Visual Analysis of Tag Co-occurrence on Nouns and Adjectives

Yuya Kohara and Keiji Yanai
Department of Informatics,
The University of Electro-Communications, Tokyo

Background

Query word:
blue & car

“red car”
“blue sky”

Irrelevant
×

Tags

North Vancouver • Canada • British Columbia • B.C. • Waterfront Park • 2012 • German Car Festival • German • car • Porsche • 911 • 993 • Porsche 911 • colourful • vibrant • vivid • red • sky • blue

Objective

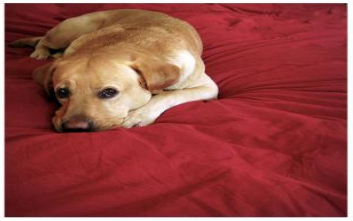
- Analyze visual relationships between nouns and adjectives

"red + flower"

visual relationships

>

"red + dog"



- Find out the tag pair with high visual relationships

Objective

- The images corresponding to the tag pairs with high visual relationships looks similar.

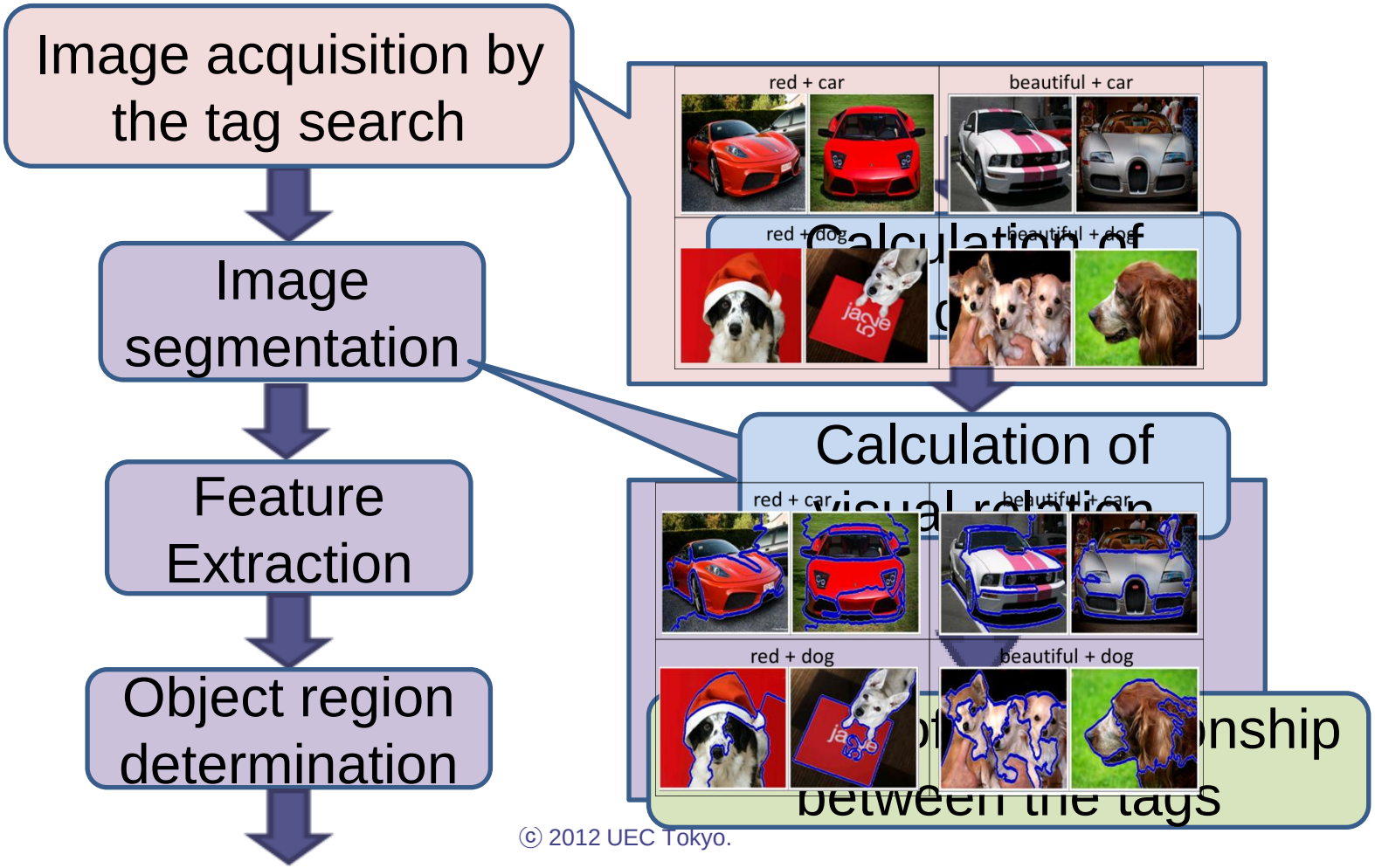
flower + red

- Help create new dataset with less noise
- Improve accuracy of simultaneous recognition of nouns and adjectives
  - There is a flower and the color of the flower is red.

