Attention-based Multi-Patch Aggregation for Image Aesthetic Assessment

Kekai Sheng
NLPR, Institute of Automation, Chinese Academy of Sciences & University of Chinese Academy of Sciences
shengkekai2014@ia.ac.cn

Weiming Dong*
NLPR, Institute of Automation, Chinese Academy of Sciences
weiming.dong@ia.ac.cn

Chongyang Ma
Snap Inc.
cma@snap.com

Xing Mei
Snap Inc.
xing.mei@snap.com

Feiyue Huang
Youtu Lab, Tencent
garyhuang@tencent.com

Bao-Gang Hu
NLPR, Institute of Automation, Chinese Academy of Sciences
hubg@nlpr.ia.ac.cn

Attention-based Objective Functions:

I. MP_{avg}

II. MP_{min}

III. $MP_{adaptive}$

I. MP_{avg}

$$f\left(\frac{1}{|S|}\sum_{x\in S} x\right) \geq \frac{1}{|S|}\sum_{x\in S} f(x)$$

$$\log\left(\frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}} Pr(\tilde{y}=\hat{y}\,|\,p,\theta)\right) \geq \underbrace{\frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}} \log\left(Pr(\tilde{y}=\hat{y}\,|\,p,\theta)\right)}_{MP_{avg}} \tag{3}$$

$$\frac{\partial MP_{avg}}{\partial\theta} = \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}} \underbrace{\frac{1}{Pr(\tilde{y}=\hat{y}\,|\,p,\theta)}}_{weights} \cdot \frac{\partial Pr(\tilde{y}=\hat{y}\,|\,p,\theta)}{\partial\theta} \tag{4}$$

II. MP_{min}

$$\log\left(\frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}} Pr(\tilde{y}=\hat{y}\,|\,p,\theta)\right) \geq \min_{p\in\mathcal{P}} \frac{1}{|\mathcal{P}|}\log\left(Pr(\tilde{y}=\hat{y}\,|\,p,\theta)\right)$$

$$= \underbrace{\frac{1}{|\mathcal{P}|}\log\left(Pr(\tilde{y}=\hat{y}\,|\,p^m,\theta)\right)}_{MP_{min}}$$

$$p^m = \underset{p\in\mathcal{P}}{argmin}\, Pr(\tilde{y}=\hat{y}\,|\,p,\theta)$$

$$\frac{\partial MP_{min}}{\partial\theta} = \frac{1}{|\mathcal{P}|}\frac{1}{Pr(\tilde{y}=\hat{y}\,|\,p^m,\theta)}\cdot\frac{\partial Pr(\tilde{y}=\hat{y}\,|\,p^m,\theta)}{\partial\theta}$$

$$= \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}}\frac{\mathbb{I}(p=p^m)}{Pr(\tilde{y}=\hat{y}\,|\,p,\theta)}\cdot\frac{\partial Pr(\tilde{y}=\hat{y}\,|\,p,\theta)}{\partial\theta}$$

III. $MP_{adaptive}$

$$MP_{ada} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \omega_\beta \cdot \log \Big(Pr(\tilde{y} = \hat{y} \mid p, \theta) \Big)$$

$$\omega_\beta = \frac{Pr(\tilde{y} = \hat{y} \mid p, \theta)^{-\beta} - 1}{Pr(\tilde{y} = \hat{y} \mid p, \theta)^{-\beta}} = 1 - Pr(\tilde{y} = \hat{y} \mid p, \theta)^{\beta}$$

$$\frac{\partial MP_{ada}}{\partial \theta} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \lambda \cdot \frac{\partial Pr(\tilde{y} = \hat{y} \mid p, \theta)}{\partial \theta}$$

$$\lambda = \frac{1 - (1 + \beta \cdot \log Pr(\tilde{y} = \hat{y} \mid p, \theta)) \cdot (1 - \omega_\beta)}{Pr(\tilde{y} = \hat{y} \mid p, \theta)}$$

