Photo Aesthetics Ranking Network with Attributes and Content Adaptation

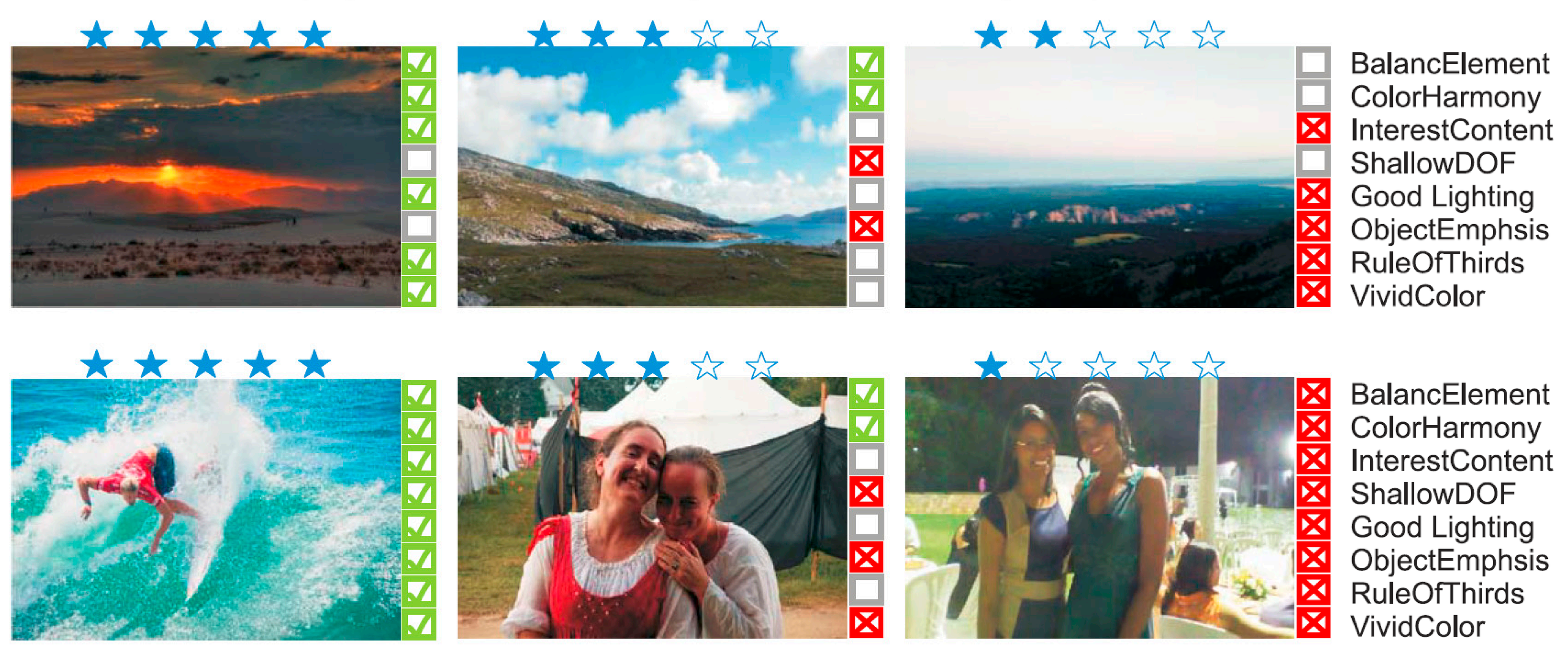
Shu Kong[1], Xiaohui Shen[2], Zhe Lin[2], Radomir Mech[2], Charless Fowlkes[1]

[1]Department of Computer Science, University of California Irvine, Irvine, California, USA

