

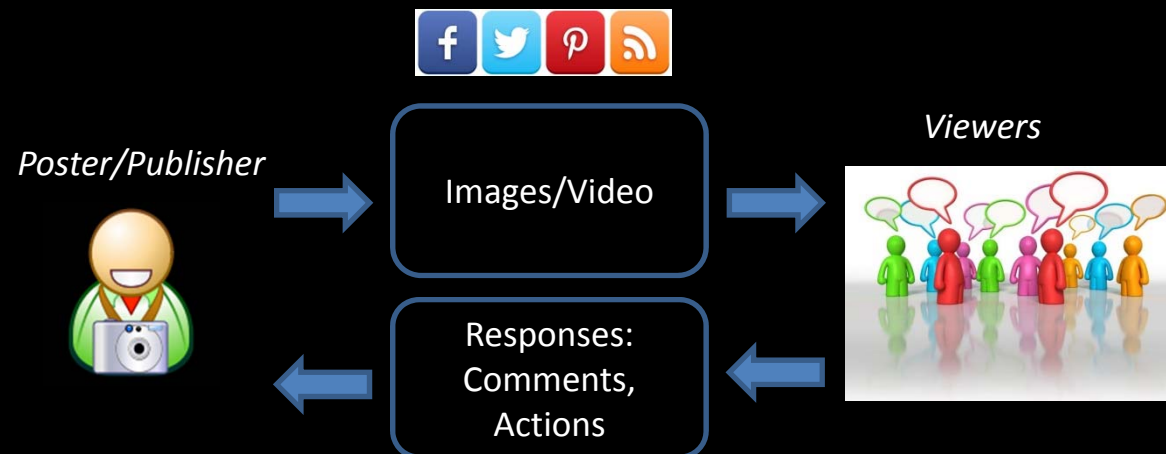
# Assistive Image Comment Robot

Teaching Machine to Suggest Social Comments  
Relevant to Image Content

*Shih-Fu Chang*

Digital Video Multimedia Lab, Columbia University

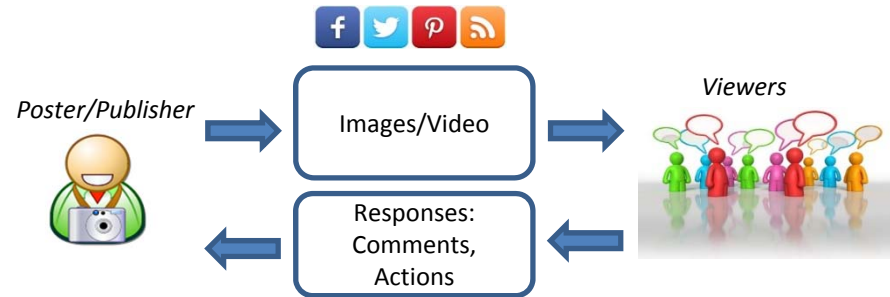
ECCV, September 2014



# Story Sharing -> Opinion Responses



# Images Evoke Emotional Responses



wow... looks **awesome** in large!!



The illumination of colors are so **natural** and **beautiful**. like this picture



lost cat posters always make me cry - I'd be **terrified** if one of my kitties went missing :-)



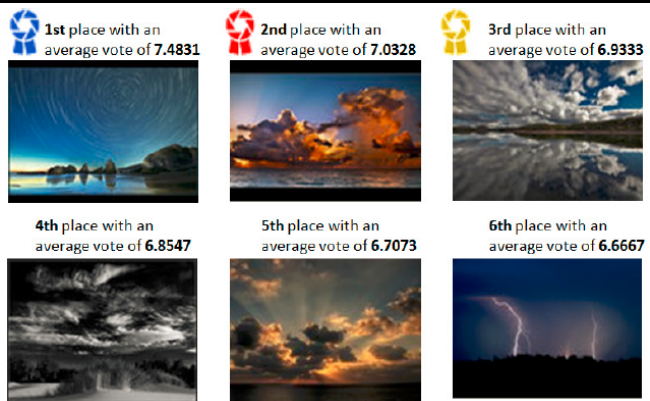
Looks **delicious**, I **love** lassi

# Goals: Asssitive Social Communication



- Facilitate Social Communication
- Validate computer abilities in
  - image content recognition
  - Detect cues of response evoking sentiment/emotion
  - comment synthesis





(CVPR 2014 Tutorial)

Aesthetics

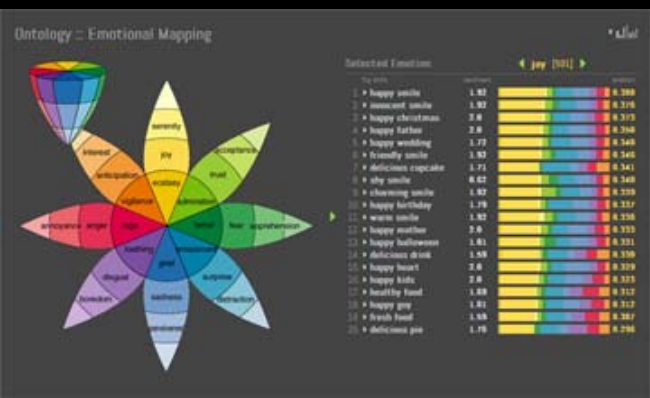
Interestingness

Beyond Semantics

Emotion

Style

Others:  
Creativity,  
Intent,  
Memorable ...



# References

## Visual Affect

- Machajdik, Jana, and Allan Hanbury. "Affective image classification using features inspired by psychology and art theory." In *ACM Multimedia*, 2010.

## Visual Emotion and Sentiment

- Borth, Damian, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. "Large-scale visual sentiment ontology and detectors using adjective noun pairs." In *ACM Multimedia*, 2013.
- Chen, Yan-Ying, Tao Chen, Winston H. Hsu, Hong-Yuan Mark Liao, and Shih-Fu Chang. "Predicting Viewer Affective Comments Based on Image Content in Social Media." In *ACM ICMR*, 2014.

## Visual Aesthetics

- Datta, Ritendra, Dhiraj Joshi, Jia Li, and James Z. Wang. "Studying aesthetics in photographic images using a computational approach." In *ECCV*, 2006.
- Naila Murray, De Barcelona, Luca Marchesotti, and Florent Perronnin. AVA: A Large-Scale Database for Aesthetic Visual Analysis. In *CVPR*, 2012.

## Visual Aesthetics, Interestingness, Memorability

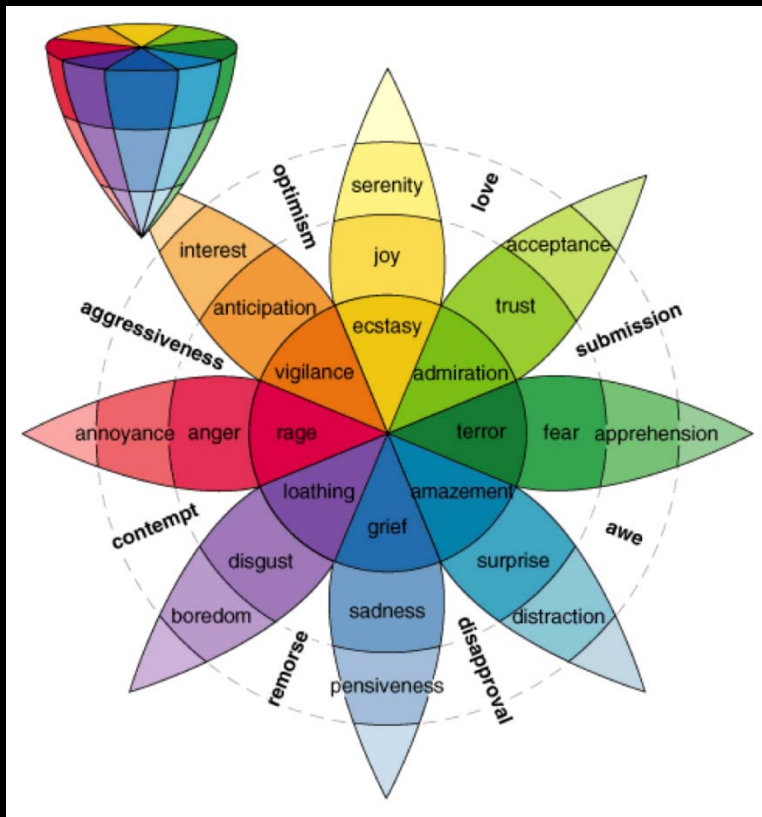
- S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *CVPR*, 2011.
- Gygli, Michael, Helmut Grabner, Hayko Riemenschneider, Fabian Nater, and Luc Van Gool. "The interestingness of images." In *ICCV*, 2013.
- Bhattacharya, Subhabrata, Behnaz Nojavanasghari, Tao Chen, Dong Liu, Shih-Fu Chang, and Mubarak Shah. "Towards a comprehensive computational model for aesthetic assessment of videos." In *ACM Multimedia*, 2013.
- Isola, Phillip, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. "What makes an image memorable?." In *CVPR*, 2011.

## Image Style

- Karayev, Sergey, Aaron Hertzmann, Holger Winnemoeller, Aseem Agarwala, and Trevor Darrell. "Recognizing Image Style." *arXiv preprint arXiv:1311.3715*(2013).

