Hierarchical Layout-Aware Graph Convolutional Network for Unified Aesthetics Assessment

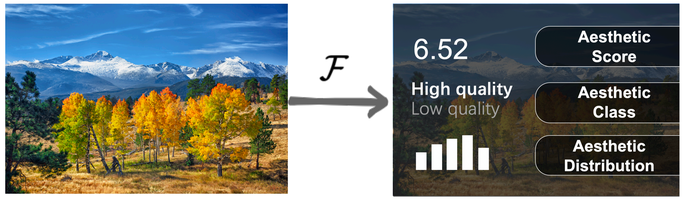
Dongyu She[1] Yu-Kun Lai[2] Gaoxiong Yi[3] Kun Xu[1*]
[1]Tsinghua University [2]Cardiff University [3]Tencent

Code & Paper

CVPR VIRTUAL JUNE 19-25

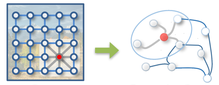
Problem

- Predicting three tasks of image **aesthetic score regression**, **binary aesthetic classification**, **and aesthetic distribution prediction** via a unified model



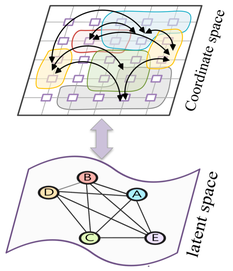
Input — Unified Aesthetics Assessment

Motivation



- The constraint of **fixed size input** compromises the **aesthetics** of original image
- Limitation of **regular receptive field** on capturing long-distance relations

Contributions



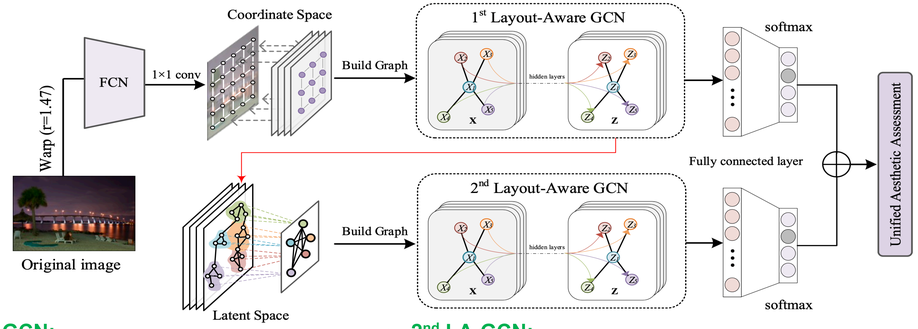
1 A layout-aware graph convolution module

- capturing layout information of different elements in images

2 Extending LA-GCN to a hierarchical architecture

- Perform graph reasoning on coordinate and latent spaces

Pipeline



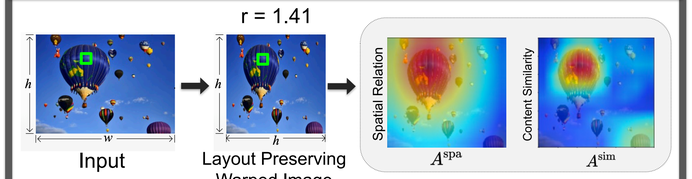
1st LA-GCN:

- preserves original **ratio information**
- uses both **position & appearance** cues
- reasons over whole **coordinate space**

2nd LA-GCN:

- aggregates spatial nodes via **projection**
- performs graph reasoning over **latent space**
- maps features back **without sacrificing efficiency**

LA-GCN module



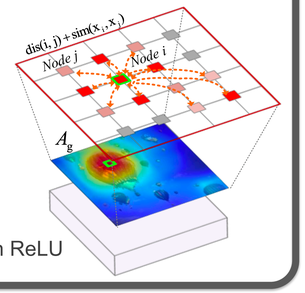
- Computing **adjacent matrix** by considering two types of relations:

$$\mathbf{A}_c = \mathbf{A}^{\text{sim}} + \mathbf{A}^{\text{spa}}$$

- **Layout-aware Graph Reasoning** via graph convolution:

$$\mathbf{H}_c^{(m+1)} = \sigma\left(\tilde{\mathbf{A}}_c^{(m)} \mathbf{H}_c^{(m)} \Theta_c^{(m)}\right)$$

$\sigma(\cdot)$ denotes the activation function ReLU

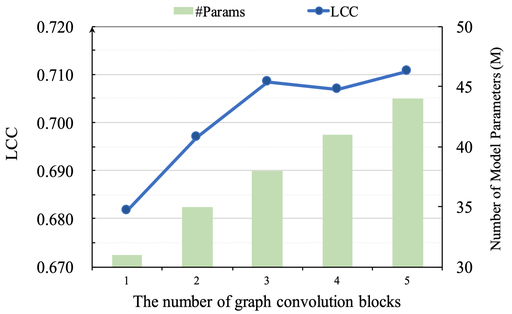


Experimental Results

Comparison with the SOTA methods on the AVA Dataset

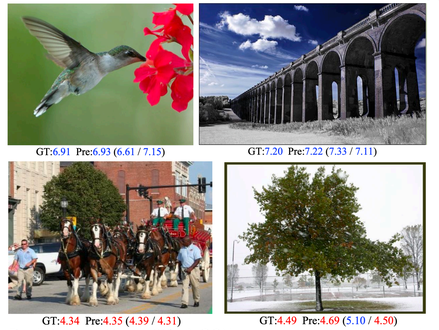
Methods	Classification	Score Regression			Distribution	
	Accuracy ↑	SRCC ↑	LCC ↑	MSE ↓	EMD_1 ↓	EMD_2 ↓
DMA-Net [27]†	75.4 %	-	-	-	-	-
MNA-CNN [29]†	77.1 %	-	-	-	-	-
Zeng *et al.* [47]	80.8 %	0.719	0.720	0.275	-	0.065
APM [32]	80.3 %	0.709		0.279	-	**0.061**
A-Lamp [28]†	82.5 %	-	-	-	-	-
MP_{ada} [41]†	83.0 %	-	-	-	-	-
RGNet [25]	82.5 %	-	-	-	-	-
Hosu *et al.* [10]	81.7 %	**0.756**	**0.757**	-	-	-
NIMA [42]	81.5 %	0.612	0.636	-	0.050	-
AFDC [3]	83.0 %	0.649	0.671	0.271	0.045	-
PA_IAA [22]‡	83.7 %	0.677	-	-	0.047	-
PA_IAA [22]‡	82.9 %	0.666	-	-	0.049	-
HLA-GCN	**84.1%**	0.656	0.678	**0.264**	**0.045**	0.065
HLA-GCN	**84.6%**	0.665	0.687	**0.255**	**0.043**	0.063

Using different number of LA-GCN blocks



- Setting block number as a tradeoff between results and model size

Qualitative results (high/low aesthetic quality)



- Predictions of 1st and 2nd LA-GCN are shown in brackets in order.