Basic idea

- Prepare many tag pairs of nouns and adjectives
  - e.g. "red + car", "blue + sky", ...
- Search web image database for the images corresponding to each of the prepared tag pairs
- Detect regions of objects for all the images
  - Eliminating of background in the images
- Evaluate the distribution of the image set of each tag pair with entropy, and calculate mutual information

Overview



Example of gathered images

red + car

beautiful + car

red + dog

beautiful + dog

Image acquisition

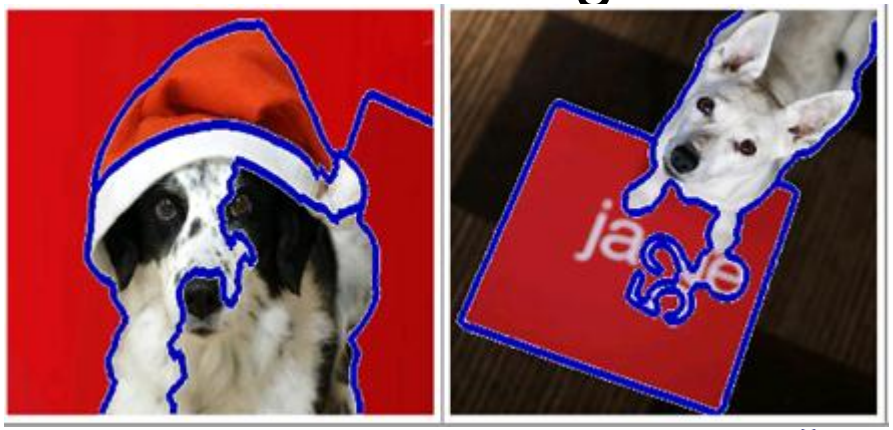
- Image acquisition from flickr
  - 20 nouns : car , dog , sky , …
  - 16 adjectives : - , red , morning , old , …
    - 20 nouns × 16 adjectives = 320 tag pairs
  - 200 positive images for each tag pair
  - 600 negative images (common to all tag pairs)
    - 64,600 images (=200 × 320+60)

