Learning Lightness from Human Judgement on Relative Reflectance

Takuya Narihira [1,3] Michael Maire [2] Stella X. Yu [1]
[1] UC Berkeley / ICSI [2] TTI Chicago [3] Sony Corporation

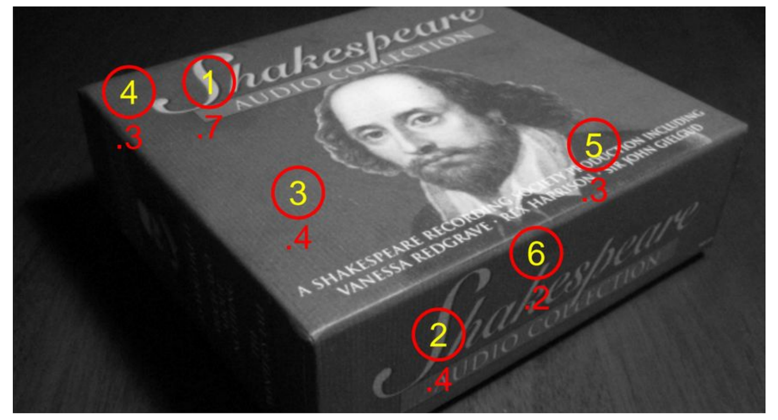
Contributions

- The first to use deep features for lightness prediction
- Simpler and direct approach to lightness
  - No hand-designed priors
  - Parameters are learned only from human judgement data
- On-par performance with the state-of-the-art intrinsic decomposition method

Goal: Compare Lightness in a Single Image

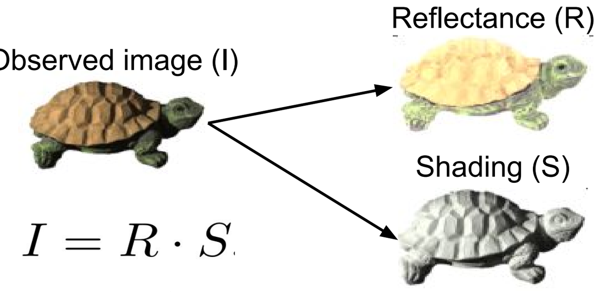
intensity = measured luminance
$I_1 > I_2 = I_3 > I_4 = I_5 > I_6$

lightness = perceived reflectance
$L_1 = L_2 > L_3 = L_4 = L_6 > L_5$



Past Work: Recover Reflectance

- Intrinsic image decomposition
- Reason about reflectance and shading together
- Reliance on assumed priors



$I = R \cdot S$

$L = R^*, \text{s.t. priors}(R, S)$

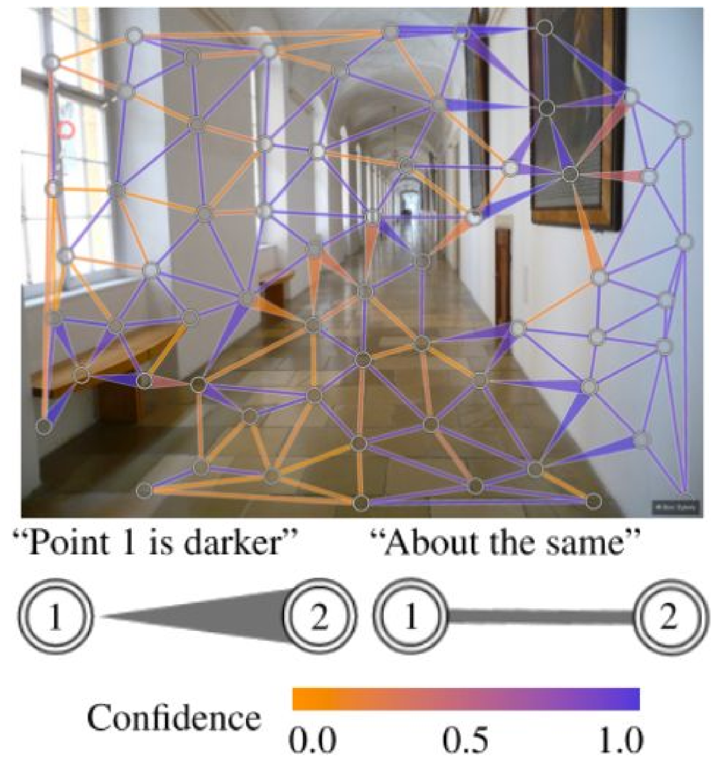
Our Work: Learn from Human Judgment Data

