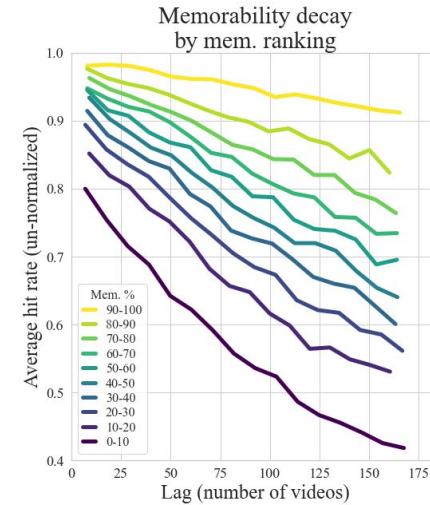
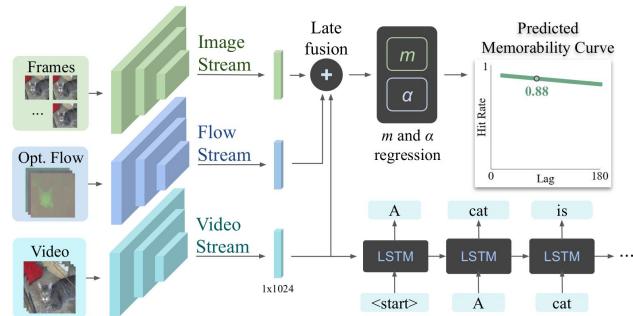


Our Contributions

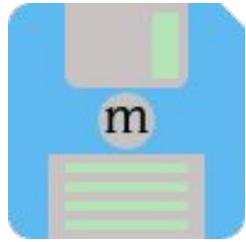


Memento10k

Modeling
Memorability Decay



SemanticMemNet



Multimodal Memorability

**Modeling Effects of Semantics and Decay
on Video Memorability**

Anelise Newman*, Camilo Fosco*, Vincent Casser, Allen Lee,
Barry McNamara, and Aude Oliva

Code, data, and models available at

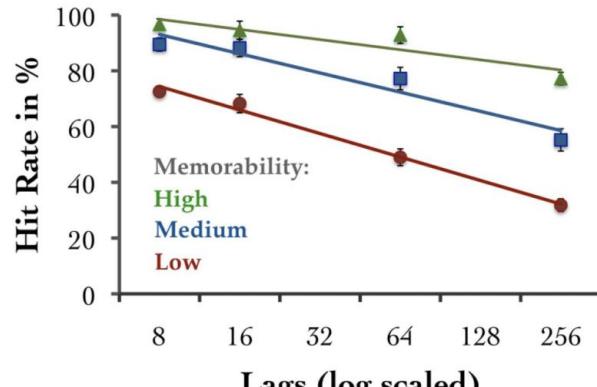
Memento.csail.mit.edu



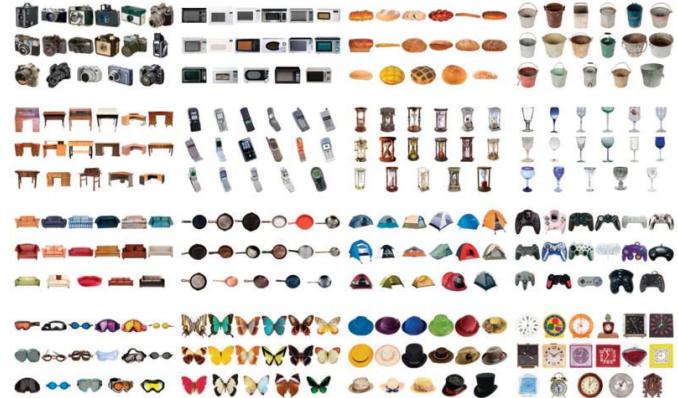
Related Work: Memorability in Cognitive Science



Isola 2011, CVPR



Vo 2017



Isola 2011, CVPR

Space for
webcam

Related Work: Memorability in Computer Vision

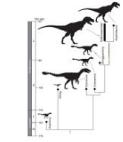


a) Predicted most memorable (87%)
Isola 2011, CVPR

original image



↑ memorability



Khosla 2013, ICCV

Borkin 2013, InfoVis

LaMem

Khosla 2015, ICCV



← Less memorable ————— More memorable →

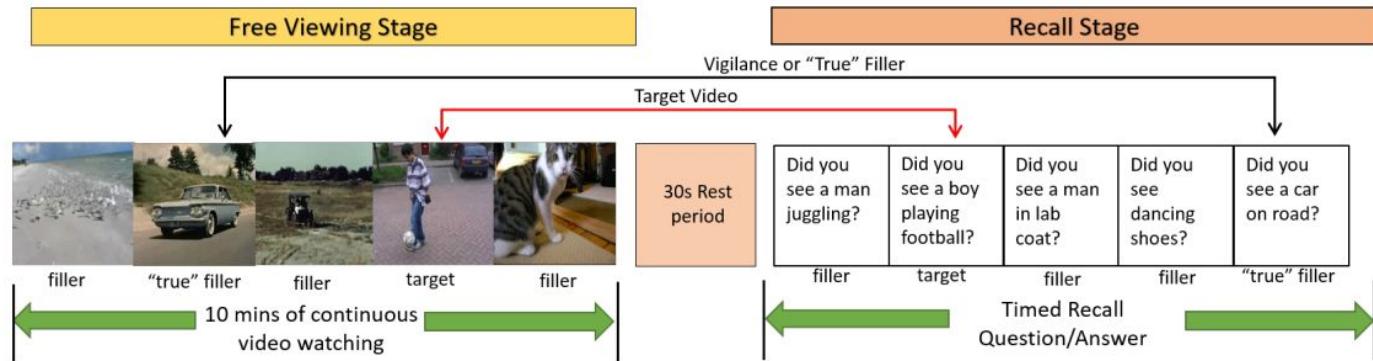
Goetschalckx 2019, ICCV

Related Work: Video Memorability



Cohendet 2019, ICCV

Shekhar
2017, ICCV





Memento10k

- 10k videos
 - Nearly 1M human responses
 - Semantic annotations: captions and action categories
 - Variety of delays

Memorability score: **0.91** Decay rate: **-0.001**



Caption: A panda bear holds and munches on several shoots of bamboo while seated on rock.

Actions: dining, snapping, feeding, biting, holding, combing

The Memento Game

Memento: The Video Memory Game

Level 1

Instructions

Press SPACEBAR (or tap the video on mobile) when you see a video we've shown before.

You will watch a stream of 3-second videos. Some videos will be shown twice. Your task is to press SPACEBAR (or tap the video on mobile) when you see a video we've already shown. If you are correct, you will see a green flash around the video, and you will skip to the next one. If you are wrong, you will see a red flash. The task should take around 10 minutes to complete.

If you do well in the task, you'll be able to advance to the next level! If you don't, you will lose a life. You start with one life, but gain one after completing the first level. If you lose all your lives, you will not be able to complete more levels. Warning: try to play with the best possible internet connection, as the game needs to load several videos. You will have loading problems if your connection is slow. If you encounter multiple loading errors, you might unfortunately be blocked, although you will not be rejected, your approval score will not be affected, and you will still be paid. We are unfortunately forced to block workers with unstable connections because loading errors on videos affect the quality of the data we're collecting. We want to make sure that turkers with stable connections only can play.

You can complete more of these HITs! In fact, you are encouraged to do so. Every new Memento HIT that you accept will recognize your workerID and load your current progress. The next HIT will represent a new level.

The videos will start playing as soon as you press "Start Game". Have fun testing your memory!

Start Game!

