HANGZHOU DIANZI UNIVERSITY

UNIVERSITY OF YAMANASHI

AffectI: A Game for Diverse, Reliable, and Efficient Affective Image Annotation

Xingkun Zuo[1,2], Jiyi Li[2], Qili Zhou[1], Jianjun Li[1], Xiaoyang Mao[1,2]

1. Hangzhou Dianzi University
2. University of Yamanashi

ACM multimedia 2020

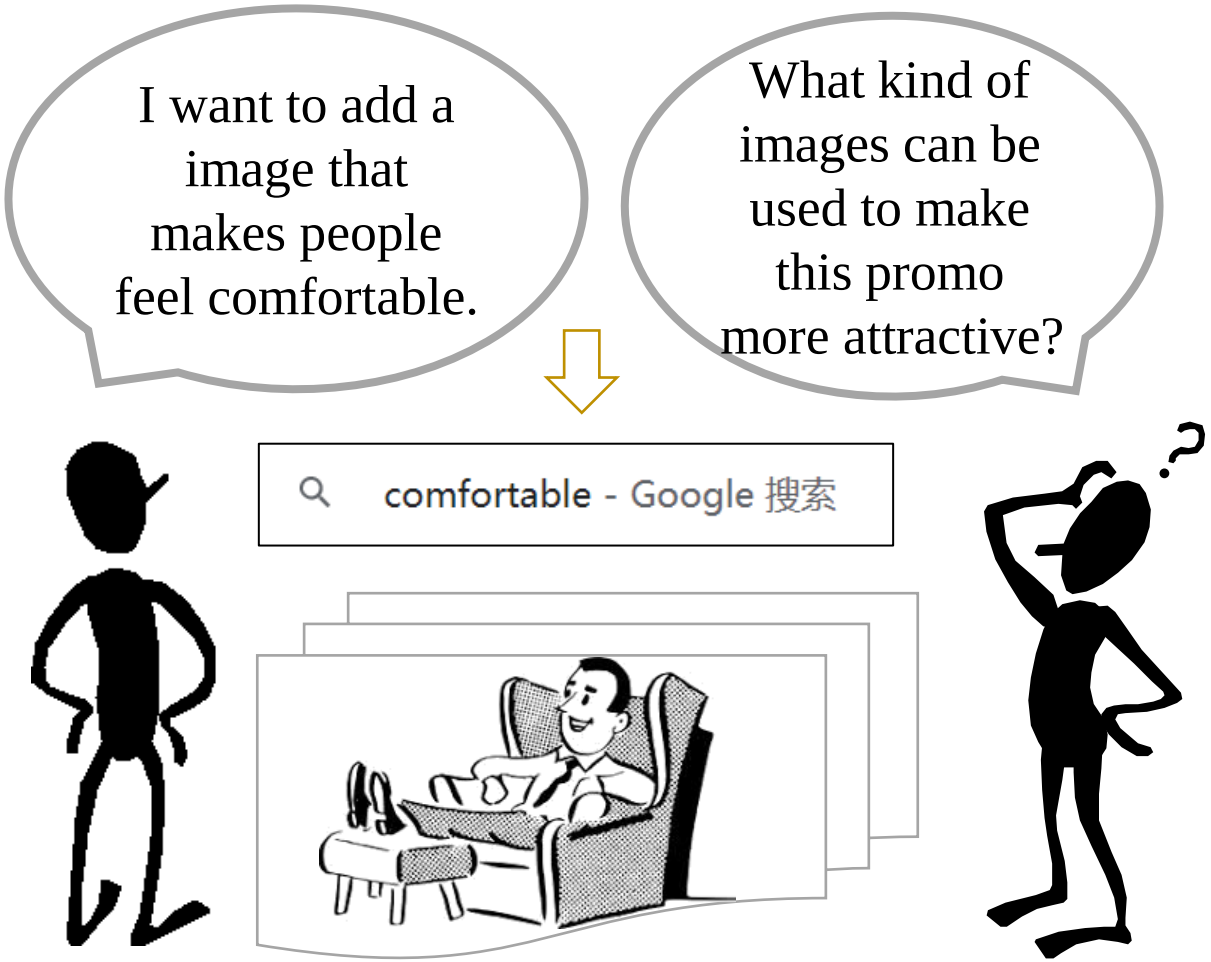
Contents

Contents

Affective Image Annotation

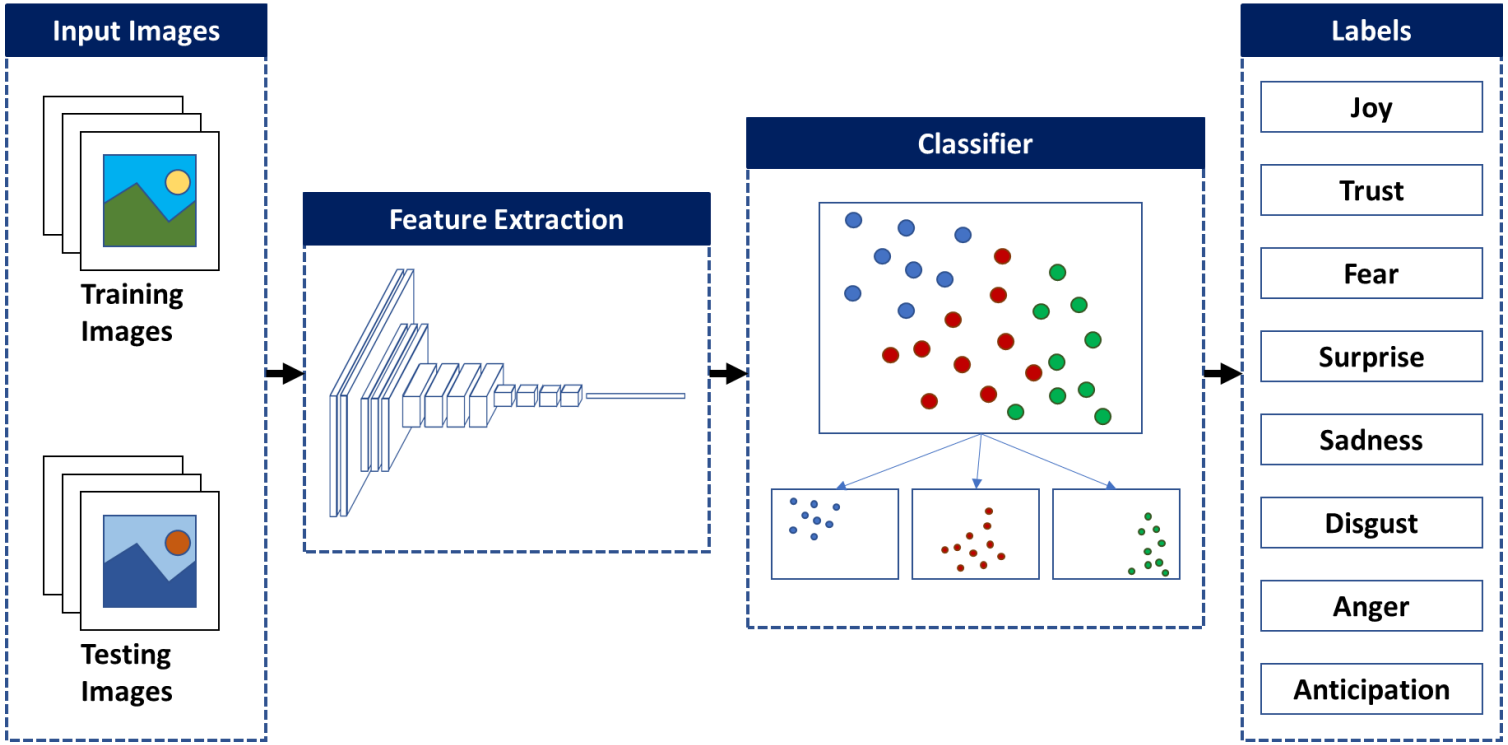
Traditional Annotation		Affective Annotation
Wedding		Happiness Beautiful …
Cat		Sad Cute …

Image Retrieval



Affective Image Content Analysis

Annotation is required for collecting training data in machine learning.



Zhao, Sicheng, et al. "Affective Image Content Analysis: A Comprehensive Survey." *IJCAI*. 2018.

