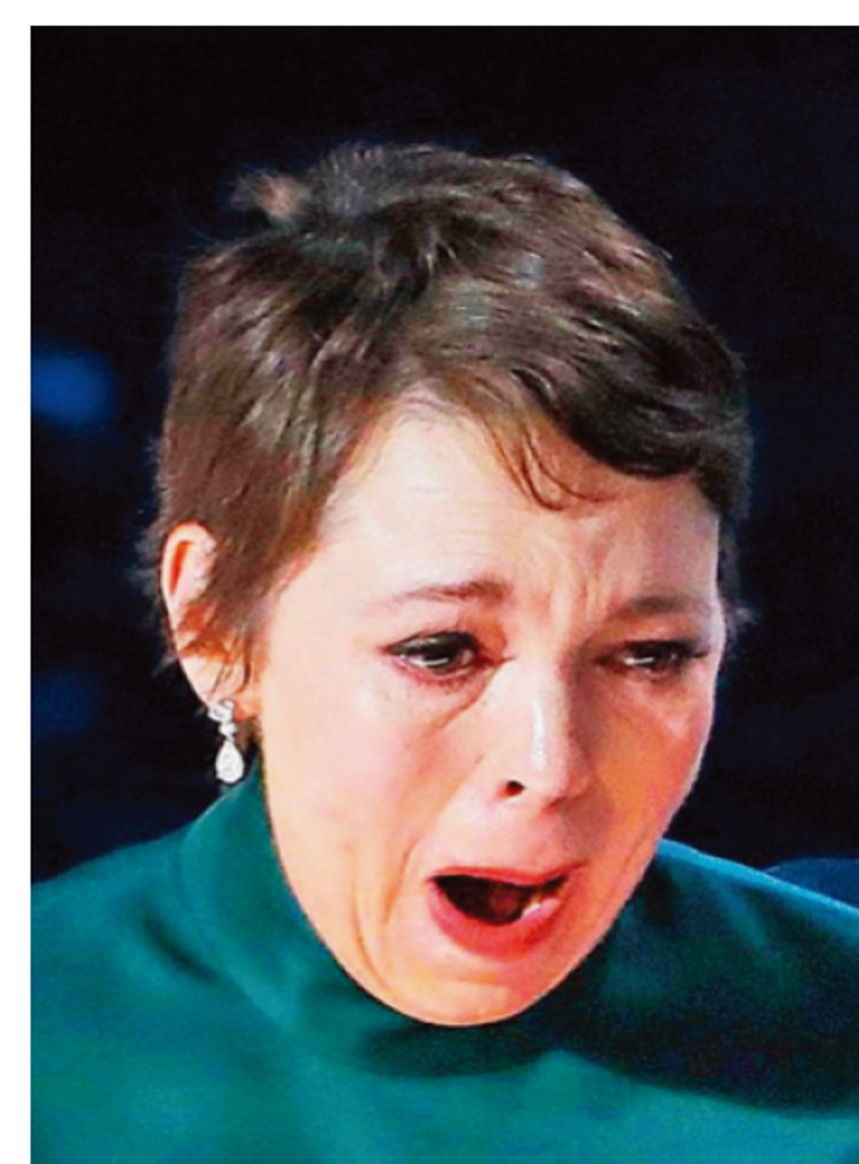# VEATIC: Video-based Emotion and Affect Tracking in Context Dataset