Image segmentation by JSEG

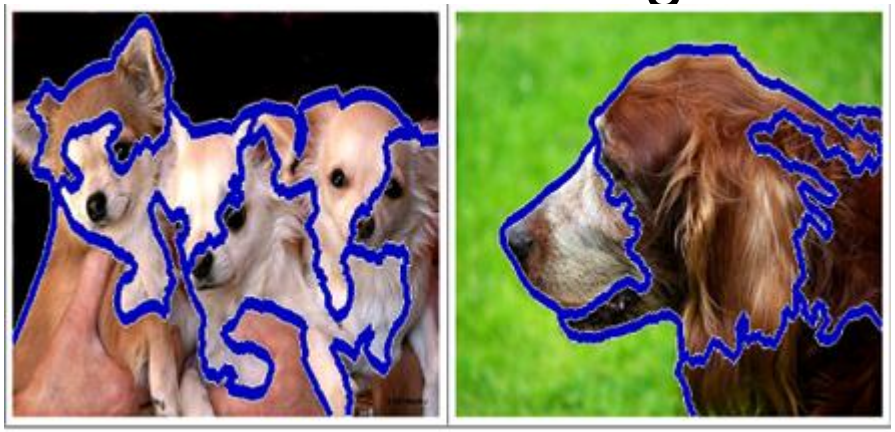
red + car

beautiful + car

red + dog



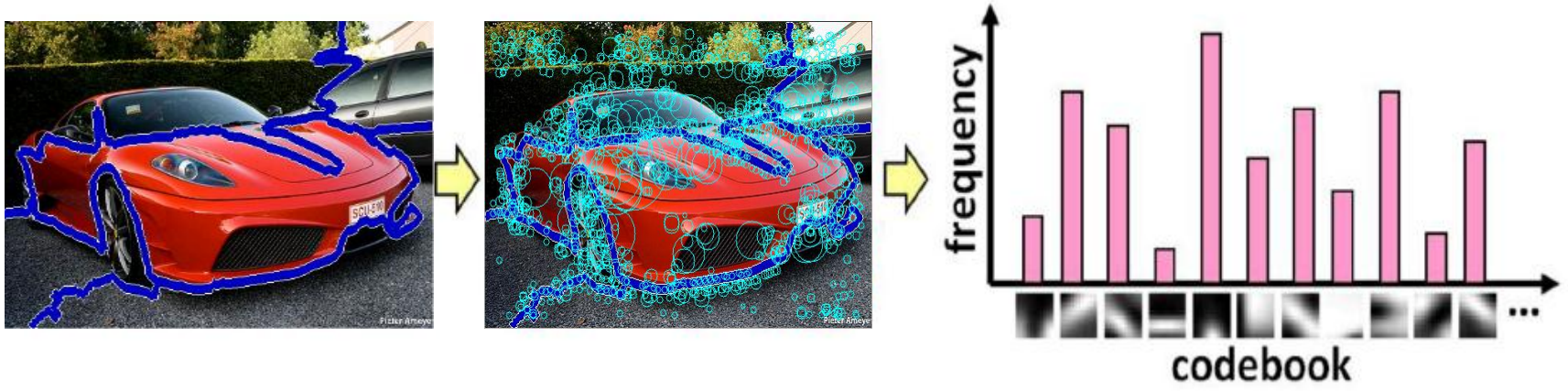
beautiful + dog



Feature extraction

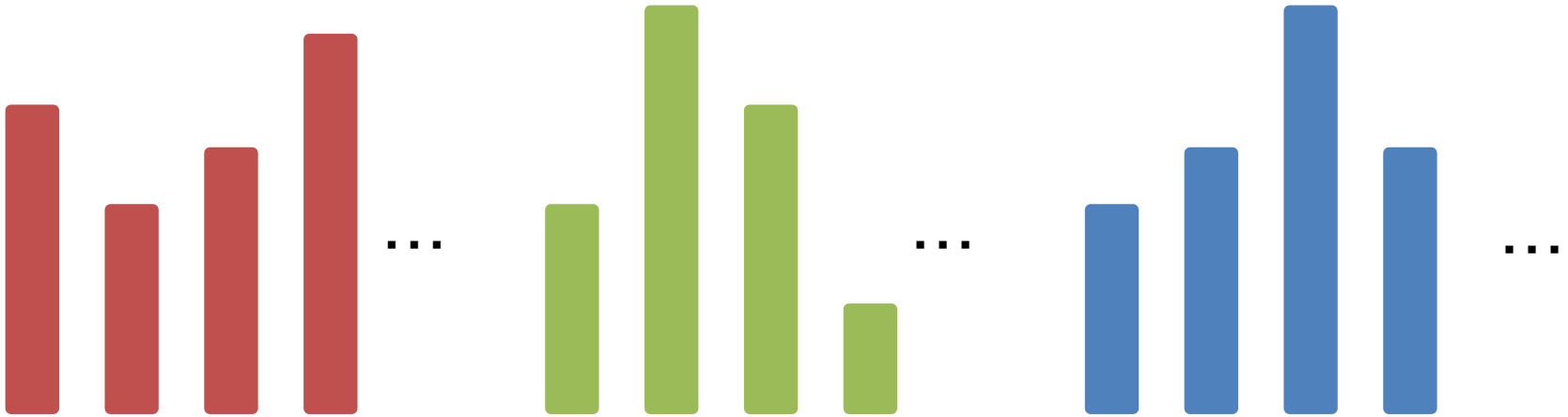
- Extract SIFT features from each region
- Build Bag of Feature vectors from a set of the Color-SIFT features

Detection of KeyPoints Frequency of representative pattern



Used Features

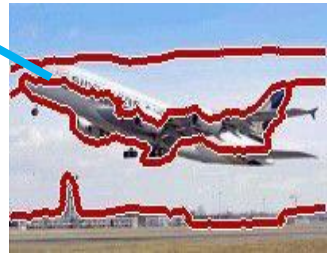
- Color-SIFT
  - Divide a picture into R,G,B and extract feature from each.
  - 384 dimensions : 128 dimensions × 3



Assume good region corresponding keyword

- Multiple instance-SVM
  - training→ evaluation → changing dataset→ training→ …… (5-loops)

High evaluated regions are positive data,
and low evaluated regions are negative data in the next loop.



