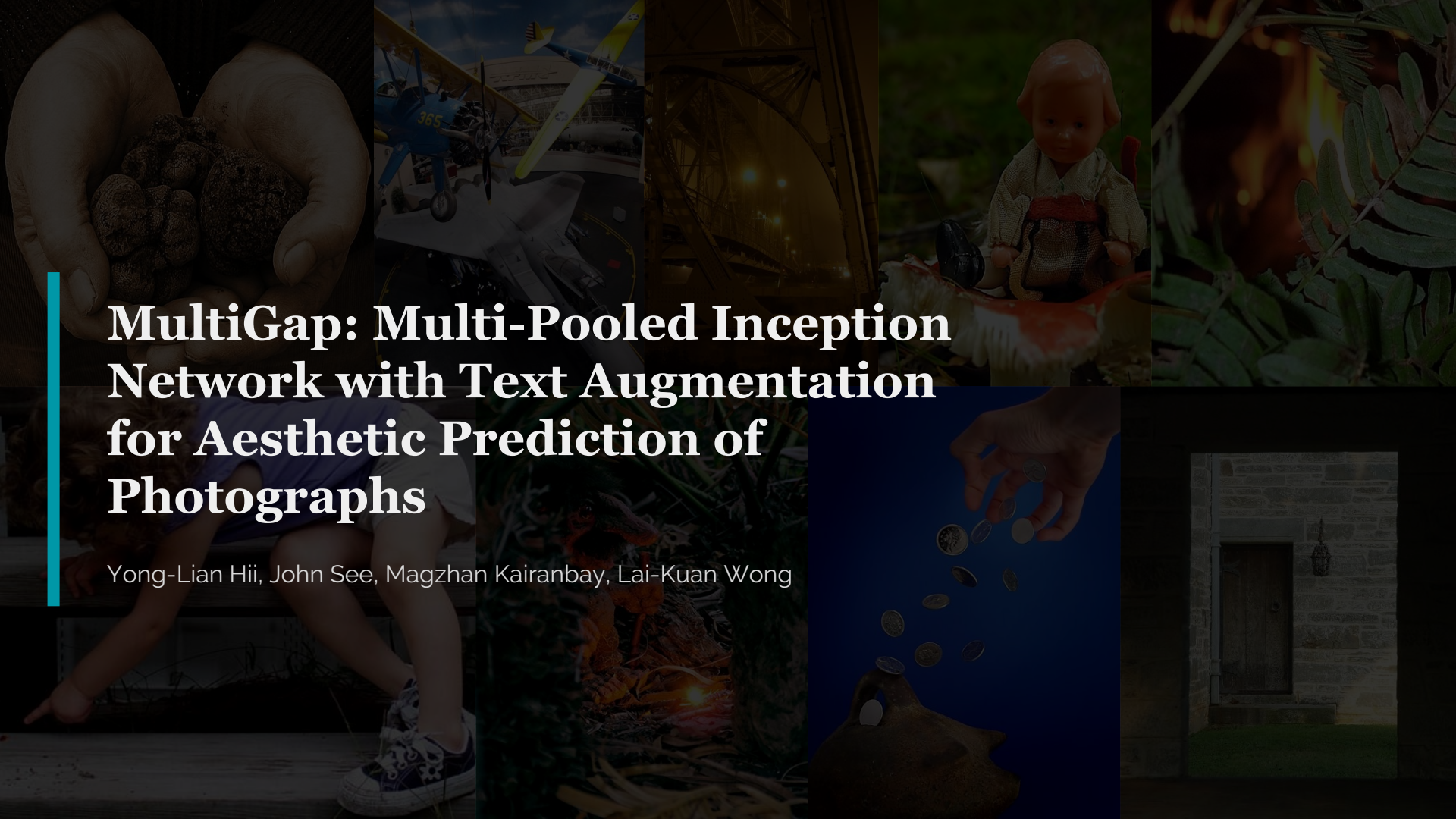
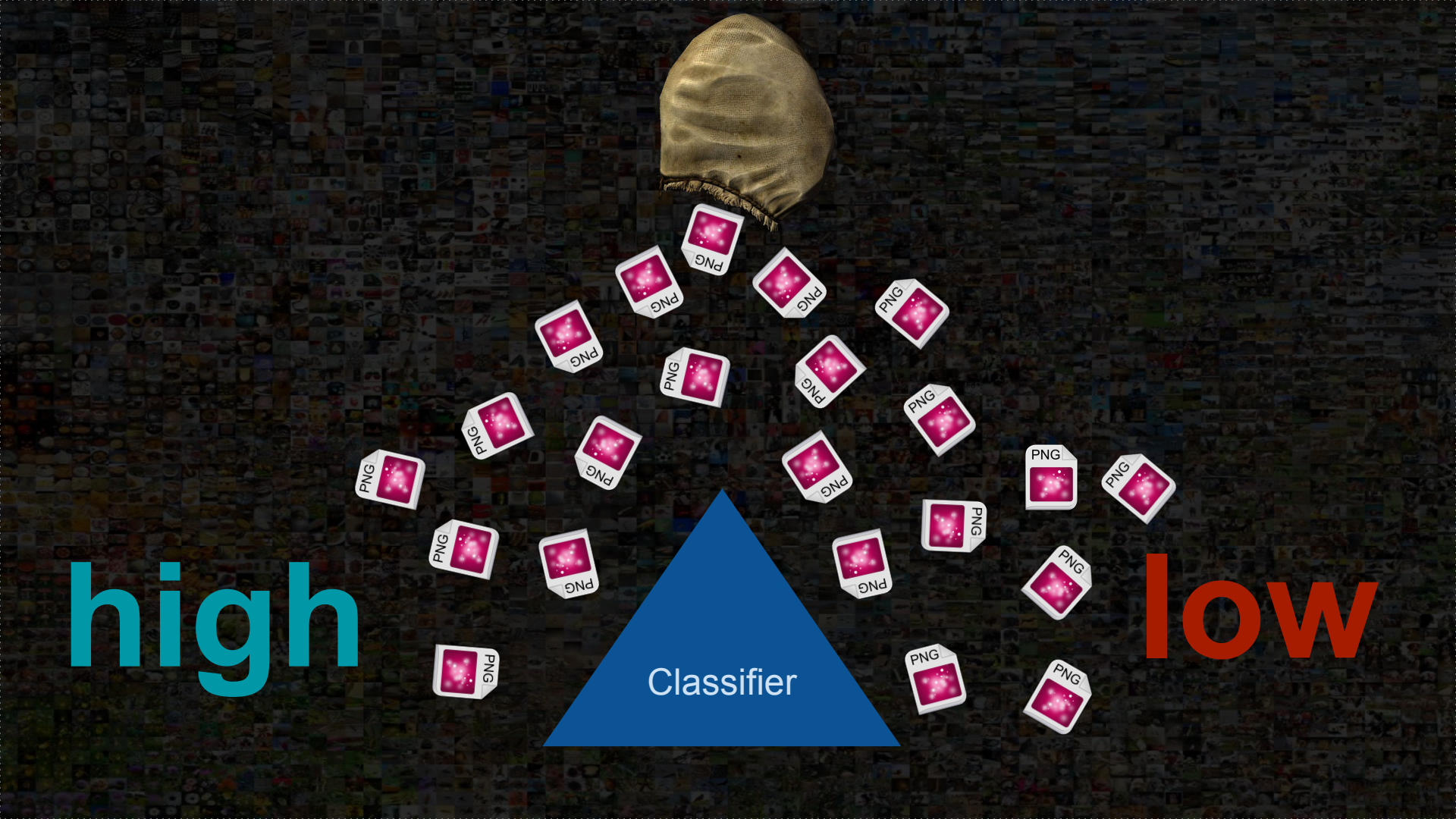