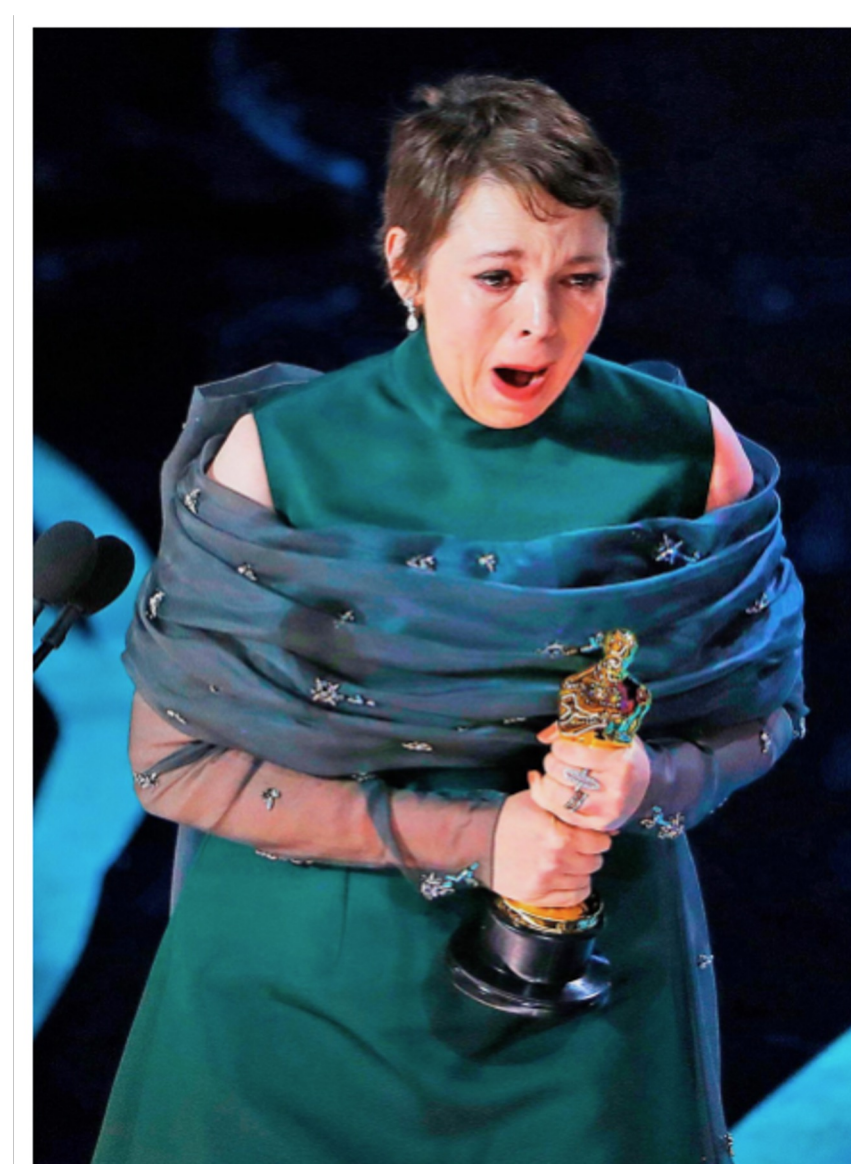
Zhihang Ren*   Jefferson Ortega*   Yifan Wang*   Zhimin Chen