Demo


AffectI
An online game for labelling images with emotional words
Round: : 20s
Score:
Start
female 25-30 English
Your name: reyaki
If you don't know how to play, please click tutorial button.
Choose player
online player
past player
Play with system
Your oppoent: -
Number of online players:1
Tutorial
Practice
Result Analysis
Emotion Wheel
Chat Box

Contents

Related Work——Affective Image Annotation

- Manual labeling
  - crowdsourcing
- Text information
- Games with a purpose

Emotion category: joy
Sentiment: positive

……..
After more rounds

Emotion category: sadness
Sentiment: negative

Long time manual labeling can easily cause fatigue and affect the reliability of the results

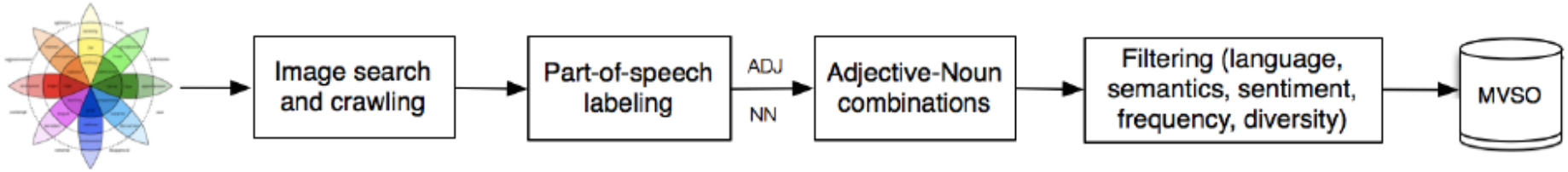
Related Work——Affective Image Annotation

- Manual labeling
- Text information
- Games with a purpose

Description: Yesterday I thought I would spend my short trip very happily.

Emotion category: joy?

Sentiment: positive?



MVSO ——A Large-Scale Multilingual Visual Sentiment Ontology

The quality of the data is usually low.

Jou, Brendan, et al. "Visual affect around the world: A large-scale multilingual visual sentiment ontology." Proceedings of the 23rd ACM international conference on Multimedia. 2015.

Related Work——Games With a Purpose

Game with A Purpose is a human-based computation technique aimed at collecting data as a side effect of game play.

- Manual labeling
- Text information
- Games with a purpose
  - ESP Game
  - KKB Game (Ho et al. 2009)
  - Karido (Bartholomus et al. 2011)

THE ESP GAME

PLAYER 1

GUESSING: CAR

GUESSING: HAT

GUESSING: KID

SUCCESS!
YOU AGREE ON CAR

PLAYER 2

GUESSING: BOY

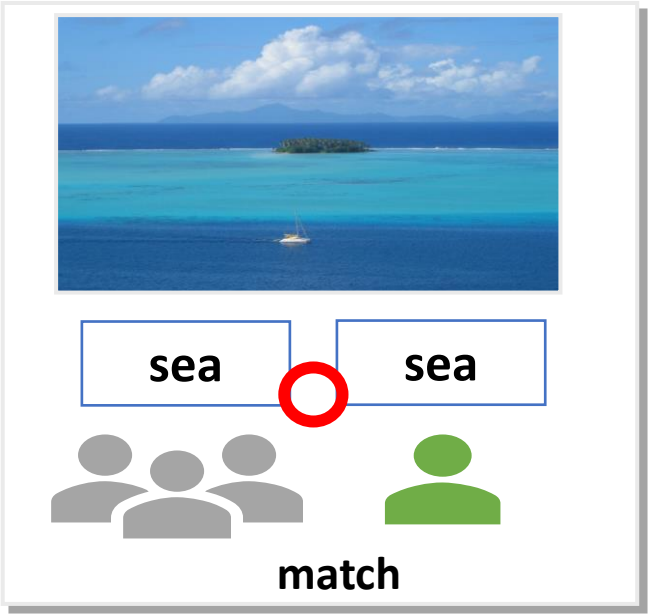
GUESSING: CAR

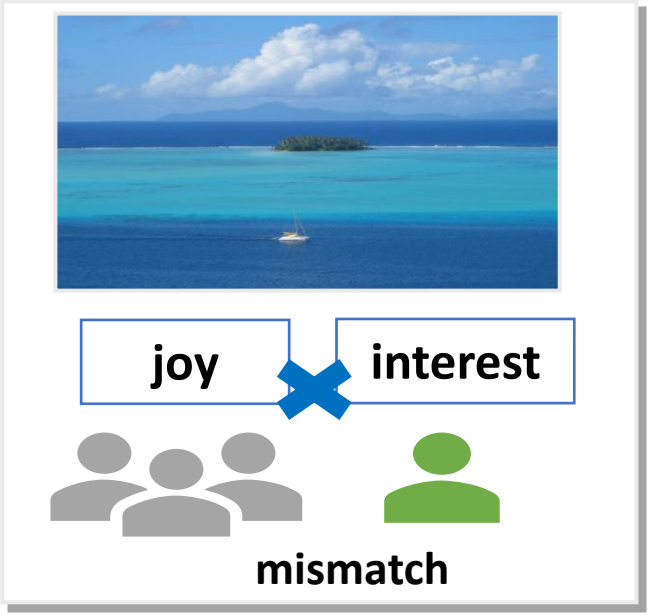
SUCCESS!
YOU AGREE ON CAR

Ahn, L. V. Ahn., and Dabbish, L. Dabbish., "Labeling Image with a Computer Game", In Proceeding of ACM Conference on Human Factors in Computing Systems (CHI), pp. 319-326, 2004.

Related Work——Games With a Purpose

- Manual labeling
- Text information
- Games with a purpose

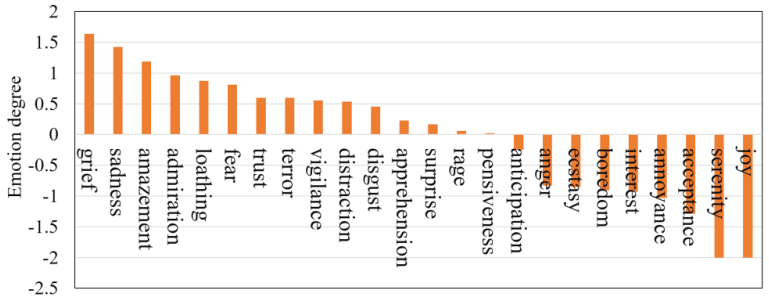




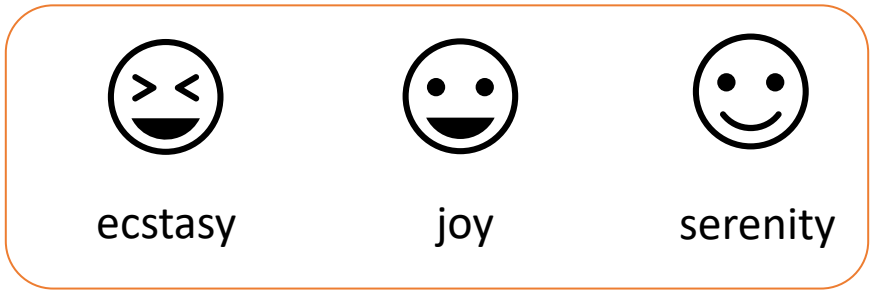
It's difficult to come up with the same emotional words by typing.

