Institute for
Infocomm Research
I²R

Predicting Event Memorability from Contextual Visual Semantics

Qianli Xu[1], Fen Fang[1], Ana Garcia del Molino[2], Vigneshwaran Subbaraju[3], Joo-Hwee Lim[1,4]

1. Institute for Infocomm Research, A*STAR, Singapore
2. ByteDance AI Lab, Singapore
3. Institute of High Performance Computing, A*STAR, Singapore
4. Nanyang Technological University

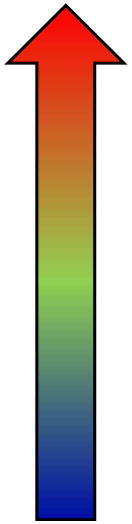
NeurIPS 2021
35th Conference on Neural Information Processing Systems
Online
December 6–14, 2021

Octboer 2021

Background

Image Memorability (Isola et al., 2011) [1]

“Memorable”



“Forgettable"

Image source [1]

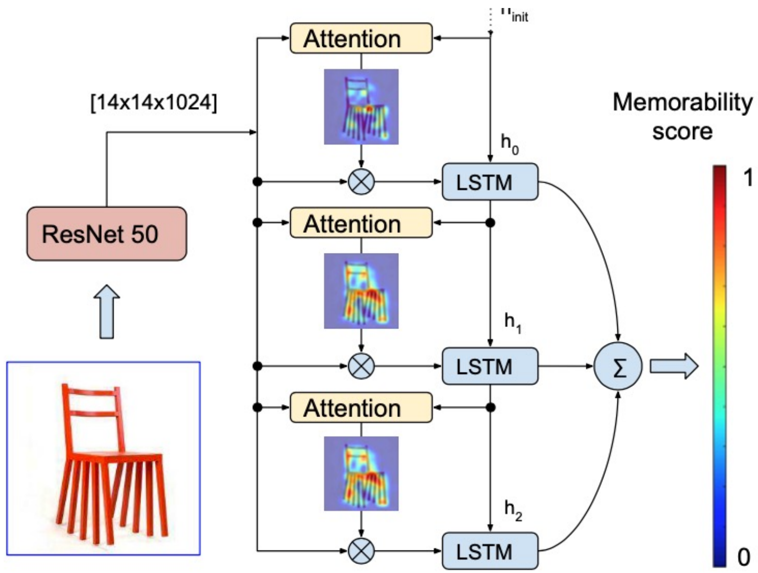


Image source [8]