# For Content to be Viral, it Needs to be Emotional

- Dan Jones



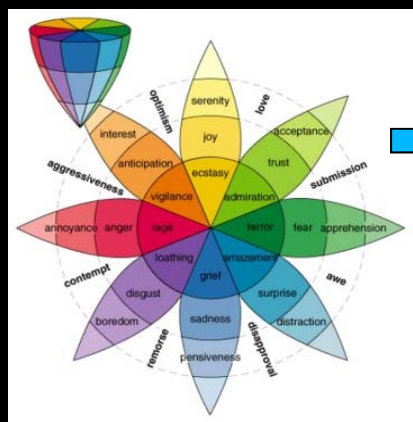
Psychology emotion wheel  
(8 emotions, by Robert  
Plutchik)



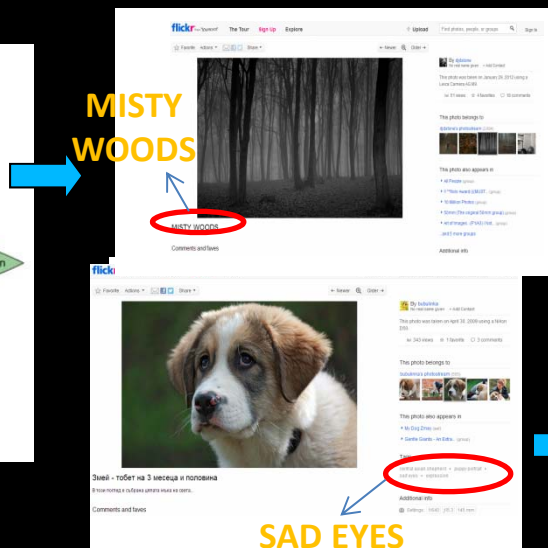
Visual concepts Evoke  
Emotions

# First, discover “emotional story” in images

-- Web mining to discover visual concepts in social images



Psychology emotion wheel (8 emotions) Robert Plutchik, '91



Analyze tags with strong sentiments

Build Sentiment Ontology



Select Concepts

Borth, Ji, Chen, Breuel, Chang, *Large-Scale Visual Sentiment Ontology*, ACM Multimedia 2013

# Concurrent tags with emotions

joy	terror	amazement	disgust
joy	terror	amazing	disgusting
happy	horror	beautiful	gross
love	zombie	nature	food
smile	fear	wonder	nasty
beautiful	dark	light	sick
flowers	street	love	dirty
light	halloween	sky	dead
nature	war	eyes	face
kids	undead	clouds	blood
christmas	bomb	landscape	insect

From 6 million tags on Flickr and YouTube  
Color code: text sentiment values



From Computer Vision Perspective:

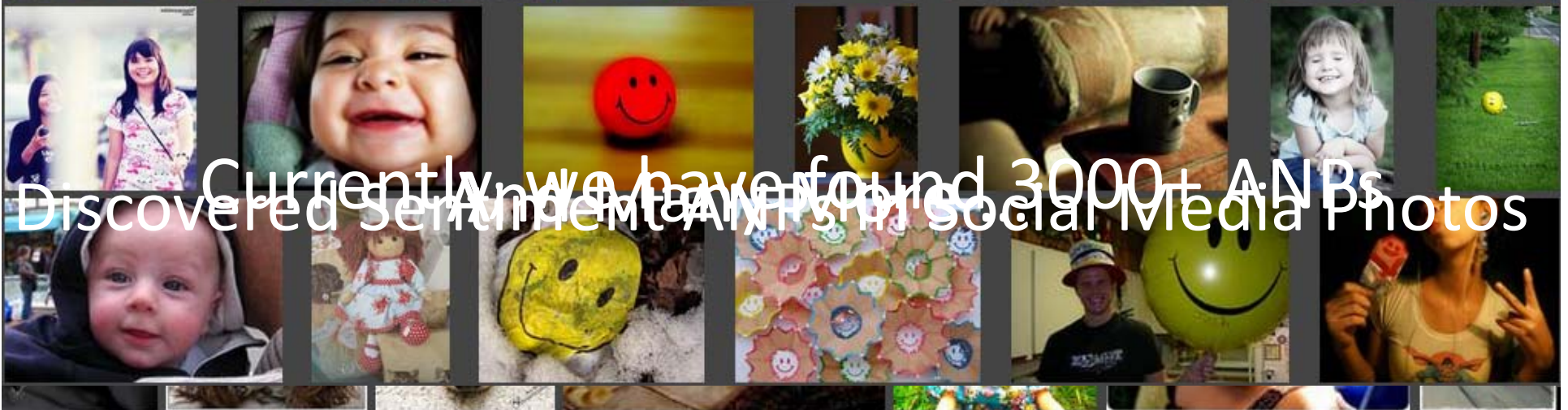
Not all concepts/entities are detectable!

-- which 1000 concepts to focus in pictures?

# Computational Focus – Adjective-Noun Pair (ANP)

- Adjective (268): express emotions
  - positive: beautiful, amazing, cute
  - negative: sad, angry, dark
- Nouns (1187): possible detection
  - people, places, animals, food, objects, weather
- Standard steps:
  - remove entities like “hot dog” via wikipedia
  - Choose sentiment rich ANP concepts by NLP tools  
“Senti-WordNet” “SentiStrength”

Slippery Bobber



# Images + Emotions in Different Languages

trust  
**interest** joy  
 fear terror surprise



(i) **english** (top: interest, joy, trust)

interés dolor miedo  
 alegría sorpresa  
 tristeza confianza



(ii) **spanish** (top: alegría, dolor, sorpresa)

joie  
**surprise**



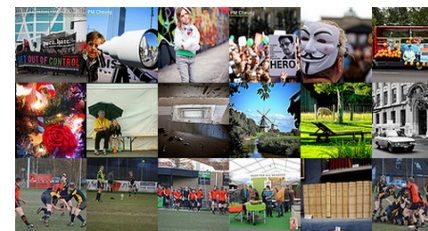
(iii) **french** (top: surprise, joie, intérêt)

興趣 憤怒  
 無聊 歡樂 悲  
 寧靜 恐怖 沉思  
 信任 驚



(iv) **chinese** (top: 恐怖, 歡樂, 无聊)

Vertrauen Langeweile  
**Freude** Terror  
 Überraschung Trauer  
 Interesse Wut



(v) **german** (top: Angst, Freude, Interesse)

الفرح  
 الحزن  
 الحزن  
 الحزن



(vi) **arabic** (top: الفرح, حزن, الحزن)

# Emotion keyword clouds (in English)

- emotion queries performed in Flickr after translating 24 emotions words from English to 15 other languages
- kept languages which have at least 100k results (search on full metadata: title, description and tags)
- word size is proportional to the number of flickr results



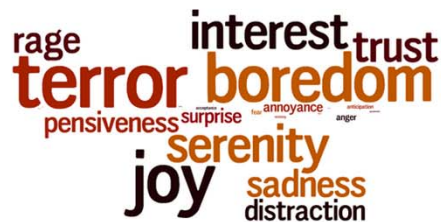
(i) **english** (top: interest, joy, trust)



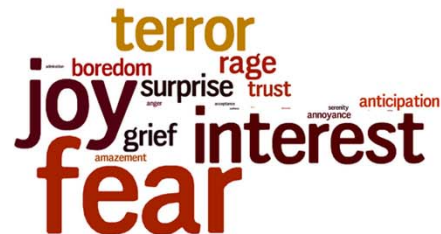
(ii) **spanish** (top: joy, grief, surprise)



(iii) **french** (top: surprise, joy, interest)



(iv) **chinese** (top: terror, joy, boredom)



(v) **german** (top: fear, joy, interest)

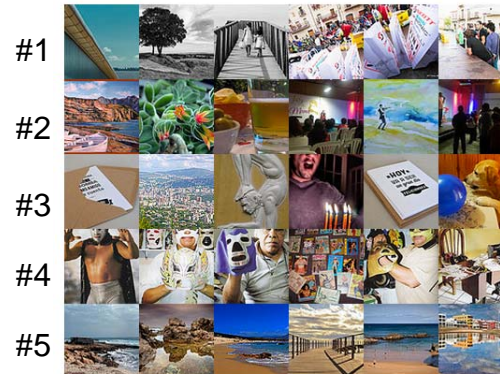
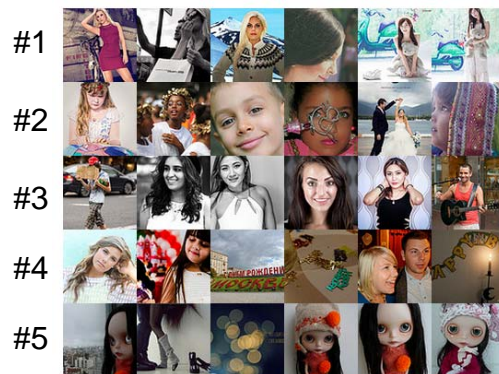


(vi) **arabic** (top: sadness, grief, joy)

# Multilingual Emotion-Related Concepts for 'joy'

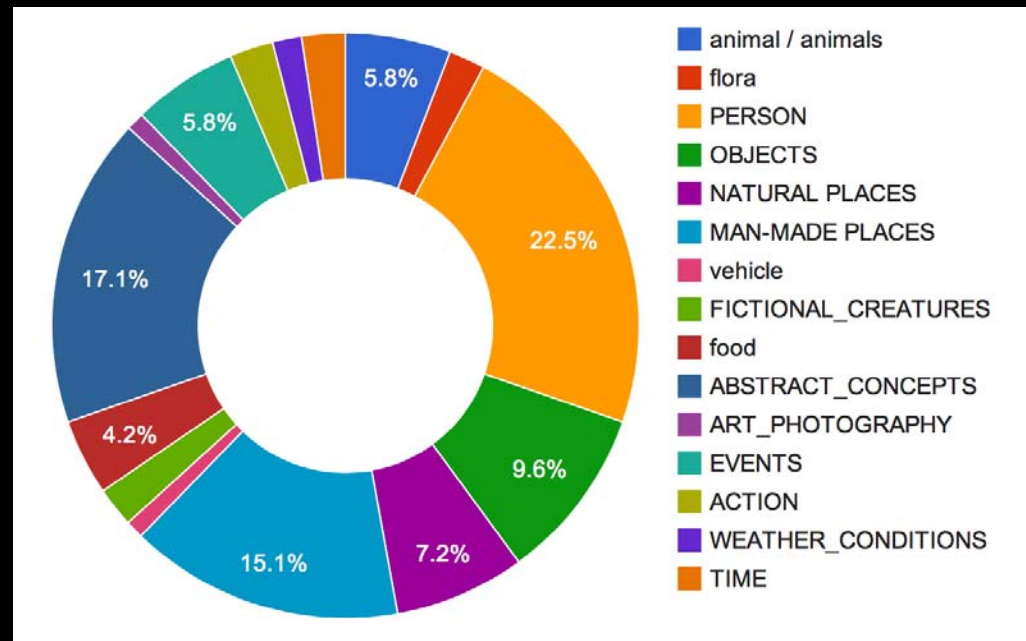
- filtering rules: non-neutral sentiment, 100 exact matches on Flickr, sorted by co-occurrence with 'joy'

	English (joy)	Spanish (alegría)	French (joie)
# 1	beautiful girl	aire libre (open air)	bonne humeur (good mood)
# 2	happy child	feliz día (happy day)	plein air (open air)
#3	beautiful smile	feliz cumpleaños (happy birthday)	belle plage (beautiful beach)
#4	happy birthday	<i>libre lucha</i> (free fight/wrestling)	bonne année (happy new year)
#5	cute love	buen tiempo (good time)	belle femme (beautiful woman)
#6	beautiful game	natural belleza (natural beauty)	belle fille (beautiful girl)
#7	happy smile	grande agua (great water)	<i>parcours ludique</i> (fun trail)
#8	happy living	simple vida (simple life)	quotidienne vie (everyday life)
#9	beautiful portrait	libre mundo (free world)	grand soleil (great sun)
#10	beautiful man	mejores amigas (best friends)	belle fleur (beautiful flower)



