PAPER PRESENTATION
LUDOVICA SCHAERF
DH-500

6 Seconds of Sound and Vision: Creativity in Micro-Videos

Miriam Redi, Neil O'Hare, Rossano Schifanella, Michele Trevisiol, Alejandro Jaimes

Published in CVPR in 2014

What are micro-videos?

Vine (2013-2017)
6 seconds of creativity

What is creativity?

unique in a
significant way

- Weisberg: "for something to be creative, it is not enough for it to be **novel**: it must have **value**, or be appropriate to the cognitive demands of the situation

specifically
aesthetic values

- Kant: judgements of aesthetic value involve **sensory**, **emotional** and **intellectual** components.

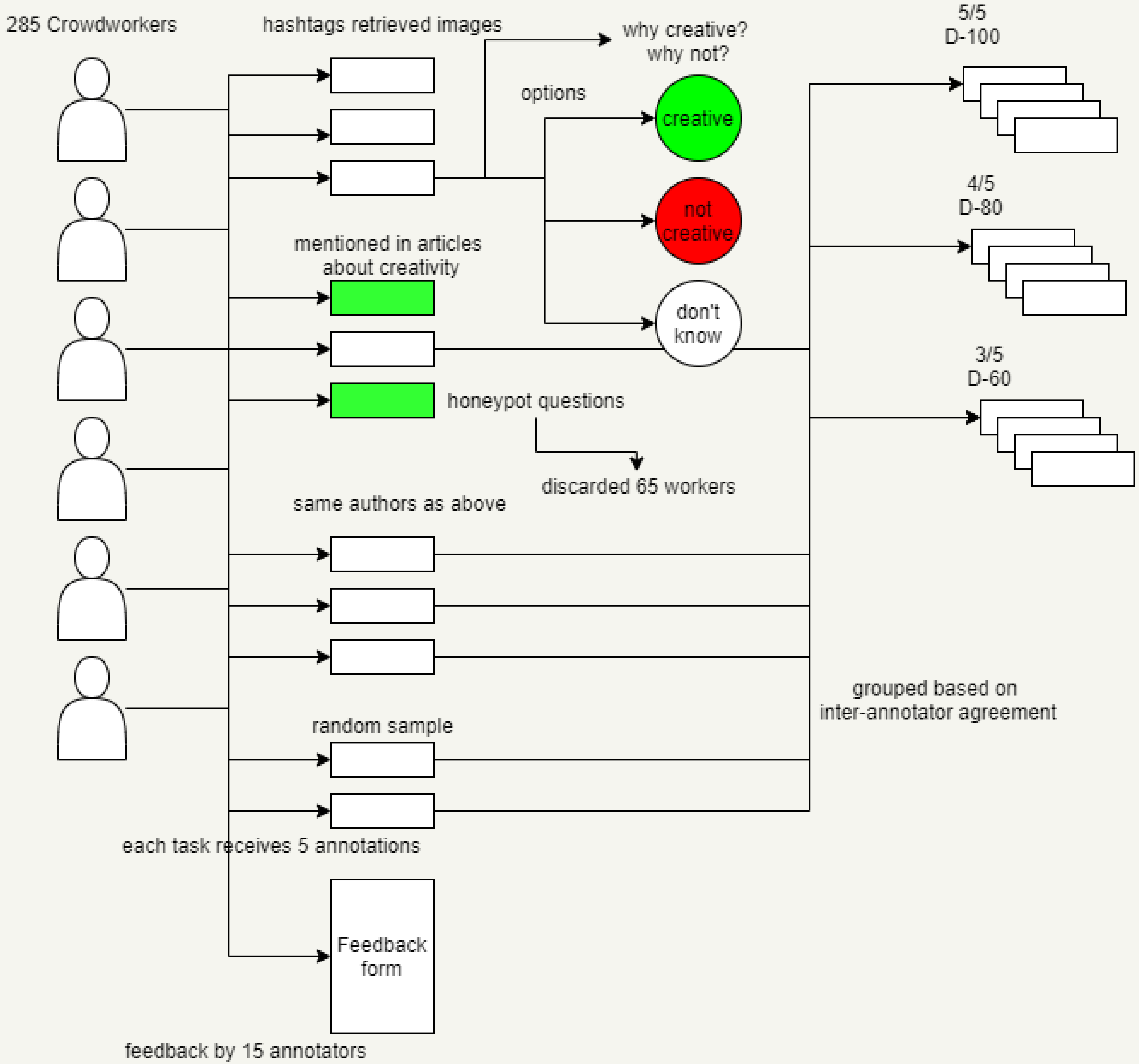
Research Question

“We study the **audio-visual features** of creative vs non-creative videos and present a computational framework to **automatically classify** these categories. In particular, we conduct a **crowdsourcing** experiment to annotate over 3,800 Vine videos, [...]. We go on to use this dataset to study creative micro-videos and to evaluate approaches to automatic detection of creative micro-videos.”

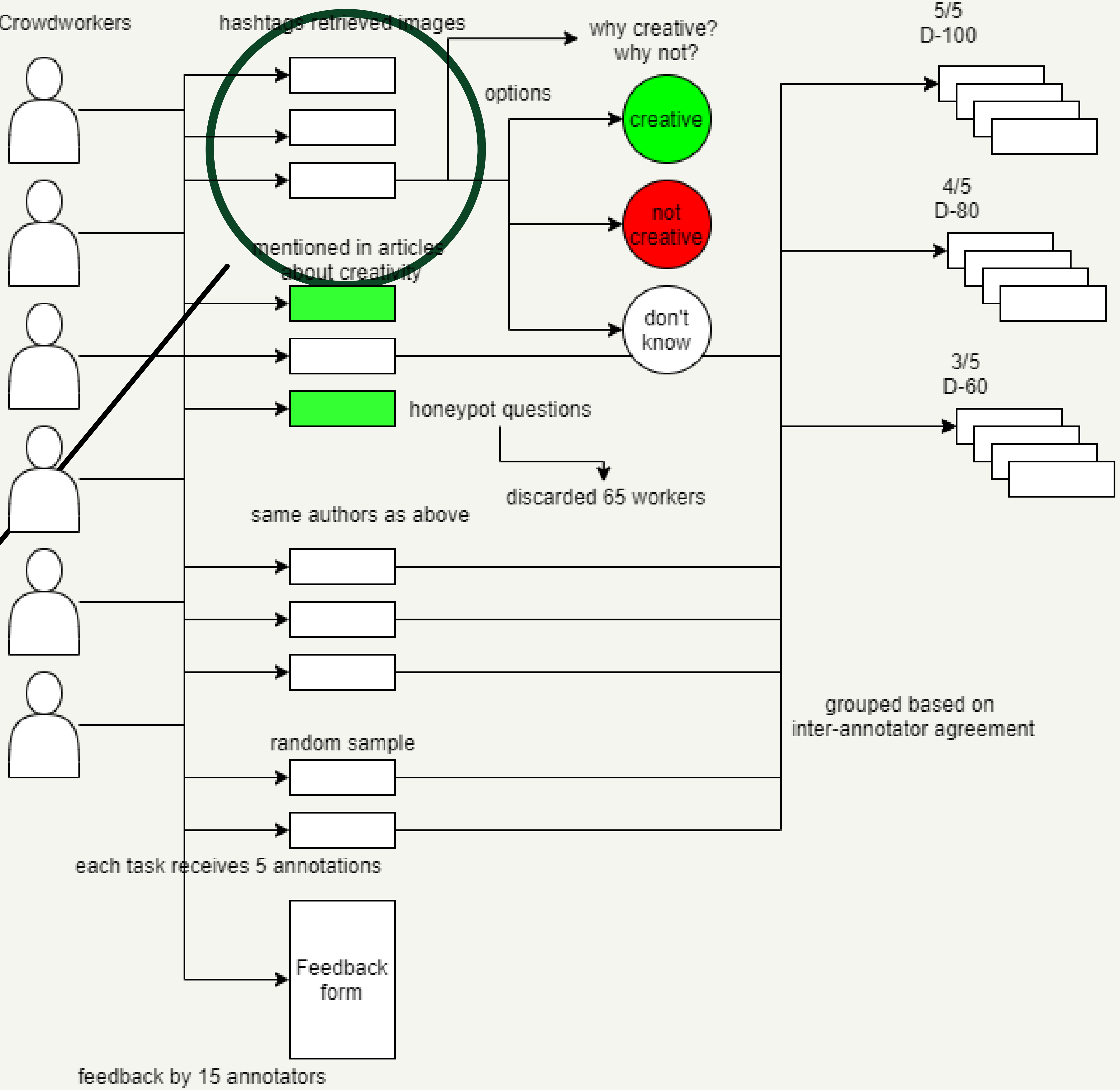
Research Question

Can we create a reliable crowdsourced dataset?
Can we extract features that identify creativity in micro-videos?
Can these be used to automatically classify a micro-video into creative and non-creative?

