Supplementary Material for Marginalized CNN: Learning Deep Invariant Representations

BMVC 2017 Submission # 470

Abstract

In this supplementary material, we present fully detailed information on 1) detailed reasons for datasets selection (*i.e.* MNIST and MS-Celeb-1M); 2) example removed outlier faces from MS-Celeb-1M with our data clean strategy; 3) results in Precision-Coverage curve and statistics for comparing the proposed mCNN and other recent state-of-the-art mothods on MS-Celeb-1M.

1 Reasons for Dataset Selection

MNIST images have clean background and sufficiently diverse transformation. Thus, they are very suitable for evaluating invariance property of the mCNN generated representations; moreover, the public Affine MNIST dataset is a good testbed for developing the transformation-invariant methods; there are well-established baselines (Jarrett *et al.*, ICCV 2009; Ciresan *et al.*, IJCAI 2011; Ciresan *et al.*, CVPR 2012) on this dataset for benchmarking.

Besides MNIST, we would like to mention that in Sec. 4 Experiments, we also test mCNN on the state-of-the-art MS-Celeb-1M validation set. MS-Celeb-1M is a large-scale face recognition dataset which consists of 1 million face images and present various poses, view-points and other transformations. On this challenging MS-Celeb-1M, the performance of mCNN is comparable to state-of-the-arts but mCNN uses a much simpler network architecture and does not require any data augmentation. This result evidently validates the strong invariance capacity of mCNN.

2 Example Removed Outlier Faces from MS-Celeb-1M

During the first stage, the outliers are effectively cleaned with the compact and discriminative deep features learned by the proposed mCNN and the clustering method, and the network parameters are initialized in an end-to-end way on the CASIA-Web Face dataset. The network parameters are further tuned during the second stage with the cleaned data in the specific target domain. The two-stage method to learn robust and invariant deep facial representations is effective for classifying celebrity faces at largescale. Some of the discarded data are visualized in Figure 1.

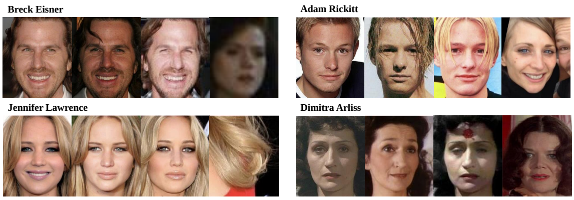
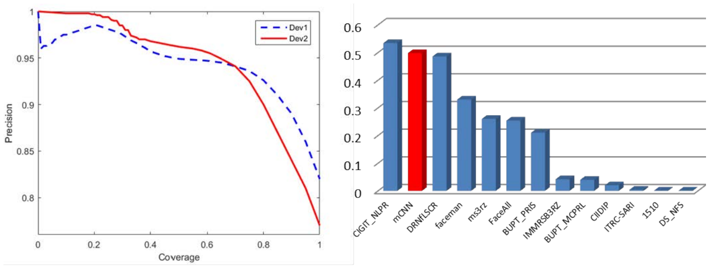


Figure 1: Example faces from MS-Celeb-1M. For eachcelebrity, the first 3 images are references and the last imageis the detected outlier. For BreckEisner, Adam Rickitt, and Jennifer Lawrence, the outliersare easily identified. For Dimitra Arliss, the outlier is actu-ally a challenging sample. Best viewed in color.



(a) (b)

Figure 2: Results on the MS-Celeb-1M dataset. (a) The Precision-Coverage curve of mCNN on the two tracks of validation set. (b) The evaluation performance of our mCNN (highlighted in red) in MS-Celeb-1M benchmark (Dev1 of the validation set) in terms of Coverage at Precision=95%. Best viewed in color.

3 Statistics for Comparing the Proposed mCNN and Other Recent State-of-The-Arts on MS-Celeb-1M

As illustrated in Figure 2 and Table 1, on Random set the proposed mCNN reaches the Coverage 65.4% when Precision=95%, and on Hard set the proposed mCNN reaches the Coverage 49.8% when Precision=95%.

Generally, mCNN shows higher performance than other recent state-of-the-art methods in terms of Coverage at Precision=95% on MS-Celeb-1M Hard set. This demonstrates that the proposed mCNN can be generalized well to other computer vision tasks, such as face recognition. Note that we only utilize a single model here for evaluation. We believe that the performance of our model can be further improved with more ensembled models specified with different loss functions, and we would like to examine this in the future.

Method	Coverage @ P = 0.95% Dev1/Dev2	Coverage @ P = 0.99% Dev1/Dev2
NII-UIT-KAORI*	-/0.001	-/0.001
BUPT_MCPRL*	0.040/0.064	0.007/0.006
CIIDIP*	0.020/0.154	0.018/0.025
IMMRSB3RZ*	0.042/0.171	0.039/0.104
BUPT_PRIS*	0.210/0.421	0.117/0.216
faceman*	0.330/0.461	**0.211**/0.339
FaceAll*	0.254/0.554	0.142/**0.417**
1510*	0.001/0.570	0.001/0.065
CIGIT_NLPR*	**0.534/0.684**	0.026/0.045
mCNN	0.498/0.654	0.136/0.316

(* indicates corresponding result is reported by MS-Celeb-1M leaderboard[1])

Table 1: Performance comparison of mCNN with state-of-the-arts on the two tracks Dev1 (Hard set)/Dev2 (Random set) of large-scale MS-Celeb-1M face recognition. Symbol "-" implies that the result is not reported for that method. A large number means better performance. The best performance is highlighted in bold.