Requirement

- Multiple labels
- Emotion distribution



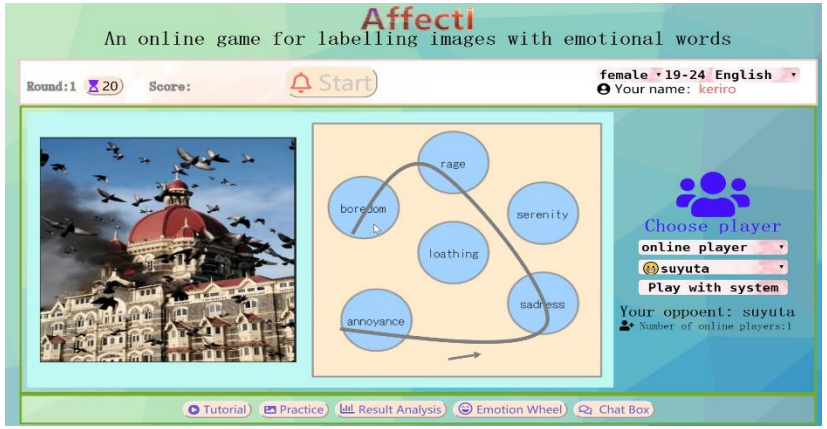
Emotion distribution

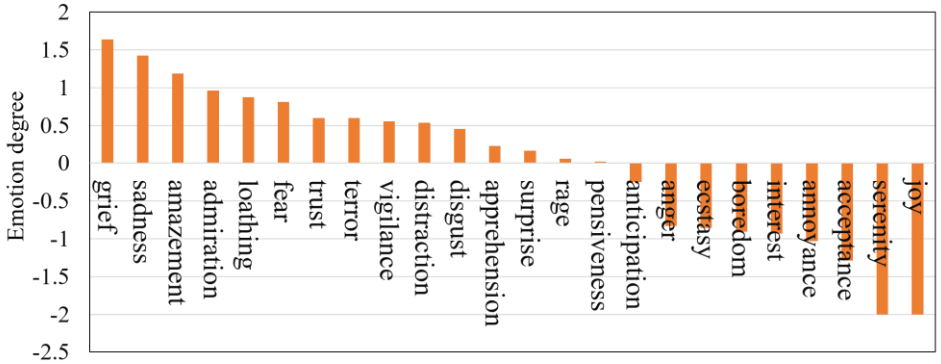


Different degree

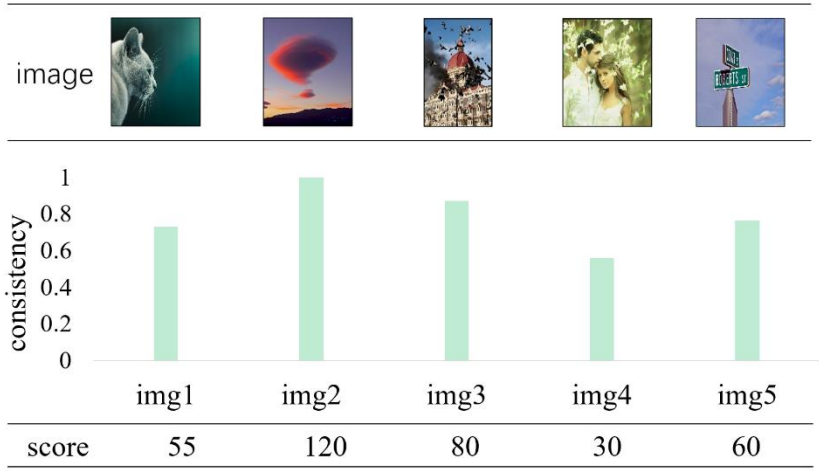
Different category

Proposed Method——Three Key Ideas

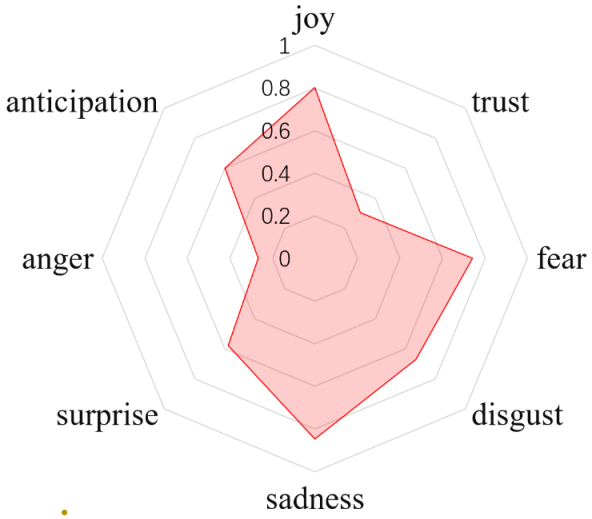




Selection mechanism



Estimation mechanism

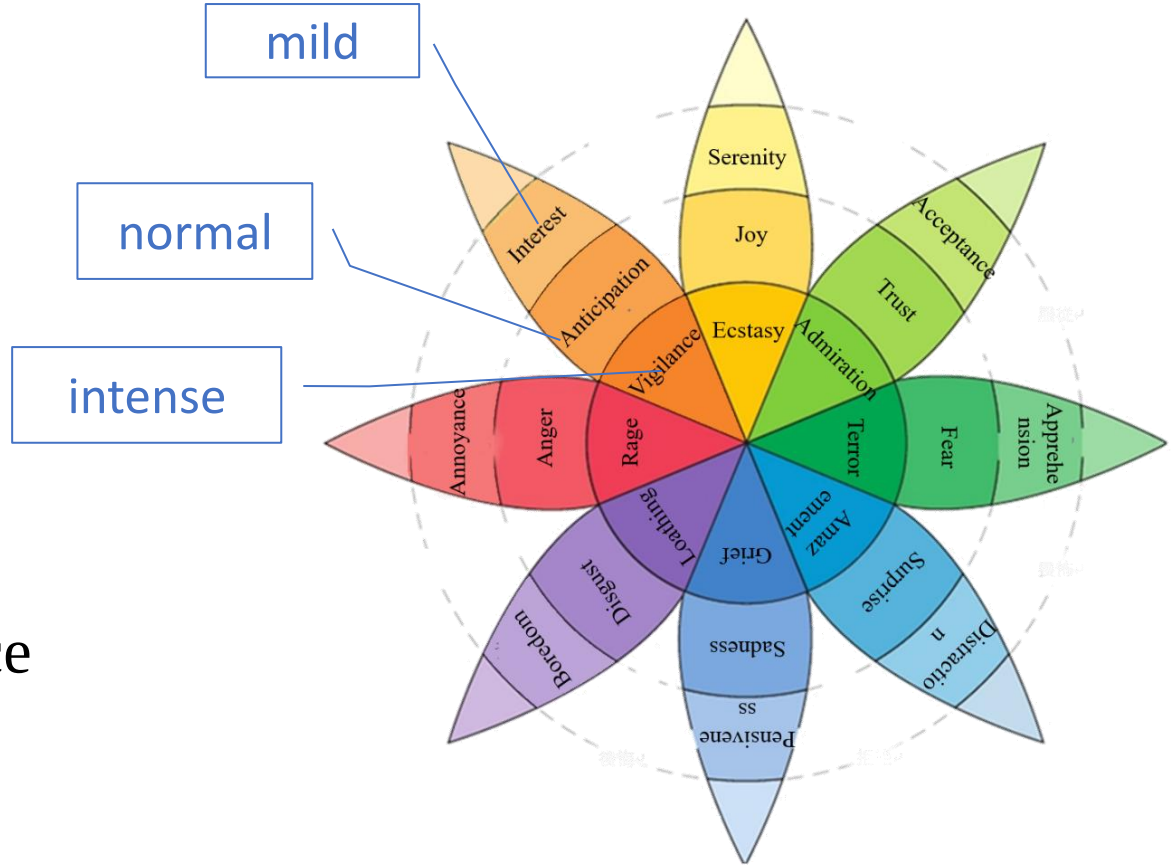


Incentive mechanism

Proposed Method——Selection Mechanism

8 categories × 3 scales

intense	**normal**	**mild**
ecstasy	joy	serenity
admiration	trust	acceptance
	.	
	.	
	.	



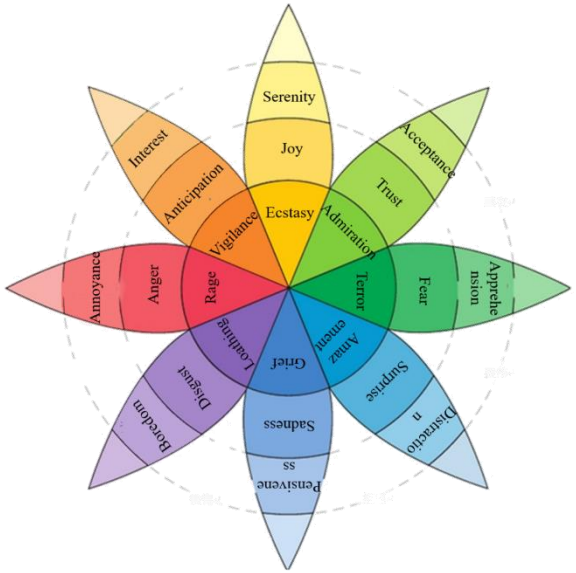
Plutchik's Wheel of emotions

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In Theories of emotion. Elsevier, 3–33.

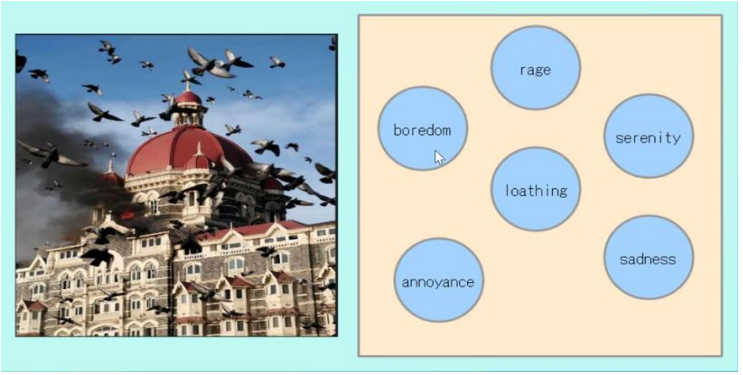
Proposed Method——Selection Mechanism

The total number of pairwise comparisons of 24 words is quite large

How to show the words?