Crowd-sourcing

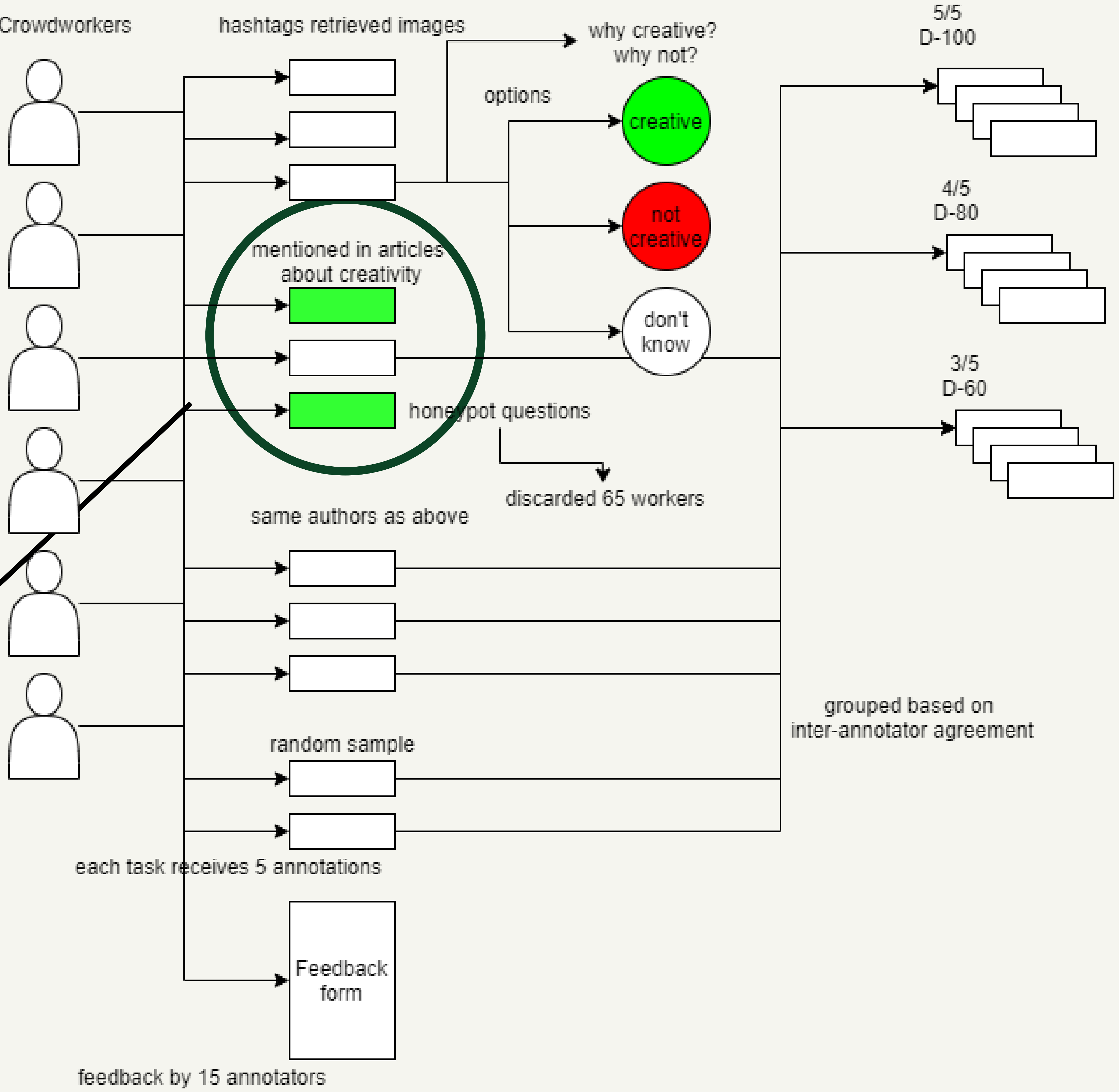


Crowd-sourcing



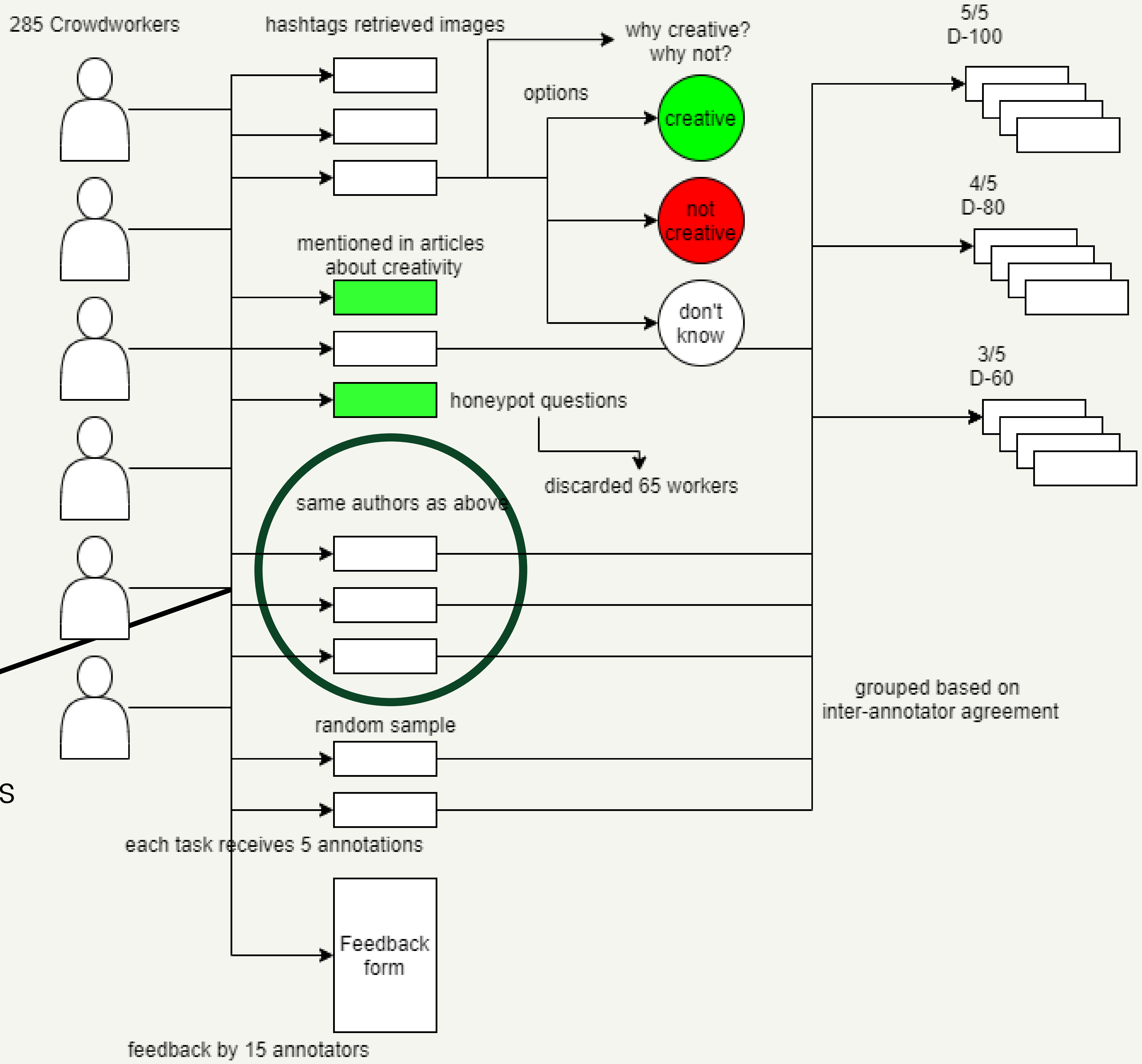
1000 videos
hashtags:
#vineart,
#vineartist,
#artwork ...

Crowd-sourcing



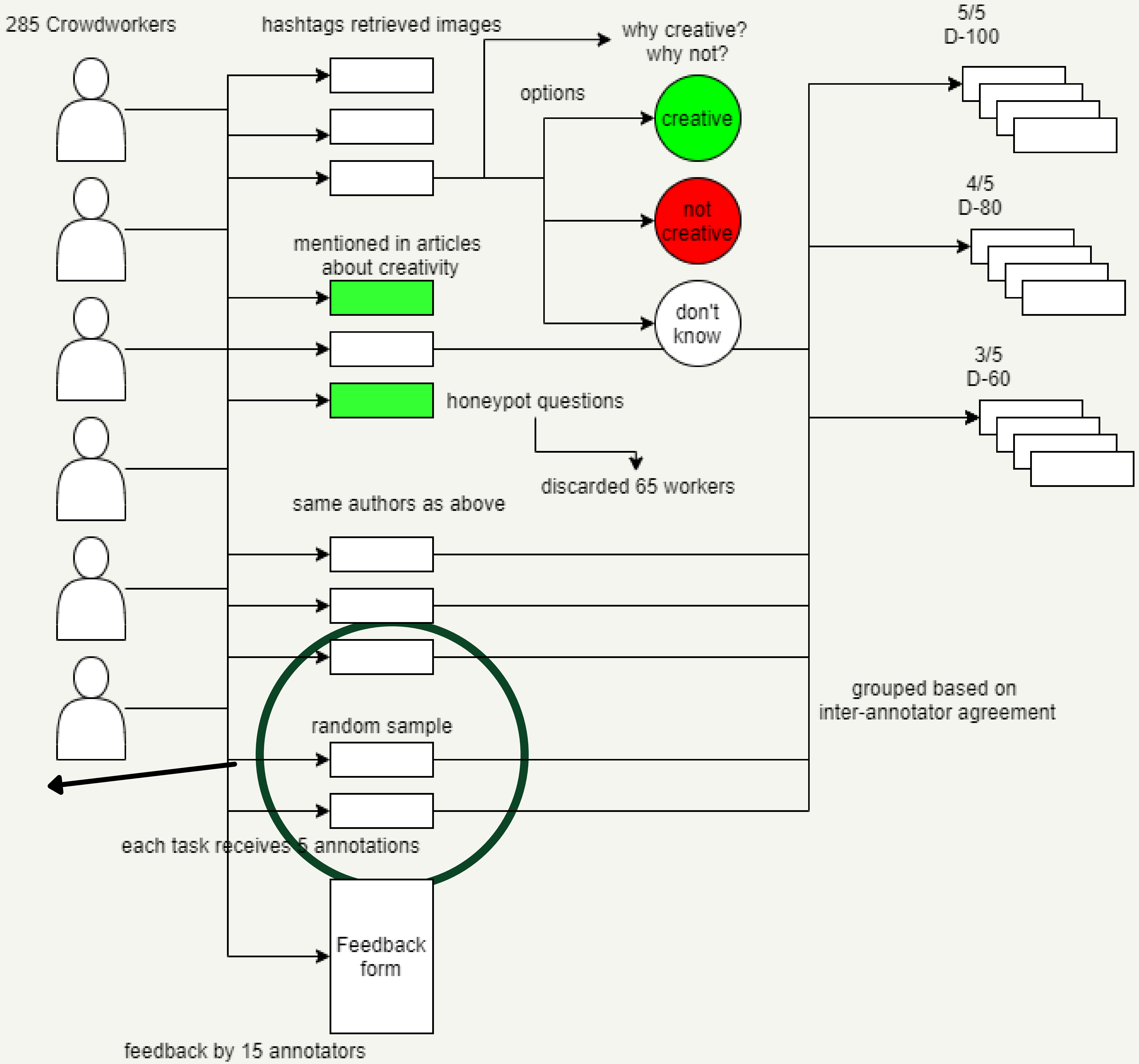
200 videos mentioned in 16 articles about creativity