Motivation

Event Memorability

Event memorability ≠ **Image** memorability

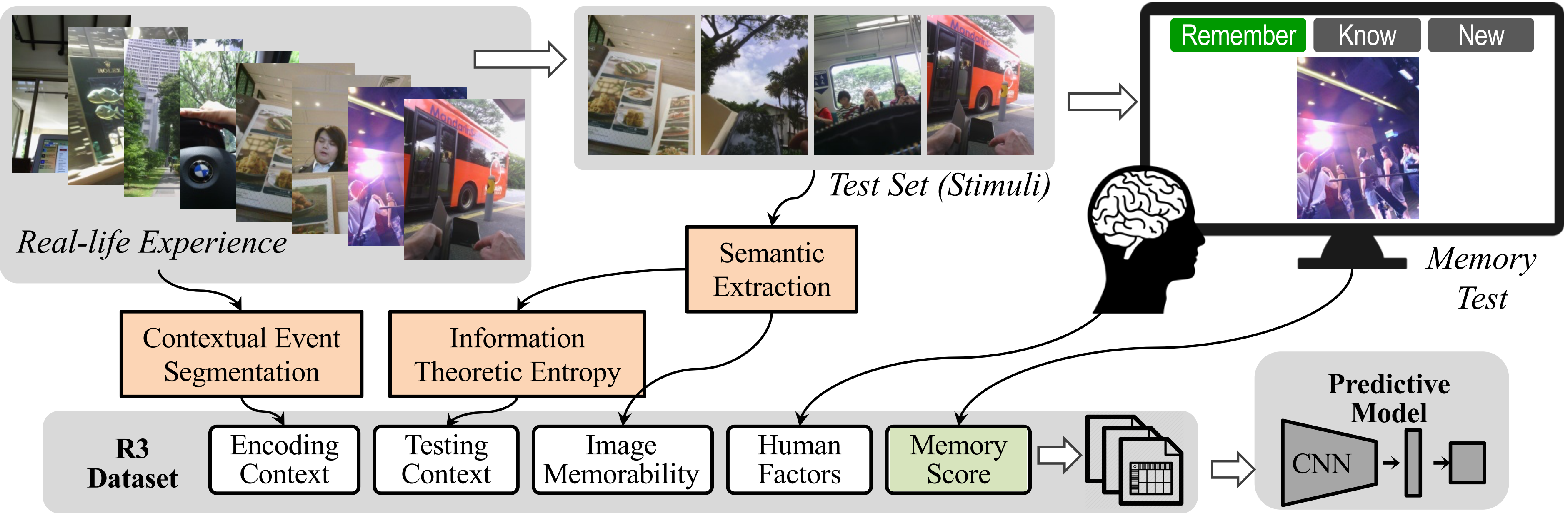
- What factors affect event memorability?
- **Can we predict a person's memory of individual events?**
- [Future work] Can we design cognitive intervention programs to enhance subjects' episodic memory by leveraging on the knowledge of event memorability?

Related works

- Image Memorability
  - What makes an [image/object/graph/scene/] memorable? [1-5]
  - Video memorability [6,7]
- Dataset
  - SUN-Mem [1], LaMem [2], FIGRIM [3], Mem-Cat [4], LNSIM [5]
- Predictive models
  - MemNet [2], AMNet [8], DeepNSM [5], ICNet [9]
- Event Memorability
  - Neuro-psychological studies [10]: only study memorability of event categories
  - Neural imaging [11,12]: intrusive and expensive, difficult to get data

CREATING GROWTH, ENHANCING LIVES

Our Approach

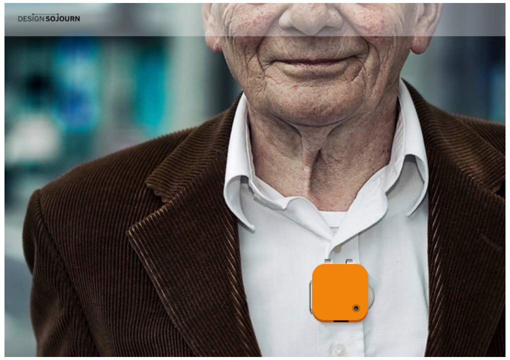


- ***An experiment protocol***: lifelogging, systematic training, standardized testing
- ***A dataset (R3)***: Sophisticated mechanisms to extract visual semantic features
- ***A predictive model***: compute item-wise event memorability

CREATING GROWTH, ENHANCING LIVES

Experiment design and dataset

Phase I			Phase II			Phase III	
14 days Recording	3 days	Test 1	14 days Recording	3 days	Test 2	7 wks	Test 3