Miller's law



The Magical Number 7 + - 2

The number of objects an average human can hold in short-term memory is 7 + - 2.

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review. 63(2): 81-97

Proposed Method——Selection Mechanism

Our selection mechanism is a hybrid strategy that makes a trade-off between exploitation and exploration.

- **Exploitation**

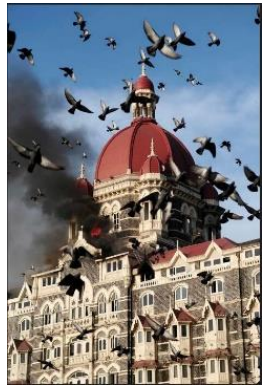
  Reconfirm the emotional words that have been labeled to the image more times to obtain a more accurate evaluation.

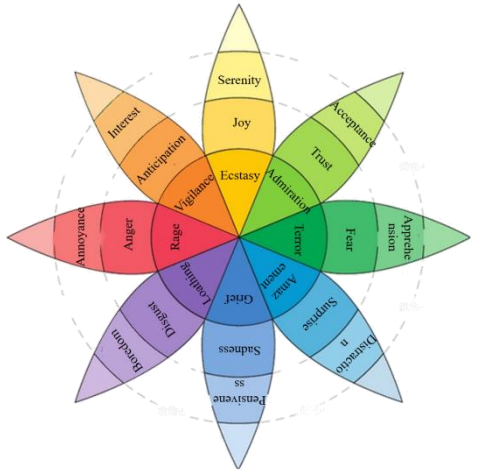
- **Exploration**

  Explore other emotional words that are labeled to the image fewer times. Maybe there have accurate evaluation words in other emotional words.

Proposed Method——Selection Mechanism

Ensures all emotion words being fairly evaluated.



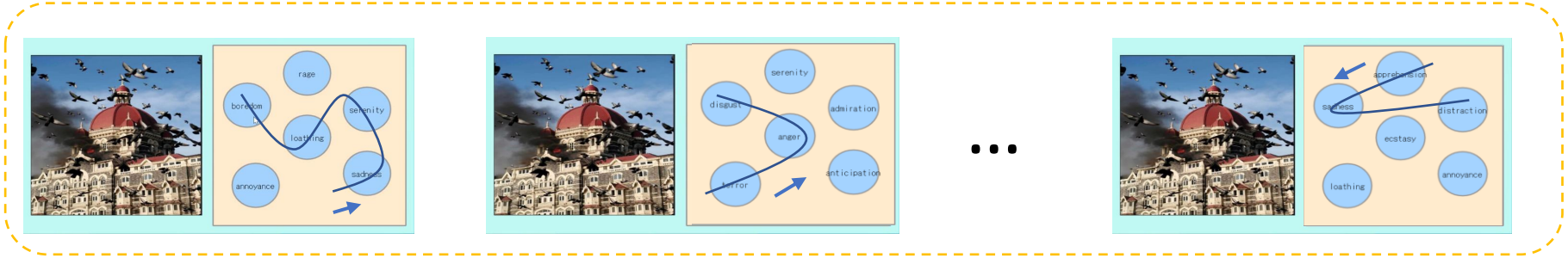


selected times	displayed times	ρ
Se_1	d_1	$\frac{Se_1}{d_1}$
…	…	…
Se_i	d_i	$\frac{Se_i}{d_i}$
…	…	…
Se_{24}	d_{24}	$\frac{Se_{24}}{d_{24}}$

Two: $p > 0.65$

Four: $\min \rho$

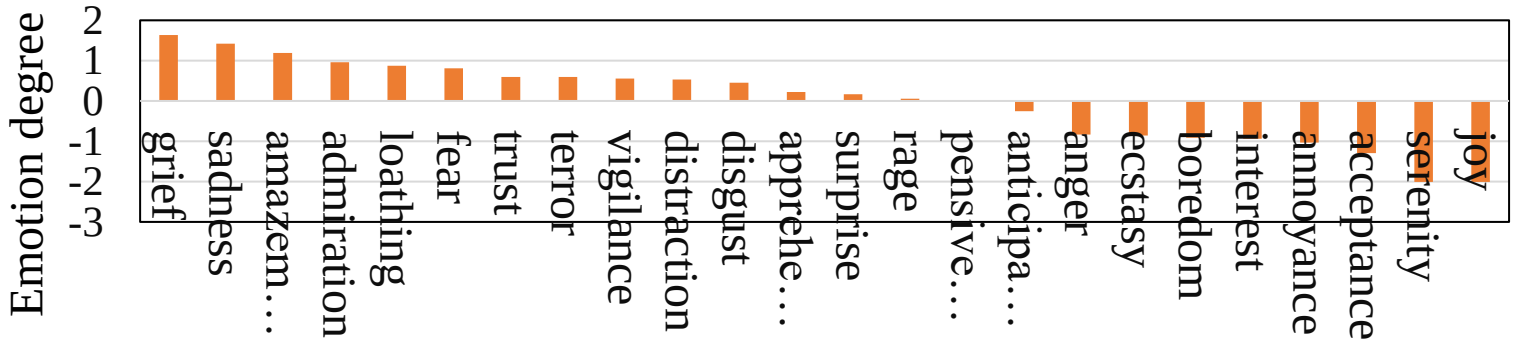
Proposed Method——Estimation Mechanism



A probability model

Relative results

Absolute results



Proposed Method——Estimation Mechanism

Bradley–Terry model

Estimate the emotion degree s of all emotional words from partial pairwise comparison labels, by minimizing the following objective function:

$$\mathcal{L} = -\sum_{ij} \log(q_{ij}p_{ij} + (1 - q_{ij})(1 - p_{ij}))$$

s_i : the emotion degree of word w_i.

q_{ij}: the observed probability that word w_i precedes word w_j

p_{ij}: the predicted possibility that word w_i precedes word w_j

$$p_{ij} = \frac{e^{s_i}}{e^{s_i} + e^{s_j}} = \frac{1}{1 + e^{-(s_i - s_j)}}$$

The objective function can be rewritten as:

$$\mathcal{L} = -\sum_{ij} \log(q_{ij}\frac{1}{1 + e^{-(s_i - s_j)}} + (1 - q_{ij})\frac{1}{1 + e^{-(s_j - s_i)}})$$

Estimate the Emotion Degree

Image I

Select result

Words: [boredom, rage, serenity, sadness, annoyance, loathing]
Order: O^1 [4, 0, 2, 1, 0, 3]

Words: [disgust, serenity, admiration, anticipation, terror, anger]
Order: O^2 [1, 0, 0, 0, 3, 2]

...

Words: [sadness, apprehension, distraction, annoyance, loathing, ecstasy]
Order: O^k [2, 1, 3, 0, 0, 0]

partial rank list

ecstasy,...,grief,...,interest

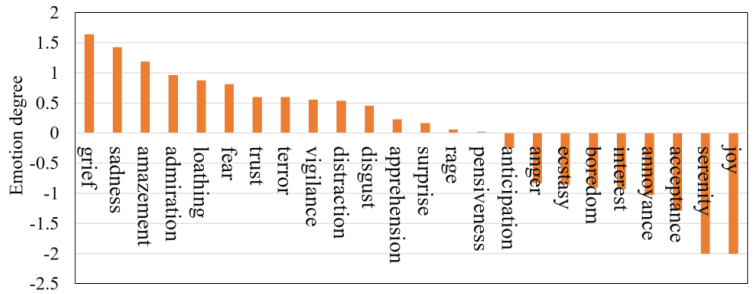
$$\begin{array}{c} \text{ecstasy,} \\ \text{...,} \\ \text{grief,} \\ \text{...,} \\ \text{interest} \end{array} \begin{bmatrix} C_{1,1} & \cdots & C_{1,j} & \cdots & C_{1,24} \\ \vdots & \ddots & \vdots & & \vdots \\ C_{i,1} & & C_{i,j} & & C_{i,24} \\ \vdots & & \vdots & \ddots & \vdots \\ C_{24,1} & \cdots & C_{24,j} & \cdots & C_{24,24} \end{bmatrix}$$

C_{ij}: the number of $w_i \succ w_j$

$w_i \succ w_j$, the order of word i is better than that of word j

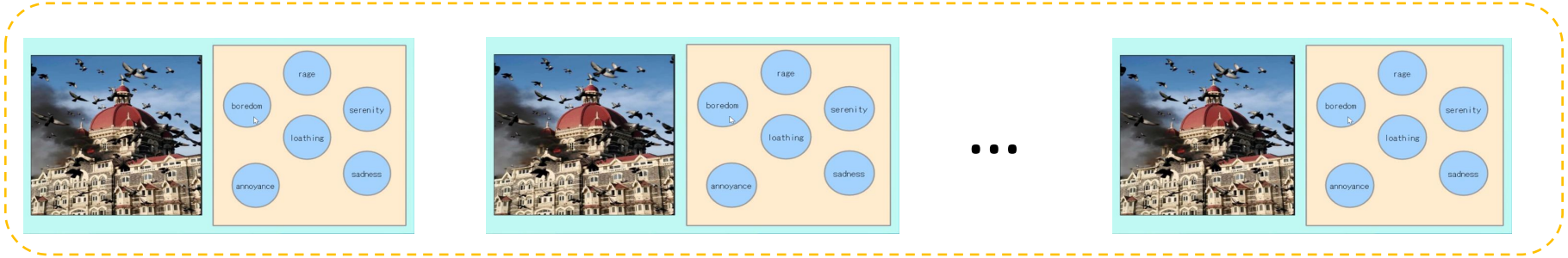
$$p(w_i \succ w_j) = \frac{e^{s_i}}{e^{s_i} + e^{s_j}} = \frac{1}{1 + e^{-(s_i - s_j)}}$$

s_i : the emotion degree of word w_i.