Crowd-sourcing



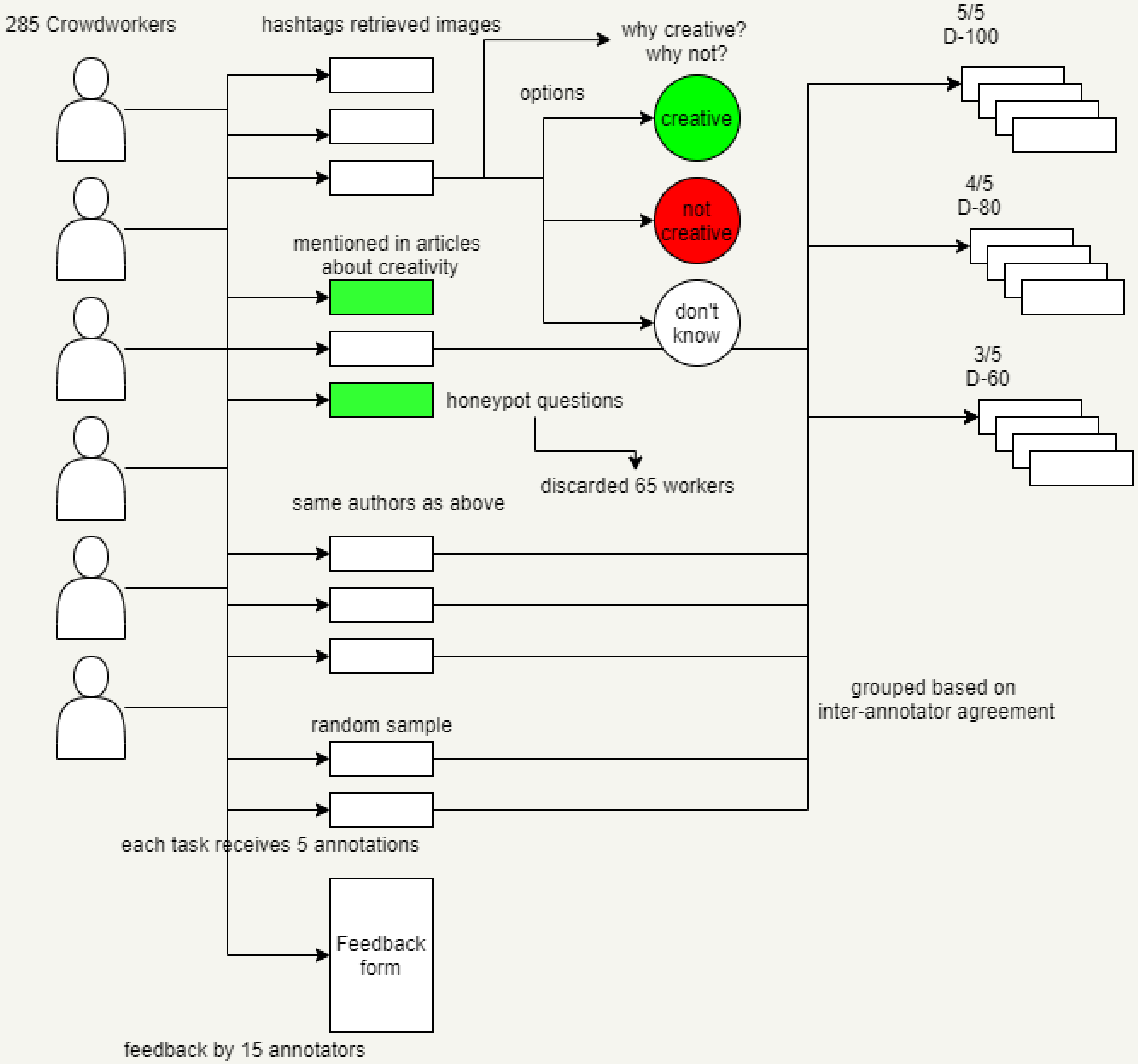
2300 videos
same authors as before

Crowd-sourcing



500 videos
from video
streamline

Crowd-sourcing



Data

Dataset	% Videos	# Creative (%)	# Non-creative (%)
D-60	100%	1141 (30%)	2708 (70%)
D-80	77%	789 (27%)	2196 (73%)
D-100	48%	471 (25%)	1382 (75%)

Table 2. Summary of the results of the labeling experiment. D-60: videos with at least 60% agreement between annotators. D-80: at least 80% agreement. D-100: 100% agreement.

	(a) Hashtags	(b) Blogs	(c) Creators	(d) Random
Creative	34.05%	79.57%	27.41%	1.88%
Non-Creative	65.95%	20.43%	72.59%	98.12%

Table 3. Creative vs non-creative videos per sampling strategy, for the D-100 dataset (100% agreement).

Data

Dataset	% Videos	# Creative (%)	# Non-creative (%)
D-60	100%	1141 (30%)	2708 (70%)
D-80	77%	789 (27%)	2196 (73%)
D-100	48%	471 (25%)	1382 (75%)

Table 2. Summary of the results of the labeling experiment. D-60: videos with at least 60% agreement between annotators. D-80: at least 80% agreement. D-100: 100% agreement.

	(a) Hashtags	(b) Blogs	(c) Creators	(d) Random
Creative	34.05%	79.57%	27.41%	1.88%
Non-Creative	65.95%	20.43%	72.59%	98.12%

Table 3. Creative vs non-creative videos per sampling strategy, for the D-100 dataset (100% agreement).

less than 2% of
the videos on
Vine are creative

Research Question

Can we create a reliable crowdsourced dataset?
Can we extract features that identify creativity in micro-videos?
Can these be used to automatically classify a micro-video into creative and non-creative?

Features

new features

Group	Feature	Dim	Description
			AESTHETIC VALUE
			Sensory Features
Scene Content	*Saliency Moments* [26]	462	Frame content is represented by summarizing the shape of the salient region
Filmmaking Technique	*General Video Properties*	2	*Number of Shots*, *Number of Frames*
	Stop Motion	1	Number of non-equal adjacent frames
	Loop	1	Distance between last and first frame
	Movement	1	Avg. distance between spectral residual [9] saliency maps of adjacent frames
	Camera Shake	1	Avg. amount of camera shake [1] per frame
Composition and Photographic Technique	*Rule of Thirds* [5]	3	HSV average value of the inner quadrant of the frame (*H(RoT),S(RoT),V(RoT)*)
	Low Depth of Field [5]	9	LDOF indicators computed using wavelet coefficients
	Contrast [6]	1	Ratio between the sum of max and min luminance values and their difference
	Symmetry [27]	1	Difference between edge histograms of left and right halves of the image
	Uniqueness [27]	1	Distance between the frame spectrum and the average image spectrum
	Image Order [28]	2	Order values obtained through Kologomorov *Complexity* and Shannon's Entropy
			Emotional Affect Features
Visual Affect	*Color Names* [17]	9	Amount of color clusters such as red, blue, green, . . .
	Graylevel Contrast Matrix Properties [17]	10	*Entropy, Dissimilarity, Energy, Homogeneity* and *Contrast* of the GLCM matrix
	HSV statistics [17]	3	Average *Hue, Saturation and Brightness* in the frame
	Pleasure, Arousal, Dominance [30]	3	Affective dimensions computed by mapping HSV values
Audio Affect	*Loudness* [15]	2	Overall *Energy* of signal and avg *Short-Time Energy* in a 2-seconds window
	Mode [15]	1	Sums of key strength differences between major keys and their relative minor keys
	Roughness [15]	1	Avg of the dissonance values between all pairs of peak in the sound track spectrum
	Rythmical Features [15]	2	*Onset Rate* and *Zero-Crossing Rate*
			NOVELTY
Novelty	*Audio Novelty*	10	Distance between the audio features and the audio space
	Visual Novelty	40	Distance between the visual features and each visual feature space