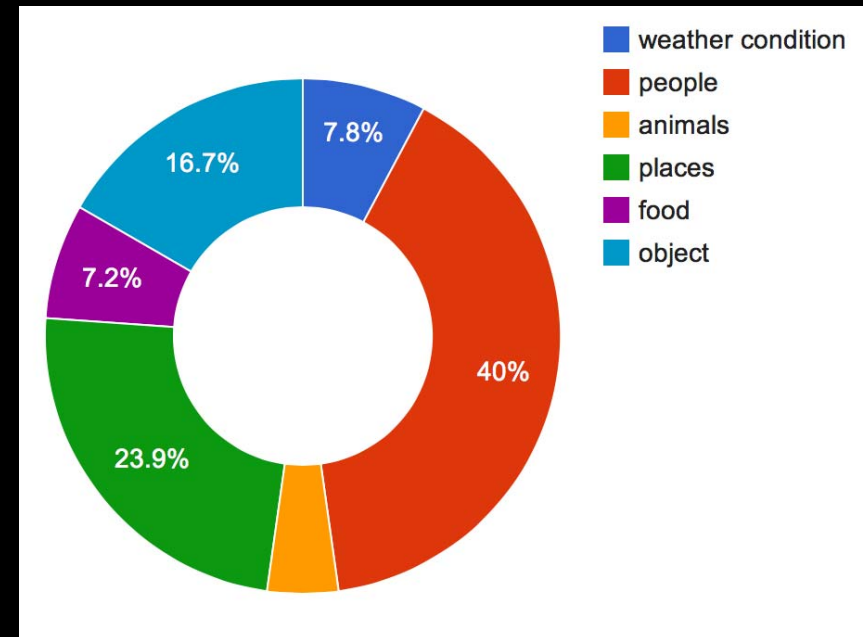
# ANP Ontology (noun)

- 6 levels
  - ANIMALS
  - FLORA
  - PERSON
  - OBJECTS
  - NATURAL PLACES
  - MAN-MADE PLACES
  - VEHICLE
  - FICTIONAL\_CREATURES
  - FOOD
  - ABSTRACT\_CONCEPTS
  - ART\_PHOTOGRAPHY
  - EVENTS
  - ACTION
  - WEATHER\_CONDITIONS
  - TIME




# ANP Ontology (adjective)

- adjectives (2 levels)
  - weather (stormy, cold, sunny)
  - people (young, attractive)
  - animals (cute, fluffy)
  - places (haunted, misty)
  - food (yummy, salty)
  - object (colorful, beautiful)

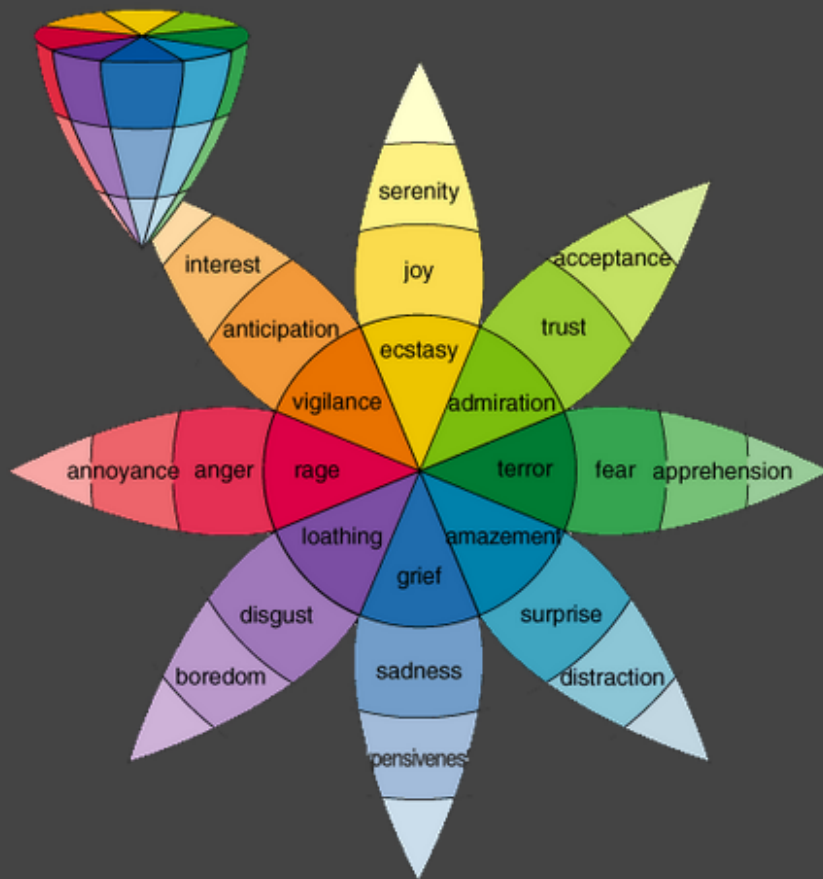


# Visual Sentiment Ontology (Browser)

Home :: Ontology :: Adjective Noun Pairs :: Downloads :: About

Visual Sentiment Ontology 

## Ontology :: Emotional Mapping




Selected Emotion:

◀ joy [591] ▶

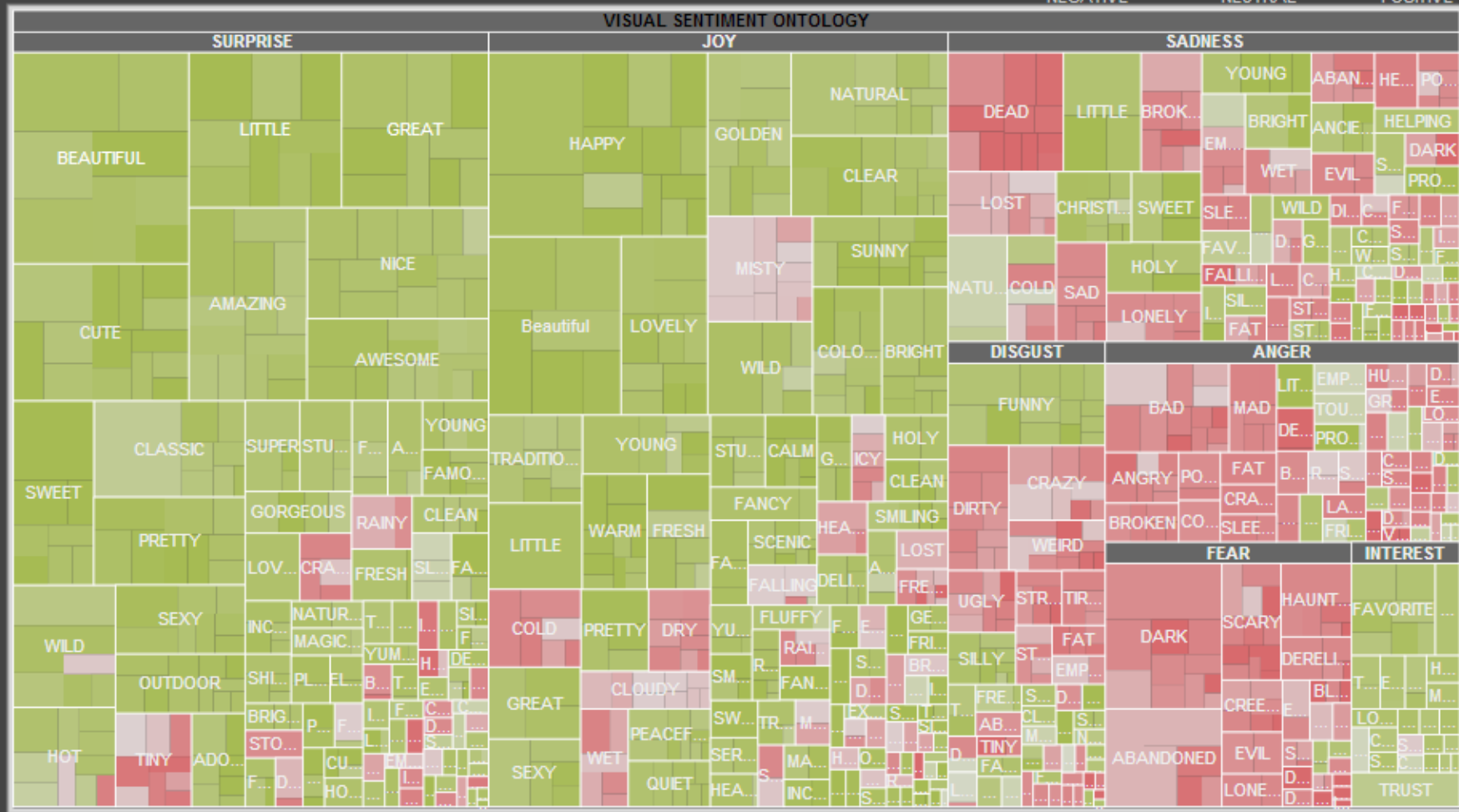
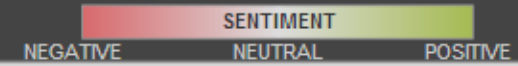
Top ANPs	sentiment	emotion
1. ▶ happy smile	1.92	0.388
2. ▶ innocent smile	1.92	0.376
3. ▶ happy christmas	2.0	0.373
4. ▶ happy father	2.0	0.358
5. ▶ happy wedding	1.72	0.348
6. ▶ friendly smile	1.92	0.346
7. ▶ delicious cupcake	1.71	0.341
8. ▶ shy smile	0.62	0.340
9. ▶ charming smile	1.92	0.339
10. ▶ happy birthday	1.79	0.337
11. ▶ warm smile	1.92	0.336
12. ▶ happy mother	2.0	0.333
13. ▶ happy halloween	1.81	0.331
14. ▶ delicious drink	1.59	0.330
15. ▶ happy heart	2.0	0.329
16. ▶ happy kids	2.0	0.323
17. ▶ healthy food	1.69	0.312
18. ▶ happy guy	1.61	0.312
19. ▶ fresh food	1.59	0.307
20. ▶ delicious pie	1.76	0.296