What is the Matrix C?

All partial rank lists will be transformed into pairwise preference comparison.



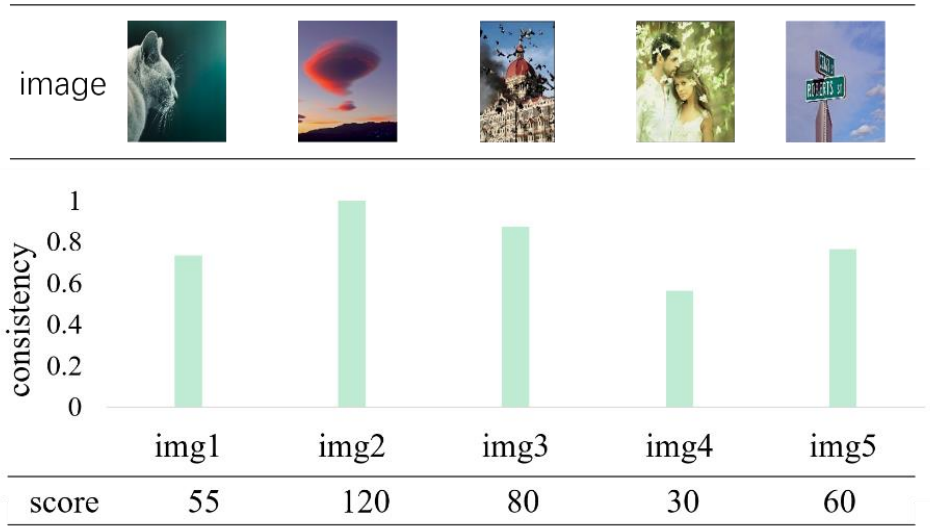
The pairwise preference comparison $w_i \succ w_j$ means the order of the word i is better than the word j.

For each subset, if the word i is labeled for the image, but the word j is not labeled, it also means the order of word i is better than word j.

Proposed Method——Incentive Mechanism

Comparison between the current player and her opponent

To encourage the player to provide labels with high quality.



Opponent-based Incentive (OI)

