MediaEval'2018 – Task overview talk

Predicting Media Memorability

R. Cohendet, C.-H. Demarty, N. Q. K. Duong, M. Sjöberg, B. Ionescu, & T.-T. Do

technicolor

Aalto University
School of Science

POLITEHNICA UNIVERSITATEA DIN BUCUREȘTI 1818

ARC Centre of Excellence for ROBOTIC VISION

MediaEval'2018, 29-31 October 2018, Sophia Antipolis, France

The task

Participants are expected to design systems that automatically **predict memorability scores for videos**, which reflect the probability of a video being remembered.

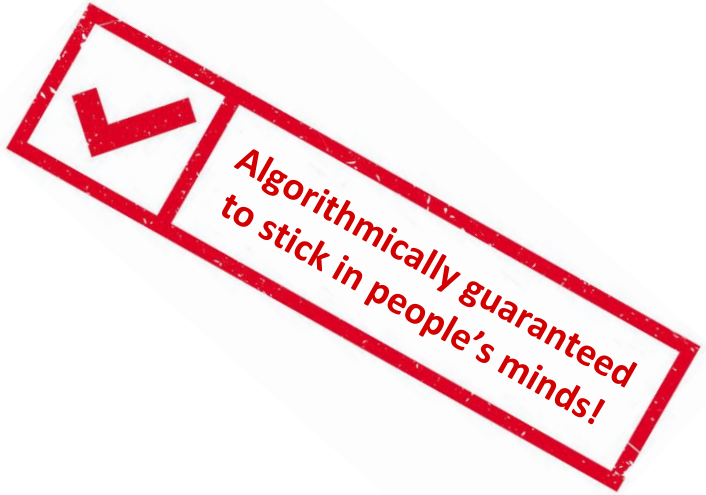
Memorability *(after ~2 days):* ***97%***

Memorability *(after ~2 days):* ***18%***

Human **share a strong tendency** to memorize/forget the same visual information. [*Isola et al., 2014*]

Automatic Prediction of Video Memorability

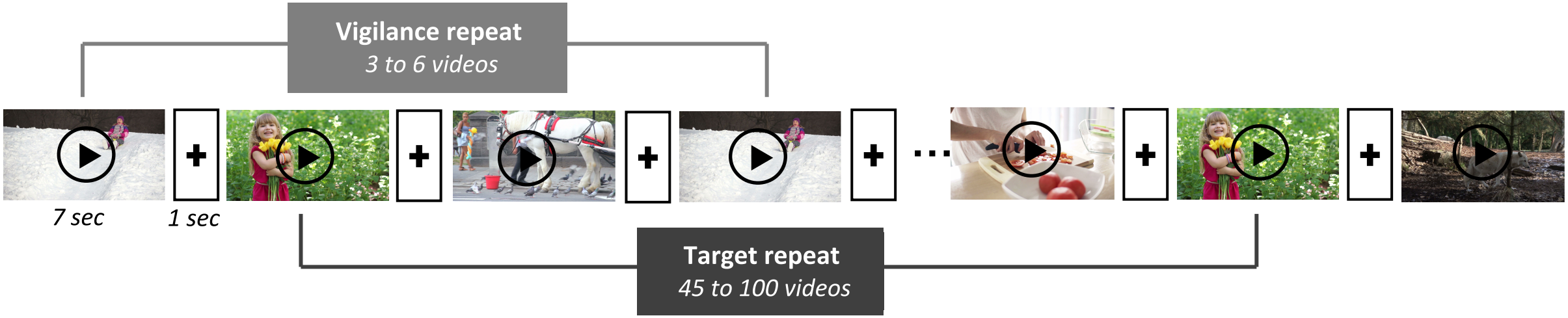
- Content retrieval and search
- Education & learning
- Health care
- Content summarization
- Content recommendation and filtering
- etc.



- Relative *short-term* memorability (consolidation process) [*McGaugh, 2000*]

Protocol to measure human memory

Short-term memorability measurement



Long-term memorability measurement (after 24-72 hours)

10,000 soundless videos with memorability scores

	dev-set	test-set
Videos	8,000	2,000 (including the 500 most annotated videos)
Ground truth	✓	✗
Nb of annotations	✓	✓
Titles	✓	✓
Pre-computed features	➢ Video-dedicated features (*C3D, HMP*) ➢ Frame-based features (*Color hist., HOG, LBP, ORB, InceptionV3, aesthetics*)	

TEAM	SHORT-TERM SUBTASK	LONG-TERM SUBTASK	APPROACH
GIBIS	—	0,118	HMP => SVR or *k*-NN regressor
Insight@DCU	0,056	0,039	Provided features => NN
HKBU	0,133	0,096	C3D, LBP, Color Hist. => SVR
TCNJ-CS	0,34	0,093	C3D, InceptionV3, HMP, Color Hist., Aesthetics + MemNet => NN
Show-And-Recall	0,355	0,15	C3D, InceptionV3, HMP, Color Hist., HoG + Saliency + Captions => NN
The Huskies	0,356	0,213	Captions => RNN
MMC-RUC	0,448	0,24	C3D, HMP, Aesthetics + I3D + Captions => SVR or RFR
Technicolor*	0,46*	0,219*	ResNet101 => fine-tuning (data augmentation)
AREA66	0,462	0,232	C3D, Aesthetics + MemNet + Saliency + face-based feature + Captions => weighted average/SVR
HCMUS	0,484	**0,257**	InceptionV3 + Captions => LSTM => Dense layer
ConduentLabs	**0,497**	0,253	Provided features + Captions + ResNet