# Visual Sentiment Ontology (Browser)

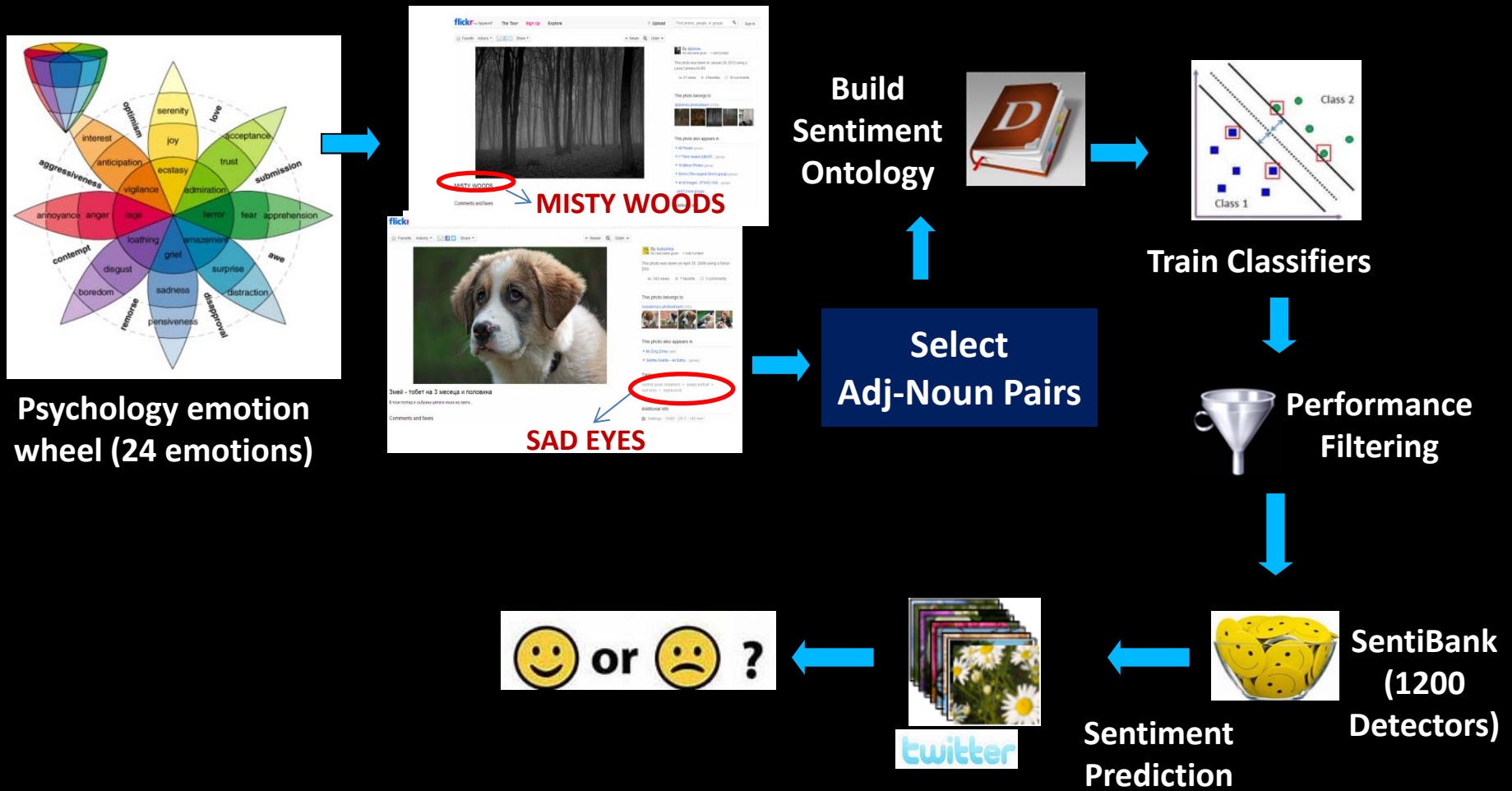
Home :: Ontology :: Adjective Noun Pairs :: Downloads :: About

Visual Sentiment Ontology 

## Ontology :: Treemap Visualization



# Next Step: Teach Machine to Recognize Visual Sentiments



# Image Features

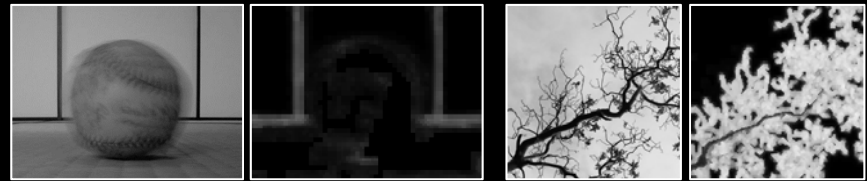
- Generic features
  - Color Histogram (3x256 dim.)
  - GIST descriptor (512 dim.)
  - Local Binary Pattern (52 dim.)
  - SIFT Bag-of-Words (1,000 codewords  
2-layer spatial pyramid, max pooling)
  - Attribute descriptor (2,000 dim.)
- Special features
  - Object detection (people, objects, etc.)
  - Aesthetics features (color schemes, layout, etc.)
  - Face and attributes
  - Improve accuracy 9%-30%

# Aesthetic Features

- Dark Channel [He et al. '09]
  - minimum of local intensity



- Sharpness [Vu et al. '11]
  - sharpness of local image regions by spectral and spatial measures



- Depth of Field
  - wavelet decomposition in HSV color space (low vs high)



- Color Harmony [Nishiyama et al. '11]
  - Using local histogram of Moon-Spencer model, which defines compatibility of two color values (example of compatible color)



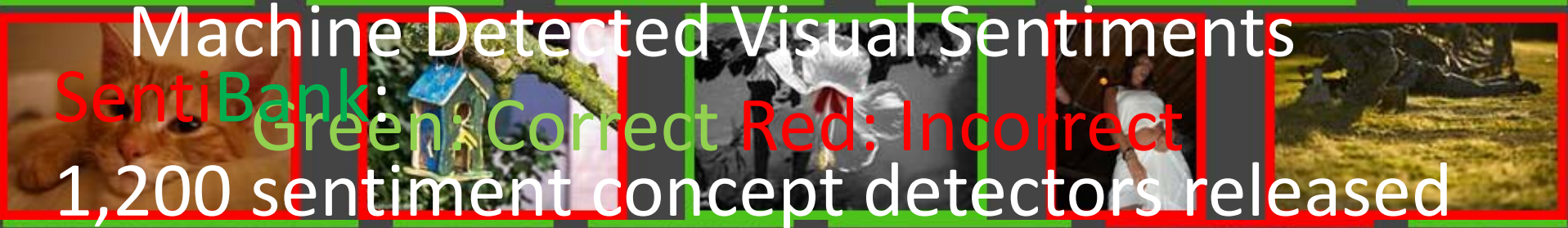
Beautiful flower  
Happy face



## Machine Detected Visual Sentiments

SentiBank: Green: Correct Red: Incorrect

1,200 sentiment concept detectors released



# What Do Humans Expect to See?

- from small annotation experiments



Jules3000@flickr



HIKARU Pan@flickr



houseofduke@flickr



masochismtango@flickr

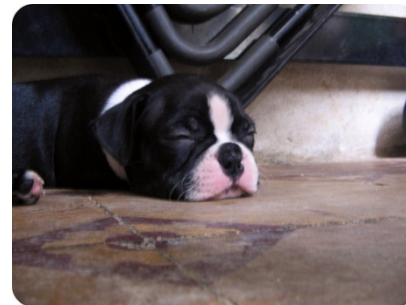
“Smiling dog”: tongue visible, mouth open, face camera, close shot, pink tongue, open mouth, frontal dog face