MultiGap: Multi-Pooled Inception Network with Text Augmentation for Aesthetic Prediction of Photographs

Yong-Lian Hii, John See, Magzhan Kairanbay, Lai-Kuan Wong


PNG
PNG
PNG
PNG
PNG
PNG
PNG
PNG
You
Tube
31
Internet


high
low
Classifier
PNG

Main contribution

1. Deep neural network architecture called MultiGAP that exploits features from multiple inception modules pooled by global average pooling (GAP), evaluated by prediction of (10-class) categorical distribution, before performing score binarization
2. The incorporation of textual features trained simultaneously with MultiGAP using a recurrent neural network (RNN) with gated recurrent unit (GRU) layers
3. The ability to leverage the GAP block for visualization of activation maps based on the aesthetic category and rating

Related work

Handcrafted low-level features

color, hue, saturation, light exposure, and also other heuristics driven by rule of thumbs used by professional photographers

Generic features

SIFT , feature encoding method such as a Fisher Vector

Deep learning models

CNN, Double Column CNN, Multi-modal CNN

Binary classification

VS

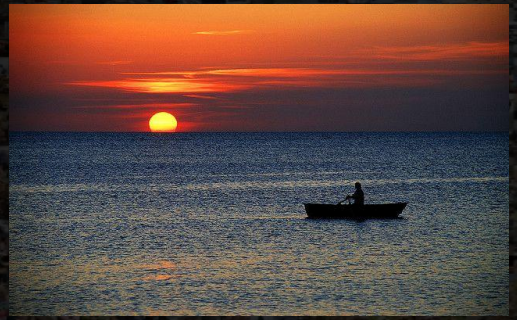
High

Low

Regression (Score)