Yunhui Guo   Stella X. Yu   David Whitney

Context is important in emotion recognition
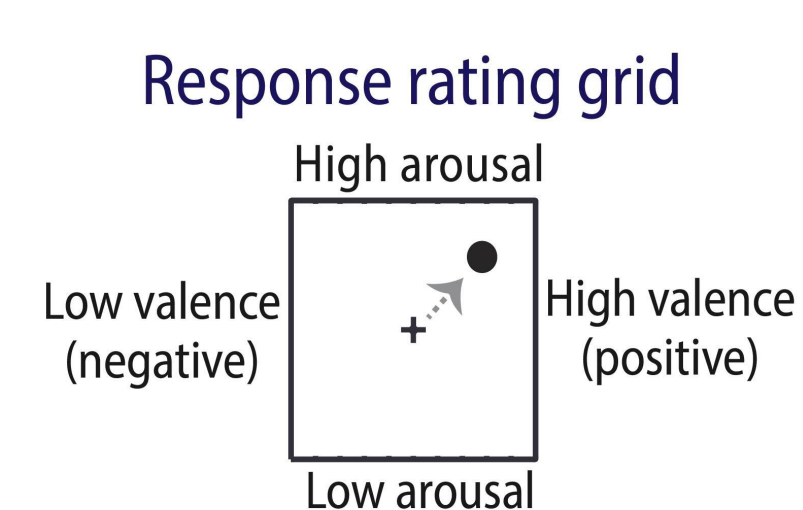

Shocked? ✖
Surprised? ✖   → Ecstatic!!
Sad? ✖   ✔

Annotation procedure for dynamic emotions

Response rating grid

Low valence (negative)   High arousal   High valence (positive)   Low arousal

Subjects rated the **valence** and **arousal** of the target character continuously