[2]Adobe Research, San Jose, USA

Abstract

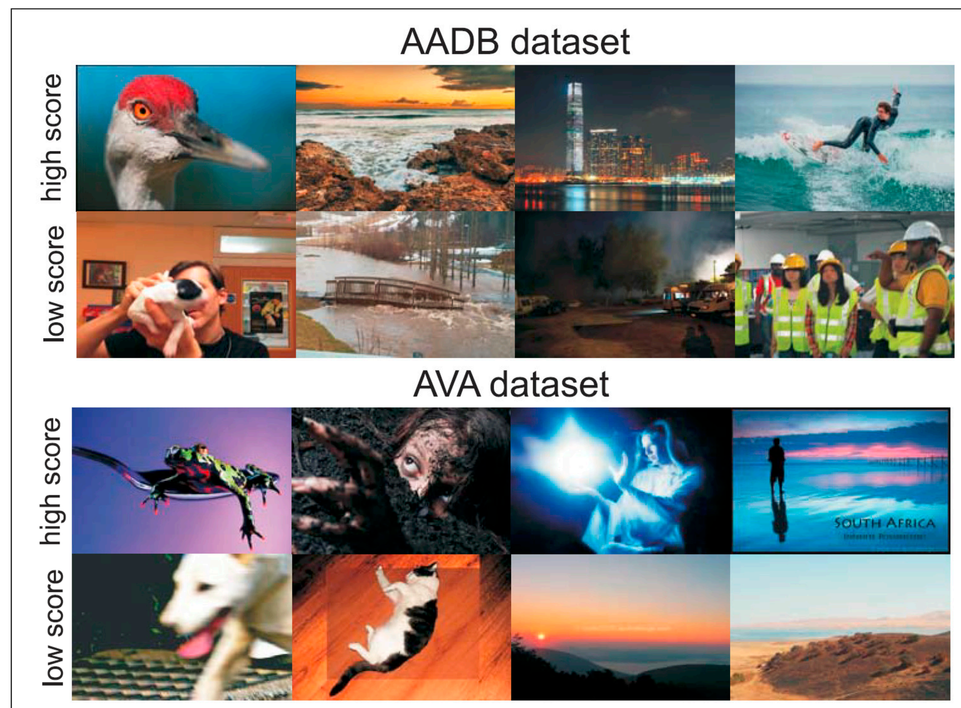
Real-world applications could benefit from the ability to automatically generate a fine-grained ranking of photo aesthetics.



Highlights:

1. A deep CNN to rank photo aesthetics with pairwse rank loss
2. Joint learning of meaningful photographic attributes and image content cues which help regularize the complicated photo aesthetics rating problem
3. A new aesthetics and attributes dataset (AADB) containing aesthetic scores and meaningful attributes assigned to each image by multiple human raters
4. Two sampling strategies for computing ranking loss of training image pairs for robustness in face of subjective judgment of image aesthetics
5. State-or-the-art classification performance on the existing AVA dataset benchmark by simply thresholding the estimated aesthetic scores

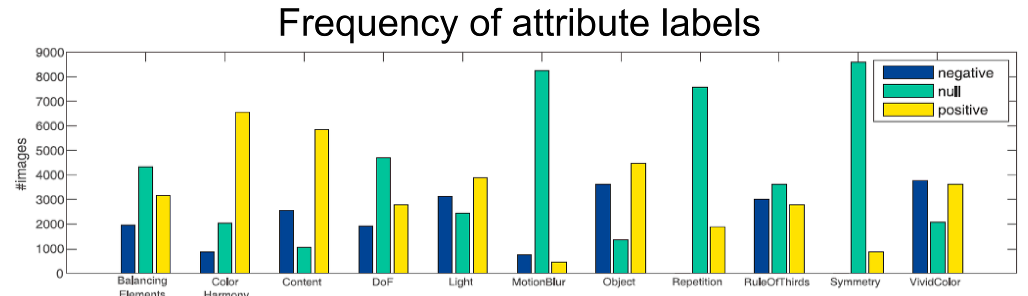
Aesthetics & Attribute Database (AADB)



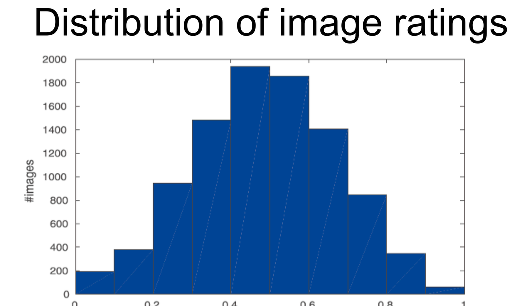
AADB images span a range of consumer and pro photos but exclude synthetic and heavily edited images.