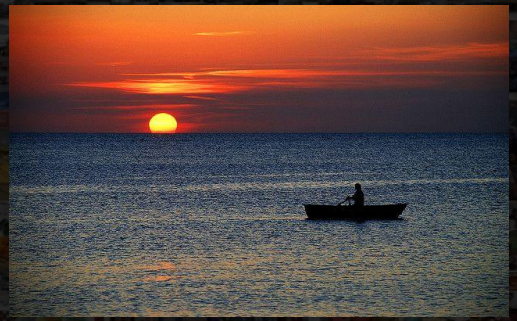
6.7

5.4

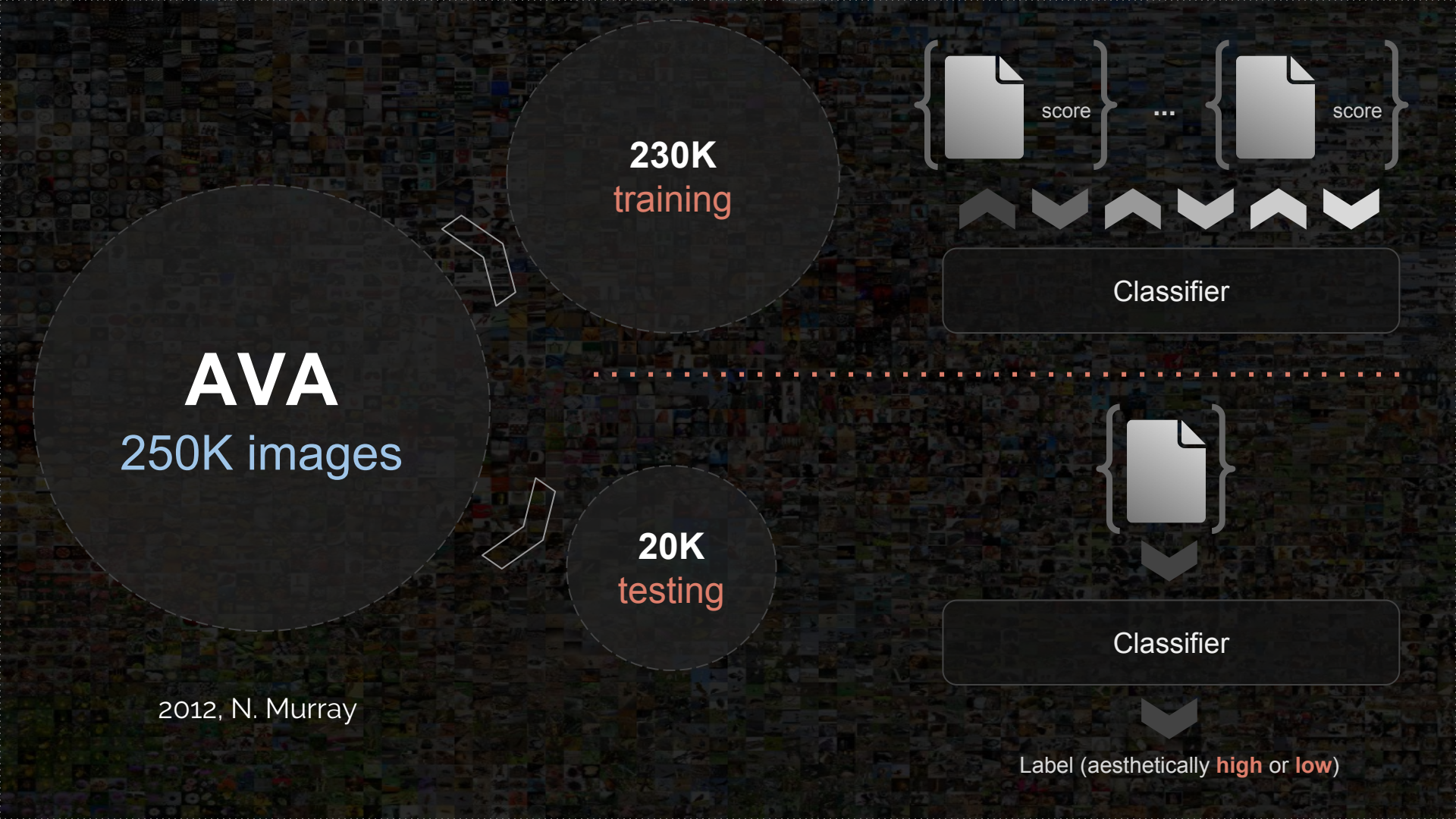


7.2

Rating distribution



1	2	3	4	5	6	7	8	9	10
0.2	0.04	0.04	0.02	0.25	0.24	0.01	0.12	0.07	0.01

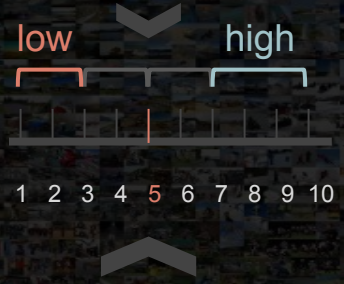

AVA
250K images
230K
training
20K
testing
score
...
score
Classifier
Classifier
Label (aesthetically high or low)
2012, N. Murray