The Memento Game



Hit Rate = # correct responses / # total responses

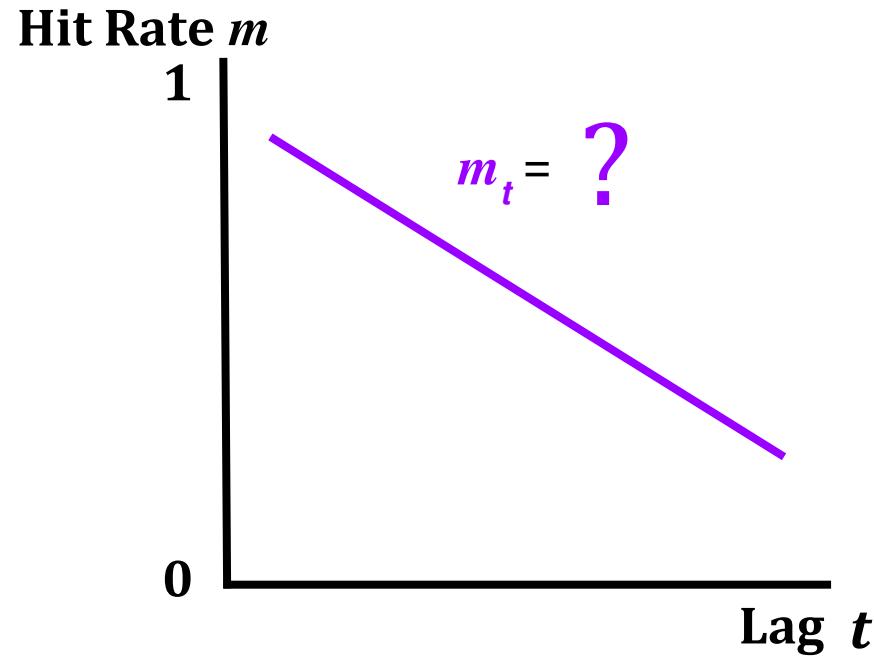
High Memorability



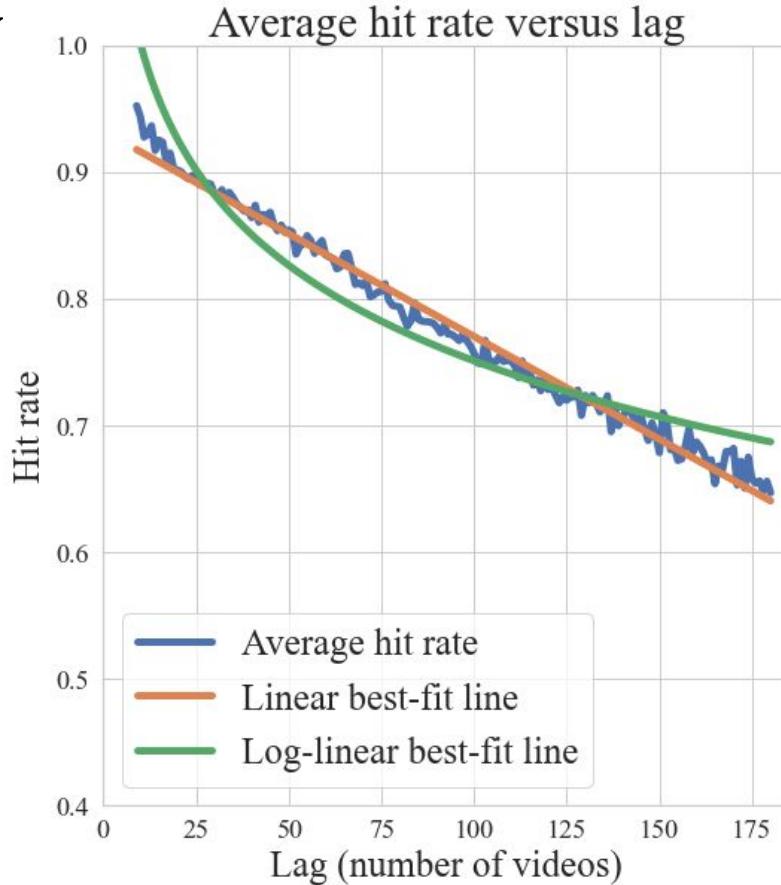
Low Memorability



Measuring Memory Decay



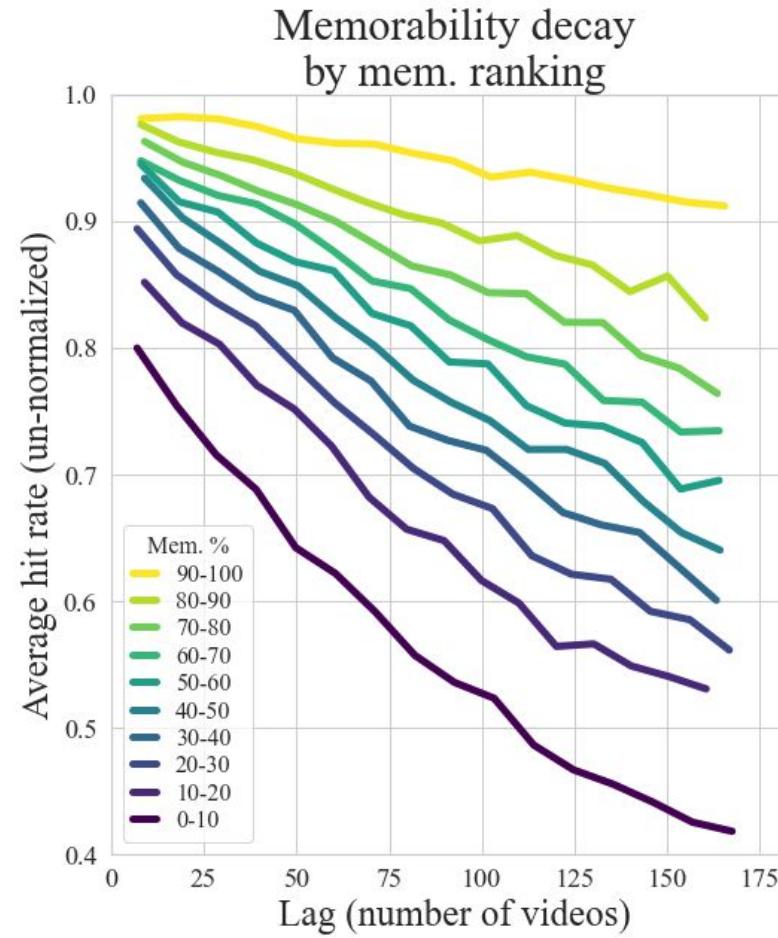
Linear Decay



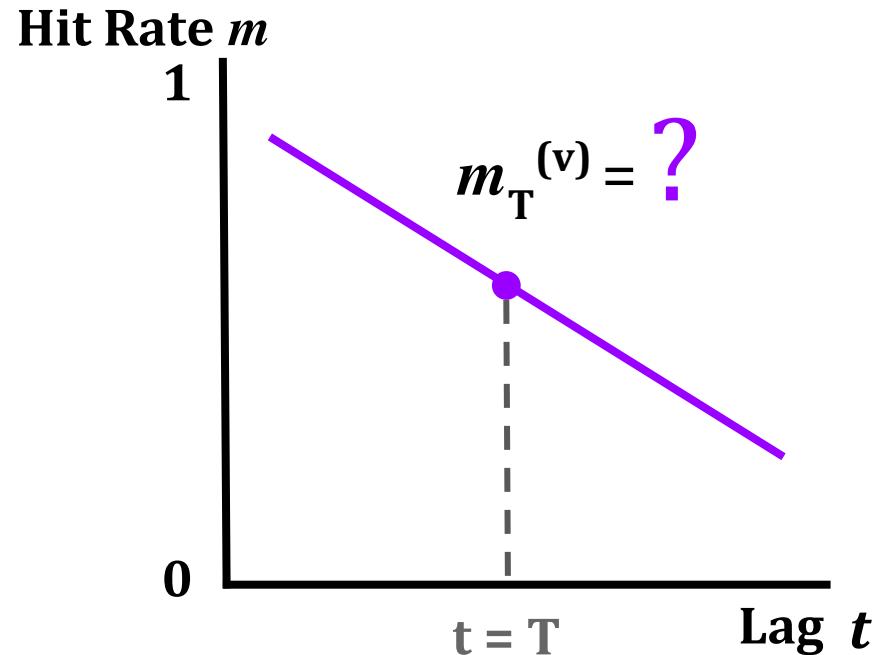
$$r = -0.953$$

$$r = -0.991$$

