AMNet: Memorability Estimation with Attention

Jiri Fajtl[1], Vasileios Argyriou[1], Dorothy Monekosso[2] , Paolo Remagnino[1]
[1]Kingston University, London, UK [2]Leeds Beckett University, Leeds, UK
{J.Fajtl, Vasileios.Argyriou, P.Remagnino}@kingston.ac.uk, D.N.Monekosso@leedsbeckett.ac.uk

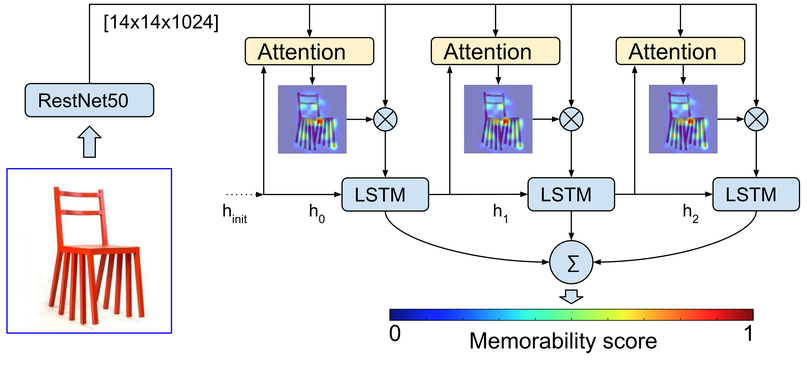
Kingston University London
Robot Vision Team (RoViT)

CVPR 2018
SALT LAKE CITY • JUNE 18-22

1. Introduction

- Deep learning model for prediction of how memorable is a still image.
- Trained on manually annotated datasets LaMem and SUN Memorability.
- Goal is to improve accuracy of ML models for memorability prediction.
- Learns memorability in a sequence similar to “eye gazes”.
- Achieves state of the art performance, on a par with humans.
- Generic architecture applicable to other tasks for regression on images, e.g. image aesthetic prediction.

2. Network Architecture



- Regression model, input is an image and output a single number between 0 and 1.
- Soft attention is used to learn spatial regions correlated with memorability given previous output state.
- LSTM applied to “gaze” on the image regions.
- Three LSTM steps experimentally found to be optimal.
- Input are CNN features extracted from 43rd layer of ResNet50 trained on ImageNet.
- Cost function encourages to explore different regions over the LSTM steps.

3. Demo & Source Code

- Demo https://amnet.kingston.ac.uk/
- Publication https://arxiv.org/abs/1804.03115
- Source code https://github.com/ok1zjf/AMNet

4. Results

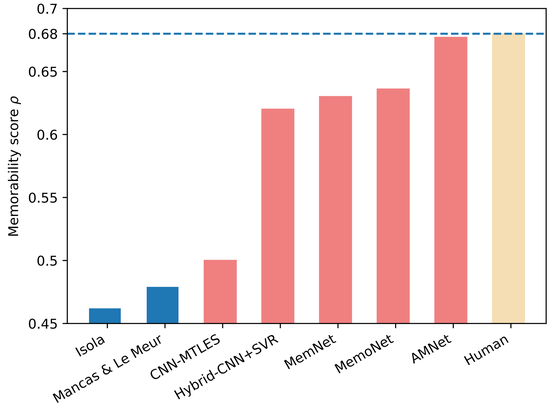


Figure 1: Comparison against the state of the art methods. In red are deep learning based methods.

Method (LaMem dataset)	ρ ↑	MSE ↓
AMNet	**0.677**	**0.0082**
AMNet (no attention)	0.663	0.0085
MemNet [1]	0.64	NA
CNN-MTLES [2]	0.5025	NA

Table 1: Average Spearman's rank correlation ρ and MSE over 5 test splits of the LaMem dataset.

Method (SUN Memorability dataset)	ρ ↑	MSE ↓
Isola [3]	0.462	0.017
AMNet	**0.649**	**0.011**
AMNet (no attention)	0.62	0.012
MemNet [1]	0.63	NA
Hybrid-CNN+SVR [4]	0.6202	0.013

Table 2: Evaluation on the SUN Memorability dataset. All models were trained and tested on the 25 train/val splits.

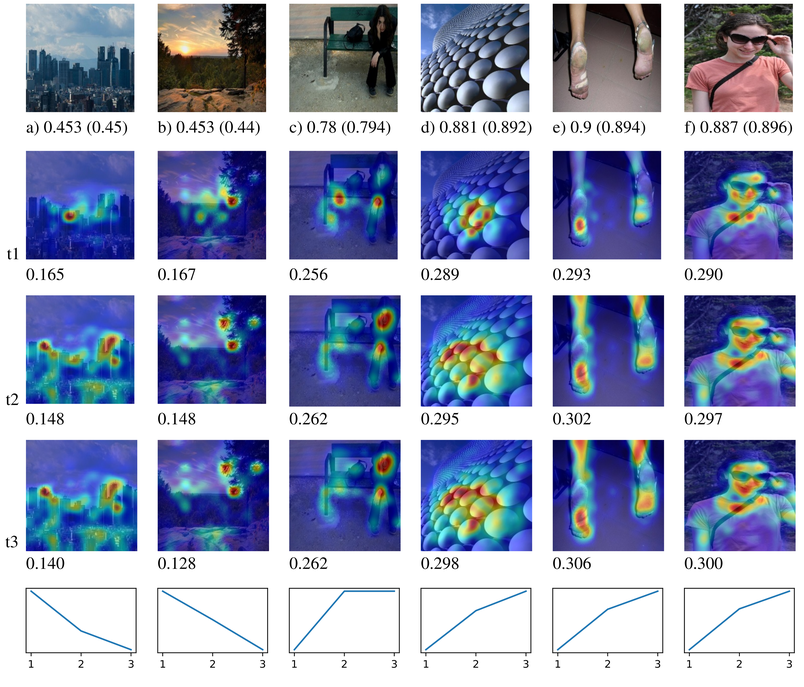


Figure 2: Examples of attention maps for low and high memorability images from LaMem test dataset. Tested images, their estimated and ground truth memorabilities (in brackets) are in the top raw. Bellow each image is a discrete memorability score estimated at the steps t1 , t2 and t3. Plots at the bottom row show gradients over three LSTM steps.

5. References

[1] A. Khosla et al. Understanding and predicting image memorability at a large scale. ICCV 2015.
[2] P. Jing et al. Predicting image memorability through adaptive transfer learning from external sources. IEEE Transactions on Multimedia 2017.
[3] P. Isola et al. What makes an image memorable? CVPR 2011.
[4] S. Zarezadeh et al. Image memorability prediction using deep features. ICEE 2017.