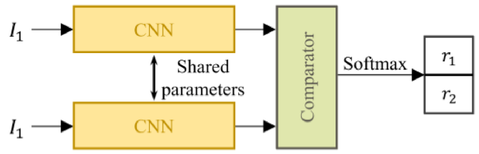
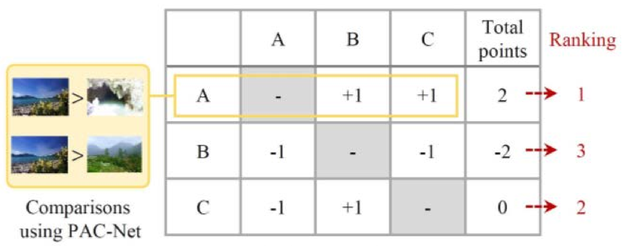
PAC-NET: PAIRWISE AESTHETIC COMPARISON NETWORK FOR IMAGE AESTHETIC ASSESSMENT

Keunsoo Ko, Jun-Tae Lee, and Chang-Su Kim

Introduction

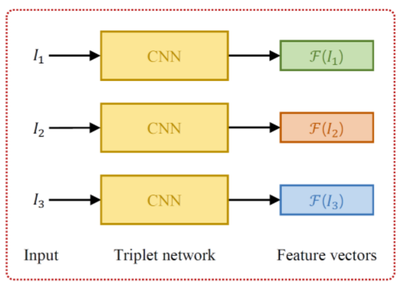
- Image Aesthetic Assessment
  - Distinguish high quality images from low quality ones
  - Classification
    - Classify quality of an image into binary classes
  - Regression
    - Quality of an image is measured by a continuous score or a rank
- Overview
  - Determine which image has a higher quality
  - Rank images using pairwise comparison method
  - Find quality of an image, comparing with reference images



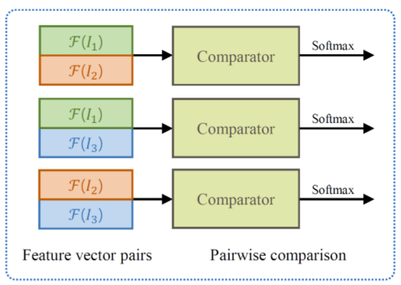


Proposed Algorithm

- PAC-Net
  - Aesthetic feature extraction
  - Pairwise feature comparison
  - Output relative rank



(a) Aesthetic feature extraction



(b) Pairwise feature comparison

- Aesthetic feature extraction
  - The network is based on the GoogLeNet
  - Employ all its nine inception modules (21 convolution layers)
  - Extracts aesthetic features $F(I_i)$ of image I_i
- Pairwise feature comparison
  - Compare the aesthetic features in pairs using comparator
  - The comparator contains two successive fully-connected layers and output 2-dimensional vector $\mathbf{r} = (r_i, r_j)$
  - Rank of two images I_i, I_j is determined according to the vector
    - If $r_i > r_j$, I_i is declared to have a higher quality than I_j
- Aesthetic-Adaptive Cross Entropy Loss
  - Pairs of images with similar aesthetic scores can confuse the network in training
  - Similar to the Focal loss function based on cross entropy loss

$$L_{aes}(\mathbf{r}, \bar{\mathbf{r}}) = -\sum_{i=1}^{2} \overline{r_i}(1 - r_i)^{\alpha(1-d)} log r_i$$

  - Adaptively determined according to the difference d
    - $\overline{s_i}$: ground truth aesthetic scores of I_i
    - D : maximum score difference within the dataset

$$d = \frac{|\bar{s}_1 - \bar{s}_2|}{D}$$

- Aesthetic Quality Assessment (Ranking)
  - PAC-Net determines relative ranks
  - The higher quality image gets point 1, while the lower quality one gets point -1
  - If the dataset contains N images, $^{N(N-1)}/_{2}$ such comparison are made
  - Summing up the points
  - Get ranking by sorting the total points
- Aesthetic Quality Assessment (Classification)
  - Select reference images
    - Aesthetic scores are close to the medium value of the dataset
  - Compute the point, by comparing an image with each reference image
  - Dichotomize into binary classes
    - Declare the test image as high quality if its average point is higher than 0, and low quality otherwise

Experimental Results

- Evaluate the performance
  - Both ranking and classification result
    - CE : adopts the cross entropy loss
    - AA : adopts the aesthetic-adaptive cross entropy loss

Algorithm	AVA dataset		AADB dataset
	Accuracy (%)	ρ (↑)	ρ (↑)
Kong *et al.* [17]	77.3	0.5581	0.6782
PAC-Net + CE	<u>81.1</u>	<u>0.8447</u>	<u>0.8122</u>
PAC-Net + AA	**82.2**	**0.8711**	**0.8371**

  - Classification result on AVA dataset

Algorithm	Accuracy (%)
AVA [8]	67.0
RDCNN [13]	74.4
DMA-Net [14]	75.4
Reg-Net [17]	77.3
MNA-Net [16]	77.4
A-Lamp [15]	**82.5**
PAC-Net+CE	81.1
PAC-Net+AA	<u>82.2</u>

- Predicted ranking

- Classification results

(a) True positives (b) False positives

(c) True negatives (d) False negatives

ksko@mcl.korea.ac.kr, jtlee@mcl.korea.ac.kr, and changsukim@korea.ac.kr