Label Denoising Adversarial Network (LDAN) for Inverse Lighting of Faces

Hao Zhou* Jin Sun* Yaser Yacoob David W. Jacobs
University of Maryland, College Park, MD, USA

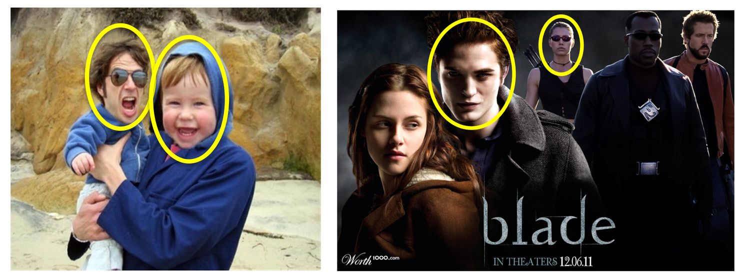
Introduction

Task: estimate lighting conditions from face images.

This has many applications:

- Face Relighting

- Image Forensics



Same lighting?

Fitting a 3DMM[1] and SIRFS[2] form expensive optimization problems for this task: they are slow and work poorly on low resolution images.

We take a deep learning approach.

Contributions

Challenge: ground truth lighting is not known for real face images.

Solution:

- Apply SIRFS to obtain noisy lighting labels for real images.
- Train a network with these labels, using synthetic data to help alleviate noise.

Main Contributions:

1. Apply deep learning for inverse lighting of face images.
2. LDAN uses synthetic data in training real face images with noisy labels.
3. LDAN increases lighting classification accuracy by 9% over baseline methods, and is thousands of times faster and more robust on low resolution images.

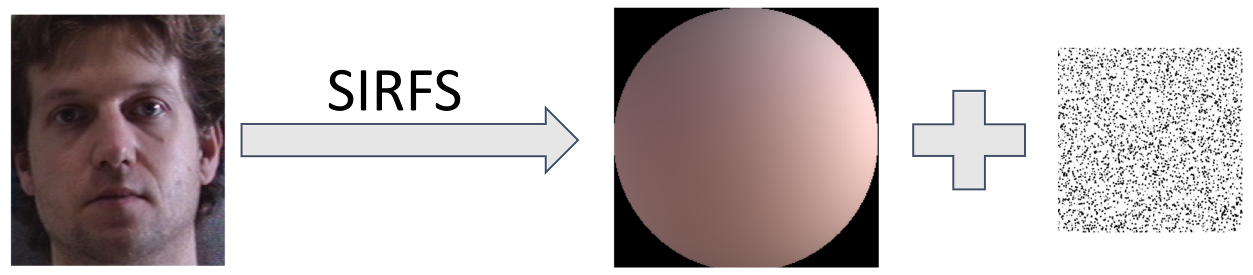
References

[1] P. Paysan, R. Knothe, B. Amberg, S. Romdhani and T. Vetter. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In *AVSS 2009*

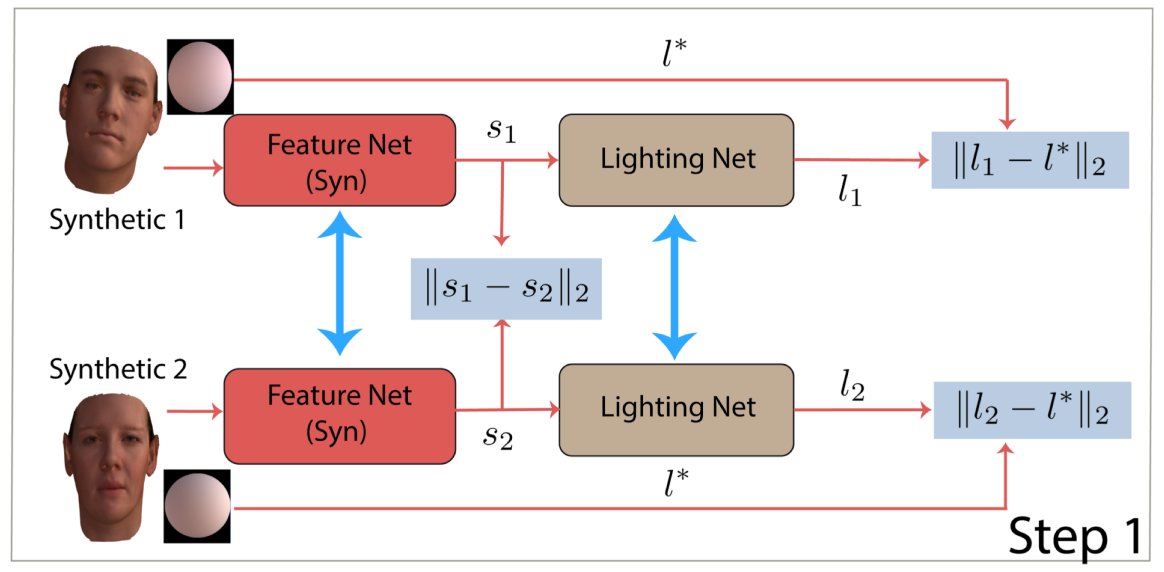
[2] J. T. Barron and J. Malik. Shape, Illumination, and Reflectance from Shading. *IEEE Transactions on PAMI, 2015*