Table 4. Audiovisual features for creativity modeling

Features

sensory features

emotion features

Group	Feature	Dim	Description
			AESTHETIC VALUE
			Sensory Features
Scene Content	*Saliency Moments* [26]	462	Frame content is represented by summarizing the shape of the salient region
Filmmaking Technique	*General Video Properties*	2	*Number of Shots*, *Number of Frames*
	Stop Motion	1	Number of non-equal adjacent frames
	Loop	1	Distance between last and first frame
	Movement	1	Avg. distance between spectral residual [9] saliency maps of adjacent frames
	Camera Shake	1	Avg. amount of camera shake [1] per frame
Composition and Photographic Technique	*Rule of Thirds* [5]	3	HSV average value of the inner quadrant of the frame (*H(RoT)*,*S(RoT)*,*V(RoT)*)
	Low Depth of Field [5]	9	LDOF indicators computed using wavelet coefficients
	Contrast [6]	1	Ratio between the sum of max and min luminance values and their difference
	Symmetry [27]	1	Difference between edge histograms of left and right halves of the image
	Uniqueness [27]	1	Distance between the frame spectrum and the average image spectrum
	Image Order [28]	2	Order values obtained through Kologomorov *Complexity* and Shannon's Entropy
			Emotional Affect Features
Visual Affect	*Color Names* [17]	9	Amount of color clusters such as red, blue, green, ...
	Graylevel Contrast Matrix Properties [17]	10	*Entropy, Dissimilarity, Energy, Homogeneity* and *Contrast* of the GLCM matrix
	HSV statistics [17]	3	Average *Hue, Saturation and Brightness* in the frame
	Pleasure, Arousal, Dominance [30]	3	Affective dimensions computed by mapping HSV values
Audio Affect	*Loudness* [15]	2	Overall *Energy* of signal and avg *Short-Time Energy* in a 2-seconds window
	Mode [15]	1	Sums of key strength differences between major keys and their relative minor keys
	Roughness [15]	1	Avg of the dissonance values between all pairs of peak in the sound track spectrum
	Rythmical Features [15]	2	*Onset Rate* and *Zero-Crossing Rate*
			NOVELTY
Novelty	*Audio Novelty*	10	Distance between the audio features and the audio space
	Visual Novelty	40	Distance between the visual features and each visual feature space

Table 4. Audiovisual features for creativity modeling

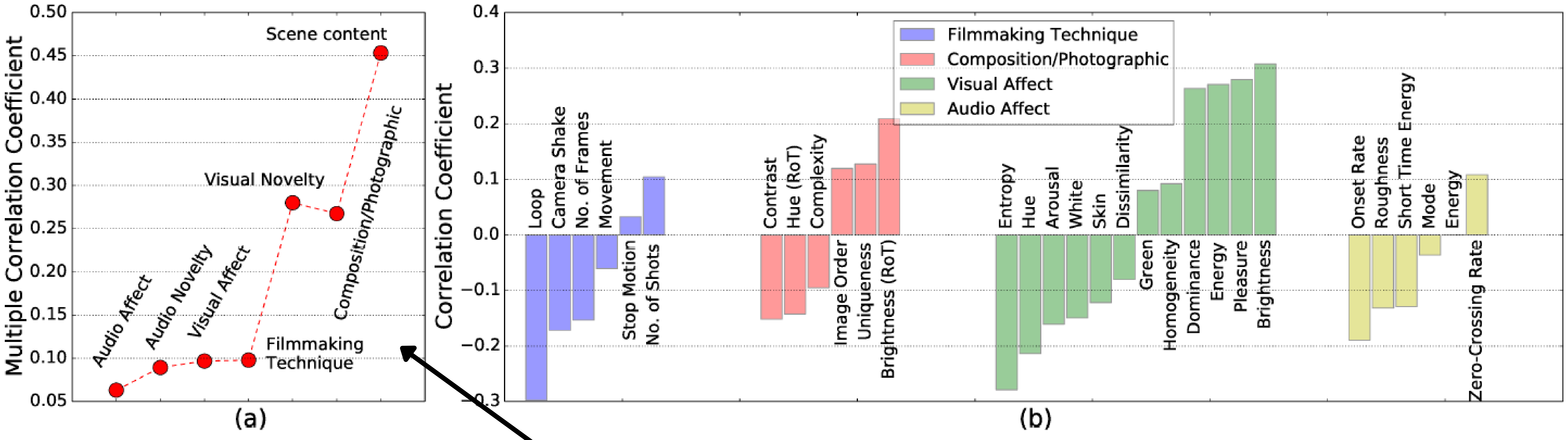


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

Data: **D100**

Features: **7 groups** of features on the left

Method: **Multiple Correlation Coefficient** (MCC)

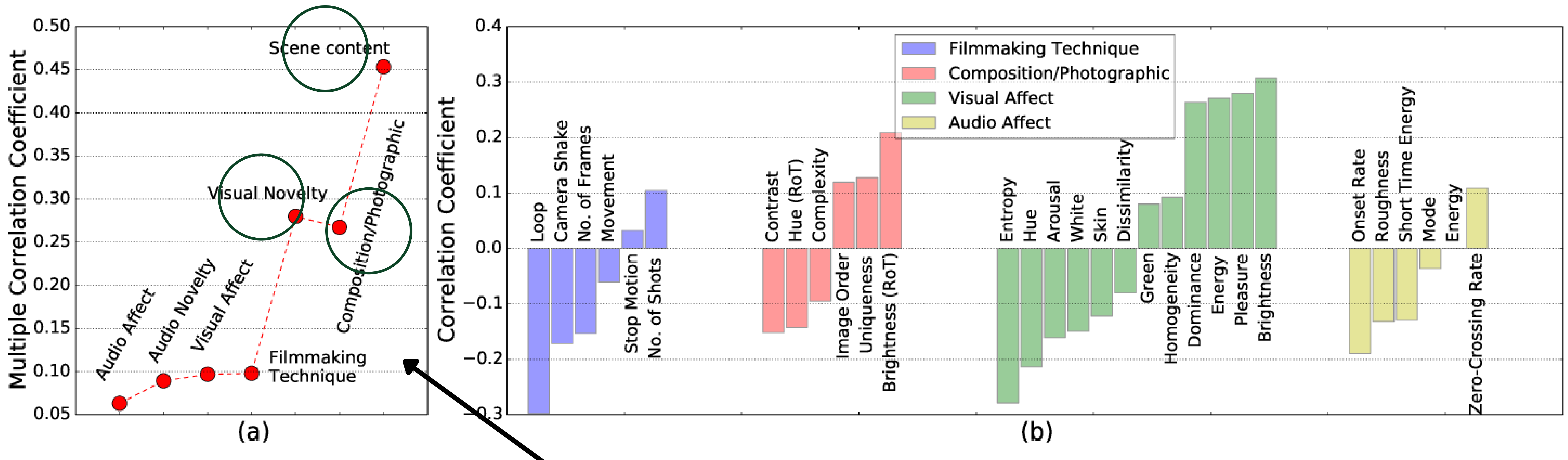


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

Both Novelty and Aesthetic features are important

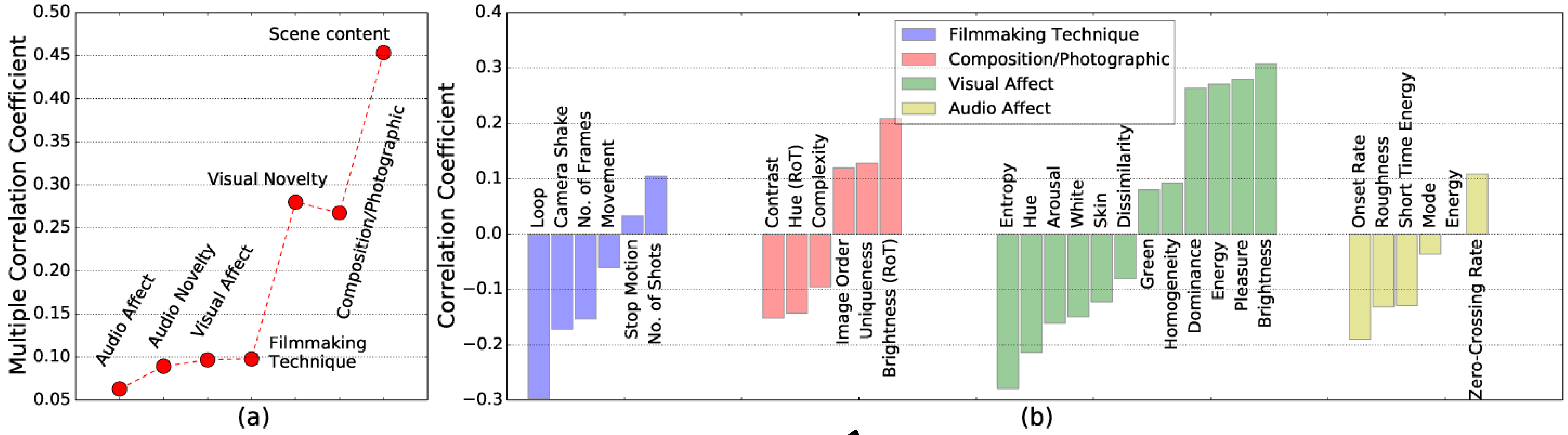


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

Data: **D100**
Features: All individual features but Scene Content and Novelty
Method: **Pearson Correlation Coefficient** (PCC)