houseofduke@flickr



hurlham@flickr



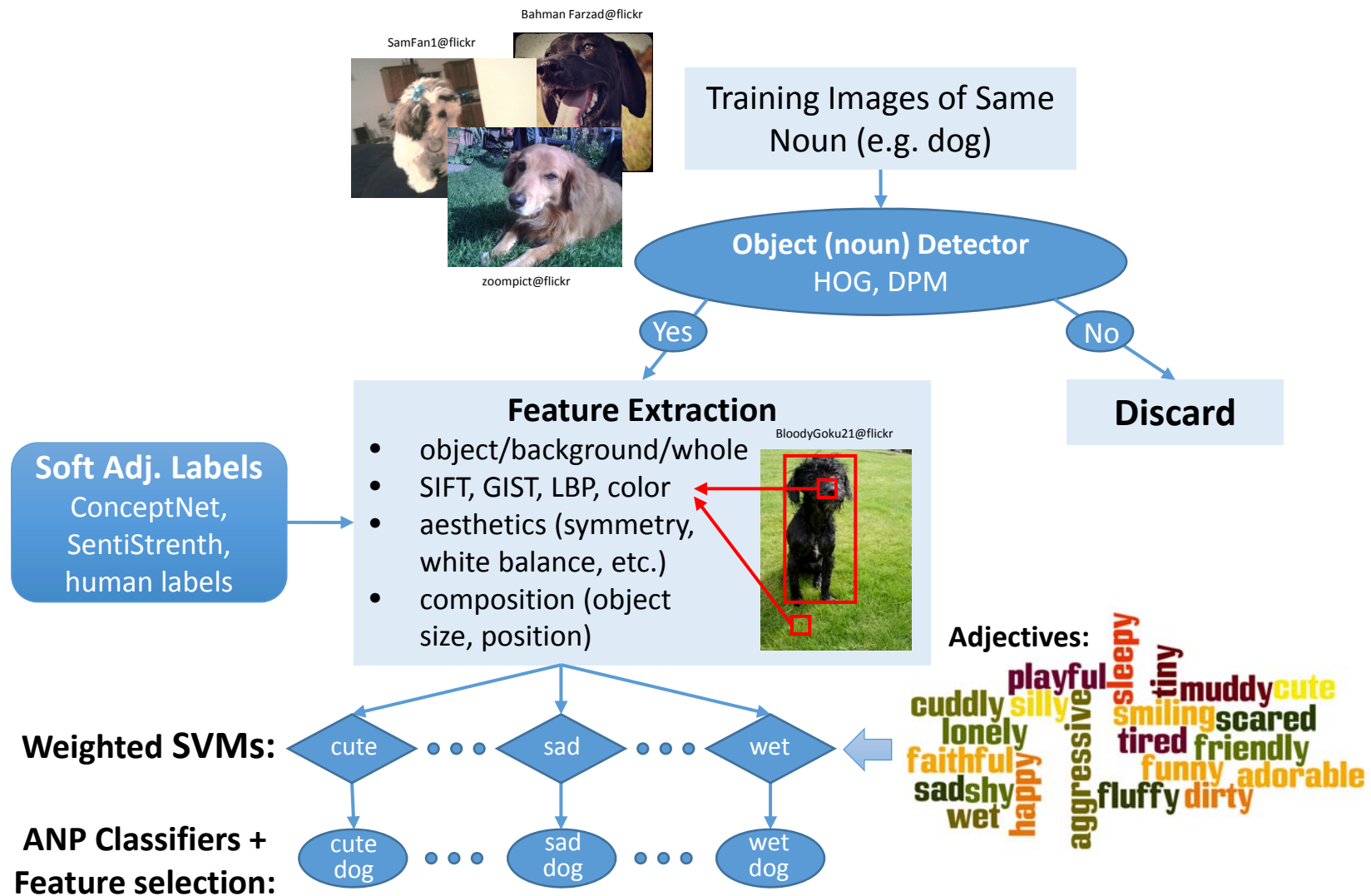
ebonique2007@flickr



springlakecake@flickr

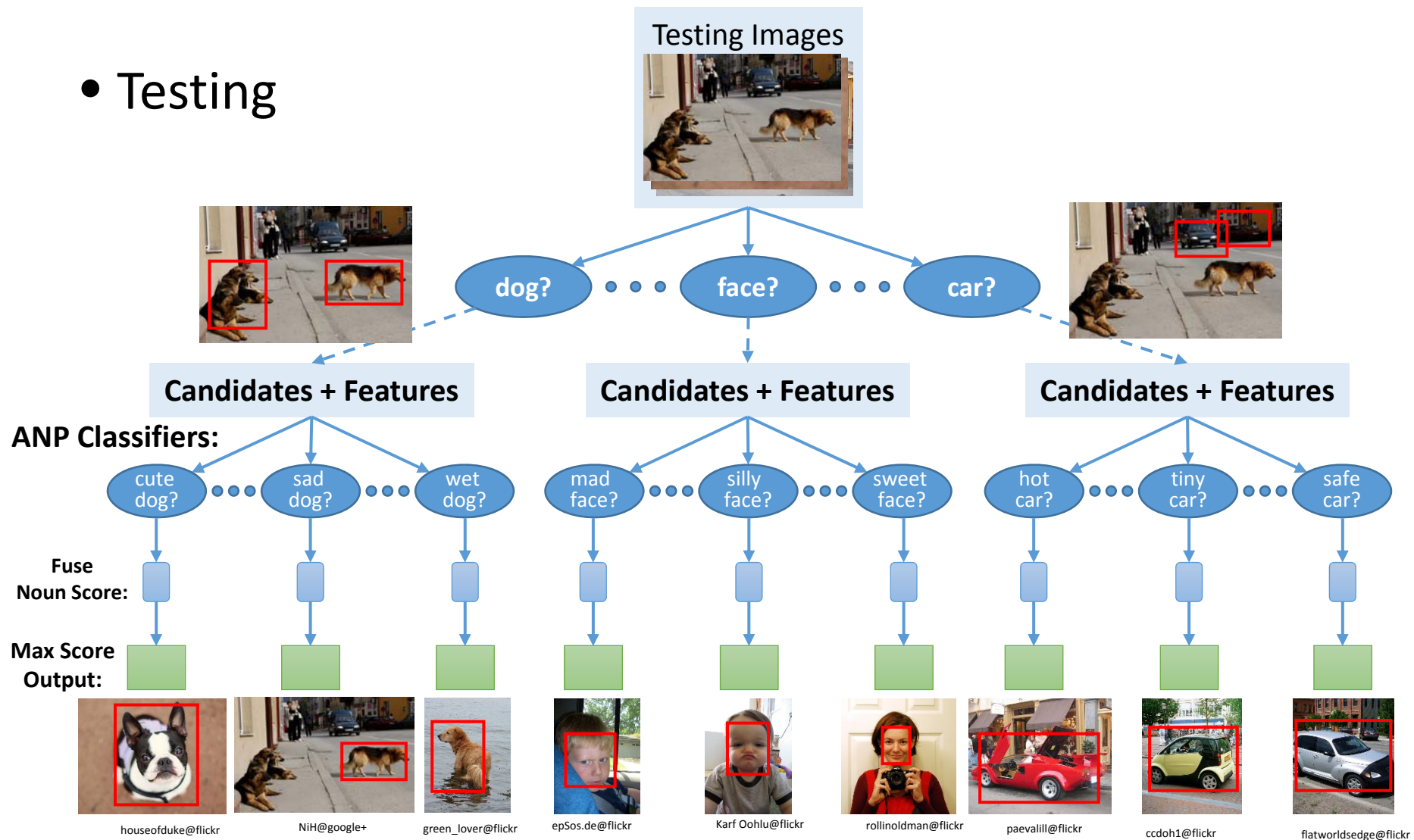
“Tired dog”: lying on floor or surface, closed eyes, yawning, resting, no action, fore legs, paws, face on floor

# So, Need to Link to Objects + Attributes



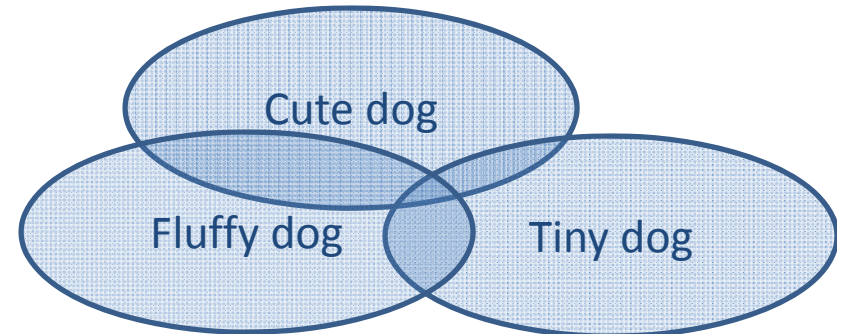
# Hierarchical: Object + Affect Attributes

- Testing



# Tricky Issue: Concept Subjectivity

- Subjective attributes are highly overlapped
  - E.g., cute dog, fluffy dog, cuddly dog
- Solution
  - Need a way to handle soft label overlap
  - Model overlap proportion in SVM



# The $\alpha$ SVM Algorithm

F. Yu; D. Liu; S. Kumar; T. Jebara; S.-F. Chang.  $\alpha$ SVM for learning with label proportions. ICML13

- Large-margin framework:

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N L(y_i, \mathbf{w}^T \varphi(\mathbf{x}_i) + b) + C_p \sum_{k=1}^K L_p(\tilde{p}_k(\mathbf{y}), p_k) \\ \text{s.t.} \quad & \forall_{i=1}^N, \quad y_i \in \{-1, 1\}. \end{aligned}$$

*Note: In the original image, red circles highlight the loss terms  $L(y_i, \mathbf{w}^T \varphi(\mathbf{x}_i) + b)$  and  $L_p(\tilde{p}_k(\mathbf{y}), p_k)$ . Red arrows point from these circles to the labels "Label prediction loss" and "proportion loss" respectively.*

- Learned with alternate optimization or a relaxed convex form

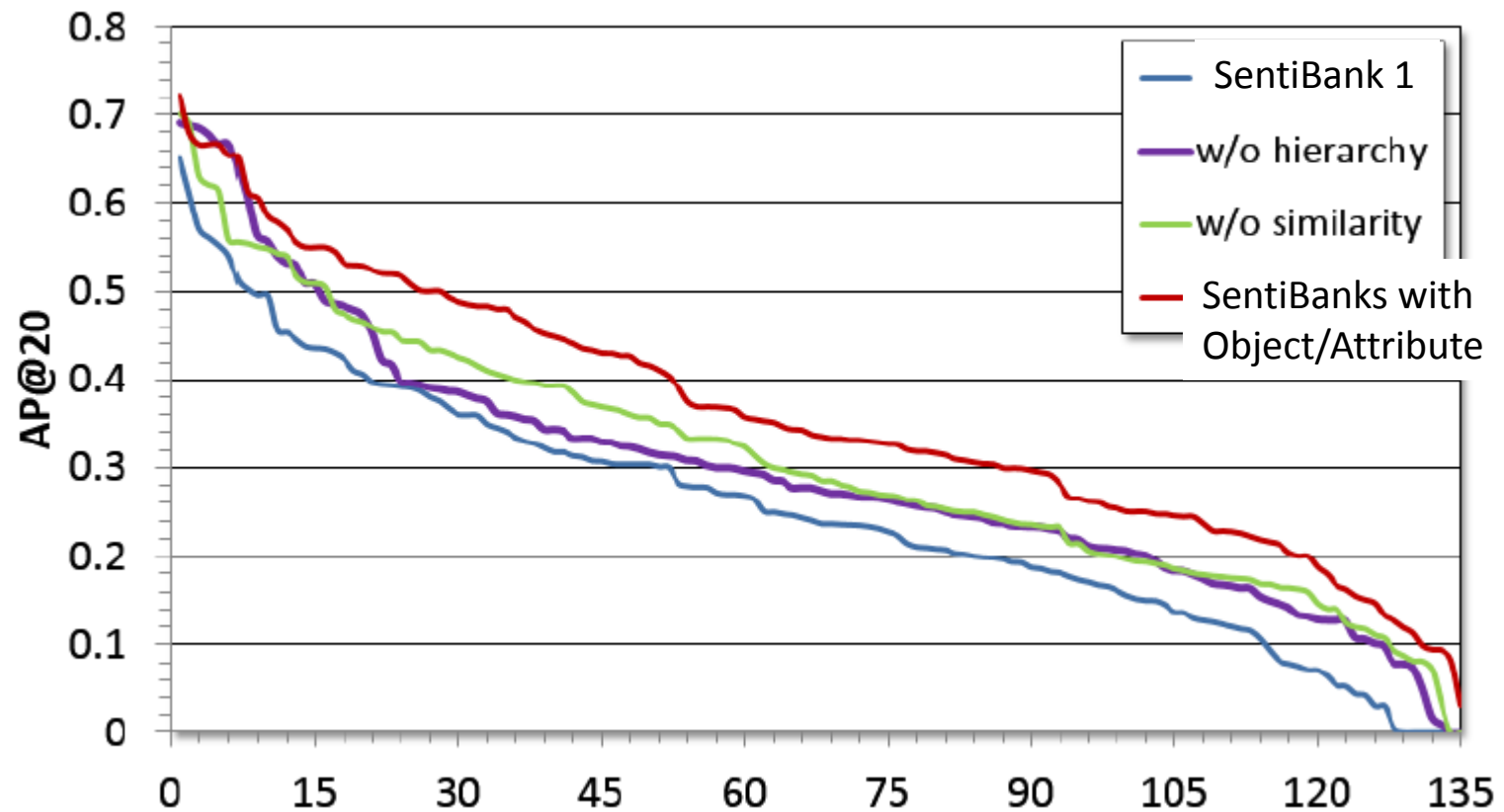
- Formulation:

Image set of “Fluffy Dog” has proportion  $p_k$  being “Cut Dog”

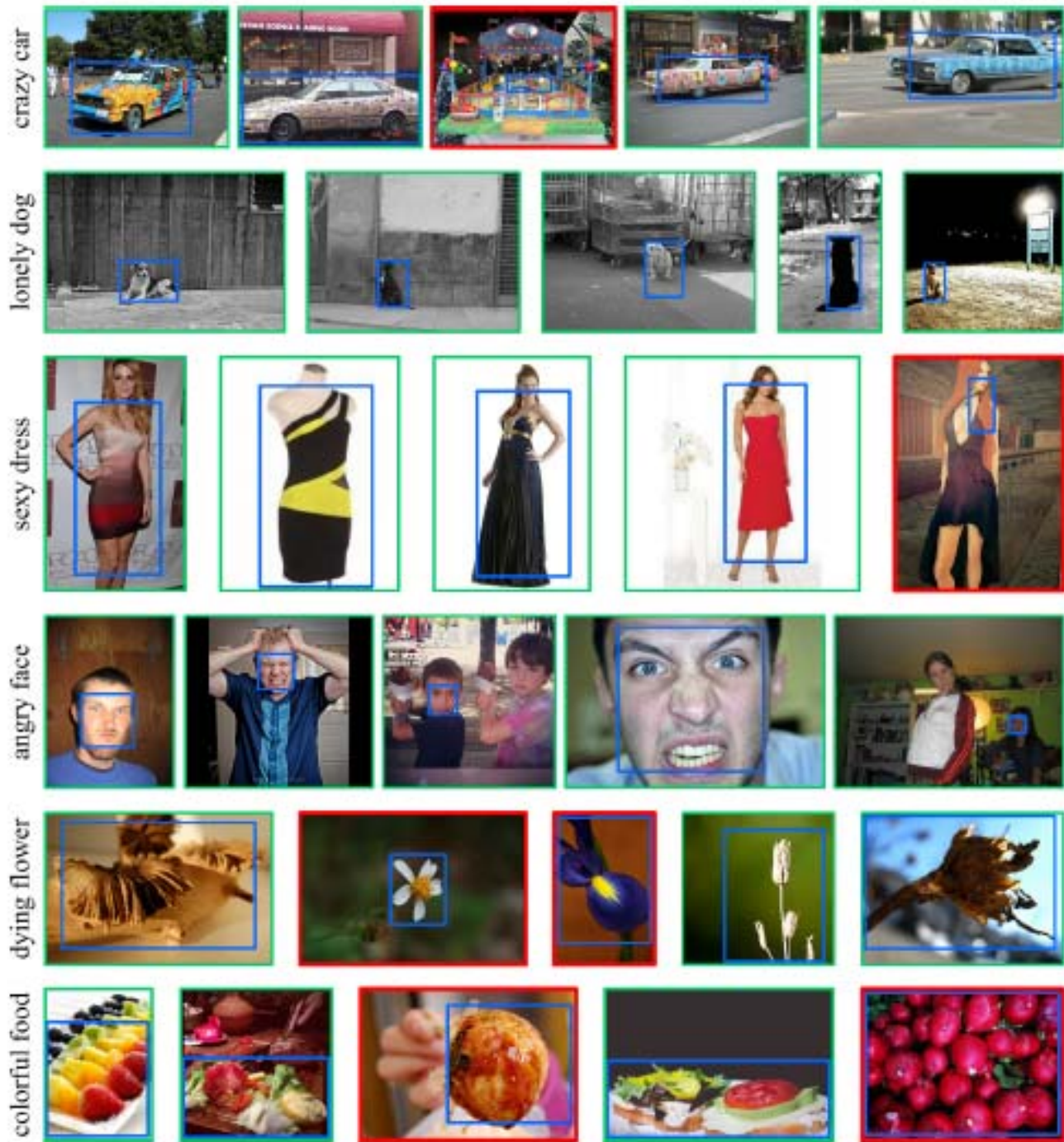
Image set of “Tiny Dog” has proportion  $p_k$  being “Cut Dog”

proportions can be approximated by ConceptNet

# SentiBank using Objects/Attributes



# Detection Examples



# SentiBank 2.0 using Convolutional NN

- Annotation Accuracy

- top-10:

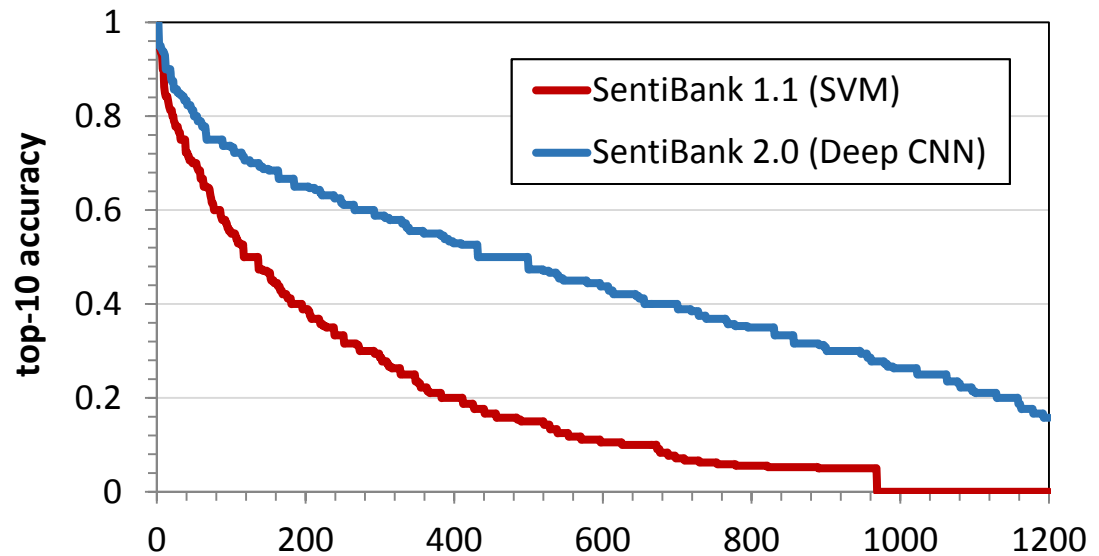
- Deep CNN: 45.5%

- Global SVM: 18.7%

- top-1:

- Deep CNN: 14.7%

- Global SVM: 3.0%



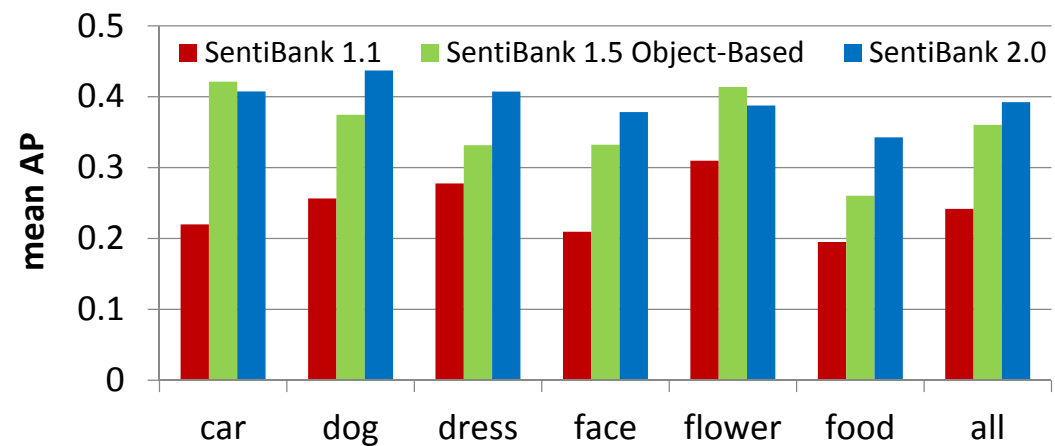
- Retrieval Performance

- mean AP

- Deep CNN: 39.2%

- Object Based: 36.0%

- SVM: 24.2%



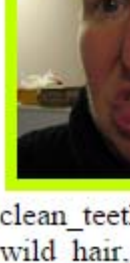
# Top K Annotations



**broken\_colorful\_fo** funny\_fac **dying\_flo** **crazy\_gr:** **empty\_tr** **dirty\_car:** crazy\_car, damaged\_car, clean\_car, filthy\_car, nice\_letter, elly\_cat, broken\_f fresh\_flowe crazy\_eye beautiful\_ awesome\_colorful\_st favorite\_street, clean\_shoes, stupid\_sign, nice\_butt, fat\_face, tradition:wild\_mushr beautiful\_ golden\_au happy\_ch abandoned cloudy\_bay, heavy\_book, abandoned\_train, shiny\_metal, empty\_box, id\_cat, creepy\_s happy\_face, bad\_teeth, colorful\_f funny\_hat clean\_roor natural\_food, bad\_graffiti, clean\_air, dead\_bug wet\_win hot\_food, ar colorful\_a scary\_sku gorgeous\_



**broken\_colorful\_fo** innocent\_f **crazy\_gr:** **empty\_tr** **dirty\_car:** damaged\_car, broken\_window, abandoned\_vehicle, \_cat, tradition:fresh\_food, mad\_face, dying\_rosi abandon golden\_su clean\_car, abandoned\_car, filthy\_car, bad\_accident, drunk\_driver, hy\_cat, tradition:happy\_birth curious\_ba fantastic\_f aweosome\_wet\_shoes scary\_house, icy\_road, shiny\_metal, expensive\_car, tiny\_car, ious\_ca melted\_c colorful\_gar little\_flow happy\_mo bloody\_ey magical\_n heavy\_train tradition:sunny\_flow elegant\_la



**broken\_colorful\_fo** dirty\_mou **crazy\_gr:** **empty\_tr** **dirty\_car:** clean\_car, tiny\_bathroom, broken\_window, evil\_cat, rotten\_w healthy\_foo **great\_hair,** tiny\_flowe bright\_lig abandoned broken\_computer, drunk\_driver, shiny\_cars, abandoned\_train, lly\_kitty fresh\_sn shiny\_hair, colorful\_eg fragile\_flow lost\_coast abandoned abandoned\_car, rotten\_wood, abandoned\_hotel, abandoned\_house, , dirty\_c creepy\_s yummy\_pie, fancy\_foo strange\_flow traditional abandoned creepy\_house, clean\_air, dead\_bug, young\_girls smelly\_c yummy\_food, sour\_ap golden\_su dangerous\_ empty buiang, goiden aragon, empty airport, warm pool, wet area weird plant, sunny autumn, wild flora, weird bug

# VSO/SentiBank Resources

Ontology and 1,200 Classifiers

<http://visual-sentiment-ontology.appspot.com/>

## Visual Sentiment Ontology

### Summary

The analysis of emotion, affect and sentiment from visual content has become an exciting area in the multimedia community allowing to build new applications for brand monitoring, advertising, and opinion mining. There exists no corpora for sentiment analysis on visual content, and therefore limits the progress in this critical area. To stimulate innovative research on this challenging issue, we constructed a new benchmark and database (you can browse the database at [► VSO Browsing Interface](#)). This database contains a Visual Sentiment Ontology (VSO) consisting of 3244 adjective noun pairs (ANP), SentiBank a set of 1200 trained visual concept detectors providing a mid-level representation of sentiment, associated training images acquired from Flickr, and a benchmark containing 603 photo tweets covering a diverse set of 21 topics.