Current player 1

Select order O^1

$[0, 2, 1, 0, 3, 4]$

Image: I

words: $[rage, serenity, sadness, annoyance, loathing, boredom]$

Current player 2

Select order O^2

$[6, 3, 2, 1, 4, 5]$

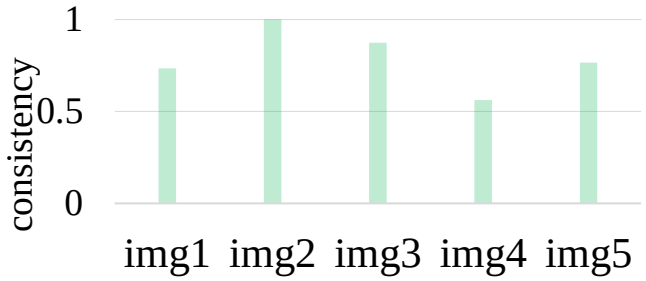
Spearman's rank correlation coefficient

$$\rho = 1 - \frac{6\sum d_l^2}{n^3 - n}$$

$$d_l^2 = \left(O_l^1 - O_l^2\right)^2$$

n: sample size $(n = 6)$

Visualization

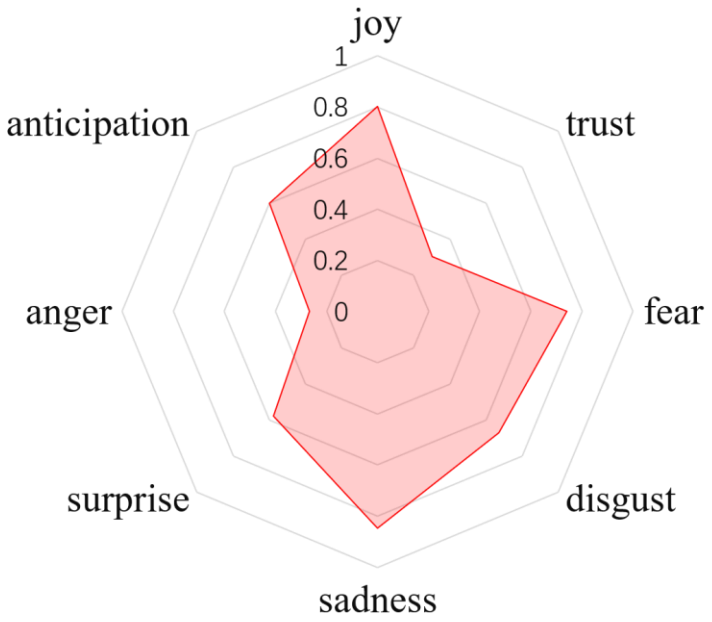


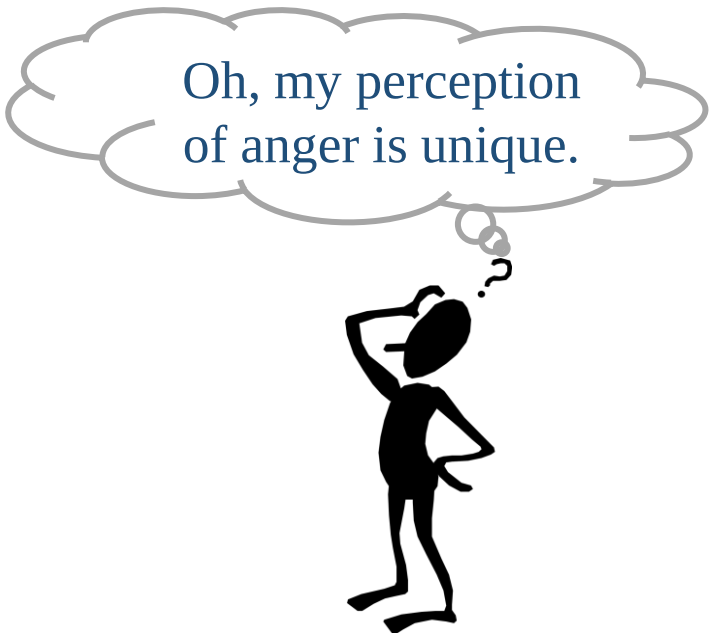
Degree of matching with players

Proposed Method——Incentive Mechanism

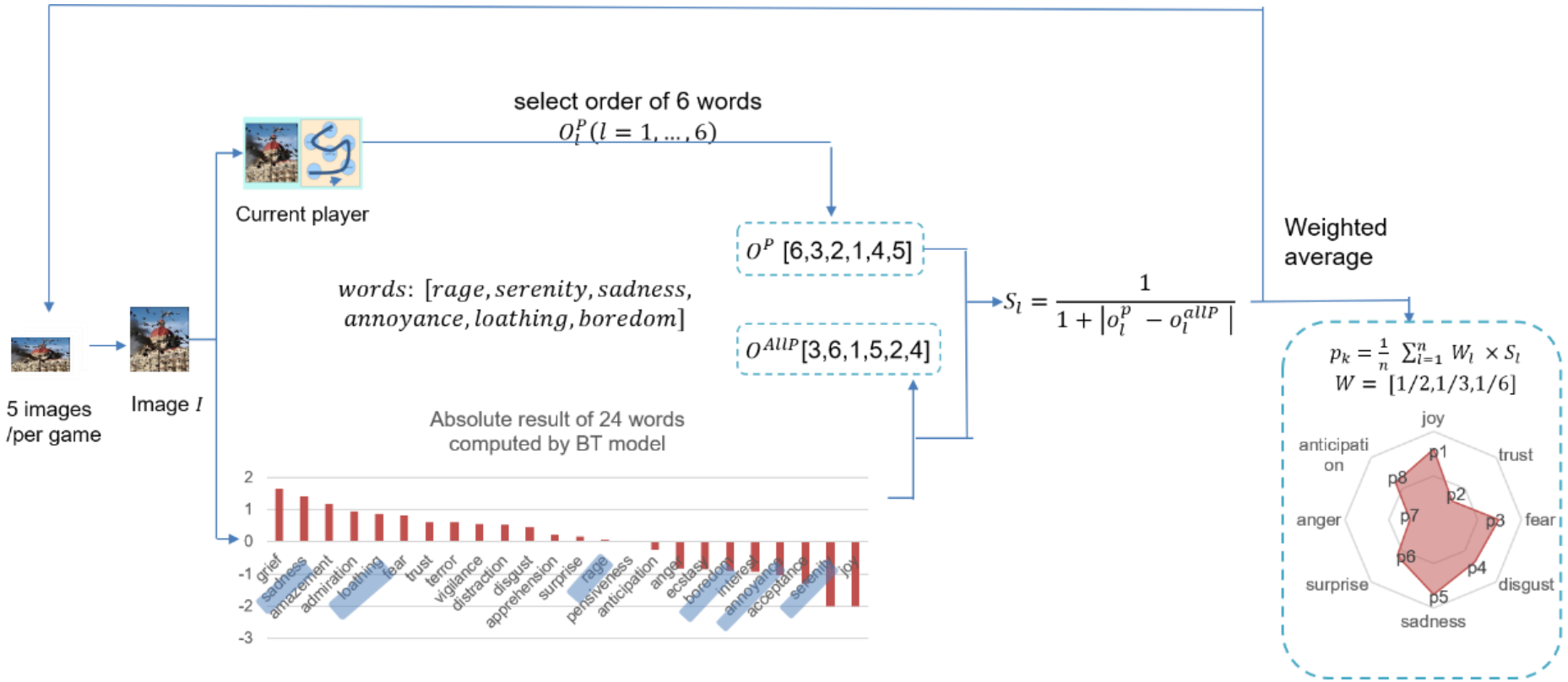
Comparison between the current player and all past players

To encourage more players to choose the labels representing their own subjective emotional perception.





Past Players-Based Incentive (PPI)



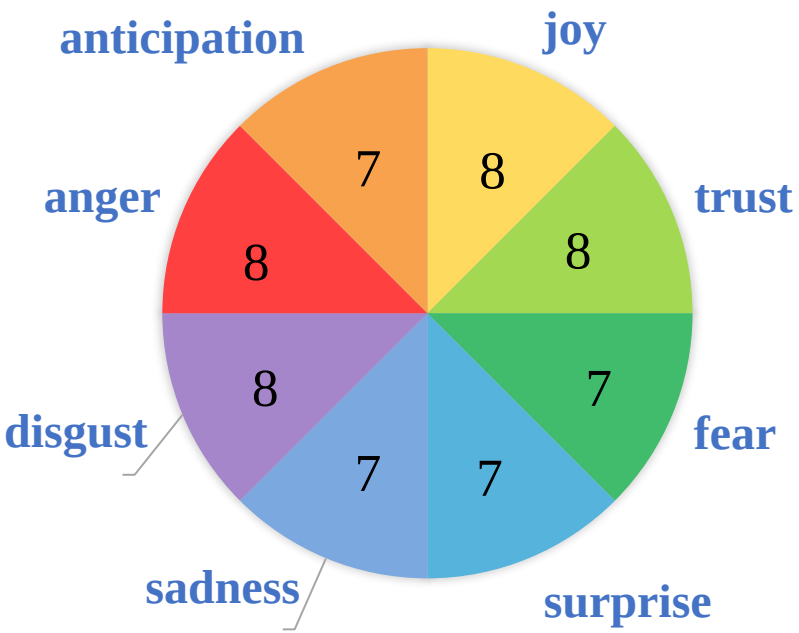
Contents

Experiment——Dataset

60 images from MVSO dataset (sub_MVSO)

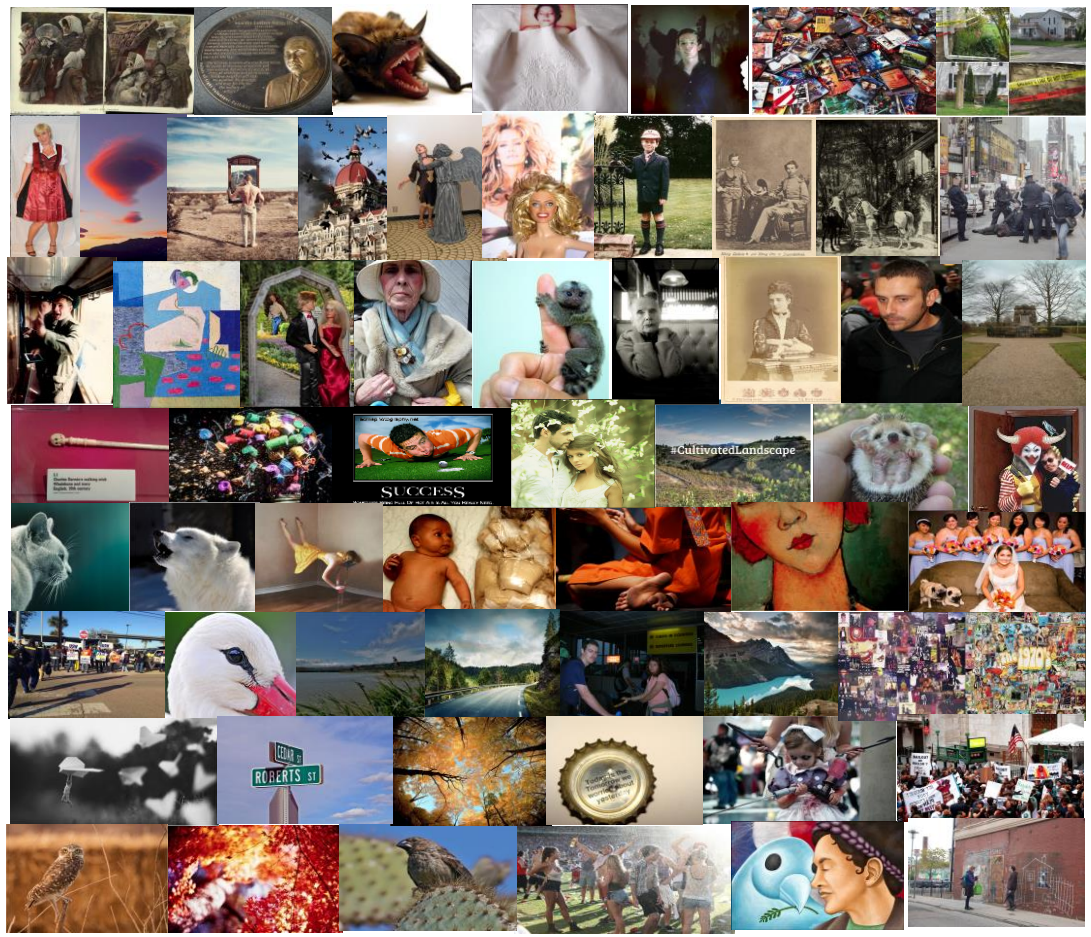
Each category: 7~8 highly viewed images on flicker

Each strength level: 1~3 images



Brendan Jou,et al. "Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology", In Proceedings of the 23rd ACM international conference on Multimedia (2015).

