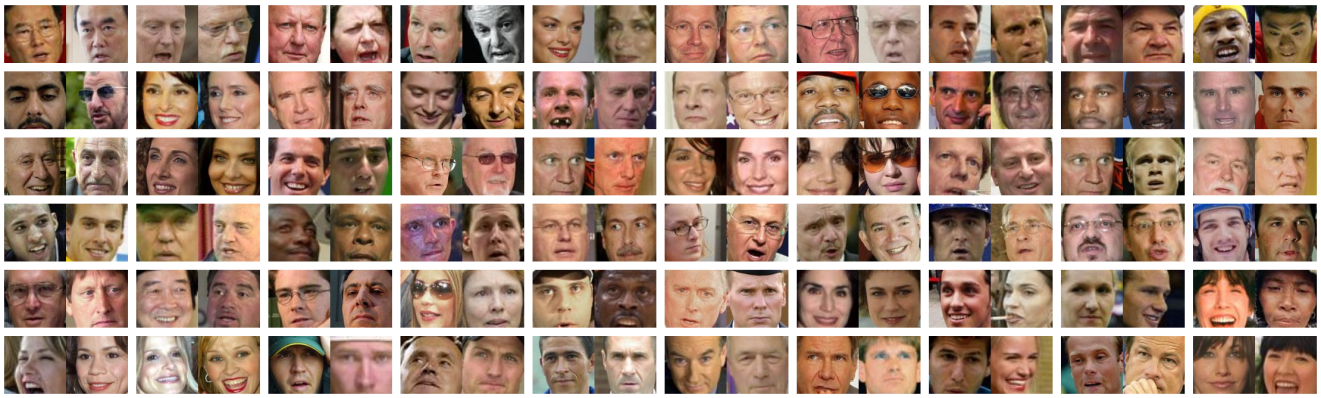

Figure 3. The **60 False-Positive** pairs (1.00%) on *LFW* by *DeepFace-ensemble*.

Error visualization on the *YTF*

For our best performing result with $\mathbf{92.5}\%$ on the *YTF* by *DeepFace-single*, there are 184 false negatives and 189 false positives. Samples are shown in Fig. 4 and Fig. 5.

Figure 4. Examples of false negatives on the *YTF* by *DeepFace-single*.

Figure 5. Examples of false positives on the *YTF* by *DeepFace-single*.

Feature vector sparsity visualization

Fig. 6 illustrates the sparsity of the face representation. The sparsity level measures the ratio of zeros out of the 4096 dimensions, i.e., a sparsity level of 80% corresponds to 3277 zeros and only 819 non-zero values.

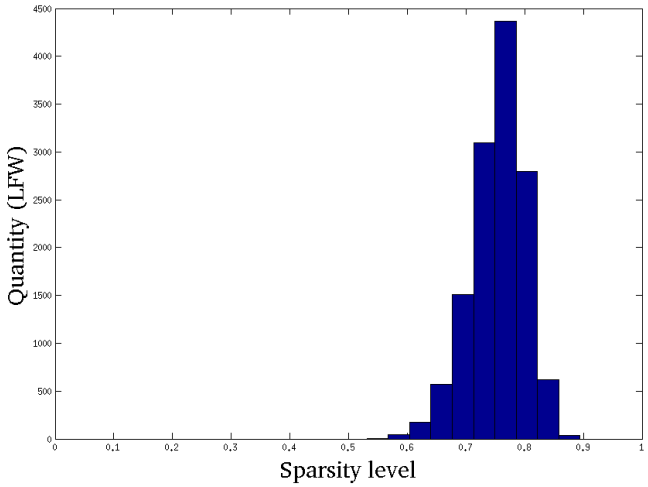

Figure 6. Histogram of the face representation sparsity level on LFW

Training on the *SFC*

Fig. 7 shows the Train and Test average (minus-) log-probability observed during training on the *SFC* dataset, in epochs.

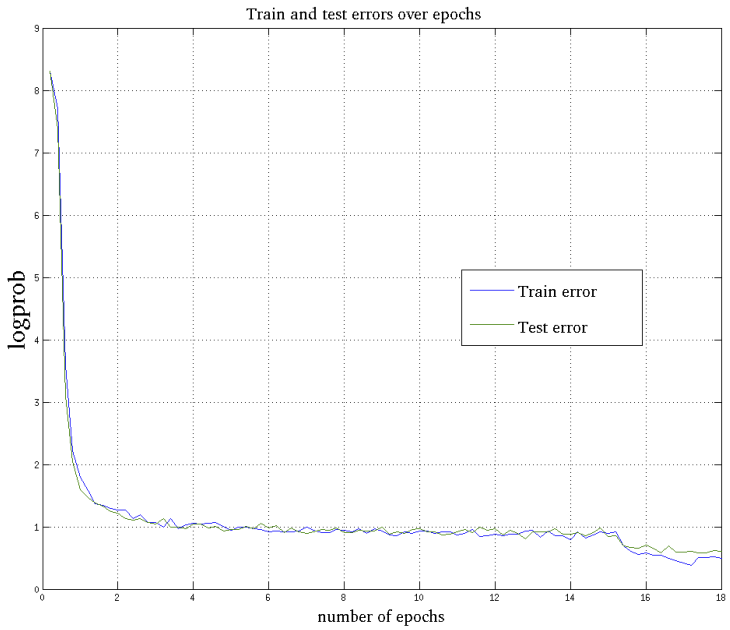

Figure 7. Train and Test errors observed during training on the *SFC*, over epochs

Source domain and Target domain distribution differences

Fig. 8 shows the age distribution of the training (*SFC*) and the testing (*LFW*) datasets. The age was estimated from faces automatically using an age estimator that incorporate an average error of 5 years.

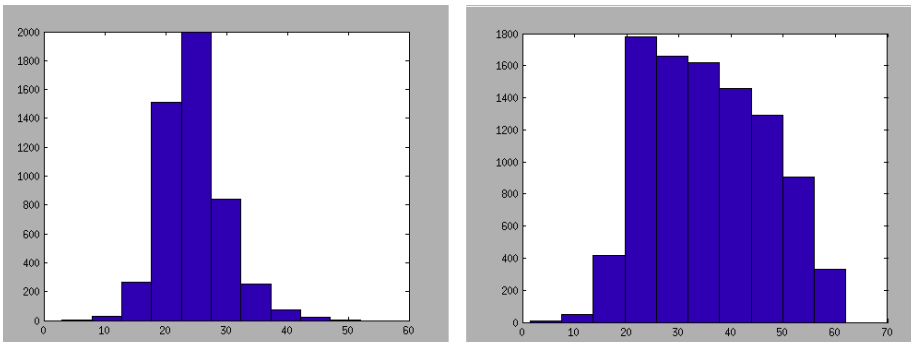

Figure 8. Left: Histogram of the estimated-age on SFC (train). Right: Histogram of the estimated-age on LFW (test)