8 s

Memory Type

- Remember
- Know
- New

3 s

Memory Level

- R: 1 2 3 4
- K: 1 2 3 4
- N: perhaps vs. certain

3 s

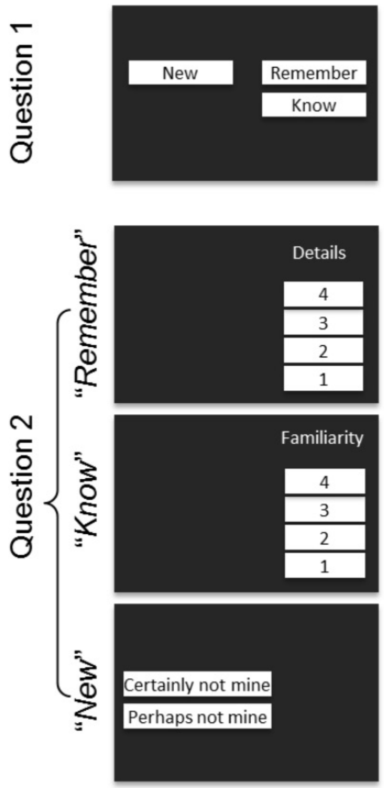
- Pilot study (8 young + 5 old); formal study (47 old, reported in this study)
- 1 month lifelog recording from each subject; >13K hours, >1.5M photos with meta-data
- >10K samples of event recall (reported in this study; privacy sensitive information removed)