Comparison of VEATIC with other datasets

Database	Annotation Type	Condition	# videos	Videos Length	#Annotators	Context
SAL	Valence-Arousal (Continuous)	Controlled	23	SAL 0: 5min SAL 1: 30min	4	✗
SEMAINE	Mixed*	Controlled	\	Total: 6.5hours	6-8	✗
SEND	valence	Controlled	193	135s	700	✗
Belfast induced	Mixed	Controlled	37	5-60s	6-258	✗
MAHNOB-HCI	Mixed	Controlled	20	34.9-117s	50	✗
MELD	7 Emotion Categories	In-the-Wild	1,433	3.59s	3	✗
OMG Emotion	Mixed	In-the-Wild	567	1min	5	✗
RECOLA	Valence-Arousal (Continuous)	Controlled	46	5min	6	✗
AFEW	7 Basic Facial Expression	In-the-Wild	1,809	0.3s - 5.4s	3	✗
AFEW-VA	Valence-Arousal (Discrete)	In-the-Wild	600	0.5s - 4s	2	✗
Aff-Wild	Valence-Arousal (Continuous)	In-the-Wild	298	6s - 14min28s	8	✗
Aff-Wild2	Valence-Arousal (Continuous)	In-the-Wild	260	4s - 15min4s	4	✗
AM-FED	12 Action Units	In-the-Wild	242	49.69s	\	✗
DEAP	Valence-Arousal-Dominance (Discrete)	Music Videos	120	1min	14-16	✓
CAER	7 Emotion Categories	In-the-Wild	13,201	1s - 5s	6	✓
CAER-S	7 Emotion Categories	In-the-Wild Image-based	\	70,000 images	6	✓
EMOTIC	Mixed	In-the-Wild Image-based	\	18,316 images	3-5	✓
VEATIC (ours)	Valence-Arousal (Continuous)	In-the-Wild	124	10s - 2min37s	192	✓

Benchmark emotion tracking in context task
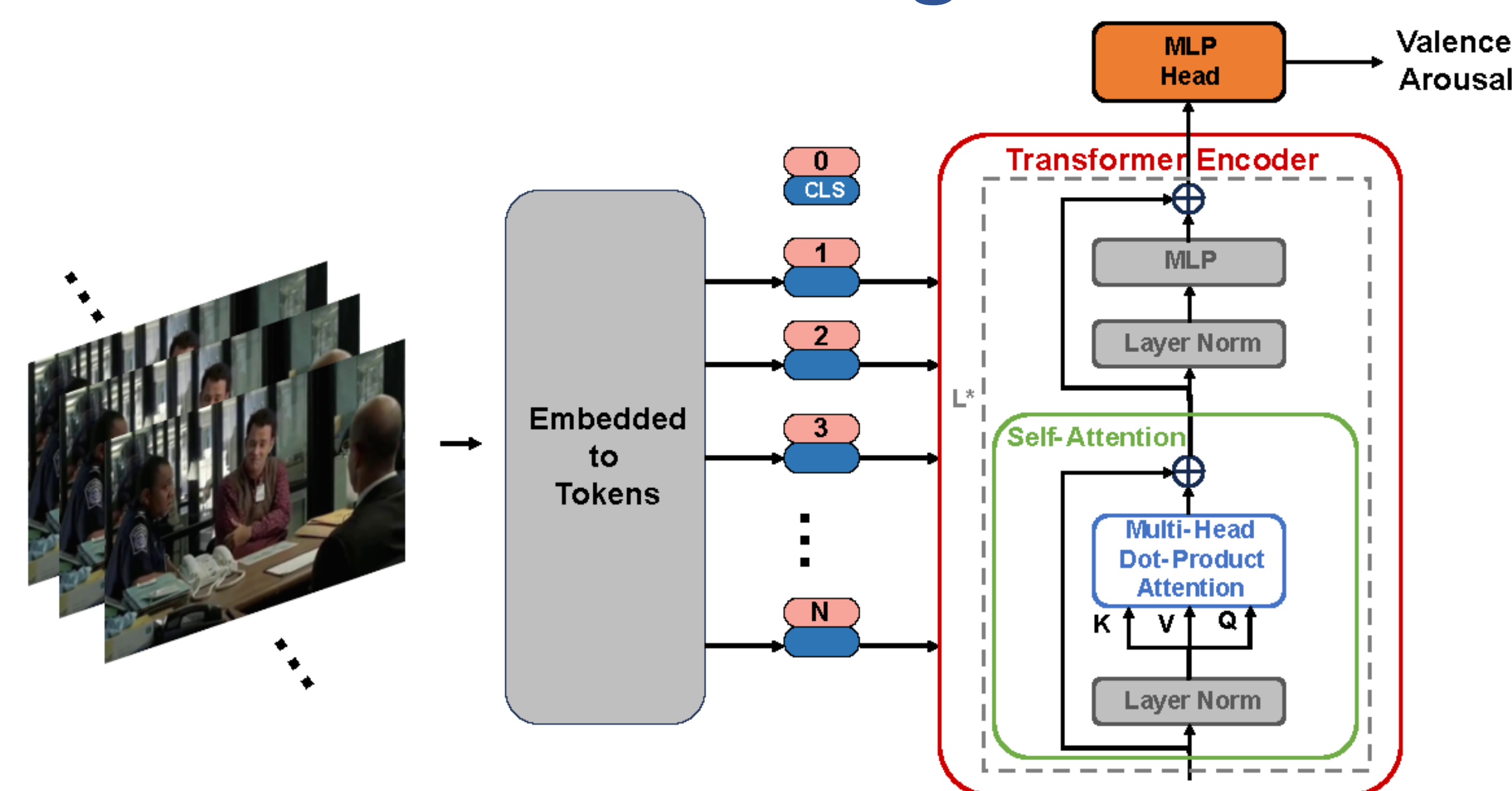
Frame Type	Valence				Arousal			
	CCC↑	PCC↑	RMSE↓	SAGR↑	CCC↑	PCC↑	RMSE↓	SAGR↑
Fully Informed	**0.6678**	**0.6967**	**0.3084**	**0.8149**	**0.6296**	**0.6584**	**0.2410**	0.7637
Character Only	0.5116	0.5609	0.3776	0.7451	0.5725	0.6247	0.2333	0.7497
Context Only	0.6185	0.6567	0.3245	0.8071	0.6088	0.6181	0.2416	**0.7828**