Experiment——Results



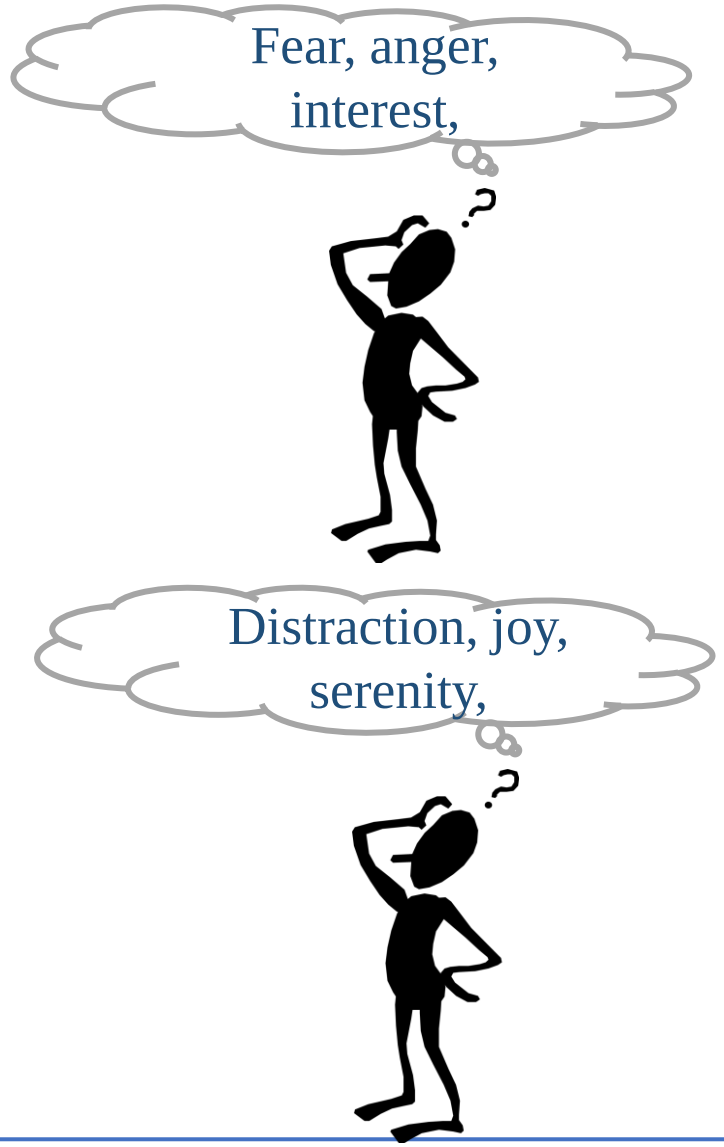
- “Our-PPI”: without PPI
- “Our+PPI”: with PPI

Number of images	60	
Participants	Age: 22-35	
	Number of unique player IDs	Our-PPI: 163
		Our+PPI: 67
Labeled times	Our-PPI: 1,892	
	Our+PPI: 710	

Evaluation——What to Evaluate

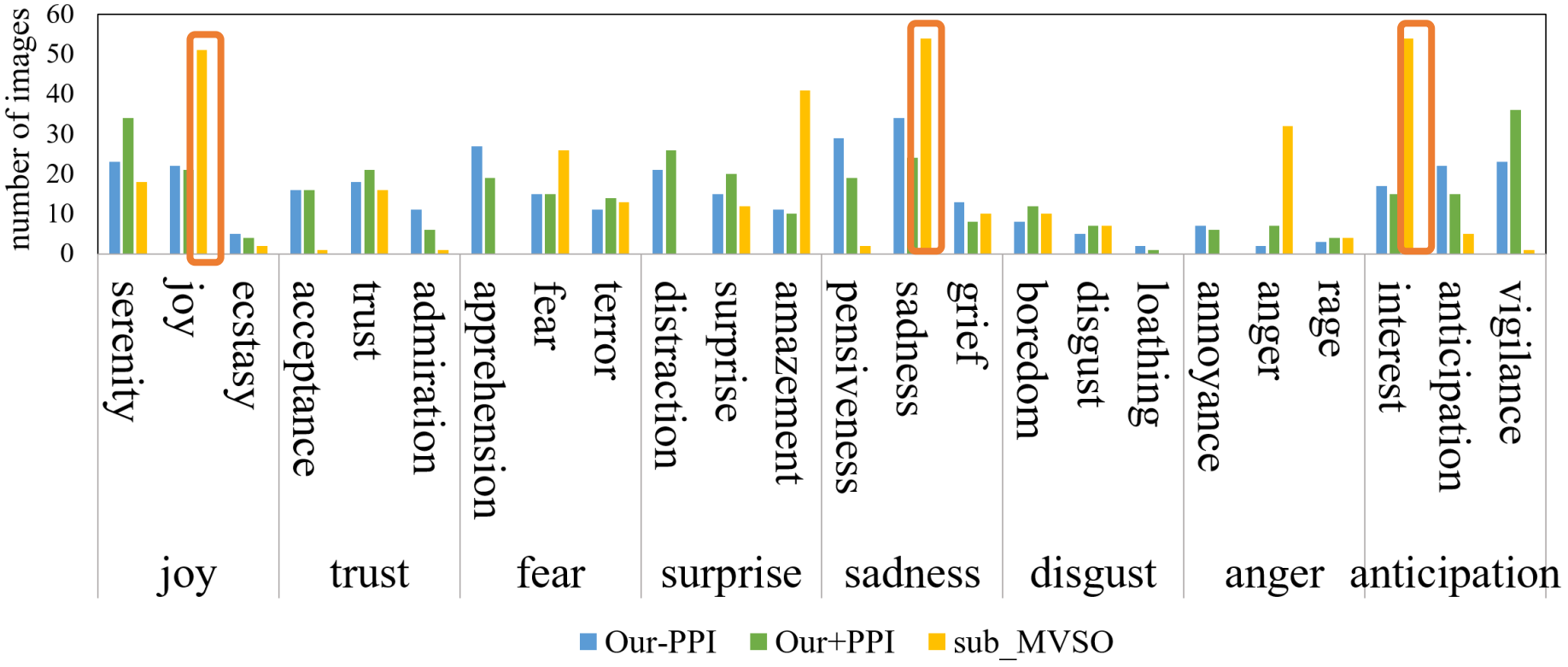
Evaluation——Diversity

Can the system successfully collect diverse emotional labels with different emotion degrees compared to the existing work?



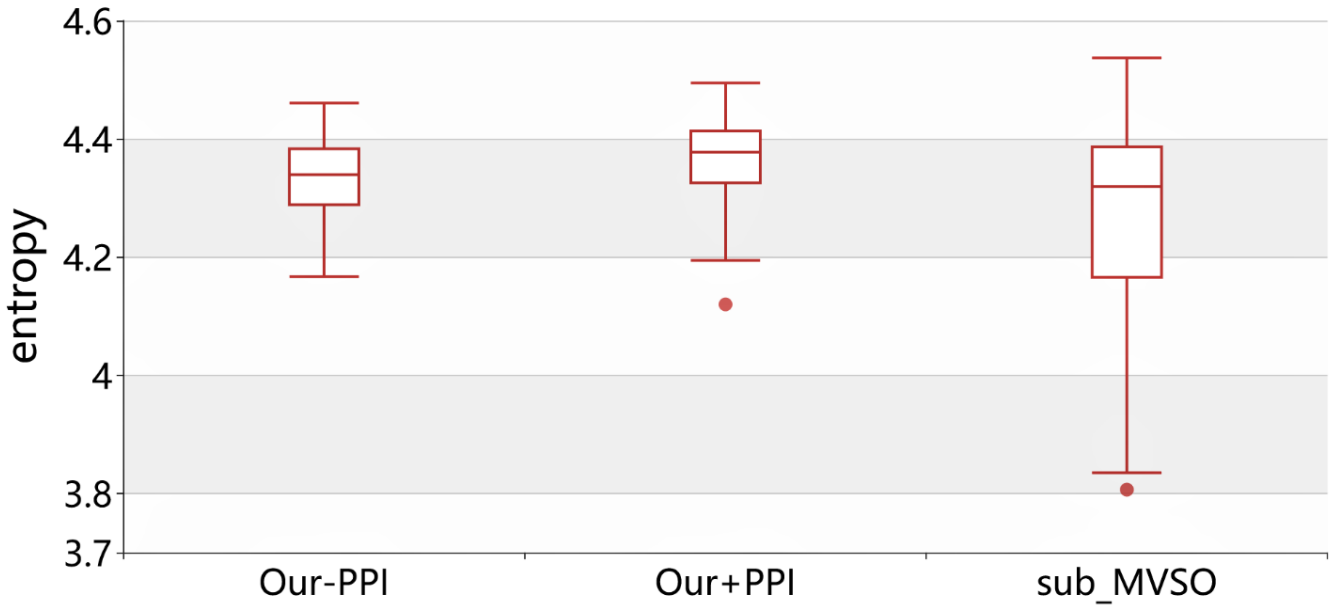
Evaluation——Diversity

Comparing the frequency of the top-6 emotion words annotated for 60 images.



Evaluation——Diversity

An entropy-based measurement.



The average entropy of the distribution of the 24 emotions for 60 images.

Evaluation——Reliability

What is the quality of the diverse labels collected by our system?