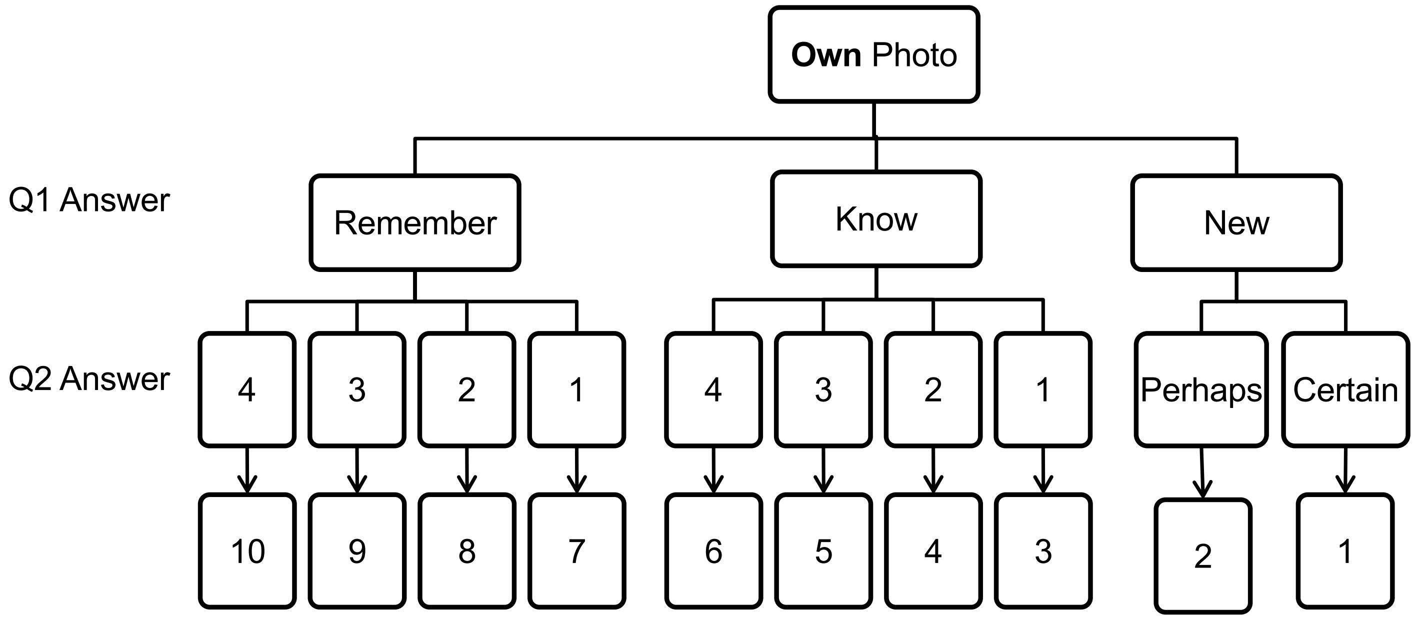
Visual Semantic Features in Context

- Intrinsic (pure visual; from image cue)
  - Saliency: image memorability, multiple benchmarking models used
  - Face: presence of human face
  - Human: presence of human body
- Extrinsic
  - Encoding context
    - Event distinctiveness: Rare events are remembered better (Hunt and Worthen, 2006); computed using CES method (del Molino et al., 2018)
    - Event boundary condition: Event segmentation theory (Gold et al. 2017)
    - Activity: Manually annotated based on local context
    - Place: Manually annotated based on local context
  - Testing context
    - Event distinctiveness: computed based on information-theoretic entropy (Bylinskii et al., 2015)
    - Encode-test interval: time between event occurrence and testing
    - Training (treatment): event has been re-consolidated before
  - Demographic
    - Age
    - gender
    - Etc.

CREATING GROWTH, ENHANCING LIVES

Memory Mnemonics





(b) Encode test outcome as 10-level graded memory types

Linear regression analysis

Hypothesis: event memorability is dependent on both intrinsic and extrinsic features

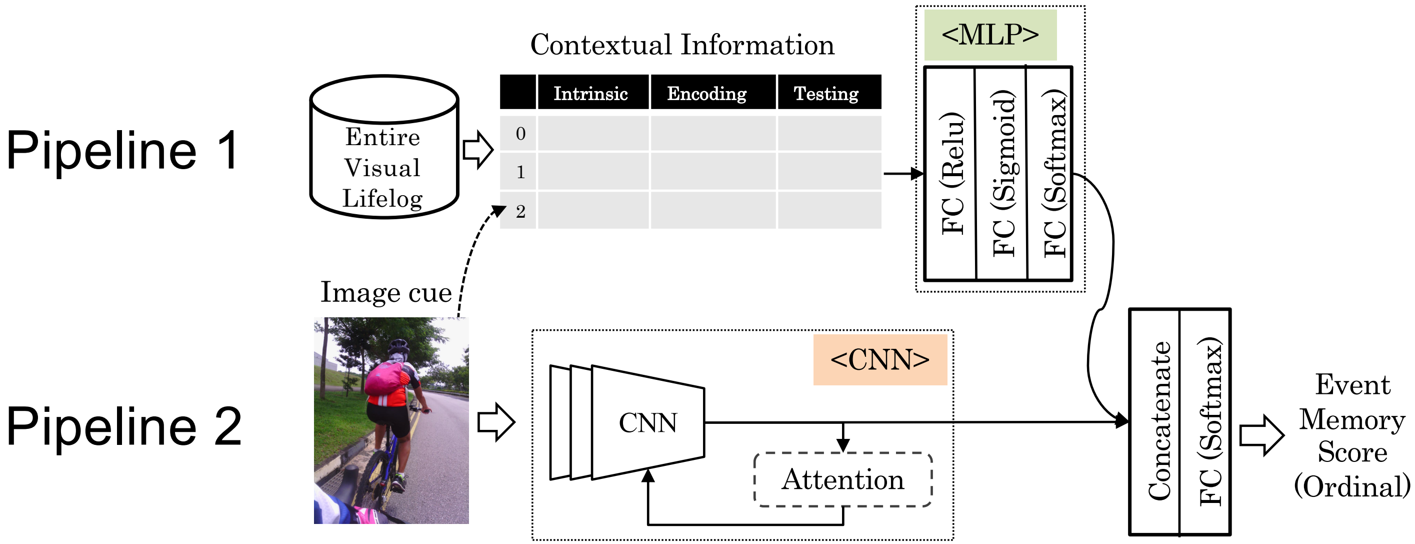
- Intrinsic feature (i.e., image memorability) predicts event memory to vary extends based on different predictive models.
  - MemNet [28]: $r = 0.02$, $p = 0.04$,
  - DeepNSM [33]: $r = 0.01$, $p = 0.54$
  - **AMNet [15]: $r = 0.19$, $p < 0.001$**

- Linear mixed-effect analysis: features are used as fixed effects and "subject" modelled as a random effect.