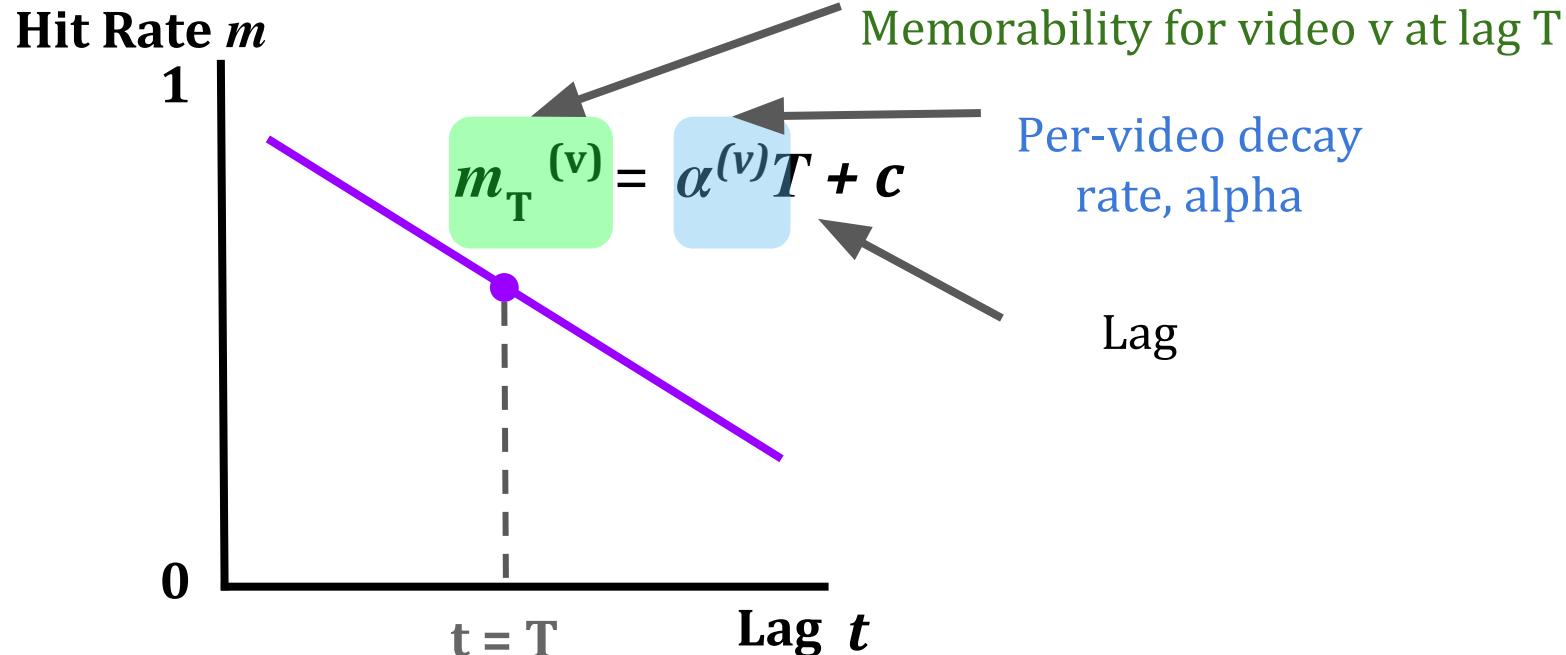
Different rates of decay



Calculating Memorability Scores

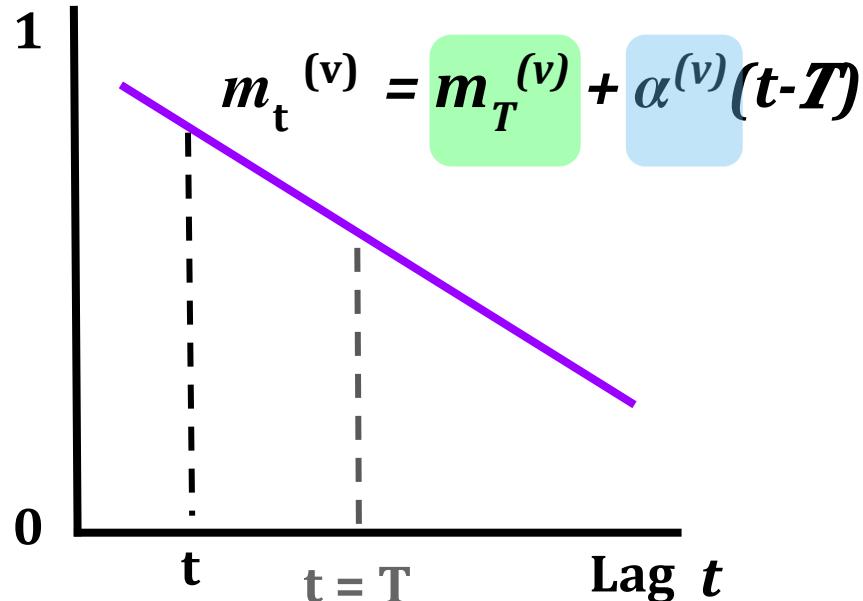


Calculating Memorability Scores



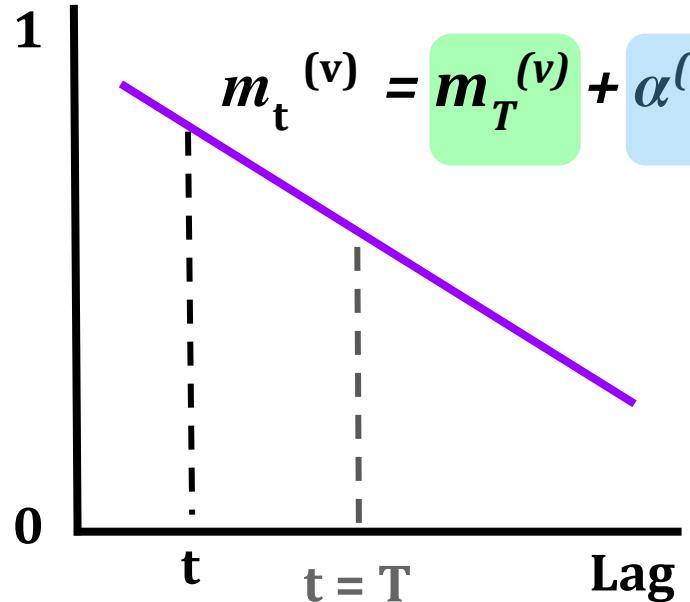
Calculating Memorability Scores

Hit Rate m



Calculating Memorability Scores

Hit Rate m



$$m_t^{(v)} = m_T^{(v)} + \alpha^{(v)}(t-T)$$

Define:

$$T = 80,$$

Memorability score

$$= m_{80}$$

Modeling Video Memorability

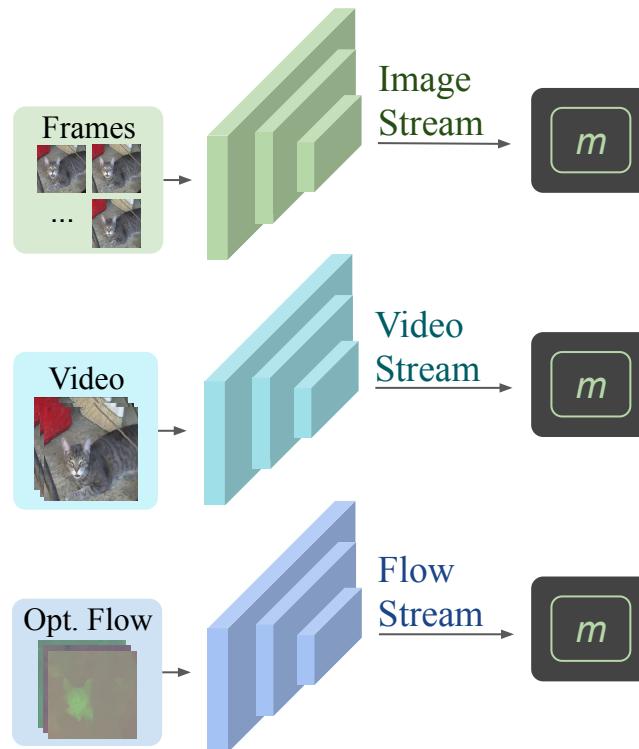
High Memorability



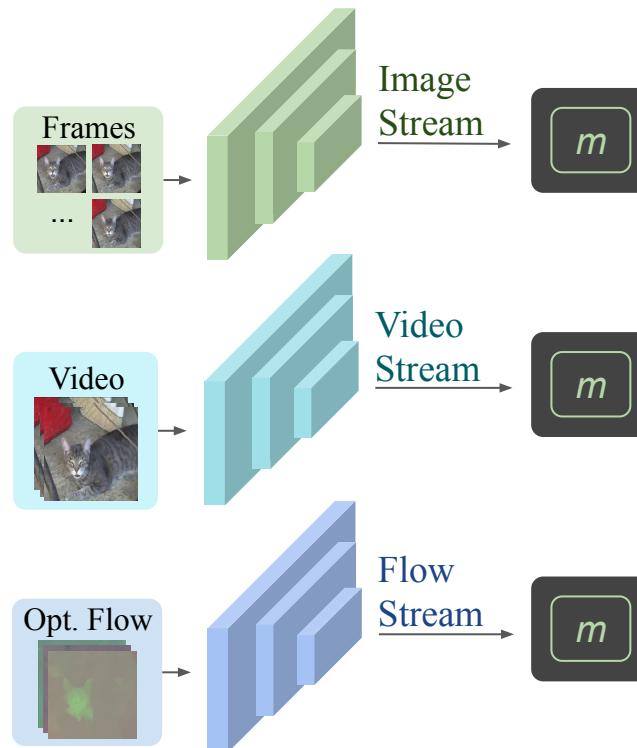
Low Memorability



Modeling Visual Features

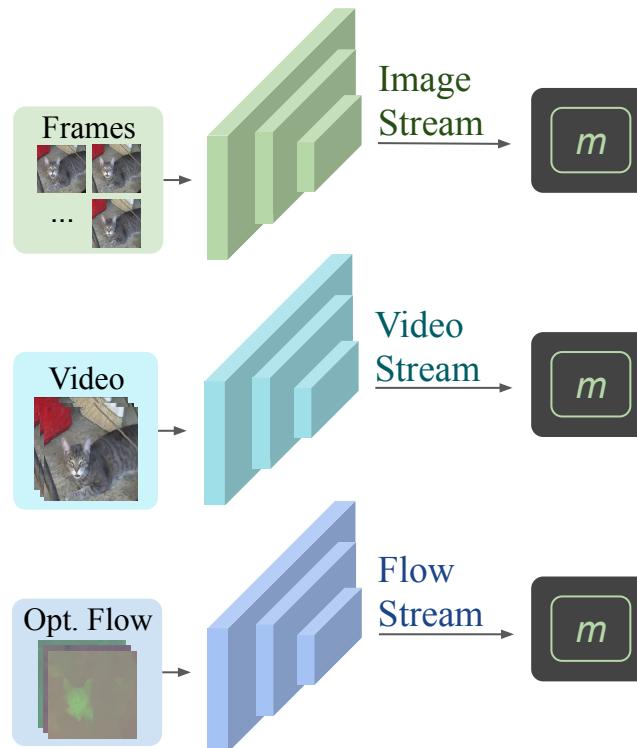


Modeling Visual Features



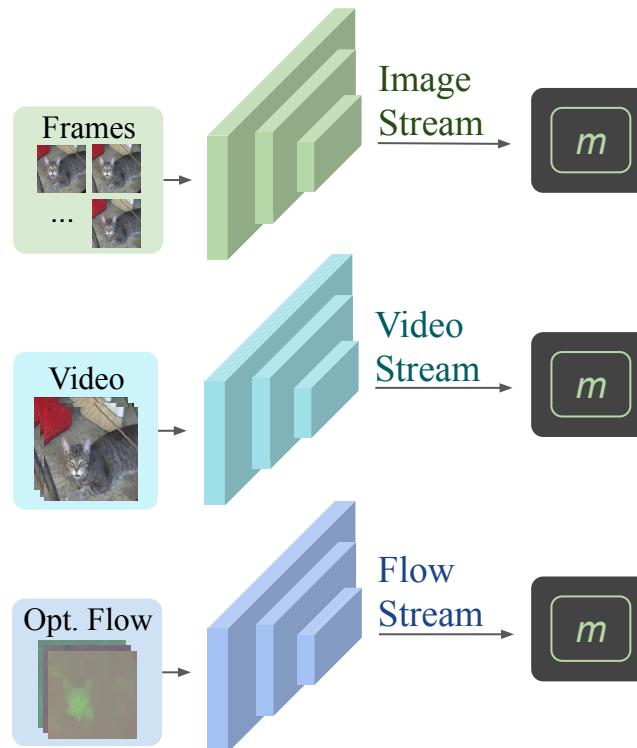
Approach	RC -- Memento10 k test set	RC -- VideoMem val set
Human consistency	0.730	0.616
Frames	0.595	0.527
Video	0.596	0.492
Flow	0.579	0.425
Frames + Video + Flow	0.659	0.555

Modeling Visual Features



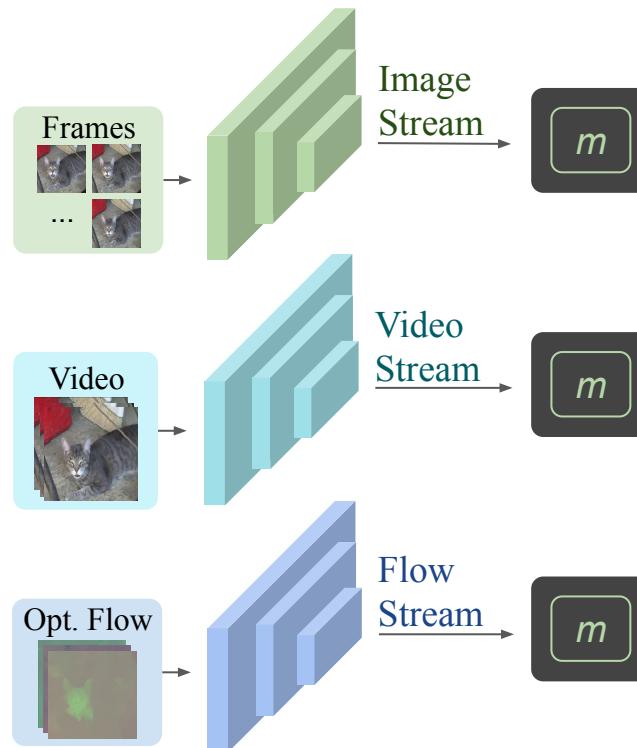
Approach	RC -- Memento10 k test set	RC -- VideoMem val set
Human consistency	0.730	0.616
Frames	0.595	0.527
Video	0.596	0.492
Flow	0.579	0.425
Frames + Video + Flow	0.659	0.555