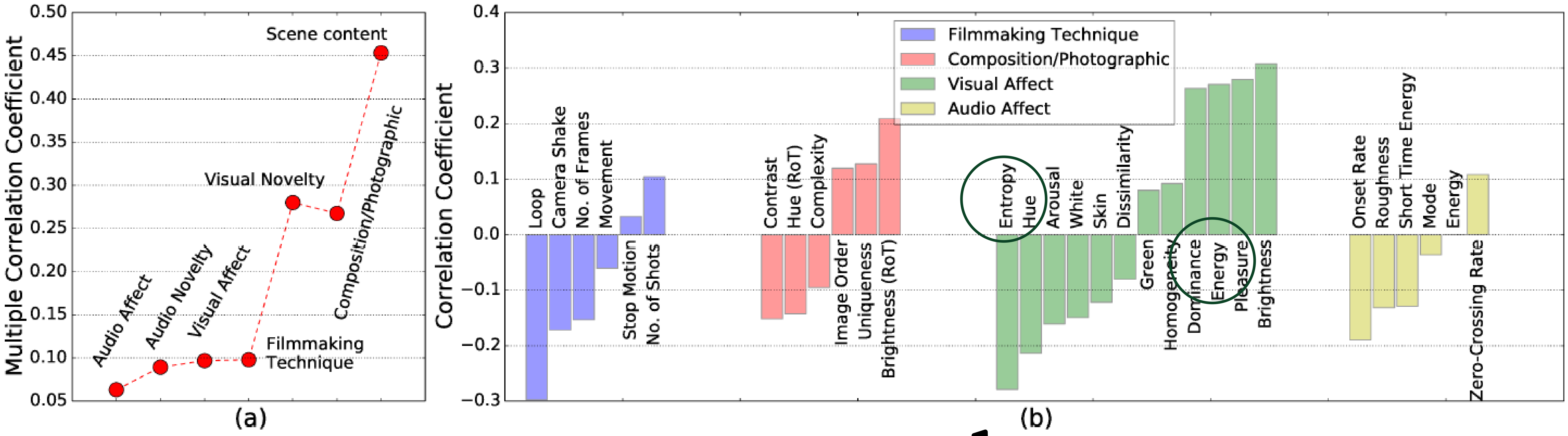


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

Favour visual uniformity

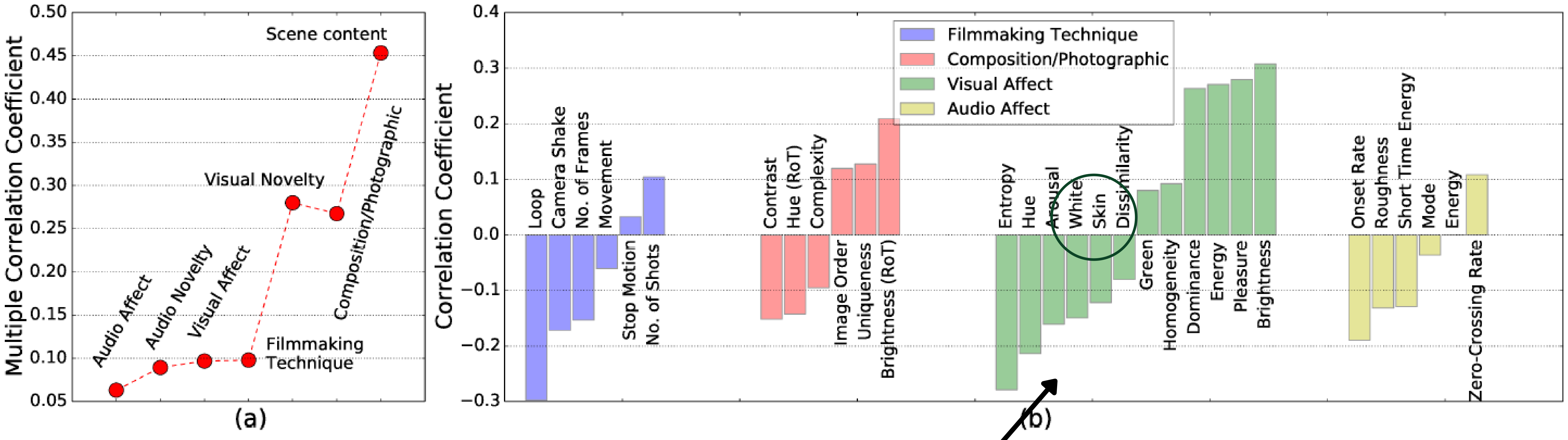


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

favour scenes without people

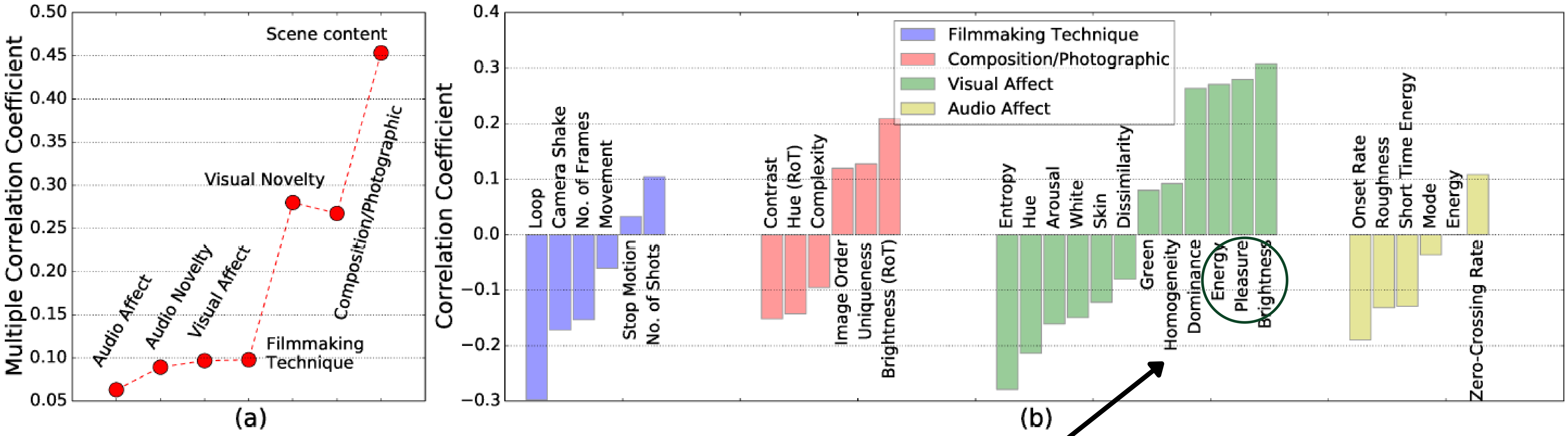


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

dominant non-overwhelming emotions

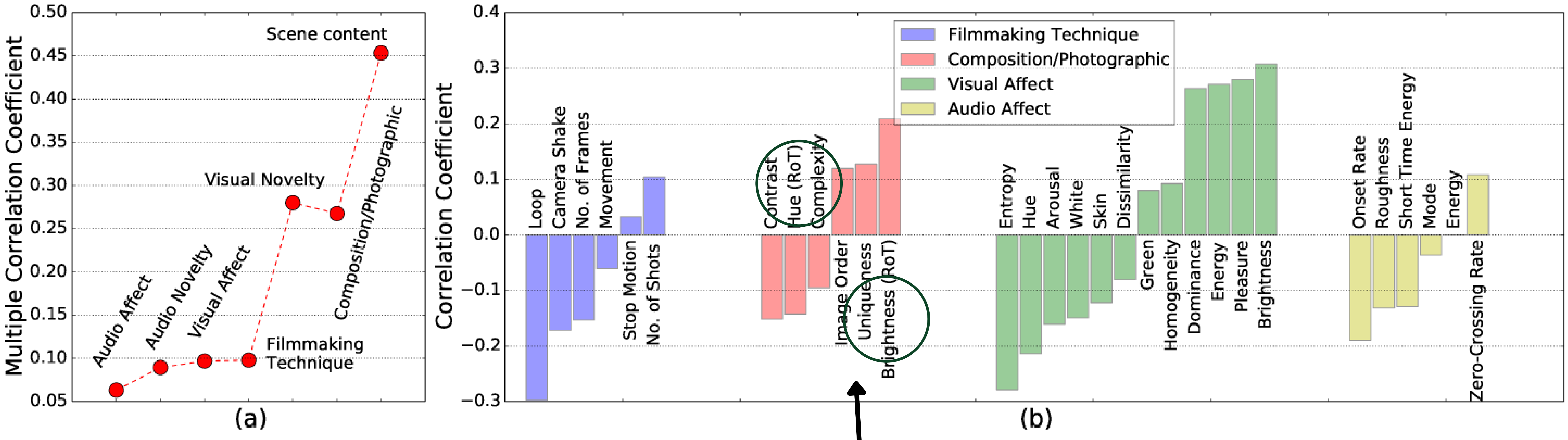


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

Warm bright colors

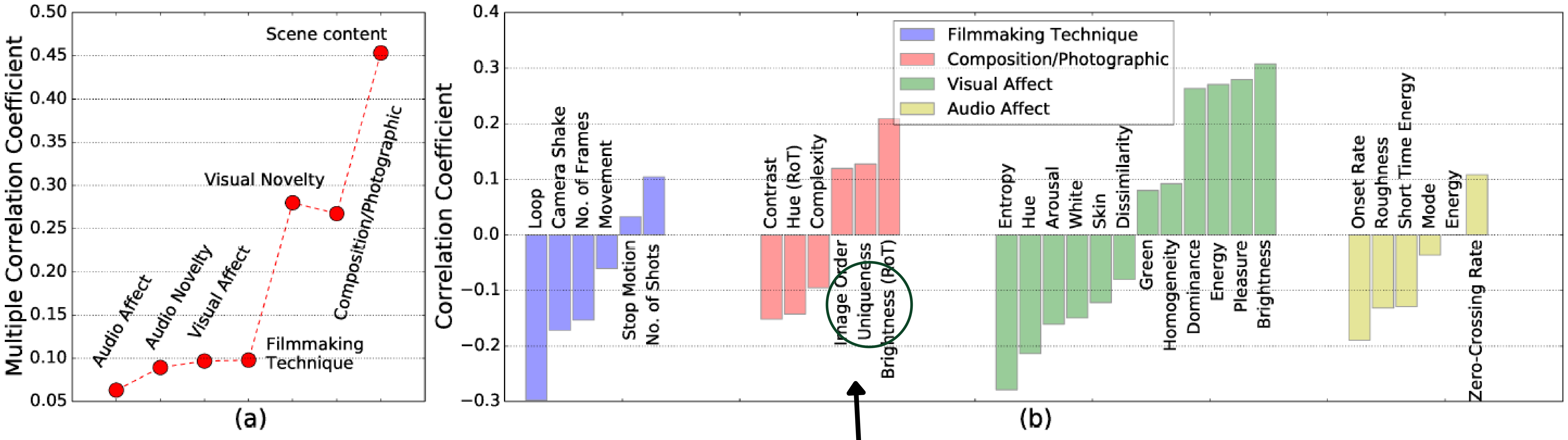


