# AMNet: Memorability Estimation with Attention

## CVPR 2018 paper

Jiri Fajtl[1], Vasileios Argyriou[1], Dorothy Monekosso[2], Paolo Remagnino[1]
[1]Kingston University, London, UK
[2]Leeds Beckett University, Leeds, UK

Kingston University, London

November 14, 2018

# What is Image Memorability?

- Ability to recall an image after viewing it
- Is an intrinsic, stable property of image and can be quantified and measured
- Is consistent among independent observers
  ...but only for images unknown to the viewer
- Some images have special meaning to individuals, then the memorability is highly subjective
- We express memorability as real number in range [0,1]. Lower numbers mean forgettable images and vice versa

# Applications

1. Memorable advertising material
2. Organizing and tagging of photos in albums
3. Real-time image memorability estimation built into digital cameras
4. Make memorable presentations and data visualizations
5. Improving memorability of key parts of a GUI
6. Helping to illustrate education material
7. Monitor a decline in memory capacity of patients affected by dementia, such as Alzheimer's and Parkinson's diseases

# Using ML to Estimate Memorability

- No clear evidence what triggers memorability
- Unsupervised learning not successful, supervised requires labeling
- Two large, annotated datasets available:
  - LaMem [KRTO15] 60k images
  - SUN Memorability [IXTO11] 2k images
- Annotation done by means of a memorability game

Memorable

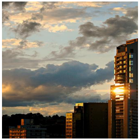
0.89

0.93

0.90

Forgettable

0.57

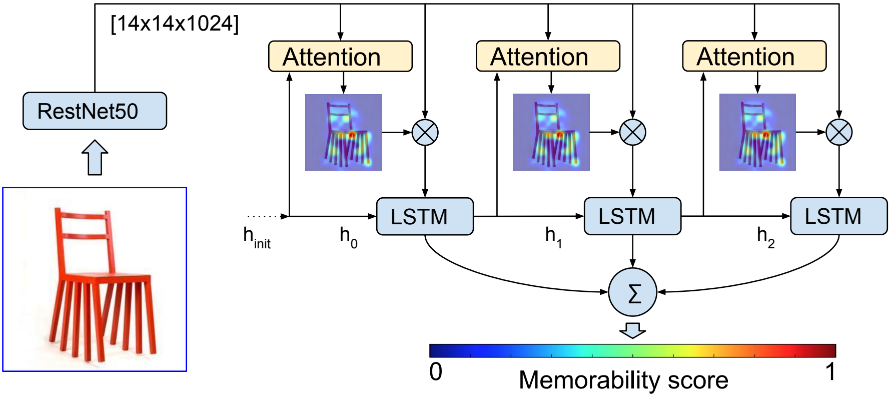


0.53

0.53

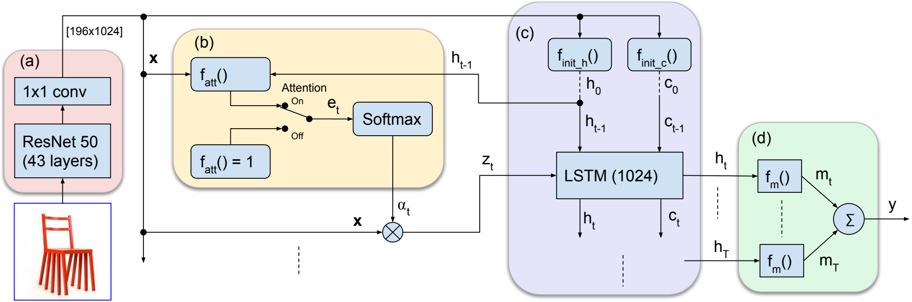
# Architecture

- Inspired by human eye saccades
- LSTM 'looks' at the image multiple times in a sequence
- At each step generates an attention map
- Attention expresses how important image regions are w.r.t. memorability



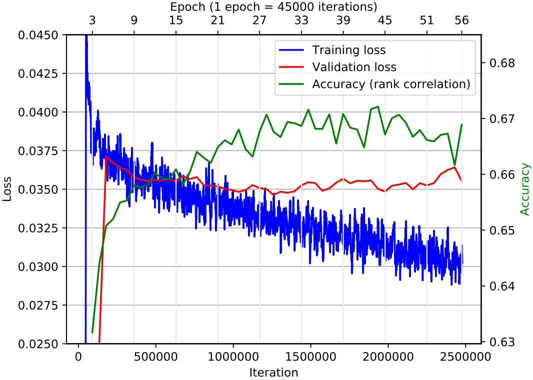
# Architecture

- Input is a ResNet50 CNN feature map 14x14x1024
- ResNet50 pre-trained on ImageNet
- Final memorability regression is done in two fully connected layers NN.