This website provides the above mentioned material for download and is structured as the following:

#### Benchmark and Dataset Citation

Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel and Shih-Fu Chang. **"Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs,"** ACM Multimedia Conference, Barcelona, Oct 2013.

- [download paper](#)
- [download material](#)

[Visual Sentiment Ontology: Ontology and Concepts](#)

[Visual Sentiment Ontology: Image Dataset](#)

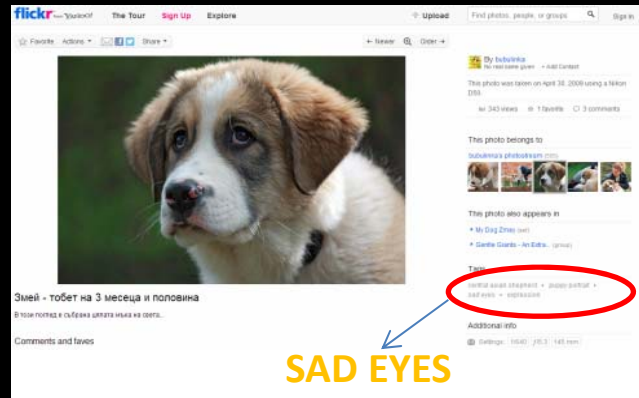
[SentiBank: Visual Sentiment Concept Classifiers](#)

[Photo Tweet Sentiment Benchmark](#)

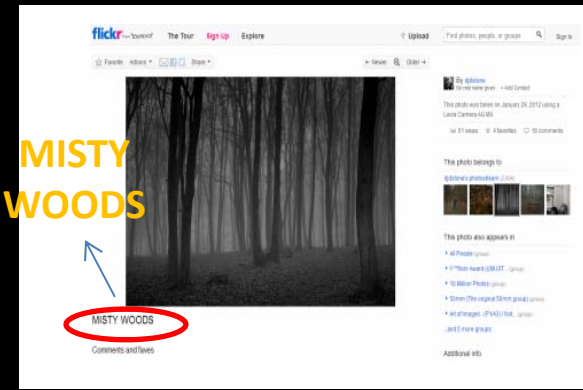
# Publisher (expressed) vs. Viewer (evoked) Affects



Publisher  
Affect  
Concept

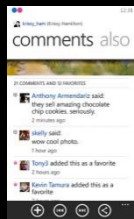


SAD EYES

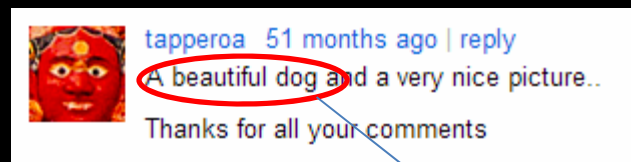


MISTY  
WOODS

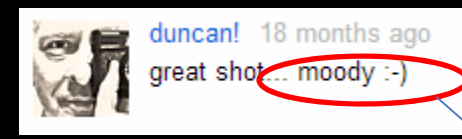
MISTY WOODS



Viewer  
Affect  
Concept



BEAUTIFUL DOG

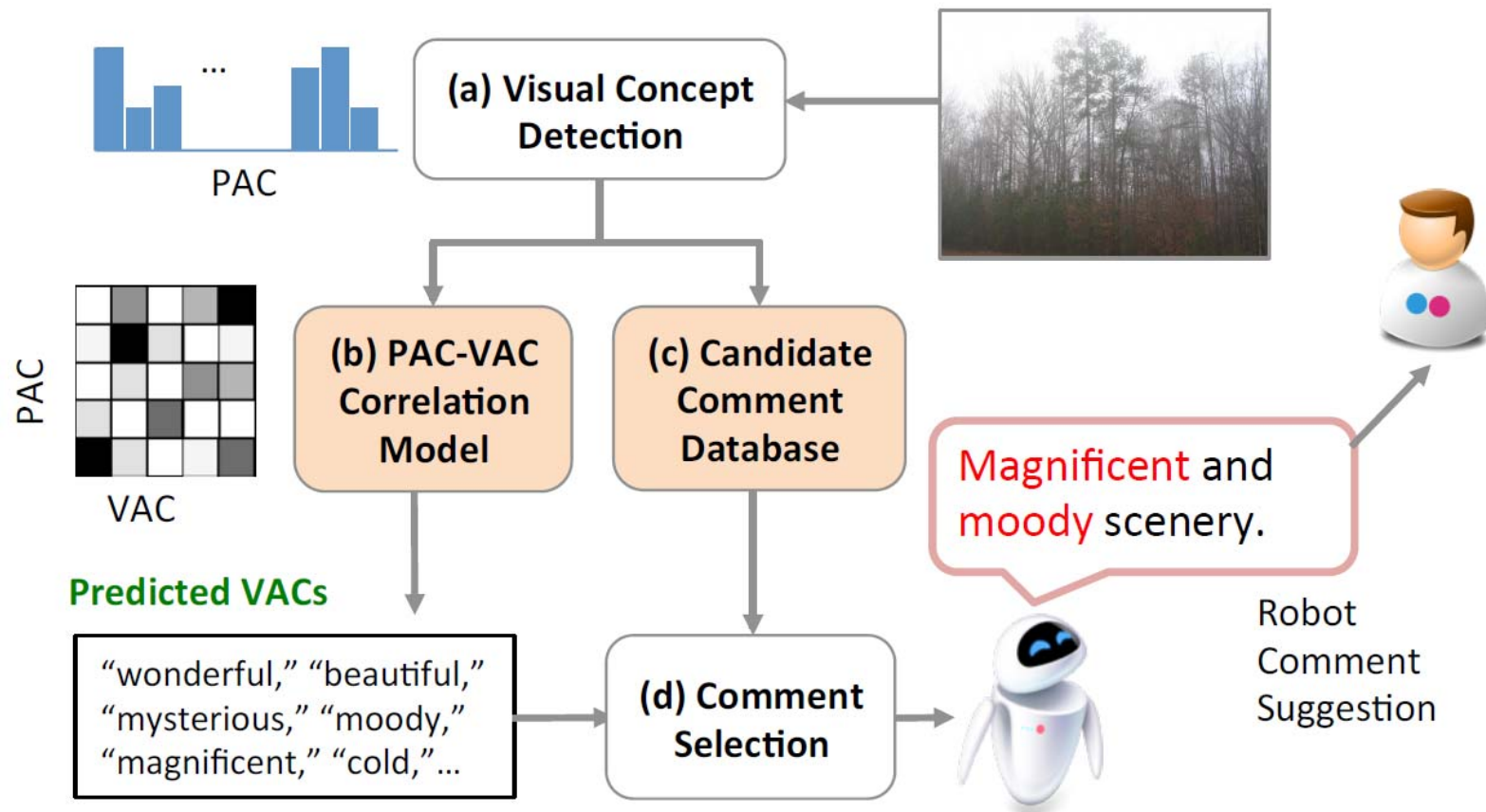


MOODY



# Assistive Image Comment Robot – A New Social Media Plug-In Tool

Demo video: <https://www.youtube.com/watch?v=RyXK-kd9Bjw>



# Probabilistic PAC-VAC Correlation Models

- Measuring **PAC-VAC Co-occurrences**:

$$P(\overline{p_k} \mid \overline{v_j}; \theta) = \frac{\sum_{i=1}^{|D|} B_{ik} P(v_j \mid d_i)}{\sum_{i=1}^{|D|} P(v_j \mid \underline{d_i})_{\text{image}}}, \quad B_{ik} : \text{presence of } p_k \text{ in image } d_i$$

- **Predicting VACs** for a given image by Bayes model:

$$P(v_j \mid d_i; \theta) = \frac{P(v_j \mid \theta) P(d_i \mid v_j; \theta)}{P(d_i \mid \theta)}$$