Modeling Visual Features



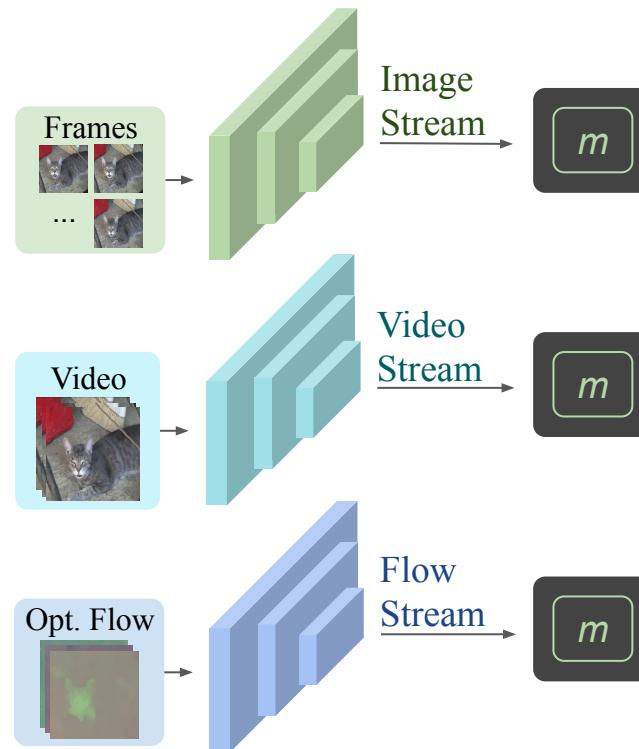
Approach	RC -- Memento10 k test set	RC -- VideoMem val set
Human consistency	0.730	0.616
Frames	0.595	0.527
Video	0.596	0.492
Flow	0.579	0.425
Frames + Video + Flow	0.659	0.555

Modeling Visual Features

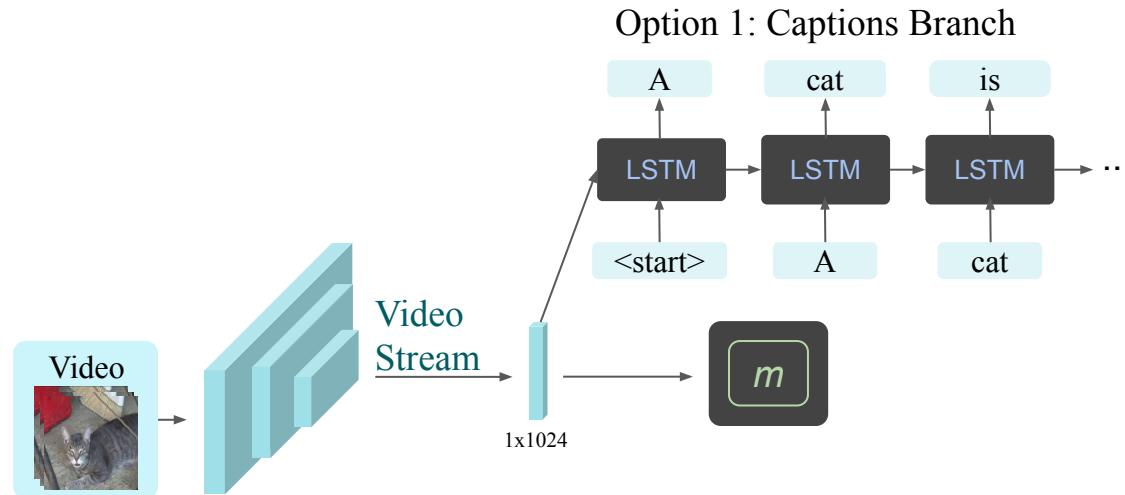


Approach	RC -- Memento10 k test set	RC -- VideoMem val set
Human consistency	0.730	0.616
Frames	0.595	0.527
Video	0.596	0.492
Flow	0.579	0.425
Frames + Video + Flow	0.659	0.555