Compared to existing datasets (e.g., AVA [23]) it is unique in having attribute labels, multiple ratings per image, and multiple images rated by each worker.

Frequency of attribute labels

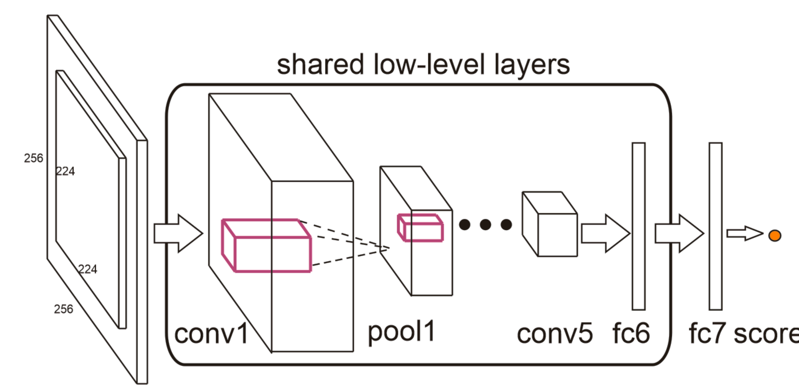


Distribution of image ratings



Fusing Attributes and Content for Aesthetics Ranking

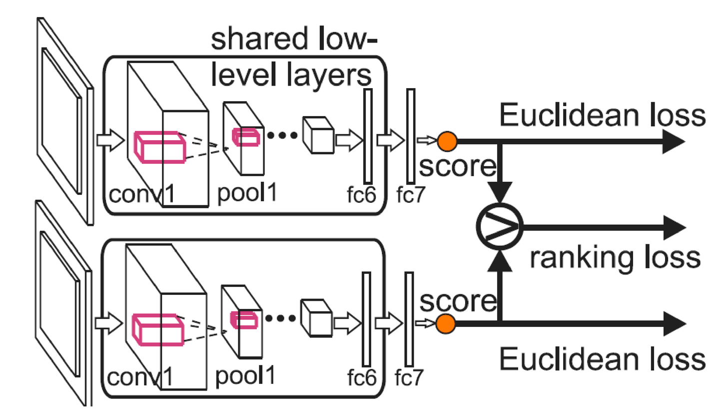
We first train a simple model with Euclidean loss for numerical rating of photo aesthetics



$$loss_{reg} = \frac{1}{2N}\sum_{i=1}^{N}\|\hat{y}_i - y_i\|_2^2$$

(a) fine-tuning with rank loss

Based on the regression net, we apply rank loss to fine-tune the network



$$loss_{reg+rank} = loss_{reg} + \omega_r loss_{rank}$$

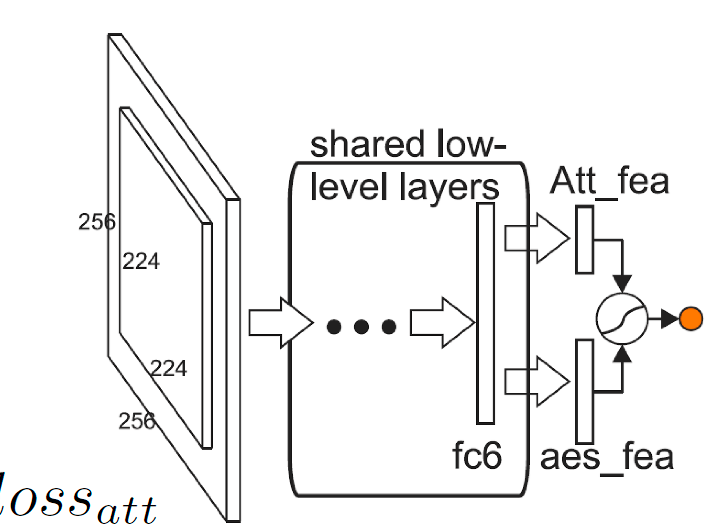
where

$$loss_{rank} = \frac{1}{2N}\sum_{i,j}\max\left(0, \alpha - \delta(y_i \geq y_j)(\hat{y}_i - \hat{y}_j)\right)$$

$$\delta(y_i \geq y_j) = \begin{cases} 1, & \text{if } y_i \geq y_j \\ -1, & \text{if } y_i < y_j \end{cases}$$