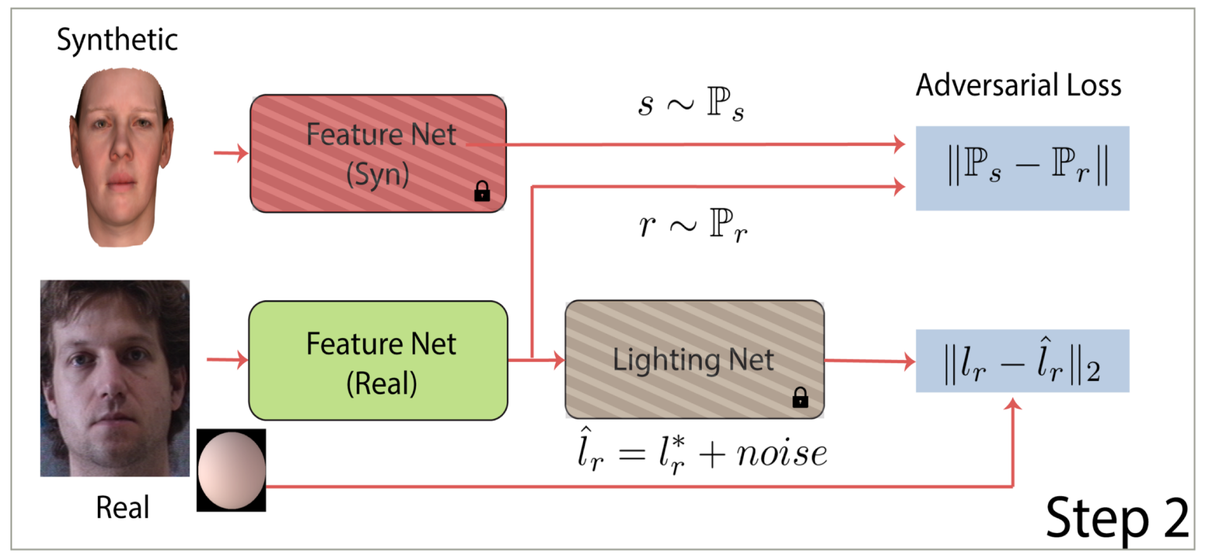
Method

We use SIRFS [2] to estimate lighting for real face images: these labels contain unknown noise.



We use synthetic data to help "denoise" these real labels in two training steps.





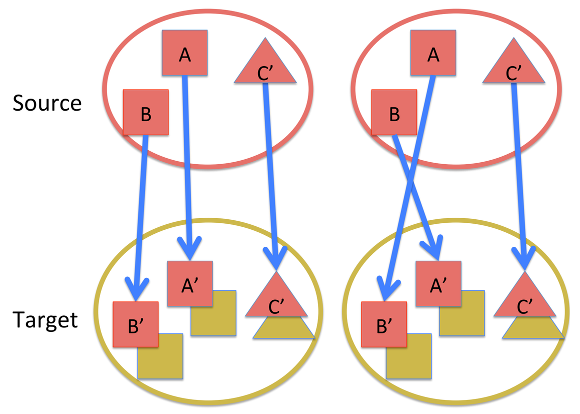
- Train a network with synthetic data.
- It has two subnetworks: feature net and lighting net.
- Since synthetic data have noise-free ground truth lighting, these two subnetworks are not affected by noise.

- Train a network with real data.
- Fix the synthetic-trained lighting network, train a feature network for real data.
- Regression loss: make estimated lighting not too far from noisy ground truth.
- Adversarial loss: map the features of real data to the noise-free feature space of synthetic data.