car & red regions

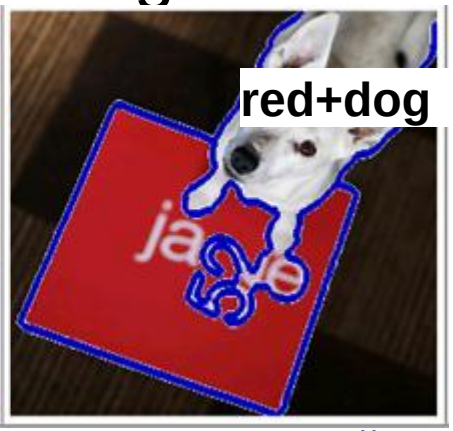
Not (car & red) regions

Results of detected regions

red + car

beautiful + car

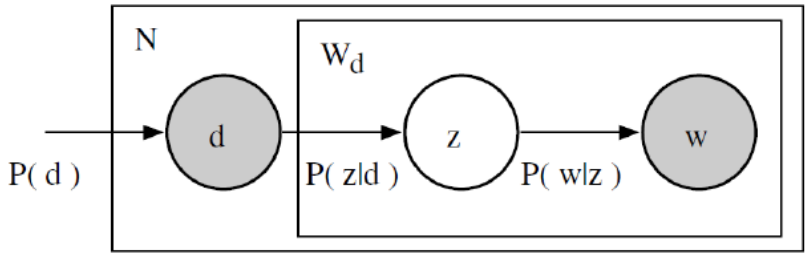
red + dog



beautiful + dog

Calculate Feature Distribution

- pLSA
  - Probabilistic Latent Semantic Analysis



$$L = \sum_{i=1}^{I} \sum_{j=1}^{J} n(d_i, w_j) \log P(d_i, w_j)$$

  - d:image, w:word, z:hidden topic
  - Calculate distribution of the features from the relationship of d, z, and w

Visual relation by mutual information

Entropy： Value of distribution of local features

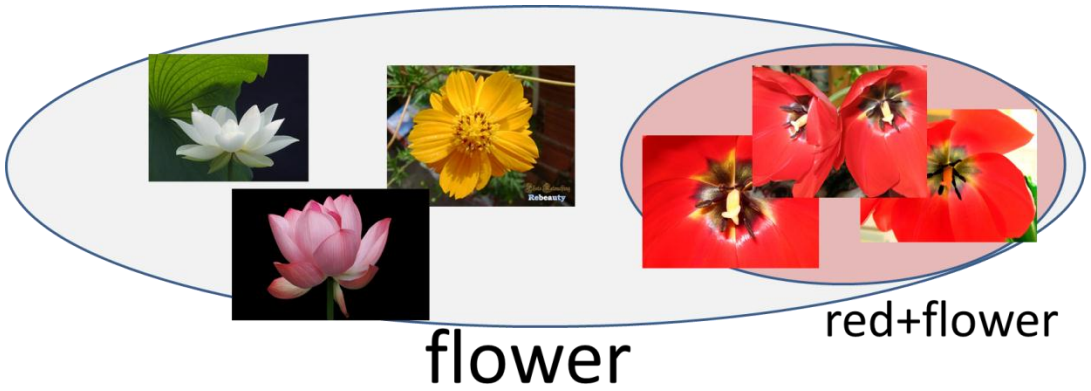
$$H(X) = -\sum P(x)\log(P(x))$$

High : Wide distribution ⇔ Low : Narrow distribution

Mutual information： Difference of entropy

$$MI(X;Y) = H(X) - H(X|Y)$$

High : High relation ⇔ Low : Low relation



Experiments (1)