(b) attibute-adaptive network

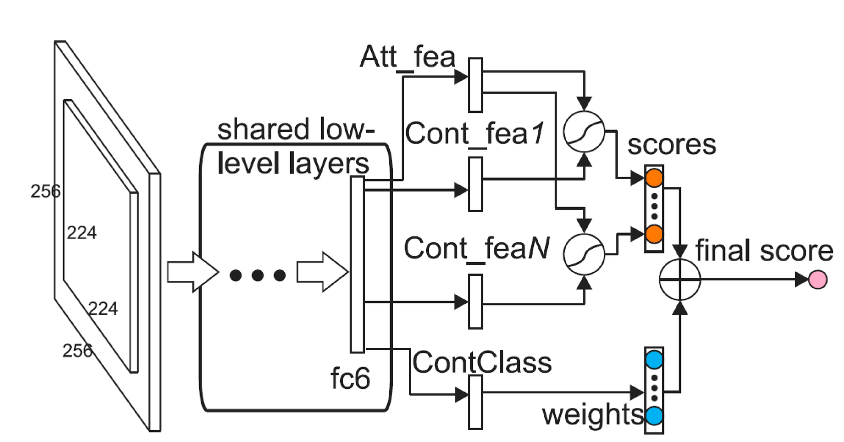
We use logistic loss to train an attribute prediction branch.



$$loss = loss_{reg} + \omega_r loss_{rank} + \omega_a loss_{att}$$

(c) attribute and content network

Similarly, we use softmax loss to train a content prediction branch whose output is used to multiplicatively gate content-specific attribute-adaptive branches. The weighted sum of scores provides the final rating.



Demo, code and model can be download through project webpage
http://www.ics.uci.edu/~skong2/aesthetics.html

References:
[8] He, K., Zhang, X., Ren, S., Sun, J., ECCV, 2014
[15] Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J., IEEE Trans. on Multimedia, 2015
[16] Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J.Z., ACMMM, 2014
[17] Lu, X., Lin, Z., Shen, X., Mech, R., Wang, J.Z., ICCV, 2015
[23] Murray, N., Marchesotti, L., Perronnin, F., CVPR, 2012

Acknowledgements: This work was supported by Adobe gift fund, NSF grants DBI-1262547 and IIS-1253538.

Experimental Results

We use Spearman's rho rank correlation (ρ) to measure ranking performance $\rho = 1 - \frac{6\sum d_i^2}{N^3 - N}$. By thresholding the rating scores, we achieve state-of-the-art classification accuracy on AVA despite never training with a classification loss.

Performance on AADB

Methods	ρ
AlexNet_FT_Conf	0.5923
Reg	0.6239
Reg+Rank (cross-rater)	0.6308
Reg+Rank (within-rater)	0.6450
Reg+Rank (within- & cross-)	0.6515
Reg+Rank+Att	0.6656
Reg+Rank+Cont	0.6737
Reg+Rank+Att+Cont	**0.6782**