AVA
250K images

230K
training

20K
testing

δ used to filter out from noisy images



threshold = 5

δ = {2, 1.5, 1, 0.5, 0}

threshold = 5

δ = {0}

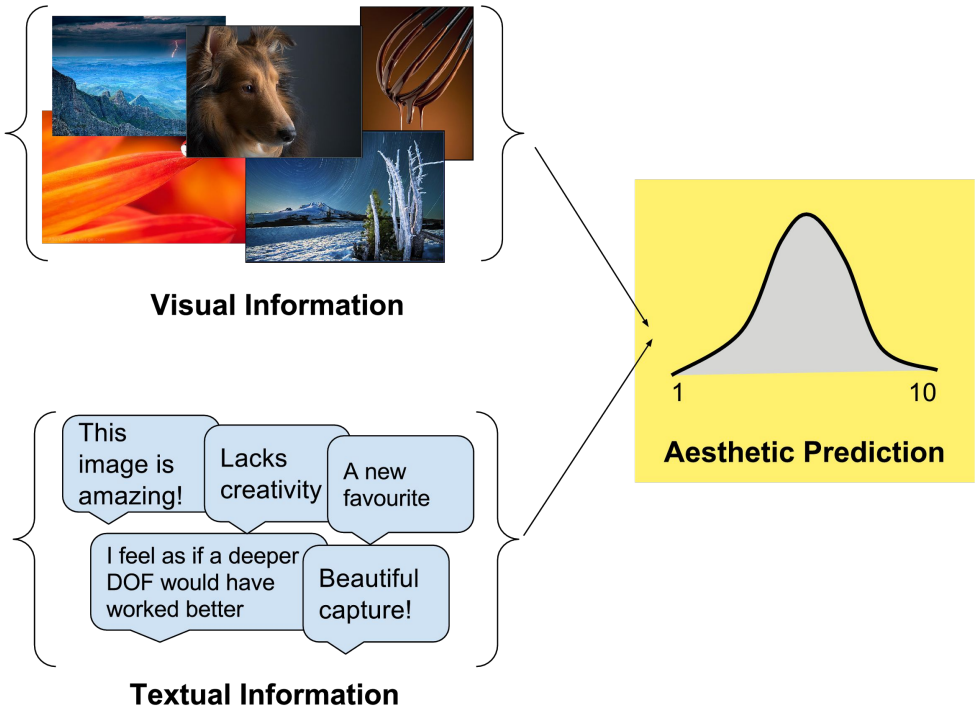
AVA
250K images

DPChallenge

AVA Comments
1.5 million comments

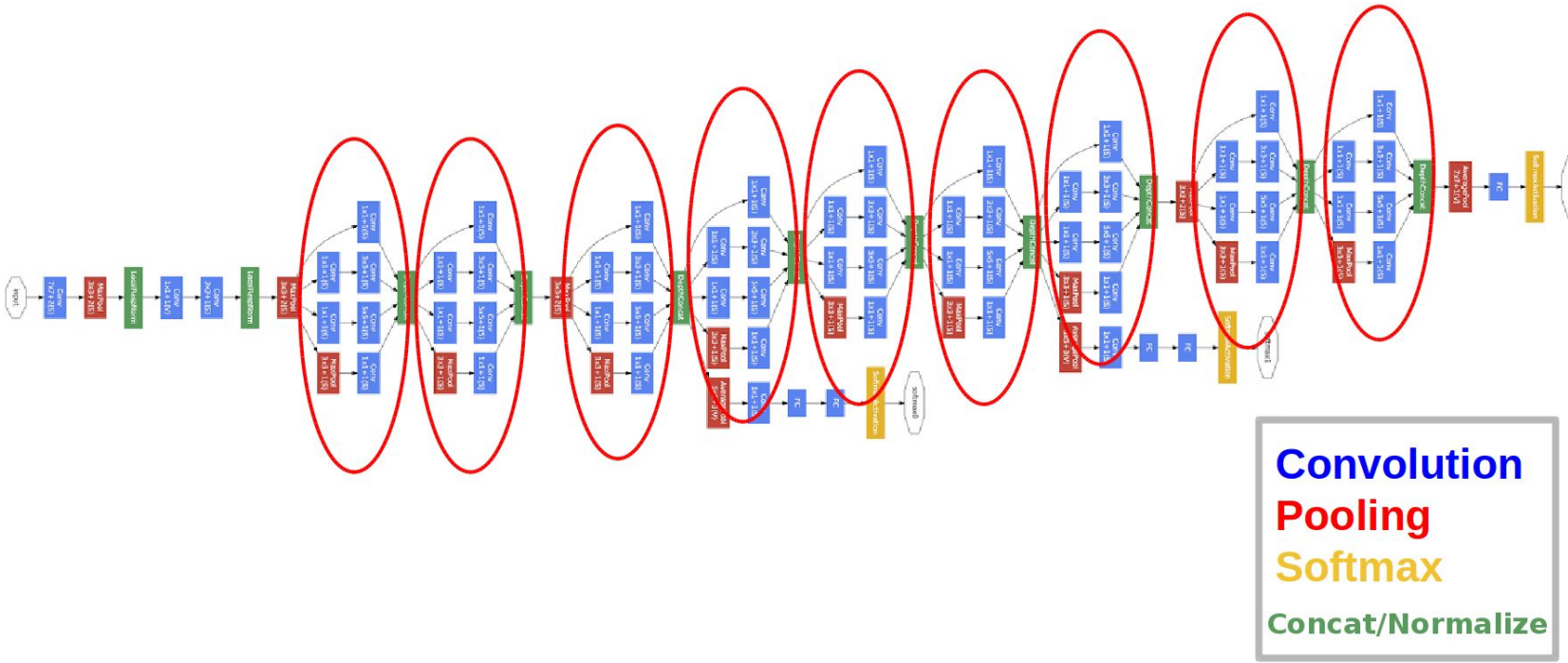
Zhou et al.