Learn pairwise lightness comparison of features z_i and z_j extracted at location i and j

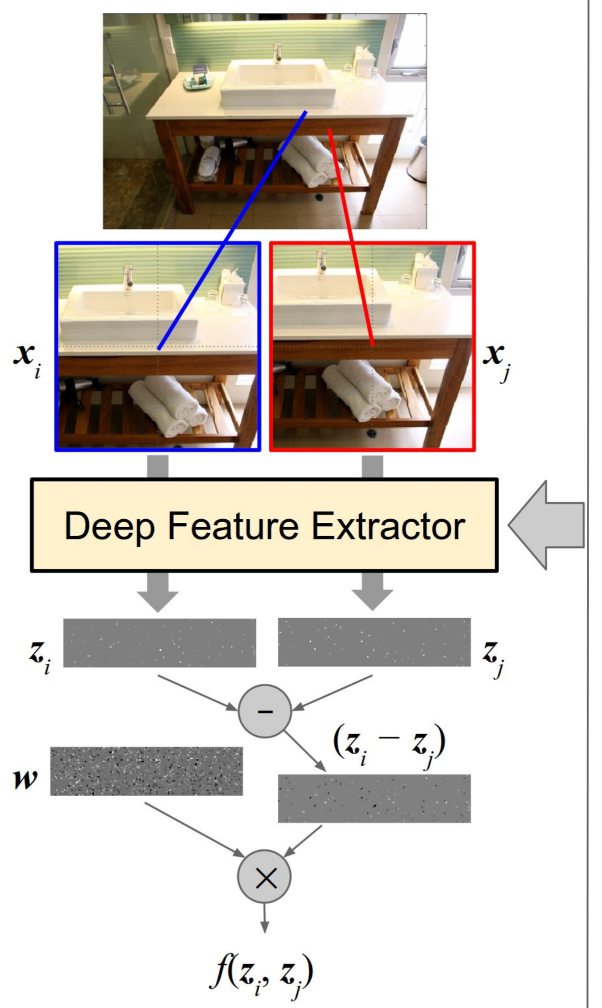
$$L_i - L_j = f(z_i, z_j) = w^T(z_i - z_j)$$

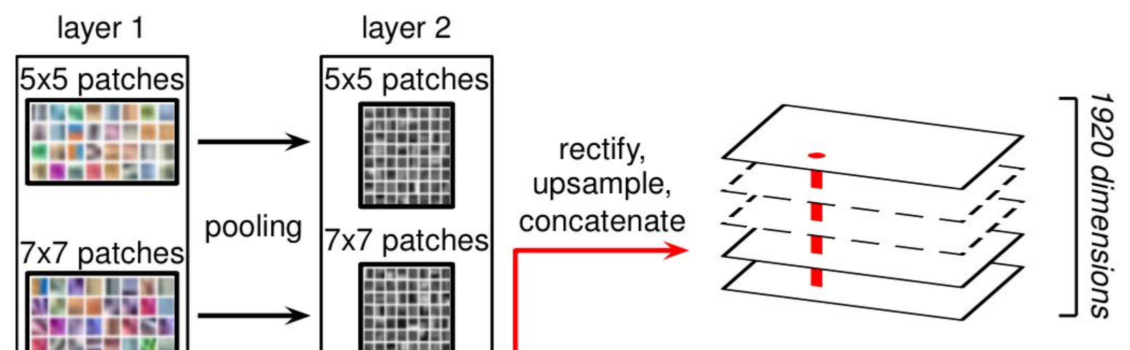
Intrinsic Images in the Wild Dataset

- 5230 interior images in the wild
- 872161 (166/image) pairwise comparisons by humans
- Pairwise label J_{ij} = 1,-1,0 for *lighter*, *darker*, *equal*
- Confidence C_{ij} for inter-observer disagreement