*Organizer

Official results (Spearman corr.) on test-set for teams' best runs

TEAM	SHORT-TERM SUBTASK	LONG-TERM SUBTASK	APPROACH
GIBIS	—	0,118	HMP => SVR or *k*-NN regressor
Insight@DCU	0,056	0,039	Provided features => NN
HKBU	0,133	0,096	C3D, LBP, Color Hist. => SVR
TCNJ-CS	0,34	0,093	C3D, InceptionV3, HMP, Color Hist., Aesthetics + MemNet => NN
Show-And-Recall	0,355	0,15	C3D, InceptionV3, HMP, Color Hist., HoG + Saliency + Captions => NN
The Huskies	0,356	0,213	Captions => RNN
MMC-RUC	0,448	0,24	C3D, HMP, Aesthetics + I3D + Captions => SVR or RFR
Technicolor	0,46	0,219	ResNet101 => fine-tuning (data augmentation)
AREA66	0,462	0,232	C3D, Aesthetics + MemNet + Saliency + face-based feature + Captions => weighted average/SVR
HCMUS	0,484	**0,257**	InceptionV3 + Captions => LSTM => Dense layer
ConduentLabs	**0,497**	0,253	Captions + ResNet

=> Results are better for *Short-term* subtask than for *Long-term* subtask (≠ nb annotations)

TEAM	SHORT-TERM SUBTASK	LONG-TERM SUBTASK	APPROACH
GIBIS	—	0,118	HMP => SVR or *k*-NN regressor
Insight@DCU	0,056	0,039	Provided features => NN
Human intuition	**~0,1**	**—**	**[*Isola et al., 2014*; *Cohendet et al., 2016*]**
HKBU	0,133	0,096	C3D, LBP, Color Hist. => SVR
TCNJ-CS	0,34	0,093	C3D, InceptionV3, HMP, Color Hist., Aesthetics + MemNet => NN
Show-And-Recall	0,355	0,15	C3D, InceptionV3, HMP, Color Hist., HoG + Saliency + Captions => NN
The Huskies	0,356	0,213	Captions => RNN
Baseline	**0,39**	**0,17**	**MemNet on 7 frames**
MMC-RUC	0,448	0,24	C3D, HMP, Aesthetics + I3D + Captions => SVR or RFR
Technicolor*	0,46*	0,219*	ResNet101 => fine-tuning (data augmentation)
AREA66	0,462	0,232	C3D, Aesthetics + MemNet + Saliency + face-based feature + Captions => weighted average/SVR
HCMUS	0,484	0,257	InceptionV3 + Captions => LSTM => Dense layer
ConduentLabs	0,497	0,253	Provided features + Captions + ResNet
Human consistency	**0,50**	**0,28**	**for videos > 15 annotations**

TEAM	SHORT-TERM SUBTASK	LONG-TERM SUBTASK	APPROACH
GIBIS	—	0,118	HMP => SVR or *k*-NN regressor
Insight@DCU	0,056	0,039	Provided features => NN
HKBU	0,133	0,096	C3D, LBP, Color Hist. => SVR
TCNJ-CS	0,34	0,093	C3D, InceptionV3, HMP, Color Hist., Aesthetics + MemNet => NN
Baseline	**0,39**	**0,17**	**MemNet on 7 frames**
Show-And-Recall	0,355	0,15	C3D, InceptionV3, HMP, Color Hist., HoG + Saliency + Captions => NN
The Huskies	0,356	0,213	Captions => RNN
MMC-RUC	0,448	0,24	C3D, HMP, Aesthetics + I3D + Captions => SVR or RFR
Technicolor	0,46	0,219	ResNet101 => fine-tuning (data augmentation)
AREA66	0,462	0,232	C3D, Aesthetics + MemNet + Saliency + face-based feature + Captions => weighted average/SVR
HCMUS	0,484	**0,257**	InceptionV3 + Captions => LSTM => Dense layer
ConduentLabs	**0,497**	0,253	Captions + ResNet

=> Provided features, as inputs of NN or SVR, does not achieve our baseline (MemNet) performance