Method	RMSE↓			Method	ACC↑
	Valence	Arousal	Overall		
EMOTIC	1.1730	1.2900	1.2315	CAER-NET-S	0.7351
Ours	1.2151	1.3213	1.2682	Ours	0.6904

CCC: Concordance Correlation Coefficient
PCC: Pearson correlation coefficient
RMSE: Root Mean Square Error
SAGR: Sign Agreement

Dataset statistics

Attribute	Description
No. of Frames	257, 601
No. of Videos	124
Total No. of Annotators	192
Avg. No. of Annotators per video	60
Length of Videos	10 s - 2 min 37 s
Mean Image Resolution	854 × 480
Hollywood movies	104
Documentaries	5
Home videos	15

Annotator Feelings		Response Standard Deviation	
Familiarity (5.0)	Enjoyment (9)	5 annotators	all annotators
1.16	4.98	0.094	0.028
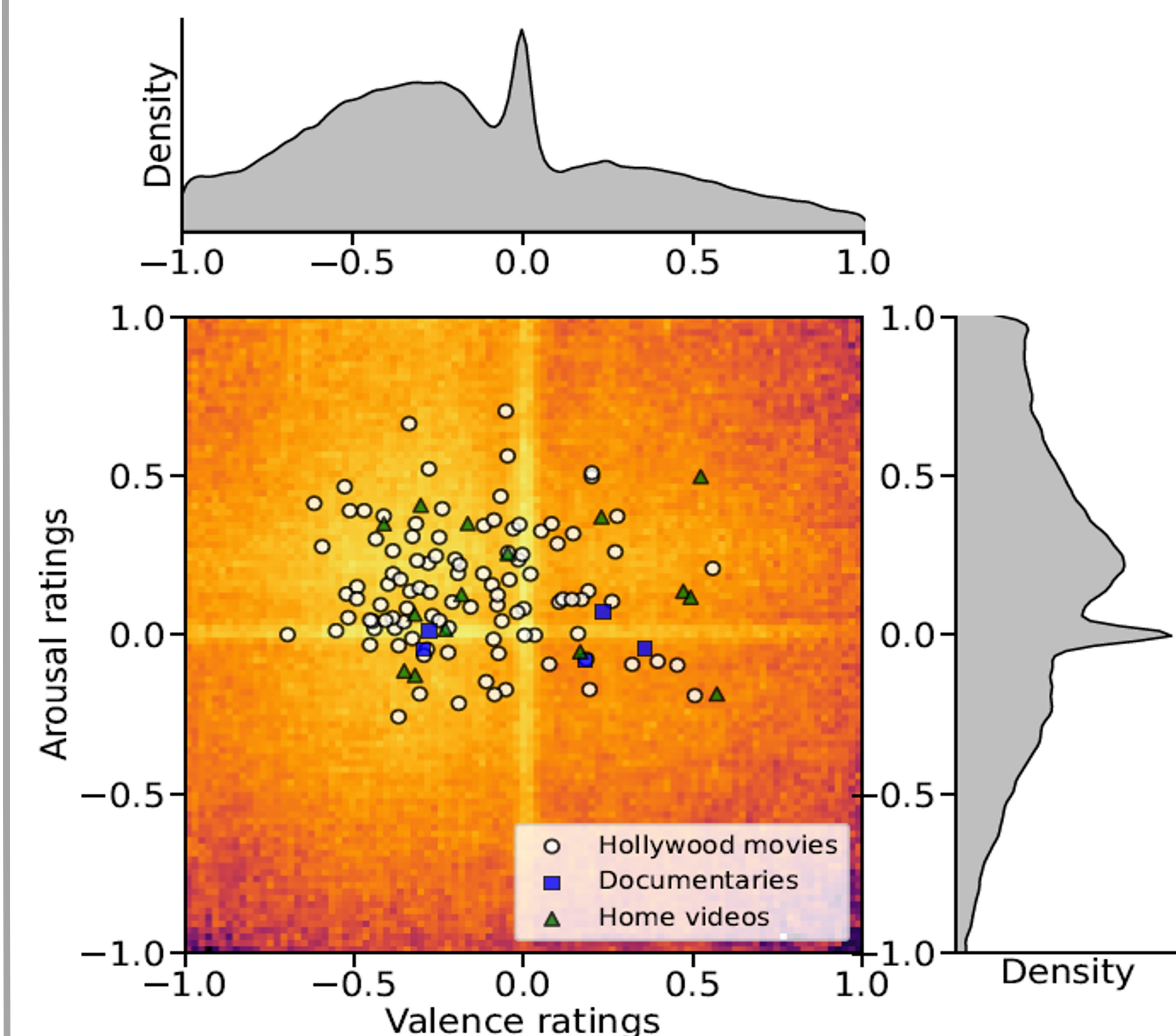
Contributions