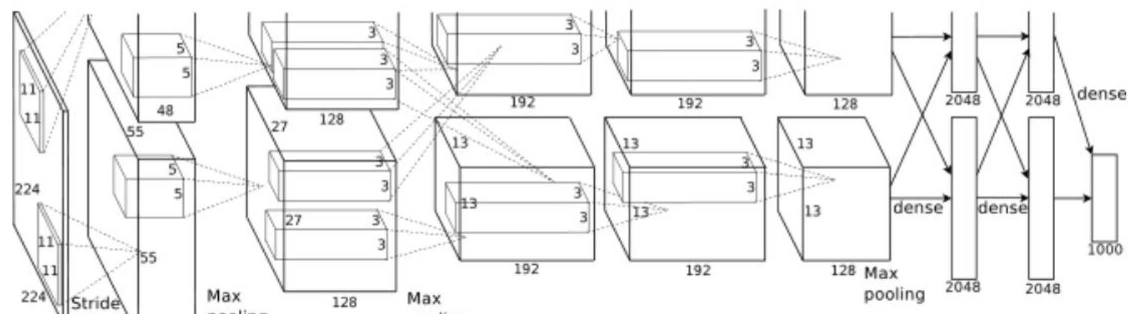
Deep Learning Framework for Pairwise Comparisons





Hierarchical Sparse Coding

- Unsupervised dictionary learning
- Sparse activations using batch orthogonal matching pursuit



Convolutional Neural Network

- End-to-end training by backprop
- Learn from scratch (CNN) or fine-tune from ImageNet
- AlexNet as distributed by BVLC/Caffe

Ranking loss between the output and the human judgement

$$\min \varepsilon(w) = \sum_{i,j} \log\left(1 + \exp(-J_{ij} w^T (z_i - z_j))\right) \qquad J_{ij} = \begin{cases} 1, & R_i^h > R_j^h \\ -1, & R_i^h < R_j^h \end{cases}$$

where R^h_i denotes perceived reflectance by human observer. For $R^h_i = R^h_j$, we create two virtual examples with both $J_{ij} = 1$ and $J_{ij} = -1$ in order to force prediction $f(z_i, z_j)$ toward zero. Also, L2 regularizer is applied. We use a standard classification loss for CNN.

Evaluation by Weighted Human Disagreement Rate (WHDR)

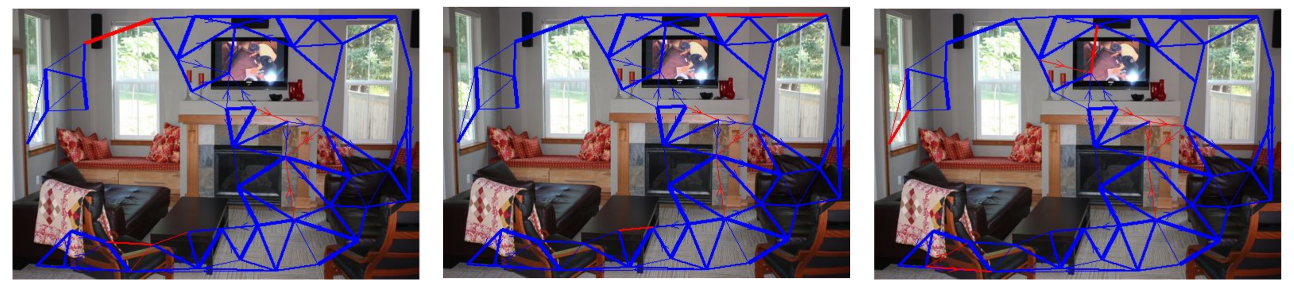
$$WHDR_\delta(J, R) = \frac{\sum_{ij} C_{ij}(J_{ij} \neq \hat{J}_{ij}(R;\delta))}{\sum_{ij} C_{ij}} \qquad \hat{J}_{ij}(R;\delta) = \begin{cases} 1, & \text{if } \frac{R_i}{R_j} > 1 + \delta \\ -1, & \text{if } \frac{R_j}{R_i} > 1 + \delta \\ 0, & \text{otherwise} \end{cases}$$

δ is chosen to minimize WHDR on training images for each model.