Proposed method



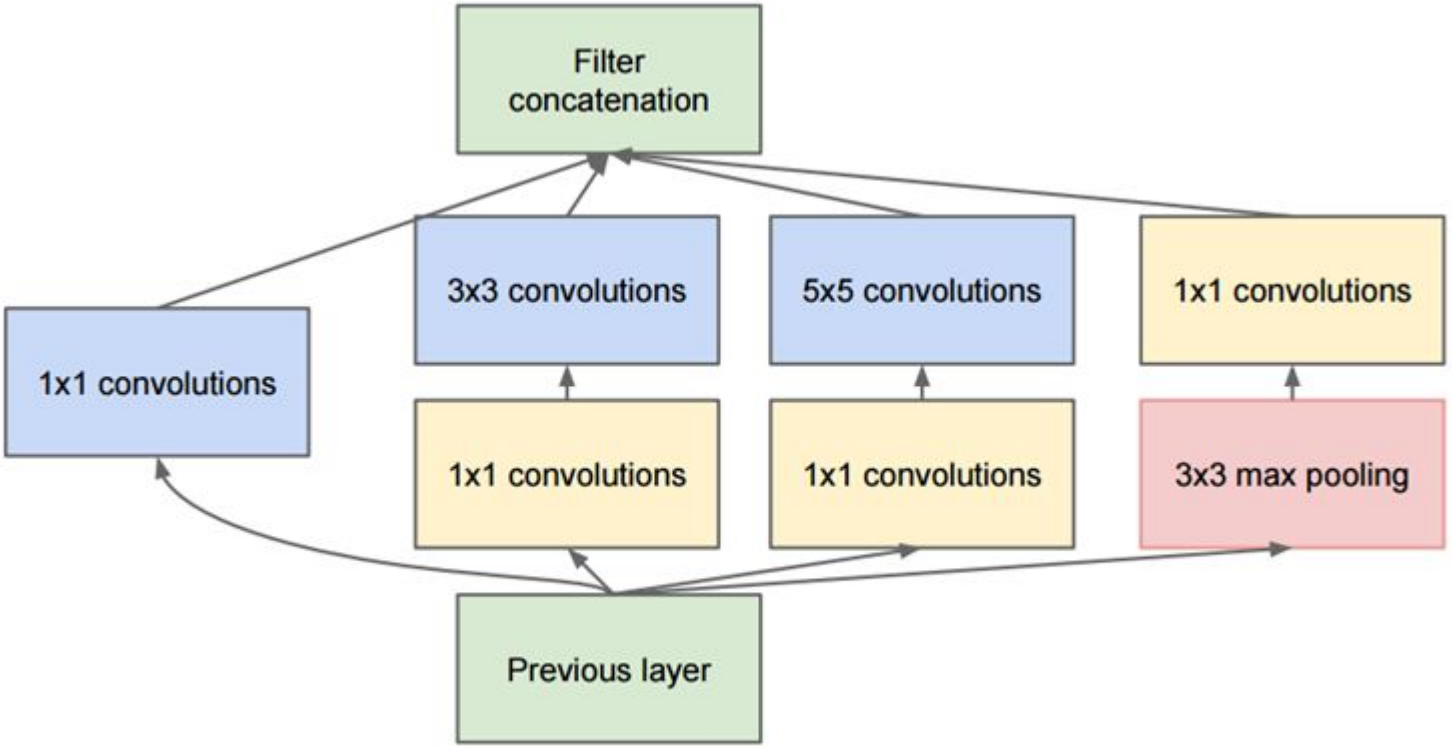
Visual features

GoogleNet



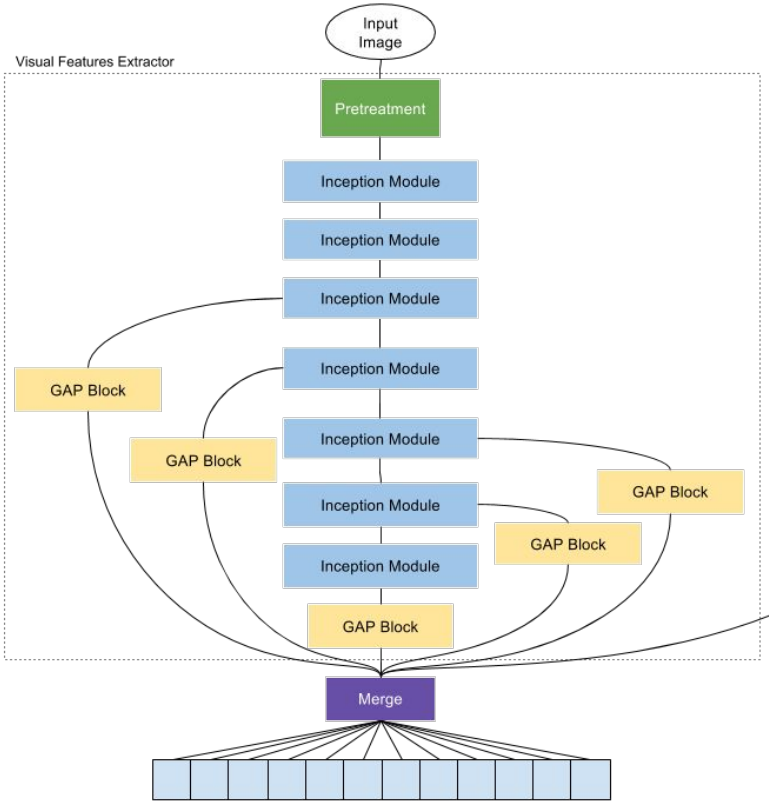
Visual features

Inception module

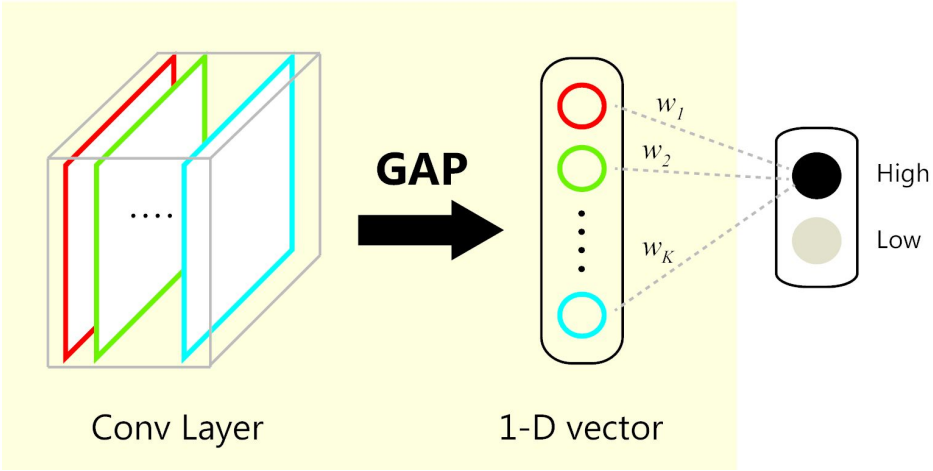


Visual features

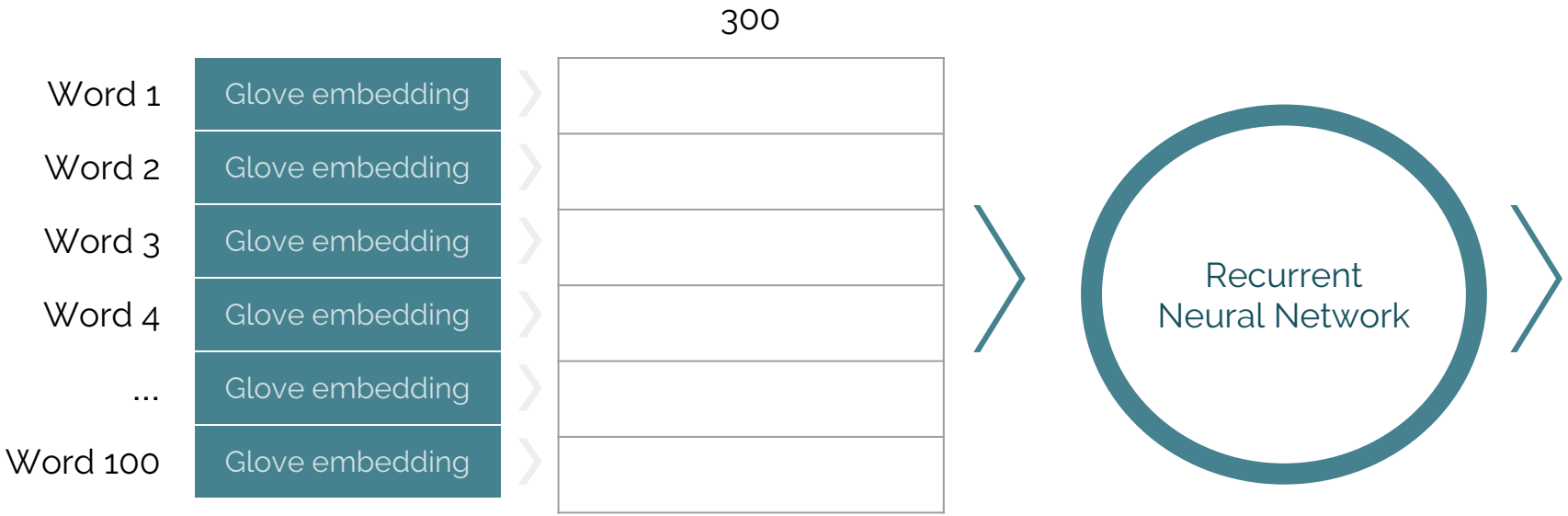
Proposed method