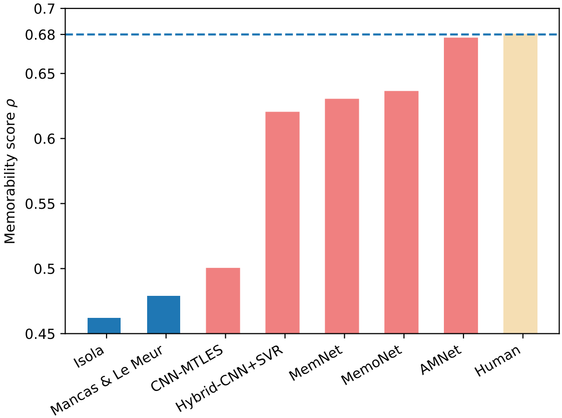
# Training

- Compound loss: MSE for memorability score + L1,L2 for attention weights
- ADAM optimizer [KB15]
- Dropout + L2 weights regularization
- Generalization attained by early stopping at highest val accuracy



# Evaluation

- 5-fold cross validation on LaMem
- 25-fold cross validation on SUN memorability
- Memorability reported as Spearman's rank correlation and MSE
- Compared with state-of-the-art (SOTA)



# Results

- Setting new SOTA on LaMem as well as SUN benchmarks
- Outperforms previous work by 5.8%
- Approaching human consistency $\rho = 0.68$

| **Method** (LaMem dataset) | $\rho \uparrow$ | MSE $\downarrow$ |
|---|---|---|
| AMNet | **0.677** | **0.0082** |
| AMNet (no attention) | 0.663 | 0.0085 |
| MemNet [KRTO15] | 0.64 | NA |
| CNN-MTLES [JSNG17] (different train/test (50/50) split) | 0.5025 | NA |

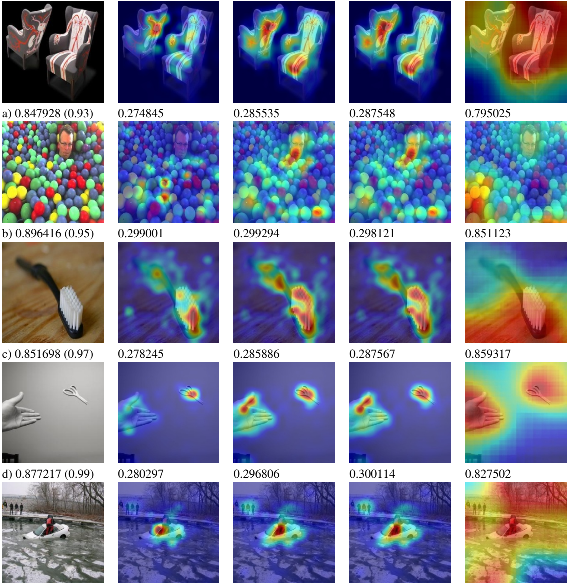
Table: Average Spearman's rank correlation $\rho$ and MSE over 5 test splits of the LaMem dataset.

# Results

| **Method** (SUN Memorability dataset) | $\rho \uparrow$ | MSE $\downarrow$ |
|---|---|---|
| Isola [IXTO11] | 0.462 | 0.017 |
| Mancas & Le Meur [MLM13] | 0.479 | NA |
| AMNet | **0.649** | **0.011** |
| AMNet (no attention) | 0.62 | 0.012 |
| MemNet [KRTO15] | 0.63 | NA |
| MemoNet 30k [BCPDSLC16] | 0.636 | 0.012 |
| Hybrid-CNN+SVR [ZRS17] | 0.6202 | 0.013 |

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

# Qualitative Results



# Qualitative Results

Legend for the previous slide

- Attention maps for $t_1, t_2$ and $t_3$ LSTM steps with discrete memorabilities bellow
- Estimated memorability followed by ground truth in brackets
- Memorability map in last column has been produced by MemNet [KRTO15]

# Demo & Code

- **Online demo**
  `https://amnet.kingston.ac.uk/`
- **Source code and trained models** (PyTorch 0.2)
  `https://github.com/ok1zjf/amnet/`

# References I


- Yoann Baveye, Romain Cohendet, Matthieu Perreira Da Silva, and Patrick Le Callet, *Deep learning for image memorability prediction: The emotional bias*, Proceedings of the ACM International Conference on Multimedia (New York, NY, USA), ACM, 2016, pp. 491–495.
- Phillip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva, *What makes an image memorable?*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2011, pp. 145–152.
- Peiguang Jing, Yuting Su, Liqiang Nie, and Huimin Gu, *Predicting image memorability through adaptive transfer learning from external sources*, IEEE Transactions on Multimedia **19** (2017), no. 5, 1050–1062.
- Diederik Kingma and Jimmy Ba, *Adam: A method for stochastic optimization*, Proceedings of the ICLR, vol. 5, 2015.

# References II


- Aditya Khosla, Akhil S. Raju, Antonio Torralba, and Aude Oliva, *Understanding and predicting image memorability at a large scale*, Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2390–2398.
- Matei Mancas and Olivier Le Meur, *Memorability of natural scenes: The role of attention*, International Conference on Image Processing, IEEE, 2013, pp. 196–200.
- Soodabeh Zarezadeh, Mehdi Rezaeian, and Mohammad T. Sadeghi, *Image memorability prediction using deep features*, Iranian Conference on Electrical Engineering (ICEE), IEEE, 2017, pp. 2176–2181.