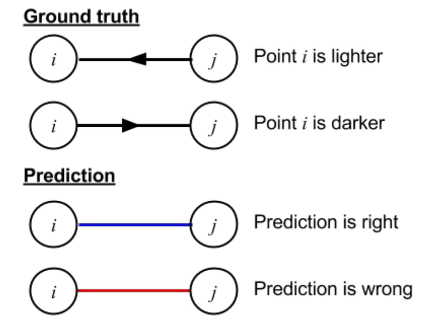
Quantitative Evaluation

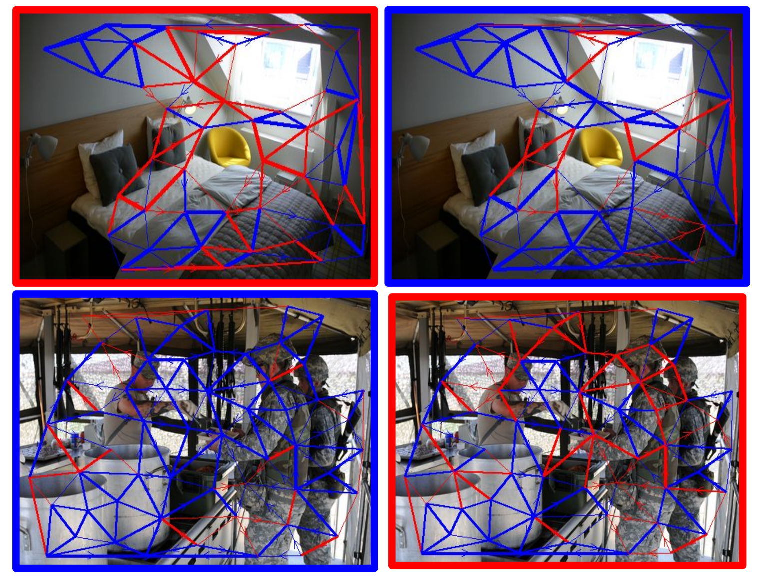
	Method	WHDR (%)	Error Rate (%)
optimized δ	Ours (HSC)	20.9	24.5
	Ours (CNN)	18.3	22.3
	Ours (CNN-ImageNet)	**18.1**	**22.0**
	CRF [4] (rescaled)	18.6	22.3
	Retinex-Color [10] (rescaled)	19.5	23.3
	Retinex-Gray [10] (rescaled)	19.8	23.8
	Shen and Yeo [22] (rescaled)	23.2	26.1
	Zhao *et al.* [26] (rescaled)	22.8	26.4
δ =0.1	CRF [4]	20.6	25.6
	Retinex-Color [10]	26.9	32.4
	Retinex-Gray [10]	26.8	32.3
	Shen and Yeo [22]	32.5	35.1
	Zhao *et al.* [26]	23.8	28.2

Sample Results: CRF vs Ours



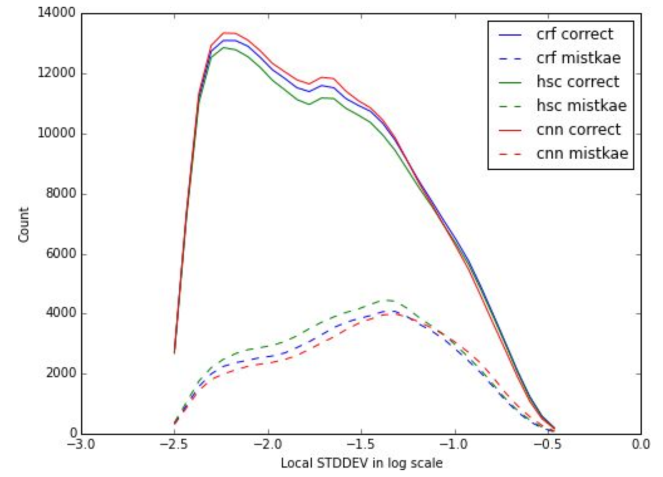
CRF HSC CNN





CRF CNN

- CNN seems to be better than CRF at smooth shading but not texture



CRF HSC

- CRF tends to err in many colors (small color palette assumption fails)
- HSC tends to err in low light and high light