- The **first** large video dataset for emotion and affect tracking in the wild with continuous valence and arousal annotations.
- A large set of annotators (192 in total) to avoid individual biases.
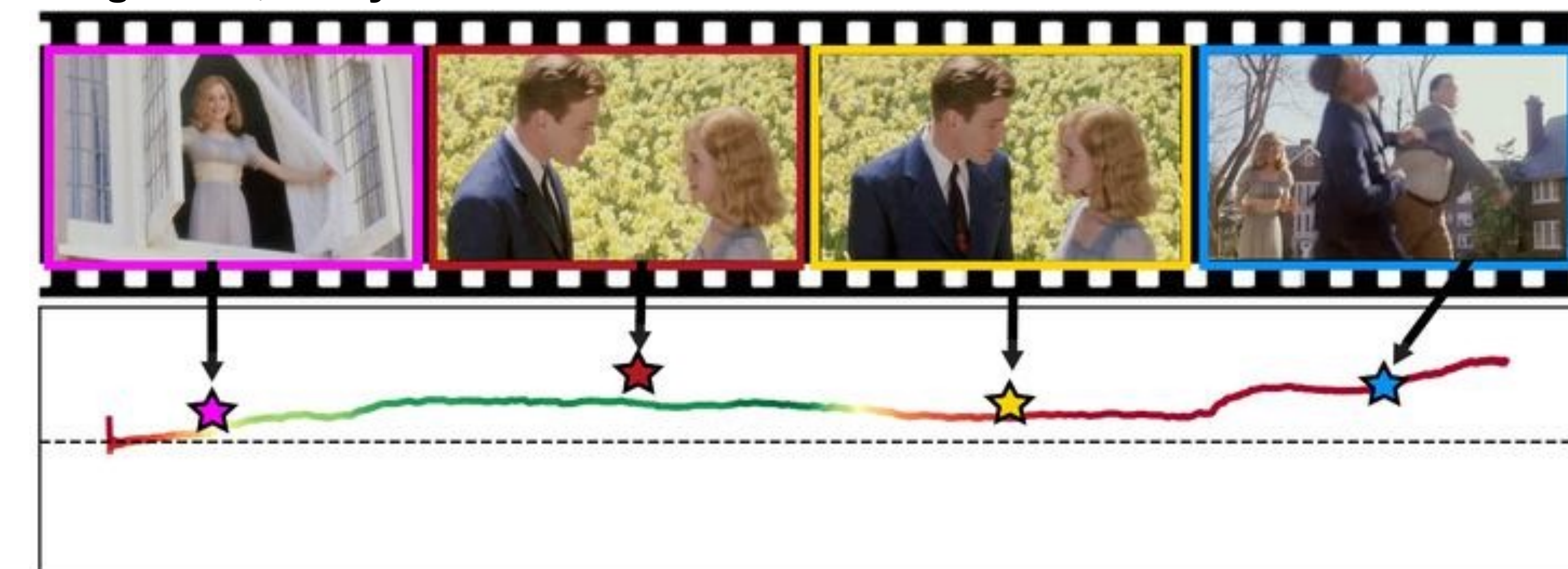- A new emotion recognition in context benchmark.
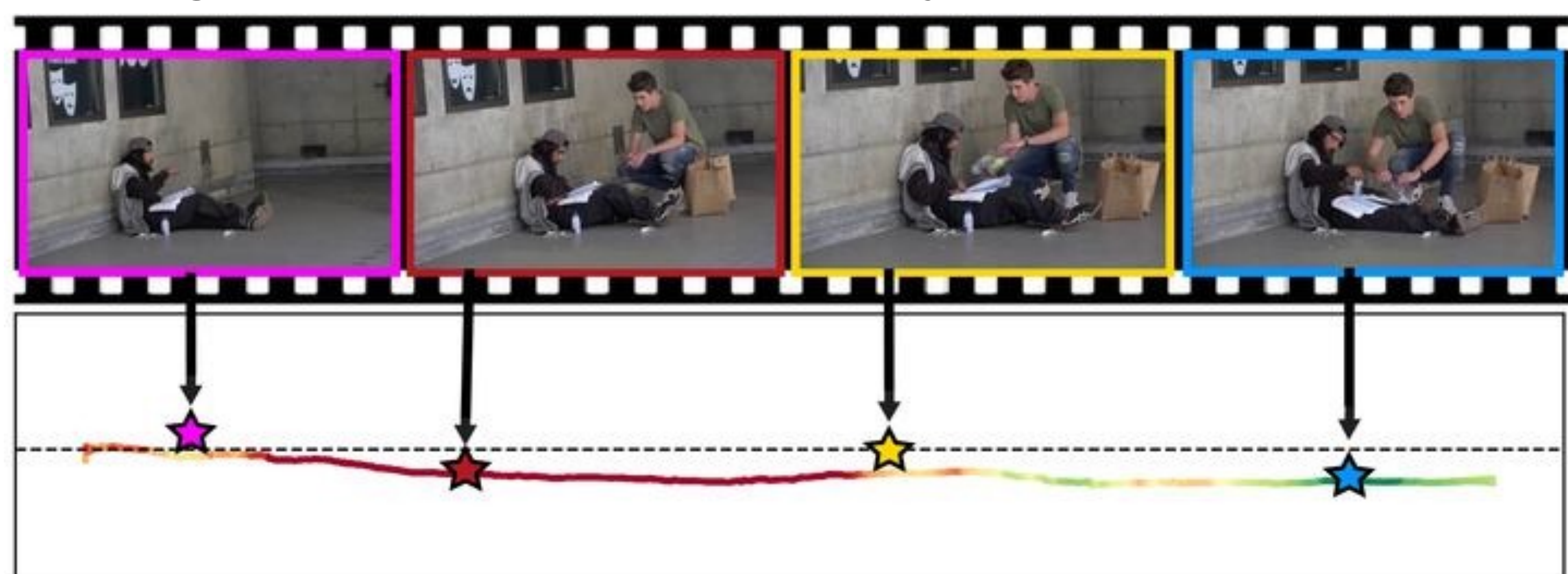

"Love Rosie", Hollywood Movie

"Testament of Youth", Hollywood Movie

"Big Fish", Hollywood Movie

"Feeding The Homeless on The 4th Of July", Home Video

"Last Holiday", Hollywood Movie

"Worlds biggest proposal fail OUCH!!!", Home Video