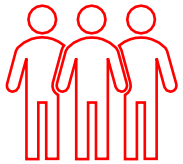
sadness
amazement
joy
interest
serenity
anger

joy
anticipation
interest
admiration
serenity
trust

Existing work

agree

Our work



agree

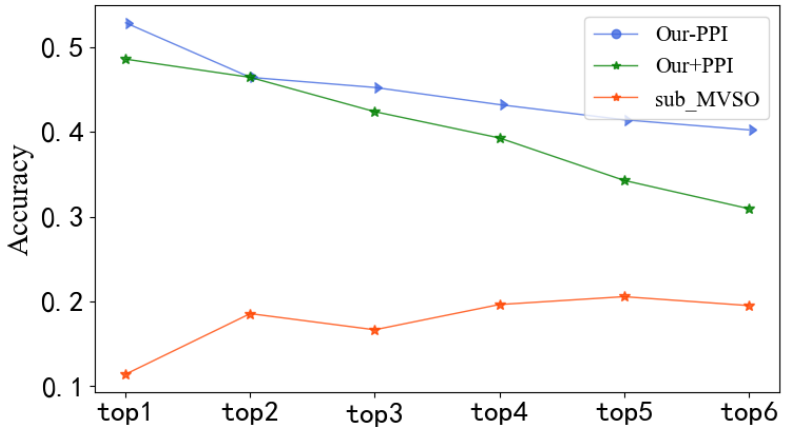
Evaluation——Reliability

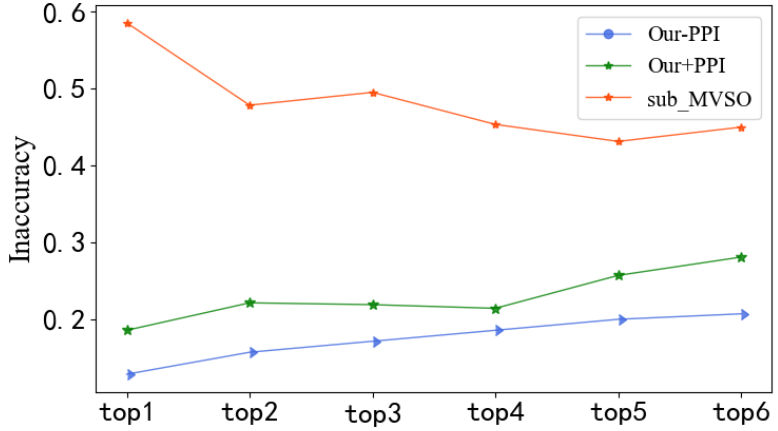
Images: 10,

Evaluator: 7 participants /per group, 3~4 images/per group,

Words: top-6 emotion words obtained by all system,

Judgments: inaccurate, neutral, and accurate





Comparison of the average votes of accurate and inaccurate for the top-k ($k \leq 6$) emotional labels.

Evaluation——User Experience

What is the player experience when using AffectI?

Evaluation——User Experience

Manual labeling experiment involving 28 participants for comparison.

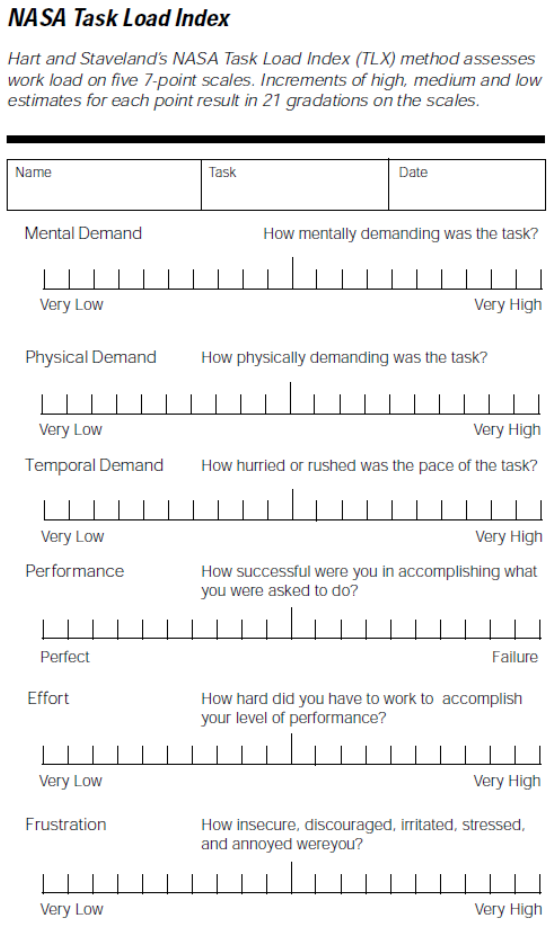
Images: 15 images / per group
Participants: 7 / per group

Q: Please select the emotional word that you think the image best represents or the emotional word caused by the image.

☐ Ecstasy	☐ Joy	☐ Serenity
☐ Admiration	☐ Trust	☐ Acceptance
☐ Terror	☐ Fear	☐ Apprehension
☐ Amazement	☐ Surprise	☐ Distraction
☐ Grief	☐ Sadness	☐ Pensiveness
☐ Loathing	☐ Disgust	☐ Boredom
☐ Rage	☐ Anger	☐ Annoyance
☐ Vigilance	☐ Anticipation	☐ Interest

Evaluation——User Experience

Comparison with the results of 15 randomly selected AffectI players



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?
Very Low — Very High

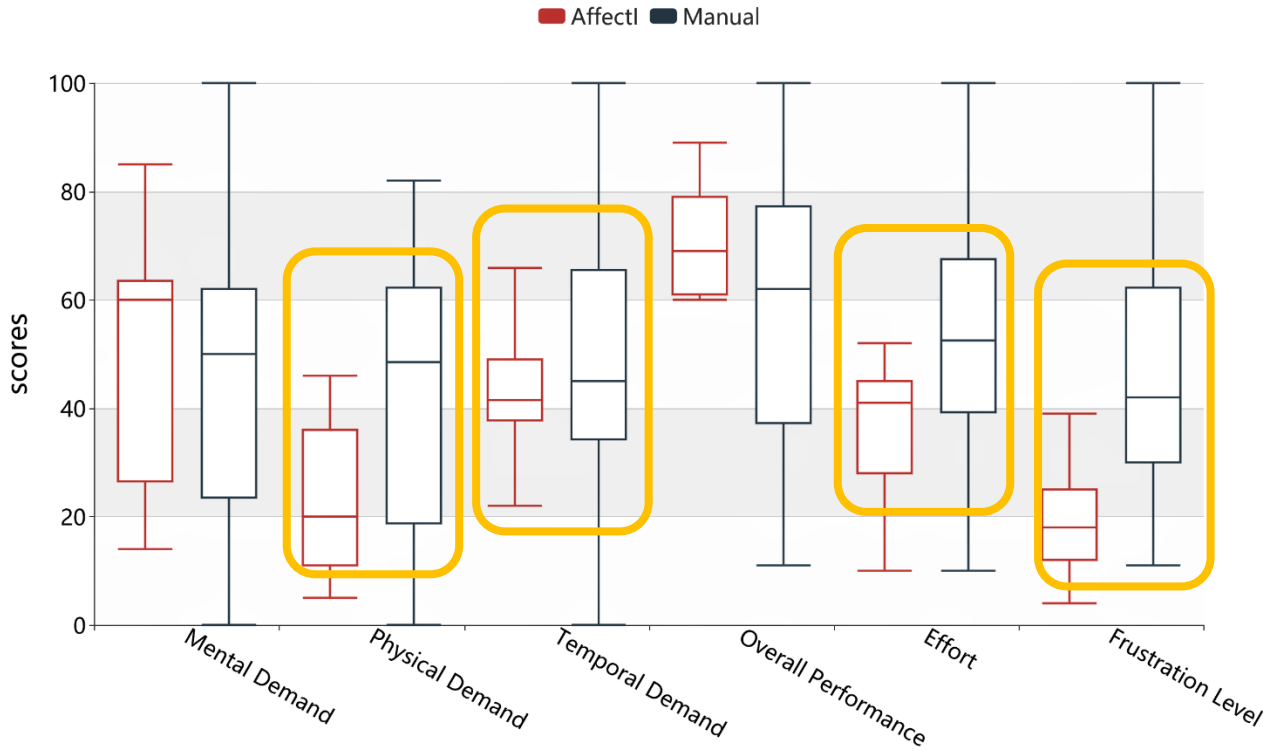
Physical Demand — How physically demanding was the task?
Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?
Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?
Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?
Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?
Very Low — Very High



NASA-TLX evaluation results

Contents

Conclusion & Future Work

Conclusion

We proposed a novel affective image annotation system, AffectI, which can efficiently collect high-quality and diverse emotional image dataset.

Future Work

- To enlarge the dataset
- To estimate the personalized emotion degrees of players
- To analyze how emotion perception on images is related to other factors such as age, gender, and language.

HANGZHOU
DIANZI
UNIVERSITY

UNIVERSITY
OF
YAMANASHI

Thanks for your attention!