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

favour less familiar layout

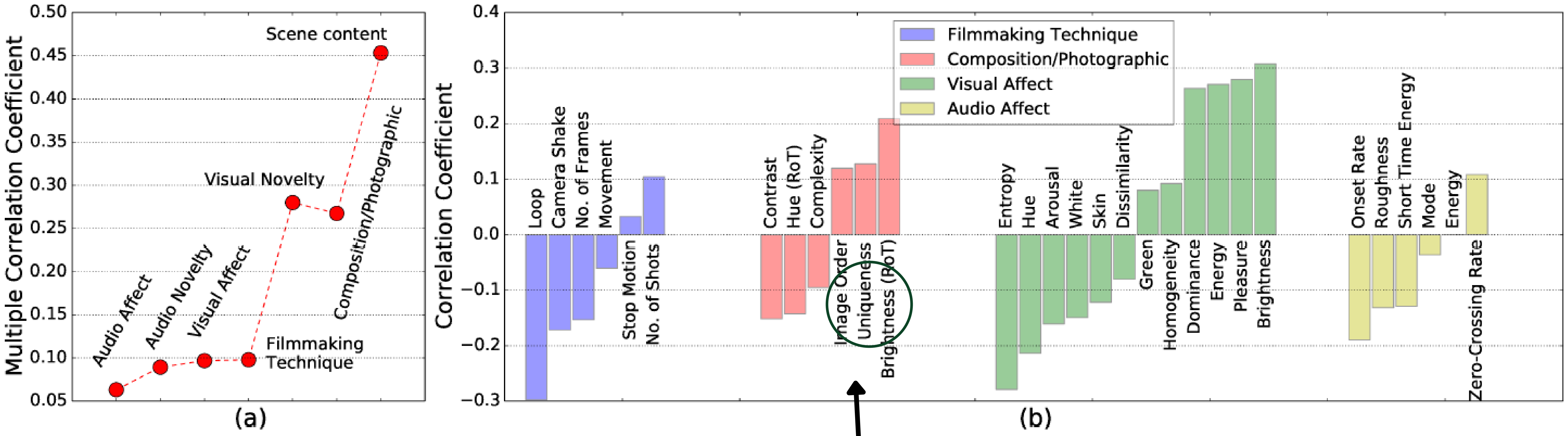


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

favour less familiar layout

but no symmetry or depth of field

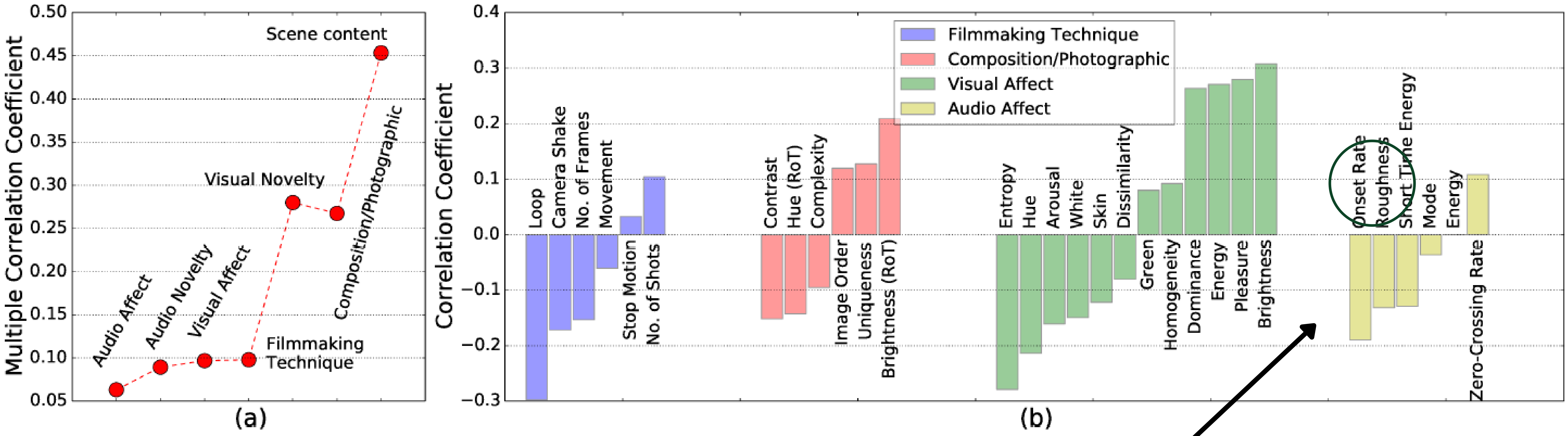


Figure 1. Analysis of the most relevant features and components for video creativity prediction

Correlation

less-frenetic, low volume

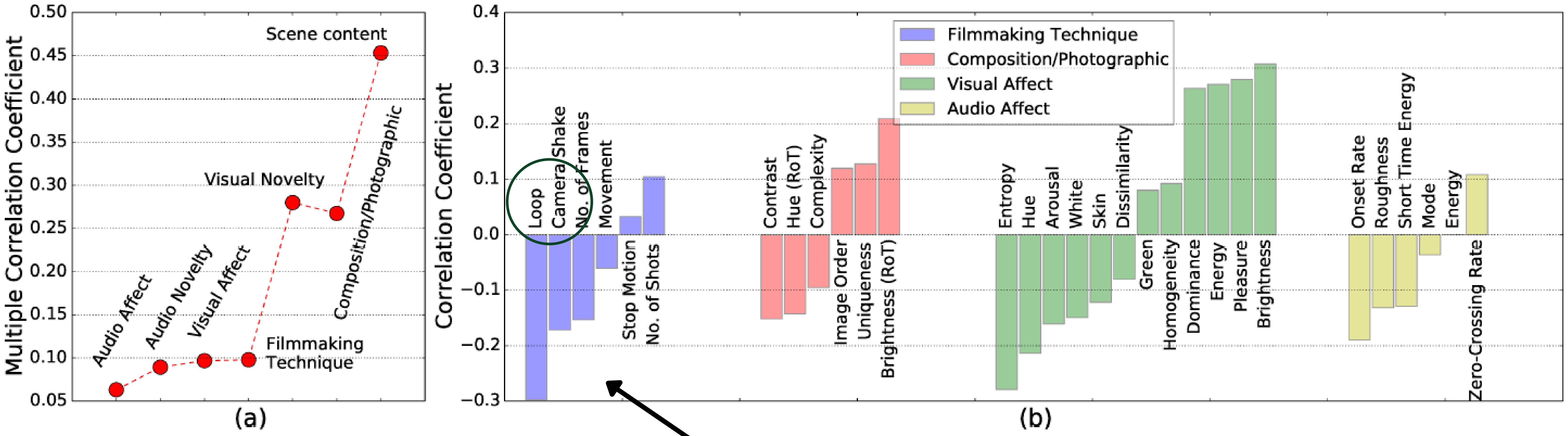


Figure 1. Analysis of the most relevant features and components for video creativity prediction

