MOON: A Mixed Objective Optimization Network for the Recognition of Facial Attributes

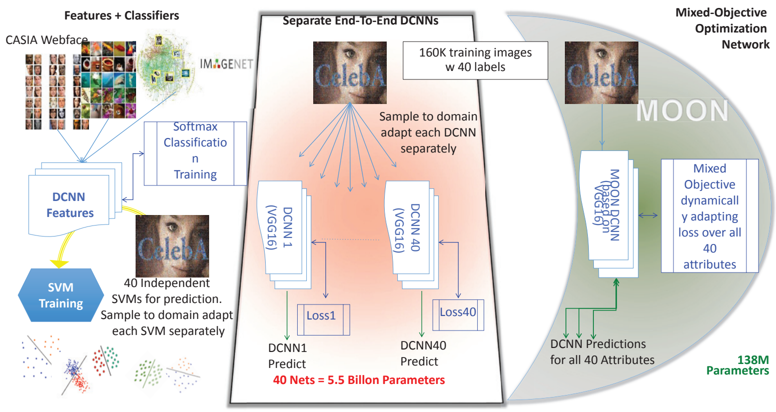
UCCS Vision And Security Technology
University of Colorado Colorado Springs

Ethan M. Rudd, Manuel Günther, and Terrance E. Boult
UCCS Vision and Security Technology (VAST) Lab

ECCV'16 European Conference on Computer Vision

Overview

- Facial attributes have implicit latent correlations – which are naturally learnt through a multi-label approach.
- Due to demographic correlations, it is difficult to obtain a balanced multi-label facial attribute training set; thus a multi-label classifier trained for one demographic might perform poorly on another due to training set biases.
  - **CelebA**, for example, contains images of celebrities where **Young** is over-represented and anti-correlated attributes like **Bald** and **Gray Hair** are under-represented.



Our Approach

- Idea: Backpropagate multi-label Euclidean loss, weighting frequency discrepancy between source distribution S_i and target distribution T_i.
- Given binary labels $\{+1, -1\}$, let S_i^+ and S_i^- be the number of occurrences in the source distribution and T_i^+ and T_i^- be the number of occurrences in the target distribution.
- The domain adaptive weights for each class then become:

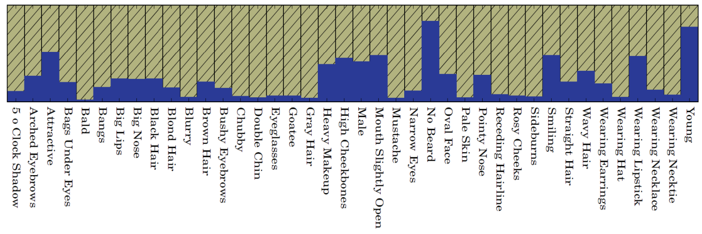
$$P(i|+1) = \begin{cases} 1 & \text{if } T_i^+ > S_i^+ \\ \frac{S_i^- T_i^+}{S_i^+ T_i^-} & \text{otherwise} \end{cases} \text{ and } P(i|-1) = \begin{cases} 1 & \text{if } T_i^- > S_i^- \\ \frac{S_i^+ T_i^-}{S_i^- T_i^+} & \text{otherwise.} \end{cases}$$

- The **mixed-objective Euclidean loss function** across an N-sample data tensor $\boldsymbol{X}$ with $N \times M$ label matrix $\boldsymbol{Y}$ and classification output $f(\cdot)$ is then:

$$L(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{j=1}^{N} \sum_{i=1}^{M} P(i|\boldsymbol{Y}_{ji}) \left\| f_i(\boldsymbol{X}_j) - \boldsymbol{Y}_{ji} \right\|^2 .$$

- Using the CelebA training partition, we trained a VGG-16 base architecture, replacing the final layer with our mixed-objective loss function.

Label Bias of the CelebA Dataset

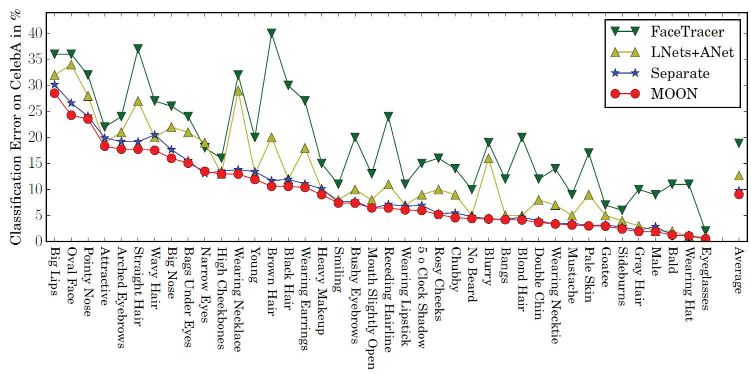


Approaches to the Balancing Problem

- **Features + Classifiers (FaceTracer, LNets+ANet):** Derive a feature space from non-attribute face data with attribute classifiers trained in that feature space.
  - \+ Easy to balance. Requires less training data.
  - \- Assumes that attribute related features will be implicitly embedded in the feature space with no explicit guidance.
- **Separate End-To-End DCNNs (Separate):** Train one network per attribute label; hope that the network learns correlations between attributes without explicitly optimizing to that end.
  - \+ More accurate.
  - \- Leads to a huge representation. Requires more data.
- **Mixed-Objective Optimization Network (MOON):** Combine the objectives of multi-label optimization with re-balancing under a domain-adaptive **mixed objective function.**
  - \+ Most accurate. Allows for a compact, domain-adaptive, multi-label representation.
  - \- Requires a new approach to balancing (this work).

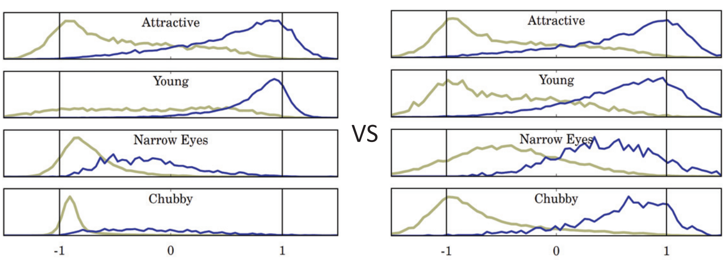
Evaluation: CelebA Dataset Un-Balanced

- For evaluation on the same dataset, we assume that $S_i \cong T_i$.
- MOON advances the state of the art over the Features+Classifiers approach with a relative reduction in classification error of 28.7%.
- Our approach also outperforms separate end-to-end networks of the same topology.



Re-Balanced CelebA: CelebAB

- Re-balancing the classifier to a uniform target distribution yields a shift in output scores distributions.



- Underperforms the unbalanced classifier on the unbalanced CelebA evaluation protocol but still advances the state of the art.
- Outperforms the unbalanced classifier when adapting the classification error measure to the target domain, i.e.:

$$E_i^B(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{j=1}^{N_{test}} \begin{cases} \dfrac{e_i(X_j, Y_{ji}) T_i^+}{N_i^+} & \text{if } Y_{ji} = +1 \\ \dfrac{e_i(X_j, Y_{ji}) T_i^+}{N_i^-} & \text{if } Y_{ji} = -1. \end{cases}$$

- Rebalancing yields superior attribute-based verification rates on the LFW dataset.