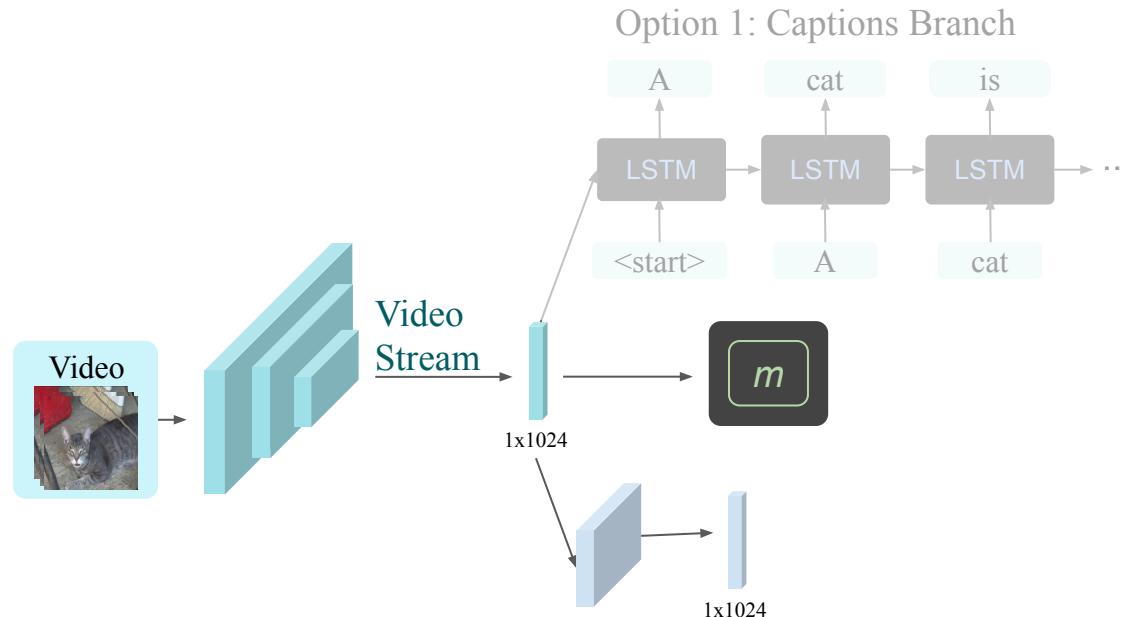
Modeling Semantic Features



Modeling Semantic Features

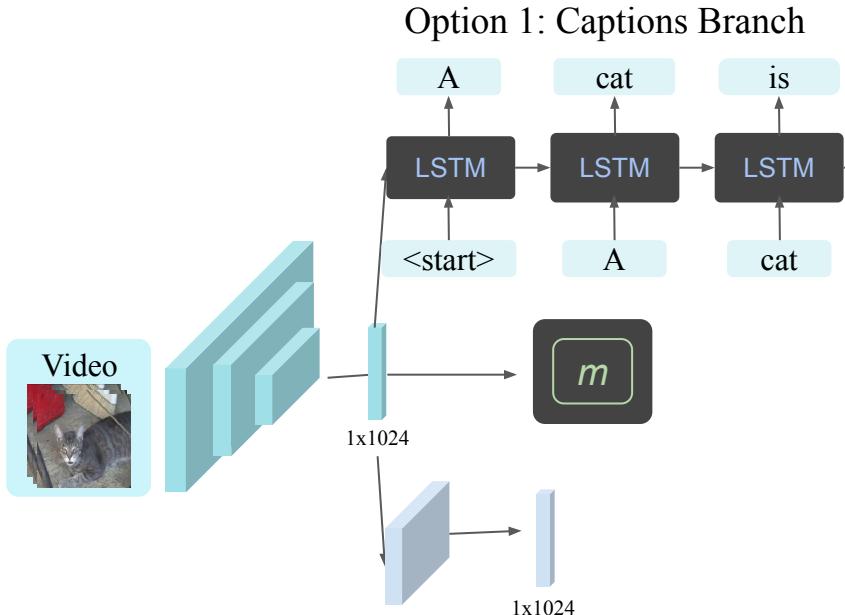


Modeling Semantic Features



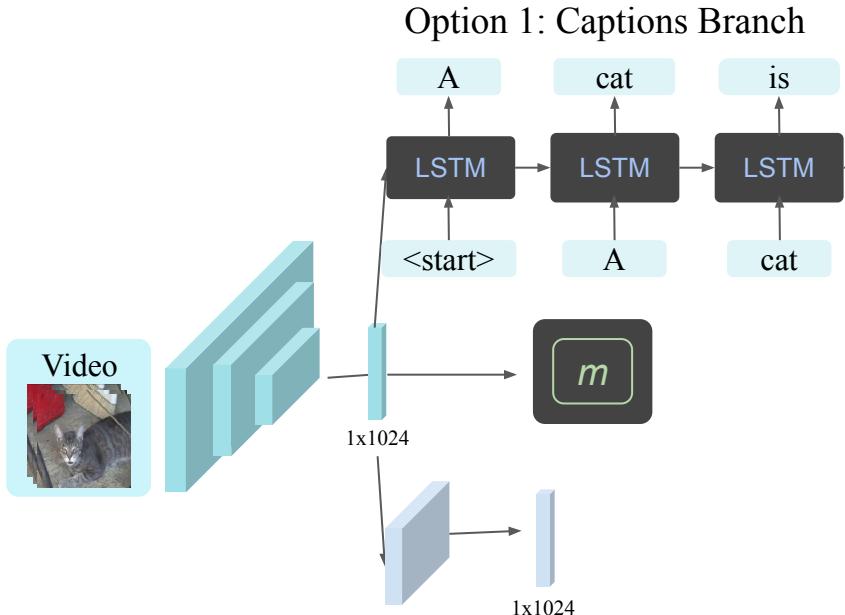
Option 2: Sentence Embedding

Modeling Semantic Features



Approach	RC -- Memento 10k test set	RC -- VideoMem val set
Human consistency	0.730	0.616
Video	0.596	0.492
Video + Captions	0.602	0.512
Video + Triplet Loss	0.599	--

Modeling Semantic Features

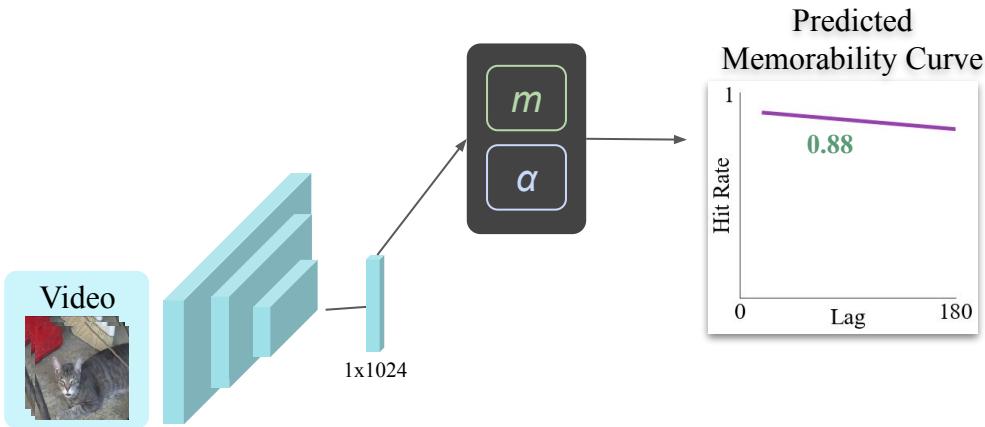


Option 2: Sentence Embedding

Approach	RC -- Memento 10k test set	RC -- VideoMem val set
Human consistency	0.730	0.616
Video	0.596	0.492
Video + Captions	0.602	0.512
Video + Triplet Loss	0.599	--