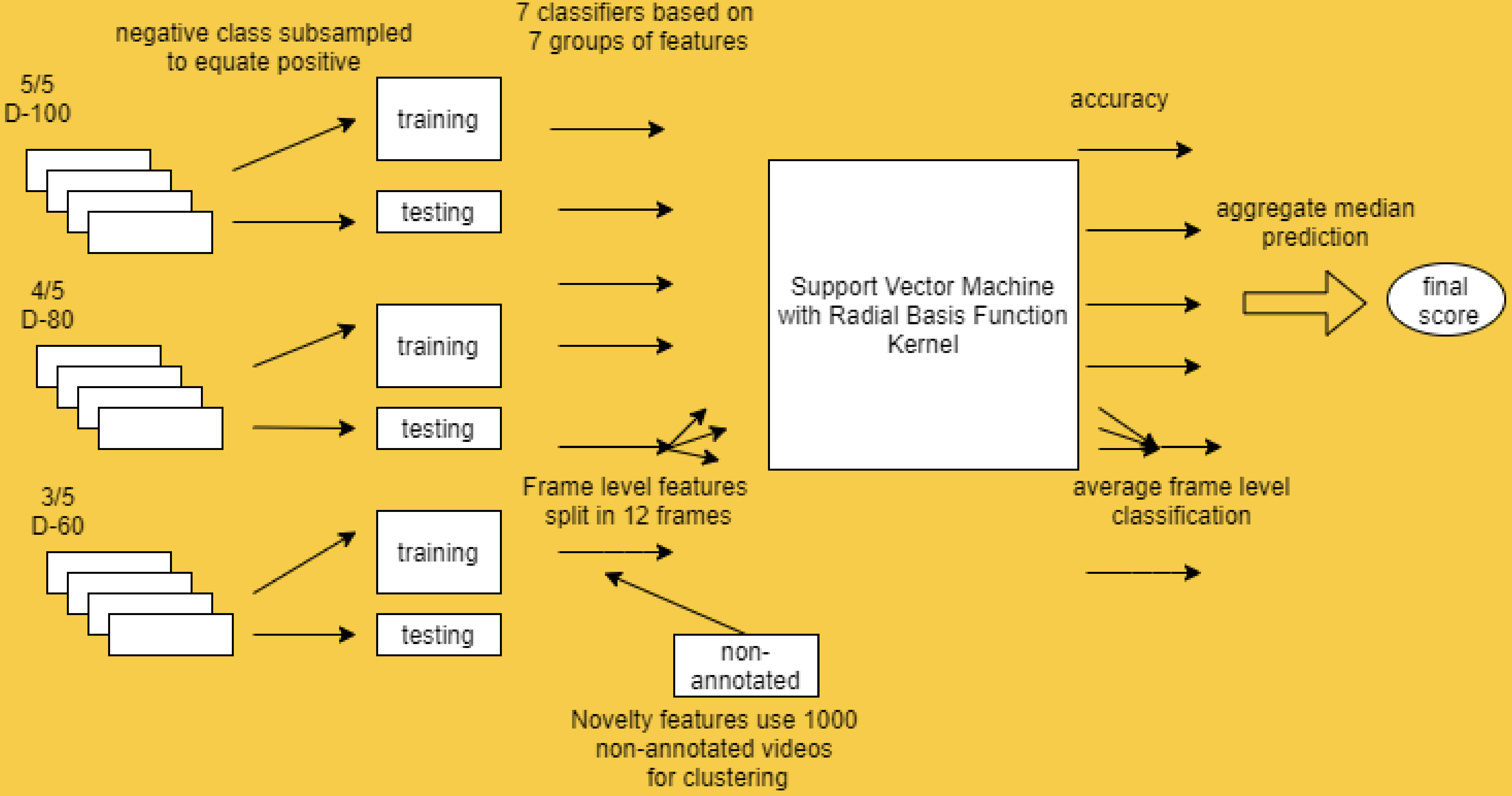
Correlation

loops are characteristic, polished videos

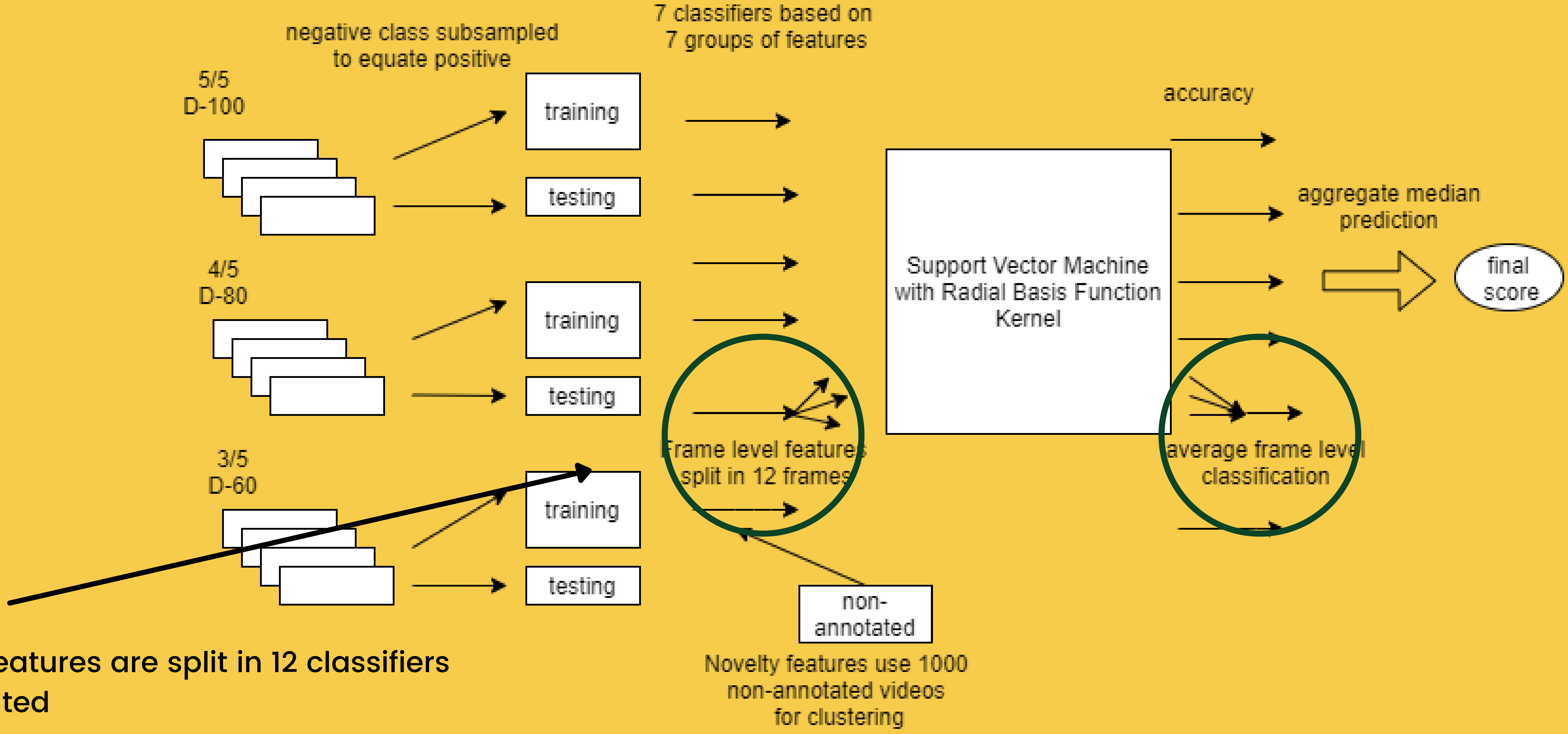
Research Question

Can we create a reliable crowdsourced dataset?
Can we extract features that identify creativity in micro-videos?
Can these be used to automatically classify a micro-video into creative and non-creative?