- Apply **matrix factorization** to smooth sparse terms

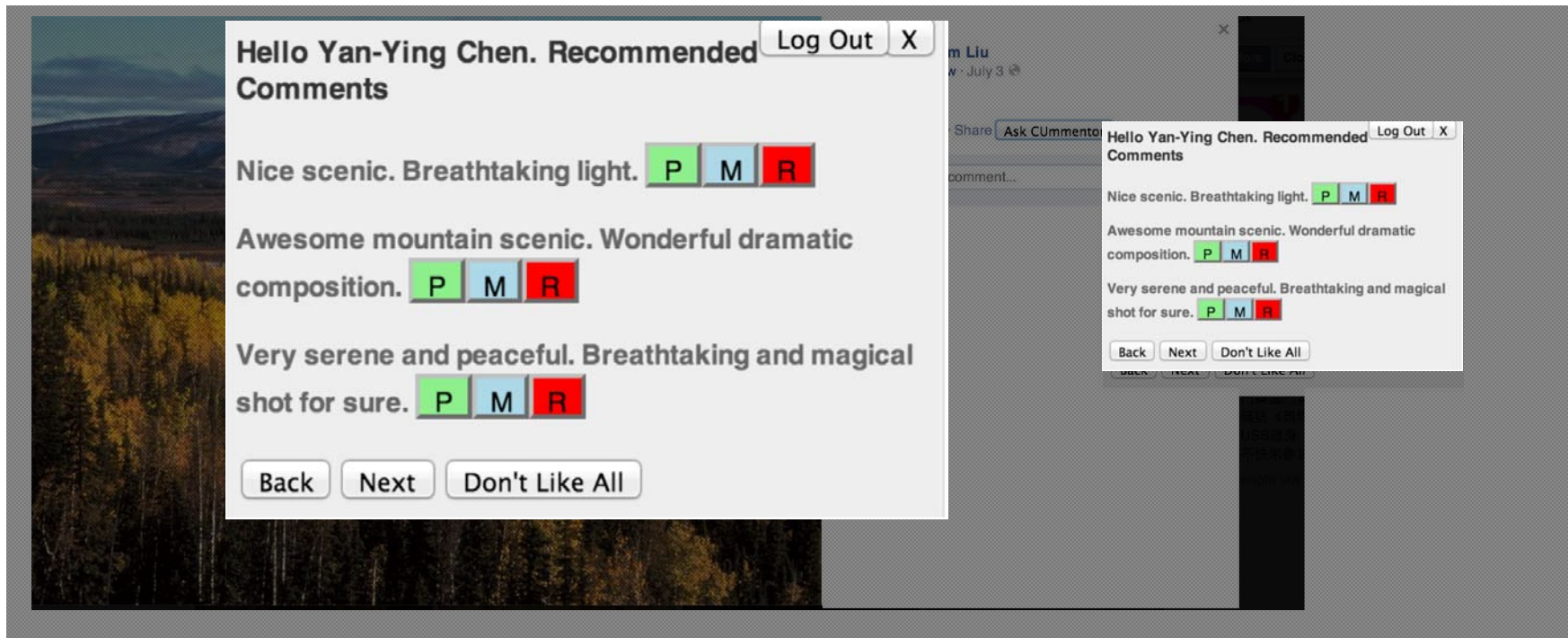
PAC-VAC pairs with strong correlations predicted by the correlation model.

PAC	#1 VAC	#2 VAC	#3 VAC
tiny dog	cute	adorable	little
weird dog	weird	funny	cute
crazy cat	hysterical	crazy	hilarious
cloudy morning	ominous	serene	dramatic
dark woods	mysterious	spooky	moody
powerful waves	dynamic	powerful	sensational
wild water	dangerous	dynamic	wild
terrible accident	terrible	tragic	awful
broken wings	fragile	poignant	poor
bright autumn	bright	delightful	lovely
creepy shadow	creepy	spooky	dark
happy halloween	spooky	festive	scary
pretty flowers	delicate	joyful	lush
fresh leaves	fresh	green	vibrant
wild horse	wild	majestic	healthy
silly girls	sick	funny	cute
mad face	mad	funny	cute
beautiful eyes	expressive	intimate	confident
sweet cake	yummy	hungry	delicious
nutritious food	healthy	yummy	delicious
shiny dress	shiny	sexy	gorgeous
colorful building	colourful	vivid	vibrant
haunted castle	spooky	mysterious	scary

- Matrix factorization used to smooth sparse correlation terms
- Markov model used in new comment sentence synthesis
- Diversity considered in generating multiple sentences

# Image Comment Robot

## – Chrome Extension FaceBook App

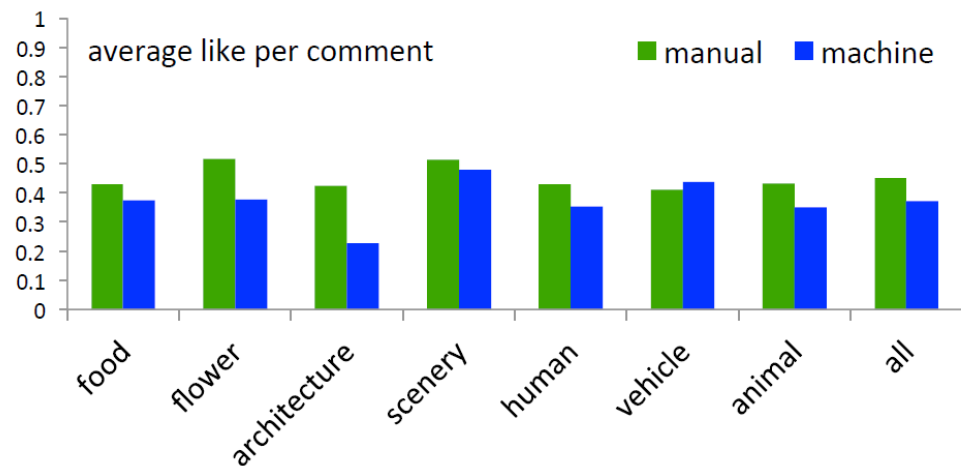


[Demo video](#)

# New User Study: Acceptance rates of robot suggested comments

image class	food	flowers	architecture	scenery	human	vehicle	animal	all
# sessions	50	52	60	51	57	74	60	405
# posts	45	51	53	50	47	63	54	363
# iterations	2.36	2.58	3.15	2.29	3.48	3.99	5.57	3.43
# clicks (next)	1.68	2.17	2.69	1.94	2.90	3.47	5.00	2.92
# clicks (more)	0.68	0.40	0.46	0.35	0.59	0.51	0.57	0.51
# clicks (reject)	1.56	0.71	2.20	0.29	2.24	2	2.83	1.75
accept rate	0.90	0.98	0.88	0.98	0.81	0.85	0.90	0.90

- The average # of likes per comment over different photo classes



- Example comments considered to be a "real".  
Top: robot-generate. Bottom: manually-generated.



Beautiful and elegant.  
Gorgeous subtle shades of purple.



Looks colourful, healthy and very very tasty. Really delightful.



What are they doing?



able shot.  
The reflection on the glass is cool.  
cent and beautiful.



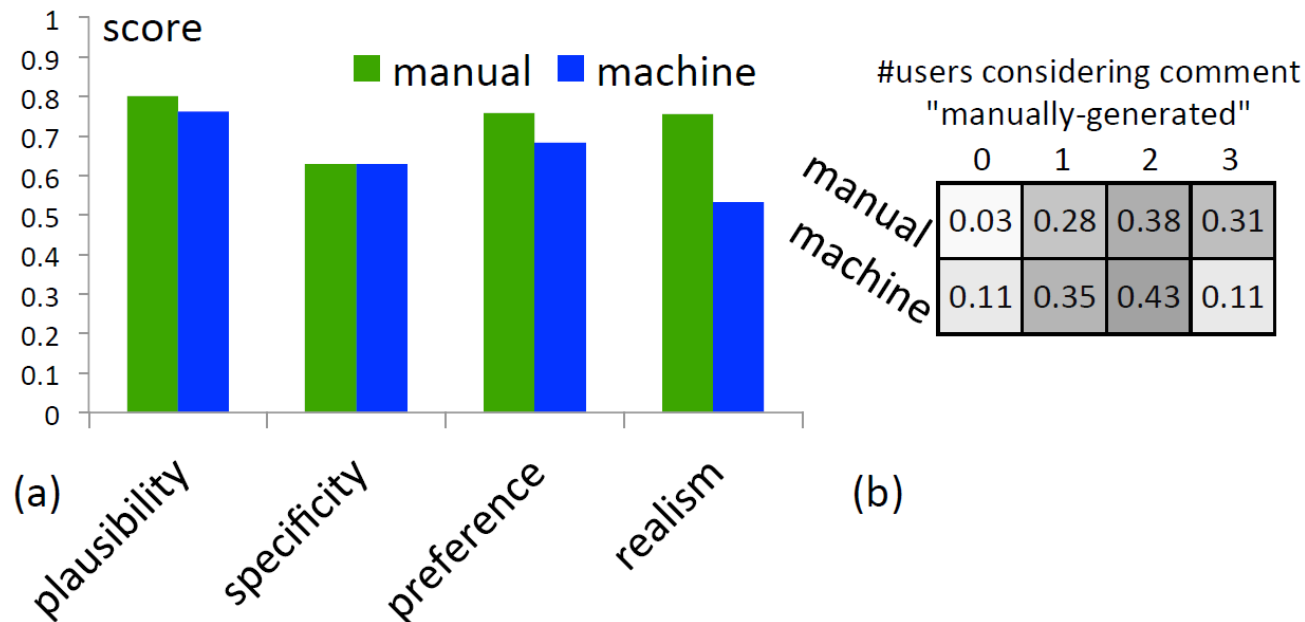
helicopters?



ate dog! What are you looking at?

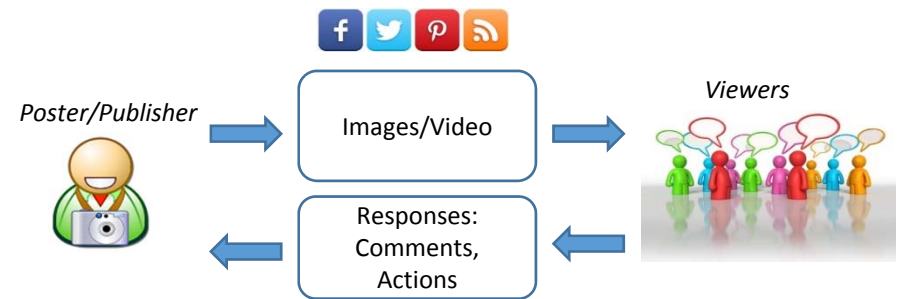
# Subjective Evaluation of Robot Suggested Comments

- 54% robot-generated comments fooled the majority of evaluators



# Summary and Open Issues

- Develop a mid-level concept representation for
  - Visual emotion analysis
  - SentiBank Tools Available
  - Viewer response modeling
  - Assistive comment robot



- Open Issues
  - Domain Differences
  - Modeling of user, context, and social network relation

# References

- Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, Shih-Fu Chang. **Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs.** In *ACM Multimedia*, October 2013.
- Yan-Ying Chen, Tao Chen, Winston H. Hsu, Hong-Yuan Mark Liao, Shih-Fu Chang. **Predicting Viewer Affective Comments Based on Image Content in Social Media.** In *ACM International Conference on Multimedia Retrieval (ICMR), full paper*, April 2014.
- Tao Chen, Felix X. Yu, Jiawei Chen, Yin Cui, Yan-Ying Chen, Shih-Fu Chang. **Object-Based Visual Sentiment Concept Analysis and Application.** In *ACM Multimedia*, November 2014.
- Yan-Ying Chen, Tao Chen, Taikun Liu, Hong-Yuan Mark Liao, Shih-Fu Chang, **“Assistive Image Comment Robot – A Novel Mid-Level Concept-Based Representation,”** invited submission to *IEEE Trans. on Affective Computing*, 2014.