Intrinsic Factors	t-statistics	**Encoding Context**	t-statistics	**Testing Context**	t-statistics
Image memorability	**11.14**	Encode distinctiveness	**7.45**	Test distinctiveness	**3.93**
Presence of faces	**10.55**	Boundary condition	-0.73	Treatment	**10.00**
Presence of human	**2.08**	Activities	**7.34**	Interval	**-14.22**
		Places	1.38		

Table 1: Factors that affects event memorability. t-value in bold font means the factor significantly correlated with memory ($p < 0.05$).

CEMNet – Predicting item-wise event memory



Pipeline	AMNet	ICNet	MLP	CEMNet wt AMNet	CEMNet wt ICNet
1	-	-	MLP	MLP	MLP
2	AMNet[8]	ICNet[9]	-	AMNet	ICNet

Code available @ https://github.com/ffzzy840304/Predicting-Event-Memorability

Experiment results

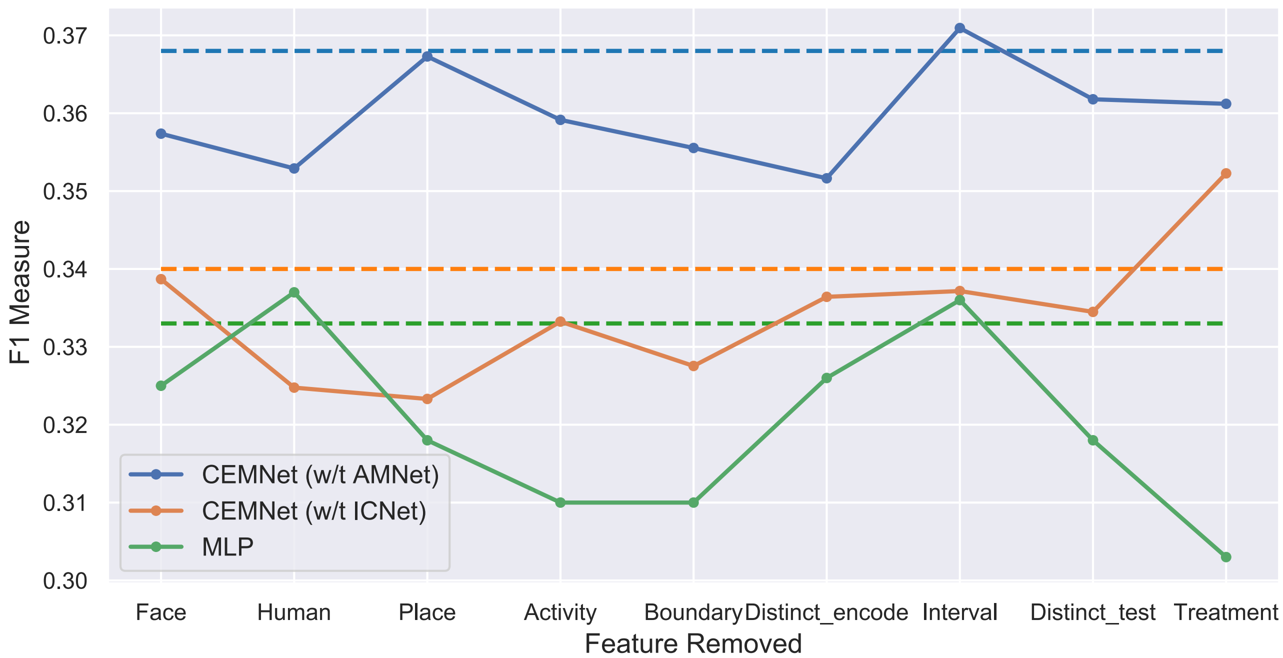
Method	Input Features	Precision↑	Recall↑	F1↑	Mean error↓
AMNet [8]	Image	0.171	0.179	0.150	3.03
ICNet[9]	Image	0.153	0.155	0.140	3.11
MLP	Extrinsic Features*	0.389	0.385	0.333	0.91
CEMNet w/t AMNet	Intrinsic + Extrinsic	**0.408**	**0.414**	**0.368**	**0.85**
CEMNet w/t ICNet	Intrinsic + Extrinsic	0.369	0.340	0.340	0.97

Table 2: Comparing performance of models. *Intrinsic features, *i.e.*, human face & body, are included.