Classification



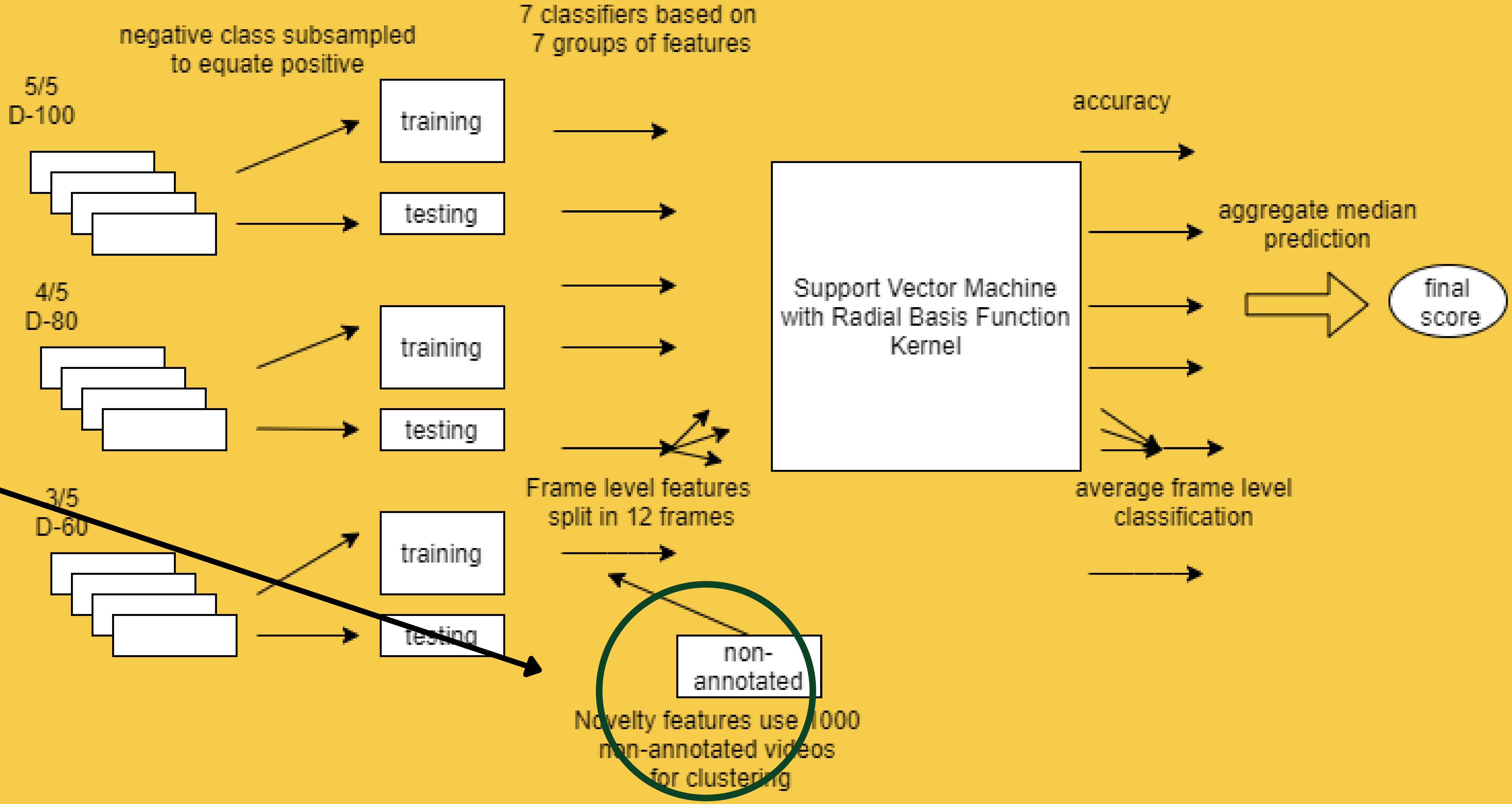
Classification



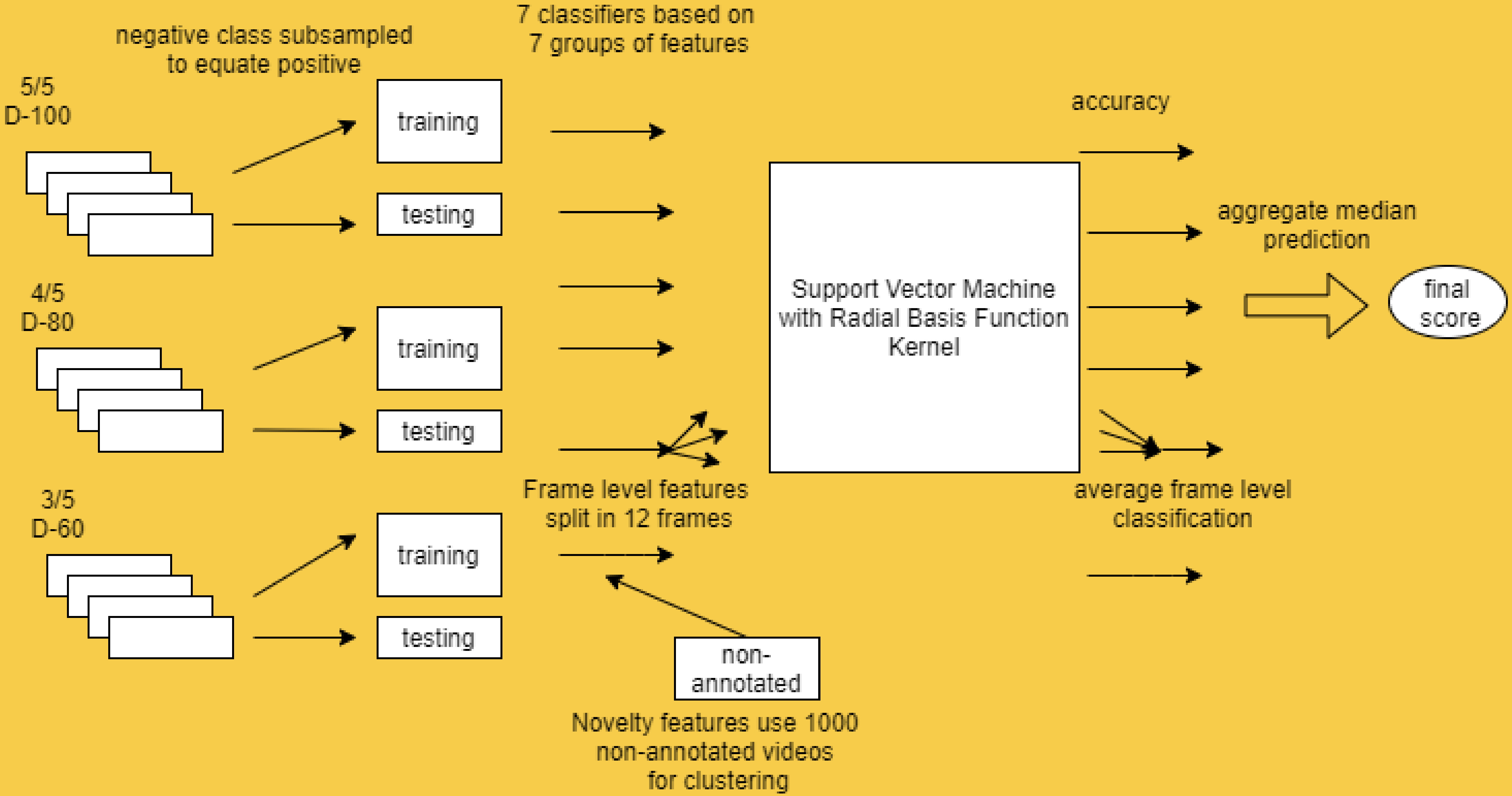
frame level features are split in 12 classifiers
and aggregated

Classification

Novelty features use random set of non-annotated images, show that size for clustering does not count



Classification



Results

Feature	Accuracy		
	D-60	D-80	D-100
Aesthetic Value			
Sensory Features			
Scene Content	0.67	0.69	0.74
Filmmaking Techniques	0.65	0.69	0.73
Composition & Photographic Technique	0.67	0.74	**0.77**
All Sensory Features	0.69	**0.75**	0.77
Emotional Affect Features			
Audio Affect	0.59	0.53	0.67
Visual Affect	0.65	0.66	0.66
All Emotional Affect Features	0.62	0.56	**0.71**
All Aesthetic Value Features	0.68	0.72	**0.79**
Novelty			
Audio	0.58	0.58	0.63
Visual	0.63	0.67	**0.74**
Audio + Visual Novelty	0.59	0.63	0.69
Novelty + Aesthetic Value	0.69	0.73	**0.80**

Table 5. Prediction results for value and novelty features

Results

Feature	Accuracy		
	D-60	D-80	D-100
Aesthetic Value			
Sensory Features			
Scene Content	0.67	0.69	0.74
Filmmaking Techniques	0.65	0.69	0.73
Composition & Photographic Technique	0.67	0.74	**0.77**
All Sensory Features	0.69	**0.75**	0.77
Emotional Affect Features			
Audio Affect	0.59	0.53	0.67
Visual Affect	0.65	0.66	0.66
All Emotional Affect Features	0.62	0.56	**0.71**
All Aesthetic Value Features	0.68	0.72	**0.79**
Novelty			
Audio	0.58	0.58	0.63
Visual	0.63	0.67	**0.74**
Audio + Visual Novelty	0.59	0.63	0.69
Novelty + Aesthetic Value	0.69	0.73	**0.80**

Table 5. Prediction results for value and novelty features

Best individual features correspond to PCC results

Results

Feature	Accuracy		
	D-60	D-80	D-100
Aesthetic Value			
Sensory Features			
Scene Content	0.67	0.69	0.74
Filmmaking Techniques	0.65	0.69	0.73
Composition & Photographic Technique	0.67	0.74	**0.77**
All Sensory Features	0.69	**0.75**	0.77
Emotional Affect Features			
Audio Affect	0.59	0.53	0.67
Visual Affect	0.65	0.66	0.66
All Emotional Affect Features	0.62	0.56	**0.71**
All Aesthetic Value Features	0.68	0.72	**0.79**
Novelty			
Audio	0.58	0.58	0.63
Visual	0.63	0.67	**0.74**
Audio + Visual Novelty	0.59	0.63	0.69
Novelty + Aesthetic Value	0.69	0.73	**0.80**

Table 5. Prediction results for value and novelty features

Combination of emotion and sensory features shows great improvement, complementarity

Results

Feature	Accuracy		
	D-60	D-80	D-100
Aesthetic Value			
Sensory Features			
Scene Content	0.67	0.69	0.74
Filmmaking Techniques	0.65	0.69	0.73
Composition & Photographic Technique	0.67	0.74	**0.77**
All Sensory Features	0.69	**0.75**	0.77
Emotional Affect Features			
Audio Affect	0.59	0.53	0.67
Visual Affect	0.65	0.66	0.66
All Emotional Affect Features	0.62	0.56	**0.71**
All Aesthetic Value Features	0.68	0.72	**0.79**
Novelty			
Audio	0.58	0.58	0.63
Visual	0.63	0.67	**0.74**
Audio + Visual Novelty	0.59	0.63	0.69
Novelty + Aesthetic Value	0.69	0.73	**0.80**

Table 5. Prediction results for value and novelty features

Mild improvement adding also novelty

Conclusion

Crowdsourcing

Good inter-annotator agreement

Three datasets.

Features

New features encoding:

- aesthetic values
- novelty

Model

Promising results,

80% accuracy

Future Work

- intellectual features
- metadata

- application to other micro-video platforms