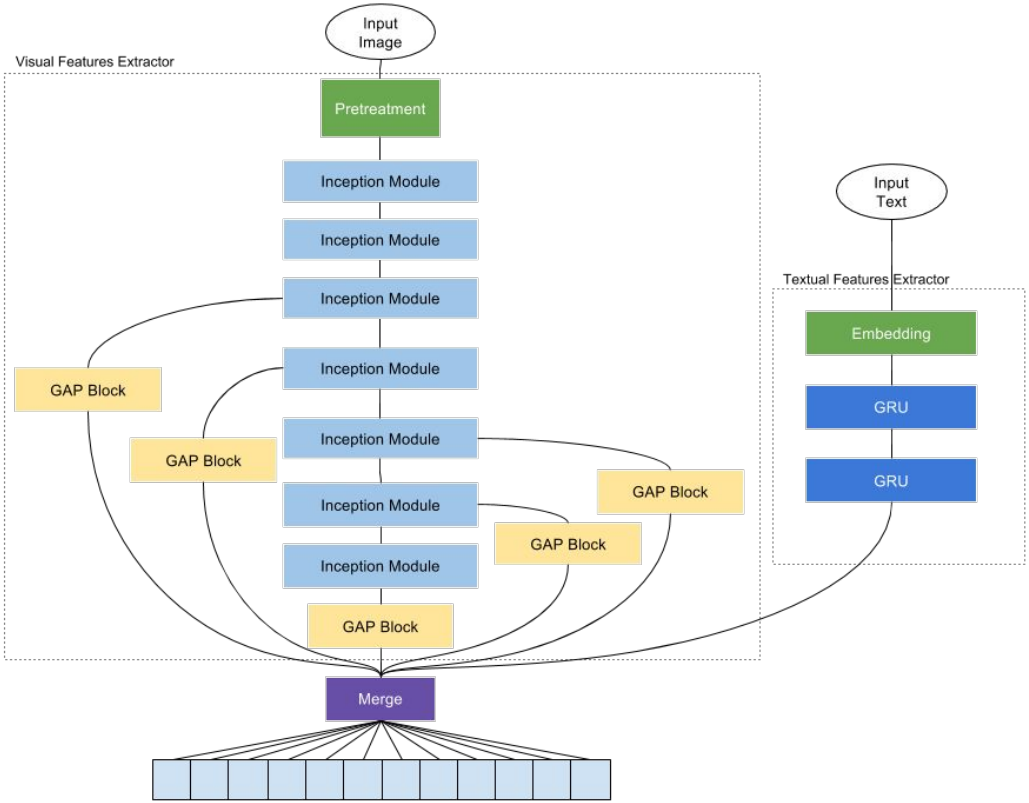
GAP (Global Average Pooling) Layer



Textual Features



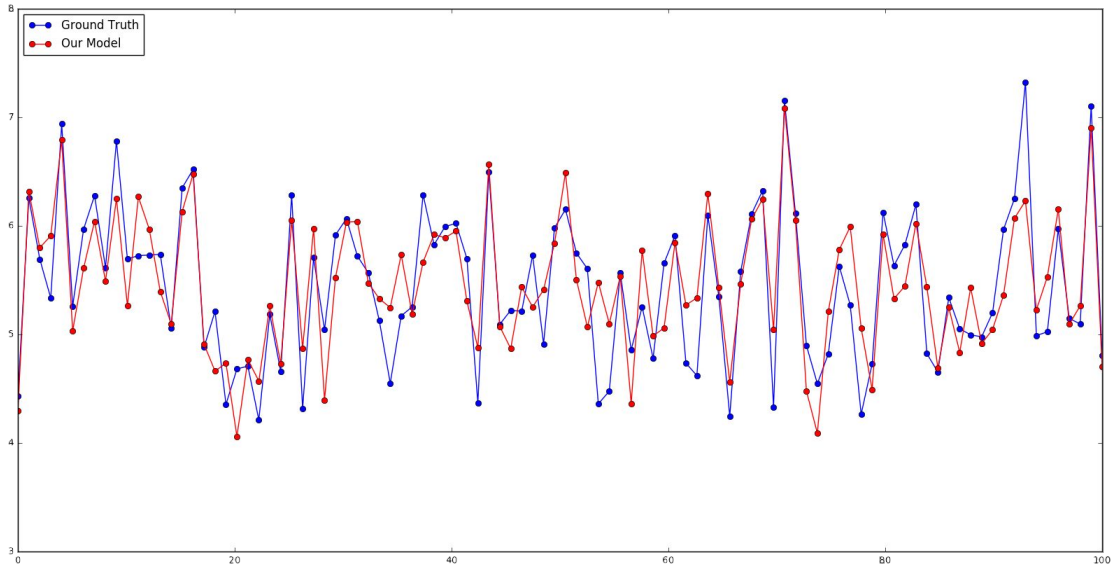
Proposed method



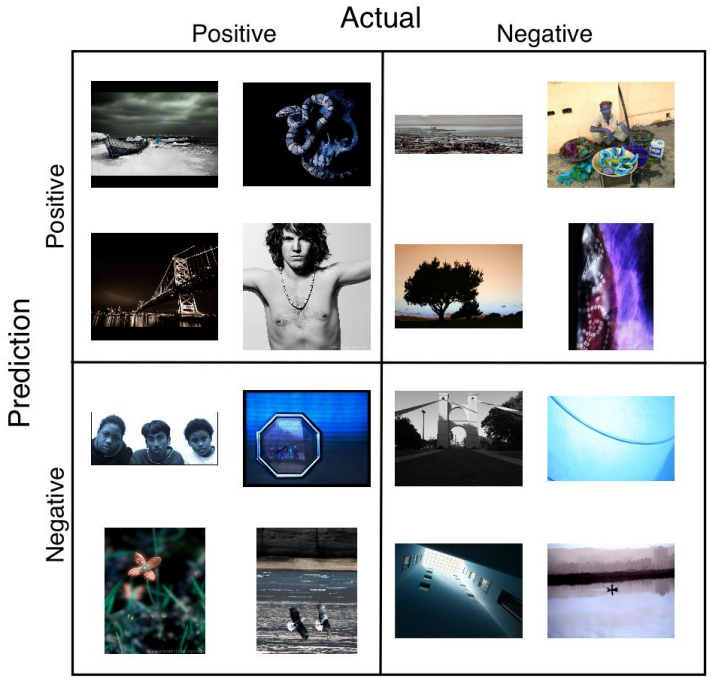
Results

	Model	Accuracy
Image	DCNN [2]	73.25
	RDCNN [2]	74.46
	Kao et al. [12]	74.51