TEAM	SHORT-TERM SUBTASK	LONG-TERM SUBTASK	APPROACH
GIBIS	—	0,118	HMP => SVR or *k*-NN regressor
Insight@DCU	0,056	0,039	Provided features => NN
HKBU	0,133	0,096	C3D, LBP, Color Hist. => SVR
TCNJ-CS	0,34	0,093	C3D, InceptionV3, HMP, Color Hist., Aesthetics + MemNet => NN
Show-And-Recall	0,355	0,15	C3D, InceptionV3, HMP, Color Hist., HoG + Saliency + Captions => NN
The Huskies	0,356	0,213	Captions => RNN
MMC-RUC	0,448	0,24	C3D, HMP, Aesthetics + I3D + Captions => SVR or RFR
Technicolor	0,46	0,219	ResNet101 => fine-tuning (data augmentation)
AREA66	0,462	0,232	C3D, Aesthetics + MemNet + Saliency + face-based feature + Captions => weighted average/SVR
HCMUS	0,484	**0,257**	InceptionV3 + Captions => LSTM => Dense layer
ConduentLabs	**0,497**	0,253	Captions + ResNet

=> Text features provide also additional cues for understanding semantics of videos

TEAM	SHORT-TERM SUBTASK	LONG-TERM SUBTASK	APPROACH
GIBIS	—	0,118	HMP => SVR or *k*-NN regressor
Insight@DCU	0,056	0,039	Provided features => NN
HKBU	0,133	0,096	C3D, LBP, Color Hist. => SVR
TCNJ-CS	0,34	0,093	C3D, InceptionV3, HMP, Color Hist., Aesthetics + MemNet => NN
Show-And-Recall	0,355	0,15	C3D, InceptionV3, HMP, Color Hist., HoG + Saliency + Captions => NN
The Huskies	0,356	0,213	Captions => RNN
MMC-RUC	0,448	0,24	C3D, HMP, Aesthetics + I3D + Captions => SVR or RFR
Technicolor	0,46	0,219	ResNet101 => fine-tuning (data augmentation)
AREA66	0,462	0,232	C3D, Aesthetics + MemNet + Saliency + face-based feature + Captions => weighted average/SVR
HCMUS	0,484	**0,257**	InceptionV3 + Captions => LSTM => Dense layer
ConduentLabs	**0,497**	0,253	Captions + ResNet

=> High level semantic features learned by CNNs trained for image classification achieve state-of-the-art performance

TEAM	SHORT-TERM OFFICIAL RESULTS	LONG-TERM OFFICIAL RESULTS
GIBIS	—	0,152 (0,118)
Insight@DCU	0,05 (Official result: 0,056)	-0,003 (0,039)
HKBU	0,154 (0,133)	0,095 (0,096)
Show-And-Recall	0,369 (0,355)	0,187 (0,15)
TCNJ-CS	0,386 (0,34)	0,126 (0,093)
The Huskies	0,433 (0,356)	0,289 (0,213)
HCMUS	0,518 (0,484)	0,341 (0,257)
AREA66	0,519 (0,462)	0,341 (0,232)
Technicolor*	0,527 (0,46)*	0,218 (0,219)*
MMC-RUC	0,527 (0,448)	**0,369** (0,24)
ConduentLabs	**0,558** (0,497)	0,302 (0,253)

*Organizer

Results on the 500 most annotated videos (Spearman corr.)
included in the test-set for teams' best runs

Independent features' performances (Spearman corr.)

	TECHNICOLOR		TCNJ-CS	
FEATURES	Short-term	Long-term	Short-term	Long-term
Aesthetic	0,28	0,13	0,28	0,13
C3D	0,28	0,13	0,29	0,13
HMP	0,28	0,11	0,22	0,07
ColorHist	0,13	0,05	0,31	0,11
InceptionV3	0,16	0,06	0,1	0,03
Caption	0,49	0,22	0,46	0,2
MemNet	0,39	0,17	0,4	0,2

=> Similar results reported by the two teams

Independent features' performances (Spearman corr.)

	TECHNICOLOR		TCNJ-CS	
FEATURES	Short-term	Long-term	Short-term	Long-term
Aesthetic	0,28	0,13	0,28	0,13
C3D	0,28	0,13	0,29	0,13
HMP	0,28	0,11	0,22	0,07
ColorHist	0,13	0,05	0,31	0,11
InceptionV3	0,16	0,06	0,1	0,03
Caption	0,49	0,22	0,46	0,2
MemNet	0,39	0,17	0,4	0,2

=> Text features alone perform well for video memorability prediction [*Cohendet et al., 2014*]

Independent features' performances (Spearman corr.)

	TECHNICOLOR		TCNJ-CS	
FEATURES	Short-term	Long-term	Short-term	Long-term
Aesthetic	0,28	0,13	0,28	0,13
C3D	0,28	0,13	0,29	0,13
HMP	0,28	0,11	0,22	0,07
ColorHist	0,13	0,05	0,31	0,11
InceptionV3	0,16	0,06	0,1	0,03
Caption	0,49	0,22	0,46	0,2
MemNet	0,39	0,17	0,4	0,2

=> Models for image memorability prediction perform well for video memorability prediction

[*Khosla et al., 2015; Squalli-Houssaini et al., 2018*]

Conclusion

- The dataset will soon be publicly released

- Future work
  - Long-term memorability
  - Beyond the average memorability

Thank you!