Rationale

1. Map lighting related features of real faces to the feature space of synthetic faces.
2. Estimated noisy lighting are treated as "anchor" points to prevent learning an arbitrary mapping.

Previous classification algorithms used adversarial loss in domain transfer. However, "anchor" points are very important in regression problems.



Data and Evaluation

Real Faces: we collect $40,000$ face images from the Internet.

Synthetic Faces: we use a 3DMM [1] to generate $40,000$ pairs of faces.

We use MultiPIE for quantitative evaluations.

Lighting classification task:

1. Estimate lighting for each face image.
2. Randomly select 90% data to compute the center of each lighting cluster.
3. Apply nearest neighbor to classify the rest of the data and report accuracy.

Results

Lighting Classification Accuracy

%	SIRFS	3DMM	REAL	LDAN
top-1	60.72	49.08	61.29	**65.73**
top-2	79.65	65.78	81.95	**84.57**
top-3	87.27	74.37	90.59	**92.43**

Ablation Study

-Ad: LDAN without adversarial loss;
-Re: LDAN without regression loss;
-Fix: LDAN without fixing lighting network.

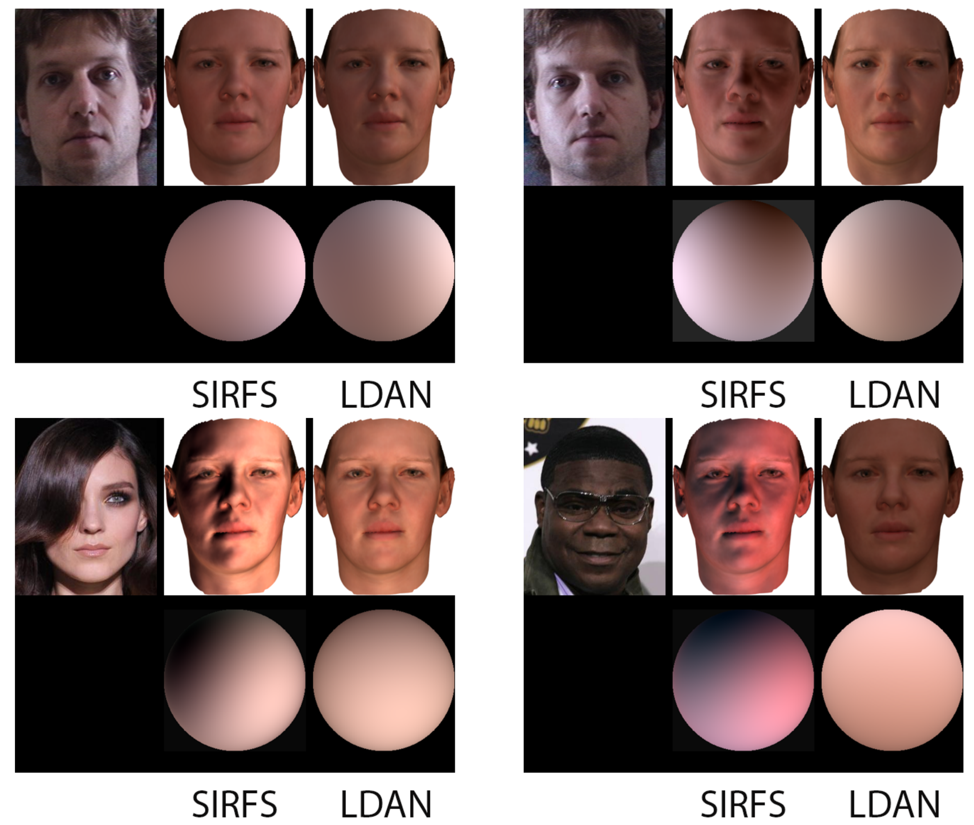
%	LDAN	-Ad	-Re	-Fix
top-1	**65.73**	63.63	30.72	63.95
top-2	**84.75**	83.44	49.12	83.97
top-3	**92.43**	91.48	61.58	92.07

On Different Resolutions

LDAN takes 64×64 images as inputs. For 32×32 and 16×16 images, we first down sample the original image to 32×32 or 16×16, then resize to 64×64.

%	64×64	32×32	16×16
top-1	65.73	64.89	61.72
top-2	84.75	84.39	82.17
top-3	92.43	92.10	90.94

Qualitative Results



Acknowledges

This work was supported by NSF grants (IIS-1302338 and 1526234) and the DARPA MediFor program under cooperative agreement FA87501620191, "Physical and Semantic Integrity Measures for Media Forensics".