- Intrinsic features (i.e., using only image cues) have limited predictive power; above chance accuracy.
- Extrinsic features (MLP model) can predict event memorability with considerable accuracy.
- Combining intrinsic and extrinsic gives best prediction outcome; Especially using more comprehensive DCNN model (e.g., AMNet)

CREATING GROWTH, ENHANCING LIVES

Ablation Study – *which feature has higher predictive value?*



- Using all features generally gives better performance
- Most features are conducive to the performance, except “human” and “interval”
- Some factors co-vary with each other, which may have caused inconsistent outcome. No causal relationship is established.

CREATING GROWTH, ENHANCING LIVES

CREATING GROWTH, ENHANCING LIVES

Summary

- Event memory can be effectively predicted with intrinsic + extrinsic factors
- Extrinsic factors are more important in event memory prediction

- R3 experiment and dataset may inspire new experiments to investigate on event memory
- We can leverage on the outcome of predicted event memory to design memory intervention programs

CREATING GROWTH, ENHANCING LIVES

References

1. P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. NIPS'11, pages 2429–2437, 2011.
2. A. Khosla, A. S. Raju, A. Torralba, and A. Oliva. Understanding and predicting image memorability at a large scale. *ICCV*, pages 2390–2398, 2015.
3. Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, and A. Oliva. Intrinsic and extrinsic effects on image memorability. *Vision Research*, 116(Pt B):165–178, 2015.
4. L. Goetschalckx and J. Wagemans. Memcat: a new category-based image set quantified on memorability. *PeerJ*, 7, 2019.
5. J. Lu, M. Xu, R. Yang, and Z. Wang. Understanding and predicting the memorability of outdoor natural scenes. *IEEE Transactions on Image Processing*, 29:4927–4941, 2020.
6. Anelise Newman, Camilo Luciano Fosco, Vincent Casser, Allen Lee, Barry, Mcnamara, and Aude Oliva. Multimodal memorability: Modeling effects of semantics and decay on video memorability. *ArXiv*, abs/2009.02568, 2020.
7. Romain Cohendet, Claire-Hélène Demarty, Ngoc Q. K. Duong, and Martin Engilberge. Videomem: Constructing, analyzing, predicting short-term and long-term video memorability. 2531–2540, 2019.
8. J. Fajtl, V. Argyriou, D. Monekosso, and P. Remagnino. Amnet: Memorability estimation with attention. *2018 CVPR*, 6363–6372, 2018.
9. H. Squalli-Houssaini, N.Q.K. Duong, G. Marquant, and C. Demarty. Deep learning for predicting image memorability. *ICASSP*, pages 2371–2375, 2018.
10. P.L. St Jacques, C. Olma, and D.L. Schacter. Neural mechanisms of reactivation-induced updating that enhance and distort memory. *PNAS*, 110(49):19671–78, 2013.
11. J. Rissman, T.E. Chow, N. Reggente, and A.D. Wagner. Decoding fMRI signatures of real-world autobiographical memory retrieval. *Journal of Cognitive Neuroscience*, 28(4):604–620, 2016.
12. W. A. Bainbridge, Daniel D. Dilks, and A. Oliva. Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage*, 149:141–152, 2017.

CREATING GROWTH, ENHANCING LIVES

Institute for Infocomm Research

I²R

THANK YOU

qxu@i2r.a-star.edu.sg

www.a-star.edu.sg

This research is partially funded by the Singapore Agency for Science, Technology and Research (A*STAR) JCO REVIVE Project (1335h0009).