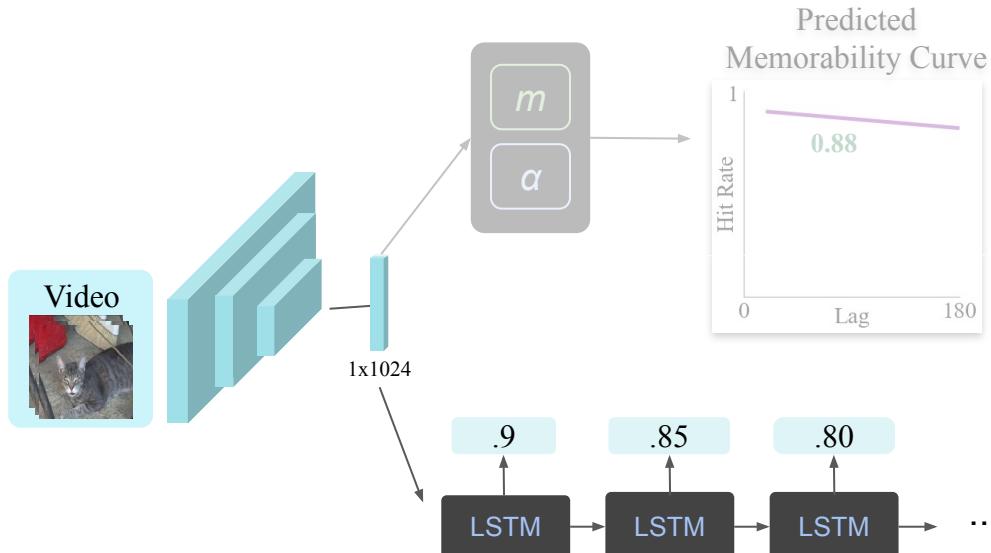
Modeling Memorability Decay

Option 1: Mem-alpha model

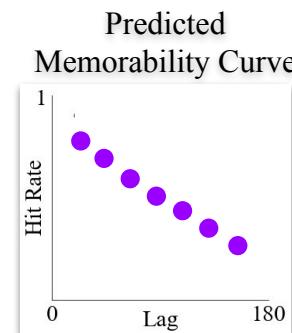


Modeling Memorability Decay

Option 1: Mem-alpha model

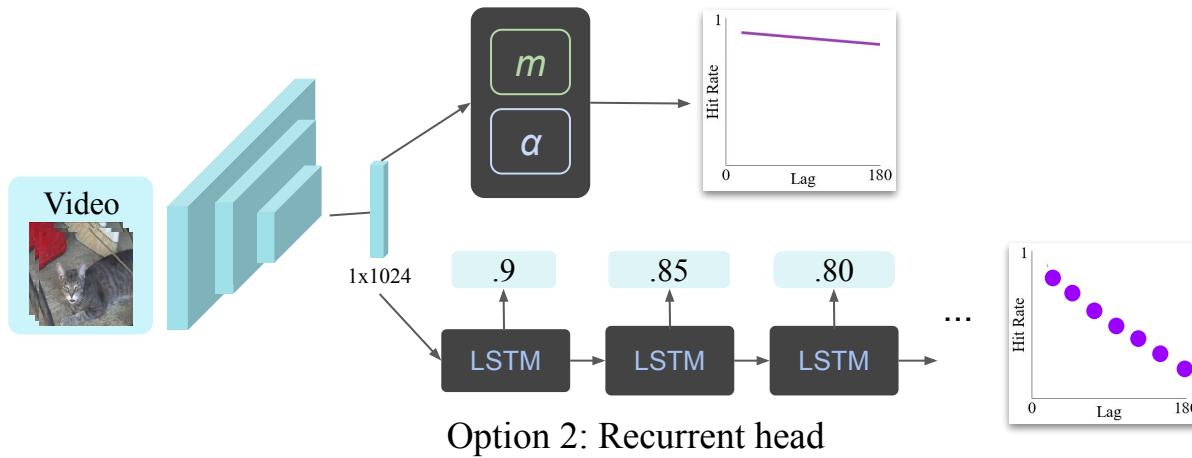


Option 2: Sentence Embedding



Modeling Memorability Decay

Option 1: Mem-alpha model

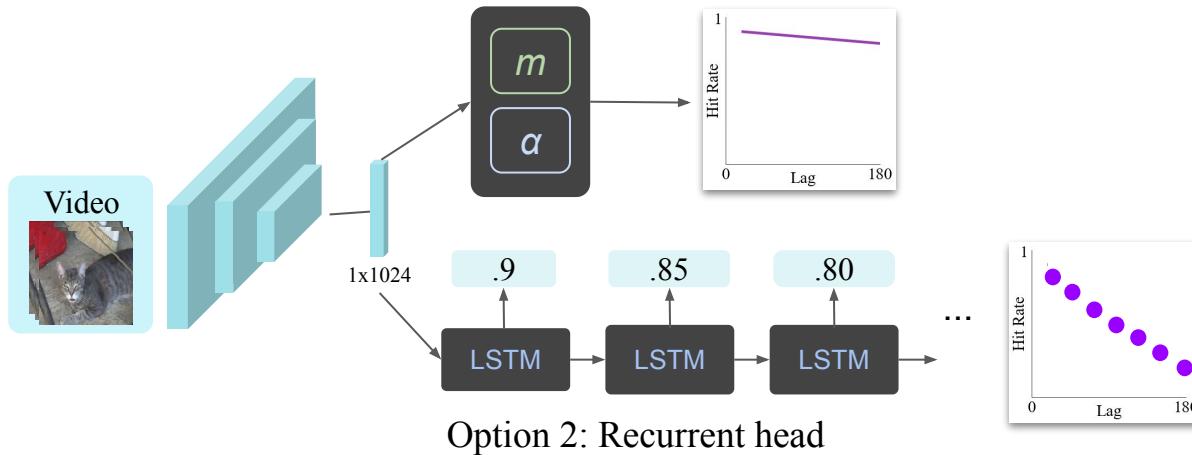


Option 2: Recurrent head

Approach	RC	R^2		
		$t=40$	$t=80$	$t=160$
Mem-alpha	0.604	0.146	0.227	0.121
Recurrent head	0.599	0.298	0.364	0.219

Modeling Memorability Decay

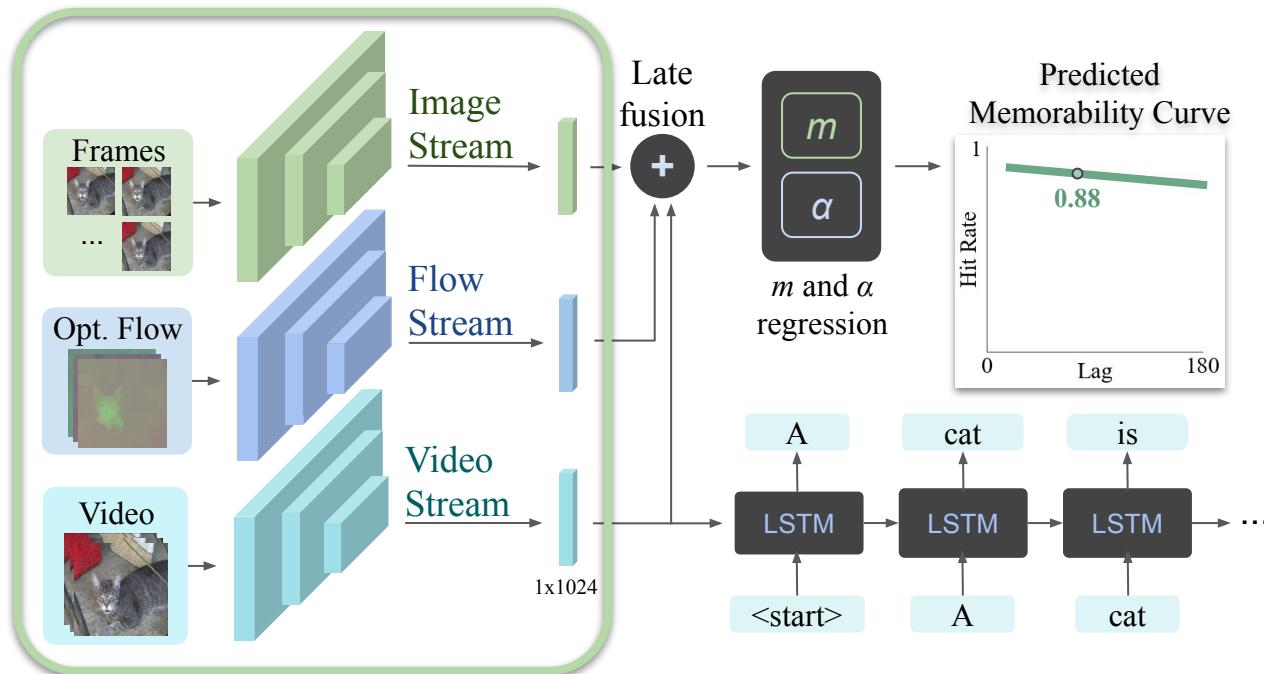
Option 1: Mem-alpha model



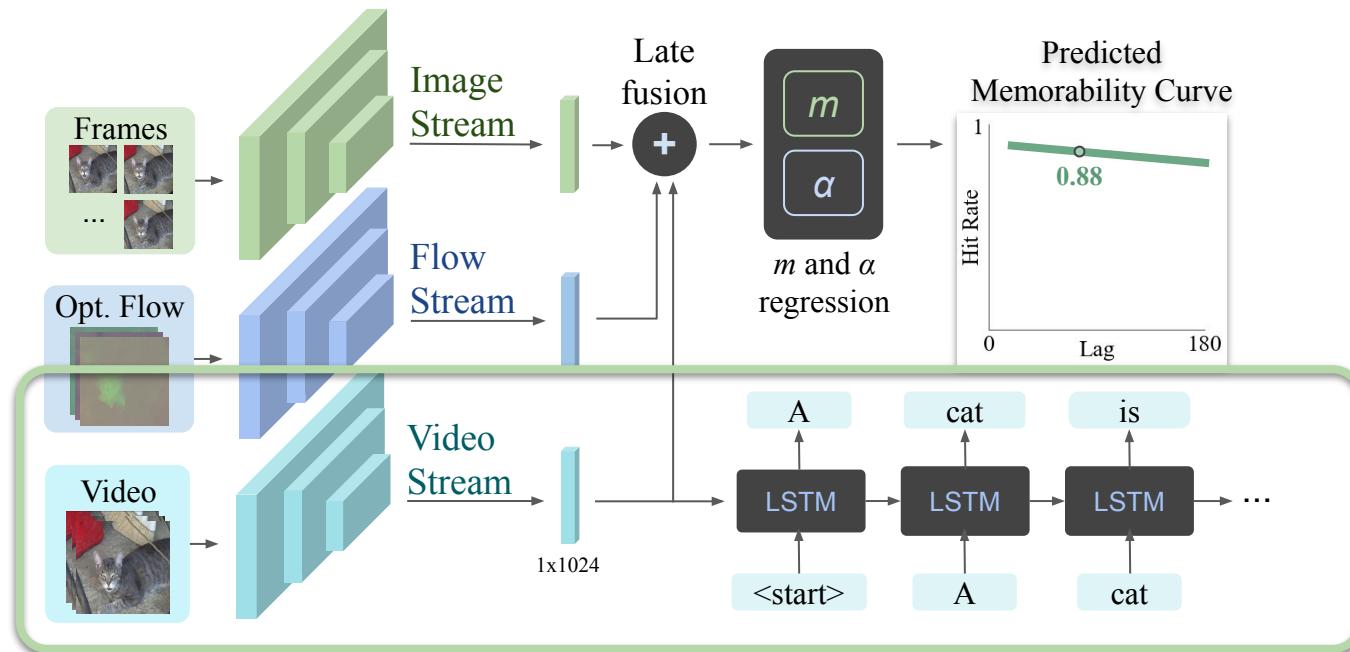
Option 2: Recurrent head

Approach	RC	R^2		
		$t=40$	$t=80$	$t=160$
Mem-alpha	0.604	0.146	0.227	0.121
Recurrent head	0.599	0.298	0.364	0.219