	AlexNet [10] – finetuned	75.11
	DMA [5]	75.41
	GoogLeNet [16] – finetuned	75.60
	MultiGAP	**75.76**
	SingleGAP	**76.31**
	BDN [13]	76.80
Text	word2vec [19]	78.40
	1D-CNN [20]	79.48
	Naive Bayes SVM [14]	80.90
	RNN (1-layer GRU)	**81.09**
	RNN (2-layer GRU)	**81.79**
Joint	Multimodal DBM [14]	78.88
	SingleGAP + RNN (2-layer GRU)	**80.54**
	MultiGAP + RNN (2-layer GRU)	**82.27**

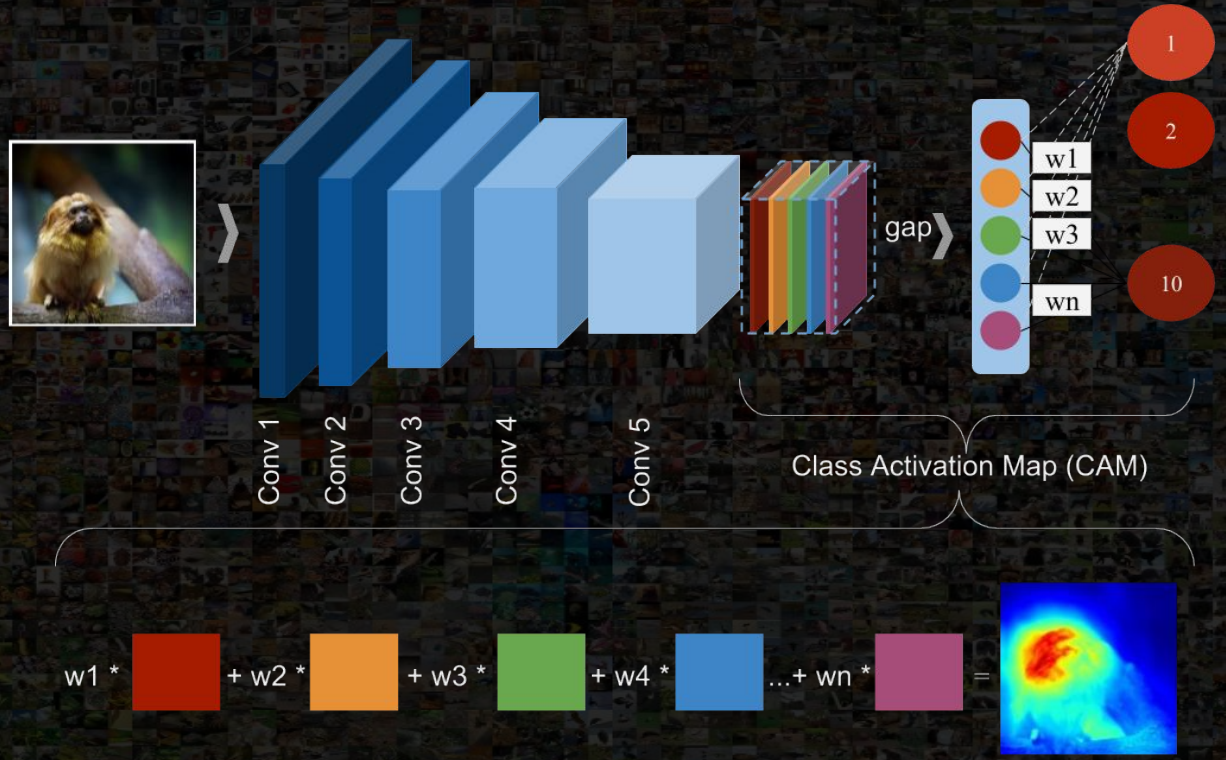
Results



Confusion matrix



Class Activation Map (CAM)



Class Activation Maps (CAM)

(a) (b)

Thank you for
your attention

Q&A