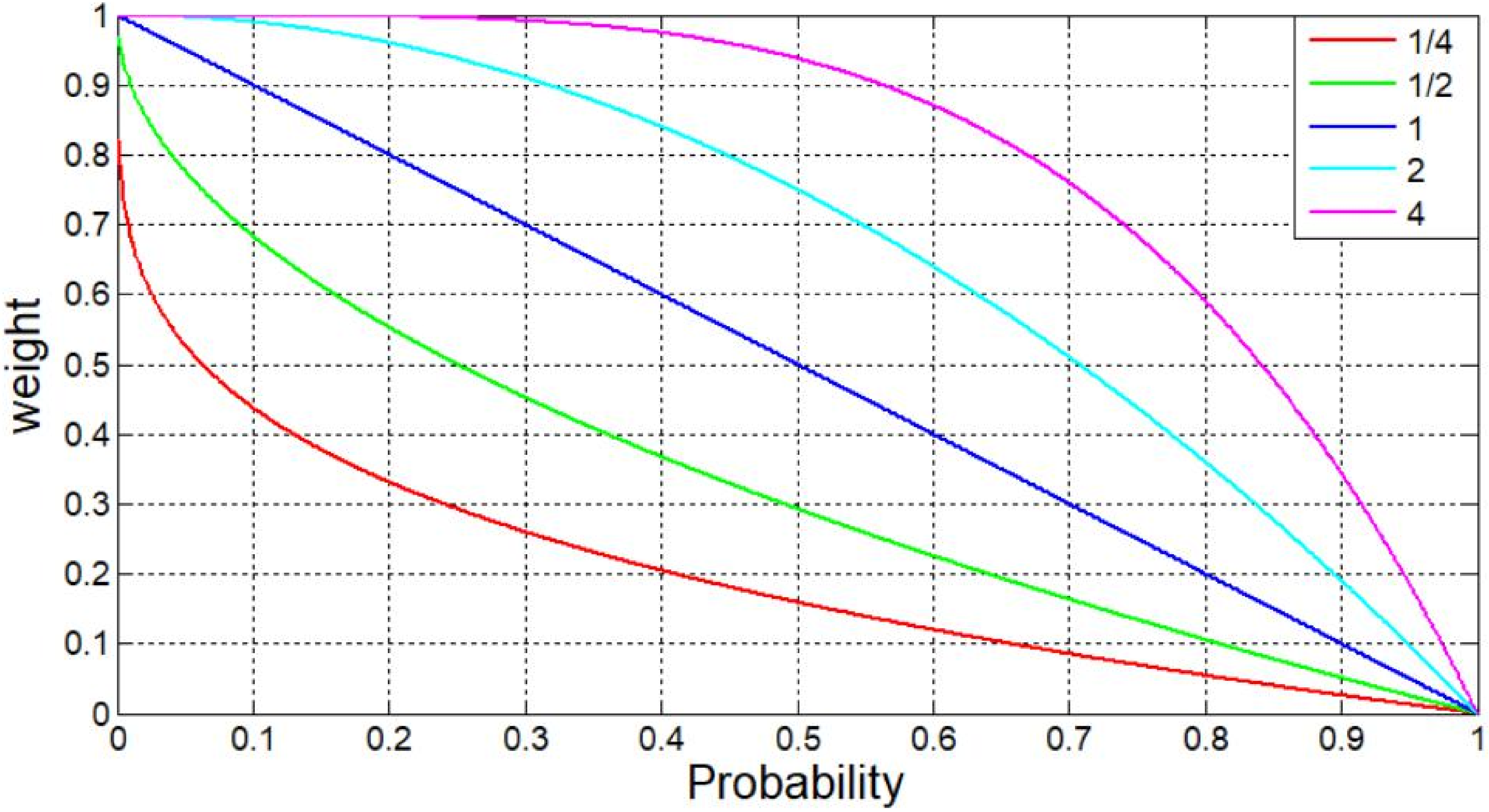

Figure 3: Curves of adaptive weights $\omega_\beta = 1 - Pr^\beta$ with different values of the hyperparameter β.

Network Architecture

Layers	Output names	Output shape
conv, 7x7, 64, stride 2	-	[112, 112, 64]
max pool, 3x3, stride 2	-	[56, 56, 64]
$\begin{bmatrix} conv, 3x3, 64 \\ conv, 3x3, 64 \end{bmatrix} x2$	0-0-BNReLU1 0-0-BNReLU2 0-0-ReLU 0-1-BNReLU1 0-1-BNReLU2 0-1-ReLU	[56, 56, 64]
$\begin{bmatrix} conv, 3x3, 128 \\ conv, 3x3, 128 \end{bmatrix} x2$	1-0-BNReLU1 1-0-BNReLU2 1-0-ReLU 1-1-BNReLU1 1-1-BNReLU2 1-1-ReLU	[28, 28, 128]
$\begin{bmatrix} conv, 3x3, 256 \\ conv, 3x3, 256 \end{bmatrix} x2$	2-0-BNReLU1 2-0-BNReLU2 2-0-ReLU 2-1-BNReLU1 2-1-BNReLU2 2-1-ReLU	[14, 14, 256]
$\begin{bmatrix} conv, 3x3, 512 \\ conv, 3x3, 512 \end{bmatrix} x2$	3-0-BNReLU1 3-0-BNReLU2 3-0-ReLU 3-1-BNReLU1 3-1-BNReLU2 3-1-ReLU	[7, 7, 512]
global average pooling	-	[512]
2d fc, softmax	-	[2]

Experimental Results:

Method	Core Features	Results
AVA [25]	handcrafted features	68.0
VGG-Scale [19]	non-uniform scaling	73.8
VGG-Pad [19]	uniform scaling + padding	72.9
SPP [22]	spatial pooling	76.0
VGG-Crop [19]	MP aggregation	71.2
DMA-Net [22]		75.41
MNA-CNN [19]		77.1
New-MP-Net [23]		81.7
DCNN [21]	multi-column aggregation	73.25
RAPID [21]		75.42
A&C CNN [12]		74.51
MTCNN [11]		78.56
MTRLCNN [11]		79.08
BDN [34]		78.08
Two-column DAN [6]		78.72
AA-Net [35]		76.9
DMA-Net-IF [22]	representation aggregation with explicit information	75.4
MNA-CNN-Scene [19]		77.4
A-Lamp [23]		82.5
NIMA [31]	distributions of human opinion scores	81.51
MP_{avg}	average weights	81.76
MP_{min}	minimum select	80.50
MP_{ada}	adaptive weights	**83.03**

Figure 5: Our aesthetic assessment results on the AVA test set predicted by the MP_{ada} scheme.


1 / 0.740 1 / 0.562
1 / 0.670 0 / 0.521
1 / 0.736 1 / 0.606
1 / 0.710 1 / 0.596
0 / 0.582 0 / 0.692
0 / 0.573 0 / 0.651
0 / 0.539 0 / 0.761
0 / 0.638 0 / 0.812
1 / 0.757 1 / 0.572

Q & A

THANK YOU