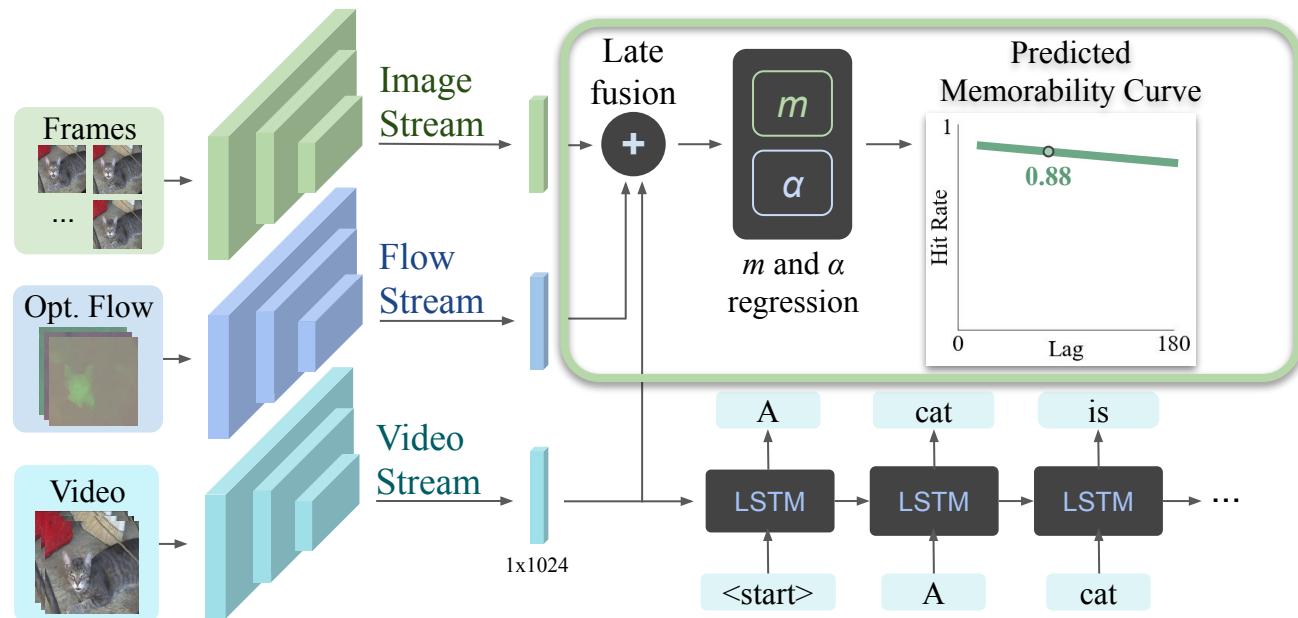
SemanticMemNet



SemanticMemNet



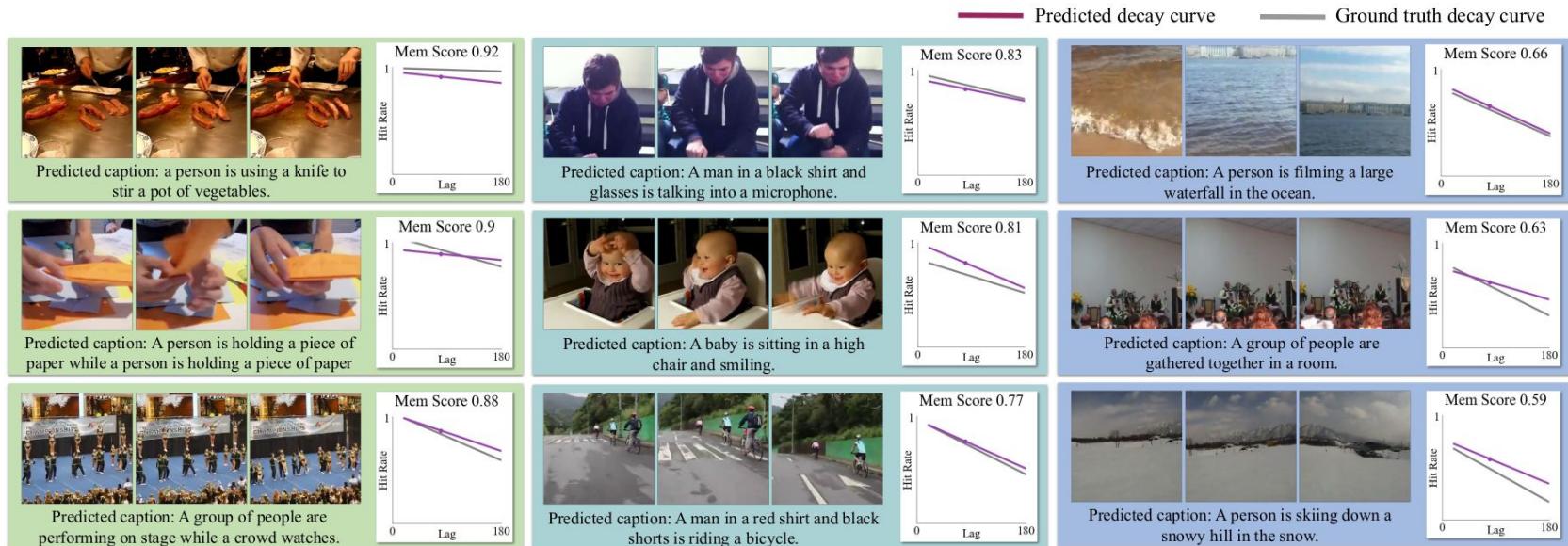
SemanticMemNet



SemanticMemNet Results

Approach	RC - Memento10k (test set)	RC - VideoMem (validation set)
Human consistency	0.730	0.616
MemNet Baseline [30]	0.485	0.425
Feature extraction + regression (as in [39])*	0.615	0.427
Cohendet et al. (ResNet3D) [13]	0.574	0.508
Cohendet et al. (Semantic)[13]	0.552	0.503
SemanticMemNet	0.663	0.556

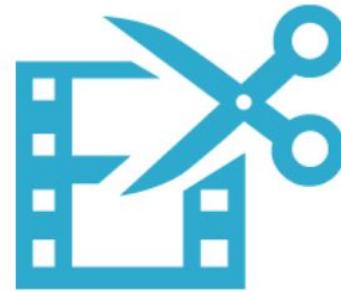
SemanticMemNet Results



Future Directions



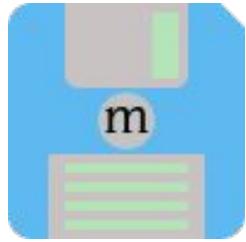
Education /
Content Creation



Video Editing /
Summarization



Video Selection



Multimodal Memorability

**Modeling Effects of Semantics and Decay
on Video Memorability**

Anelise Newman*, Camilo Fosco*, Vincent Casser, Allen Lee,
Barry McNamara, and Aude Oliva

Code, data, and models available at

Memento.csail.mit.edu