Performance on AVA

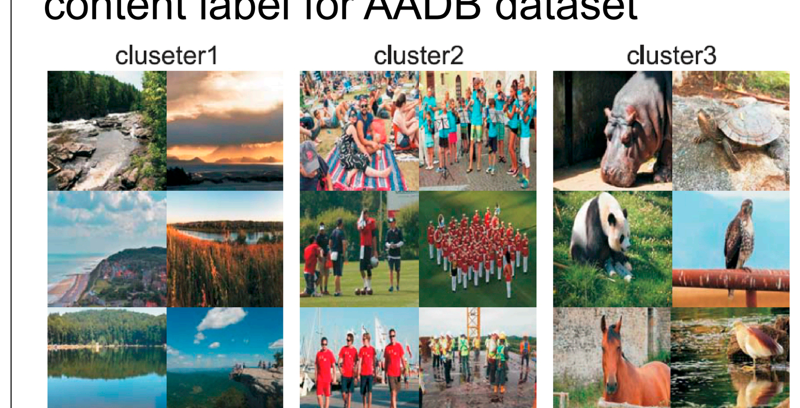
Methods	ρ	ACC (%)
Murray *et al.* [23]	-	68.00
SPP [8]	-	72.85
AlexNet_FT_Conf	0.4807	71.52
DCNN [16]	-	73.25
RDCNN [16]	-	74.46
RDCNN_semantic [15]	-	75.42
DMA [17]	-	74.46
DMA_AlexNet_FT [17]	-	75.41
Reg	0.4995	72.04
Reg+Rank	0.5126	71.50
Reg+Att	0.5331	75.32
Reg+Rank+Att	0.5445	75.48
Reg+Rank+Cont	0.5412	73.37
Reg+Rank+Att+Cont	**0.5581**	**77.33**

Analyzing Model Architecture

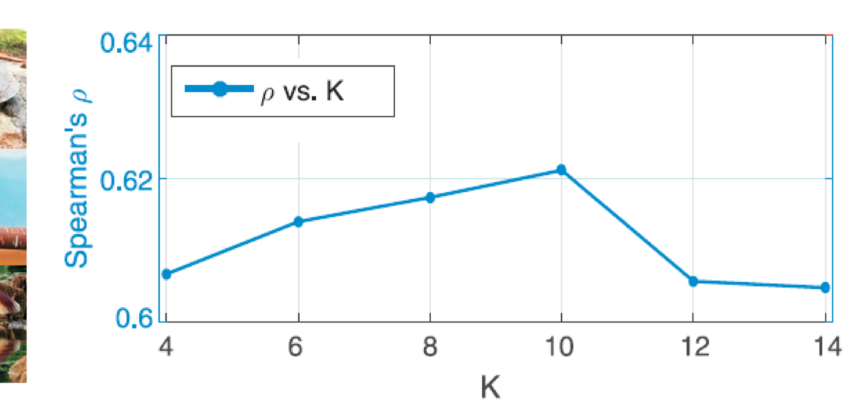
Analysis of content-aware model on AVA dataset. Confidence-weighted gating after fine tuning out-performs branch selection and simple branch averaging.

	ground-truth content label	predicted content label	branch averaging	confidence weighted gating
ρ	0.5367	0.5327	0.5336	0.5426
acc (%)	75.41	75.33	75.39	75.57

Clustering is used to generate target content label for AADB dataset



Performance as a function of the number of content branches (K)



Rank loss improves performance over pure regression loss

ω_r	0.0	0.1	1	2
AADB	0.6382	0.6442	0.6515	0.6276
AVA	0.4995	0.5126	0.4988	0.4672

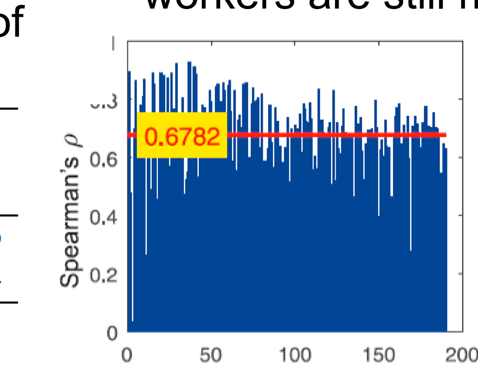
Sampling image pairs rated by the same individual helps

#ImgPairs	2 million	5 million
cross-rater	0.6346	0.6286
within-rater	0.6450	0.6448
within- & cross-rater	0.6487	0.6515

Limited model transferability indicates different taste of user groups.

Spearman's ρ	test: AADB	test: AVA
train AADB	0.6782	0.1566
train AVA	0.3191	0.5154

The model achieves similar agreement to the average inter-subject agreement but the best workers are still more consistent.



#images	#workers	ρ
>0	190	0.6738
>100	65	0.7013
>200	42	0.7112
Our best	–	0.6782