Dataset:64,600 images for 320 tagged pairs

1. Evaluate mutual information of tag pairs

2. Compare visual feature based relation with text based relation

Results of mutual information of tag pairs

	red	blue	green	black	white	circle	square
beach							
bird							
boat							
–							
morning	night	winter	summer	new	old	beautiful	cool
cat							
cloud	0.071	0.354	0.161	0.232	0.078	−0.018	0.151	0.336	0.143	0.003	0.042	0.085	−0.028	0.016	−0.022
cup	0.139	0.105	0.003	0.130	0.118	0.131	0.035	0.101	0.129	0.049	0.044	0.150	−0.003	0.018	0.039
dog	0.003	0.061	0.046	0.145	0.117	0.061	0.092	0.032	0.083	0.069	0.064	0.046	0.044	0.070	0.048
flower	0.078	0.066	−0.020	0.154	−0.024	0.030	0.217	0.220	0.135	0.063	−0.064	0.069	−0.005	0.086	0.014
fruit	0.105	0.137	0.100	0.121	0.150	0.073	0.096	0.169	0.103	0.132	0.013	−0.027	−0.060	−0.015	−0.005
house	0.027	0.024	0.069	0.120	0.124	0.144	0.086	0.137	0.211	0.069	0.048	0.050	0.038	0.066	0.008
people	0.096	0.185	0.145	0.082	0.055	−0.040	0.175	0.153	0.088	0.011	0.077	0.106	−0.128	0.018	0.030
sea	0.112	0.113	0.157	0.242	0.085	0.042	0.113	0.006	0.050	0.117	0.149	0.007	−0.046	0.061	0.018
sky	0.114	0.170	0.163	0.161	0.060	0.040	0.091	0.224	0.078	0.129	0.033	−0.011	0.093	−0.003	−0.001
snow	0.084	0.047	0.024	0.114	0.078	0.035	0.013	0.164	0.153	0.093	0.020	0.134	0.058	0.090	0.040
sun	0.211	−0.022	−0.038	0.198	−0.030	−0.032	0.108	0.439	0.237	0.198	−0.066	0.056	−0.021	0.077	−0.006
tower	0.188	0.108	0.016	0.146	0.030	−0.026	0.036	0.287	0.237	0.011	0.048	0.053	−0.022	0.006	−0.002
train	0.036	0.261	0.038	0.084	0.044	−0.014	0.167	0.279	0.159	0.054	−0.009	−0.013	0.047	0.084	0.077
tree	0.278	0.044	0.027	0.069	−0.008	0.176	0.042	0.237	0.248	0.069	0.007	−0.016	−0.067	0.179	0.013
	0.113	0.234	0.051	0.151	0.046	0.022	0.063	0.443	0.101	0.044	0.012	0.043	0.037	0.015	0.015
	0.056	0.133	0.054	0.242	0.128	0.149	0.071	0.036	0.145	0.028	0.016	0.040	0.045	0.050	0.023
	0.014	0.137	0.072	0.183	0.058	−0.022	0.173	0.376	0.186	0.164	0.056	0.046	0.056	−0.003	0.011

Some representative results

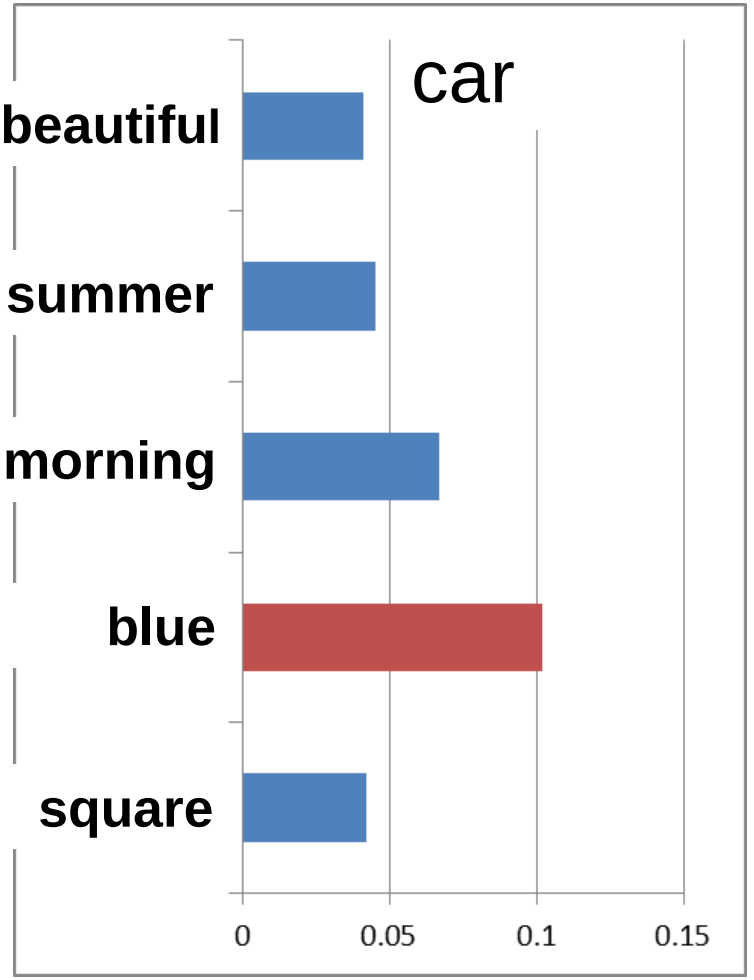
High visual relationship between pair tags

- "red+sun" , "red+car"
  - Color adjective and object noun

- "morning+sun", “night+sun"
  - Time adjective and noun related sky

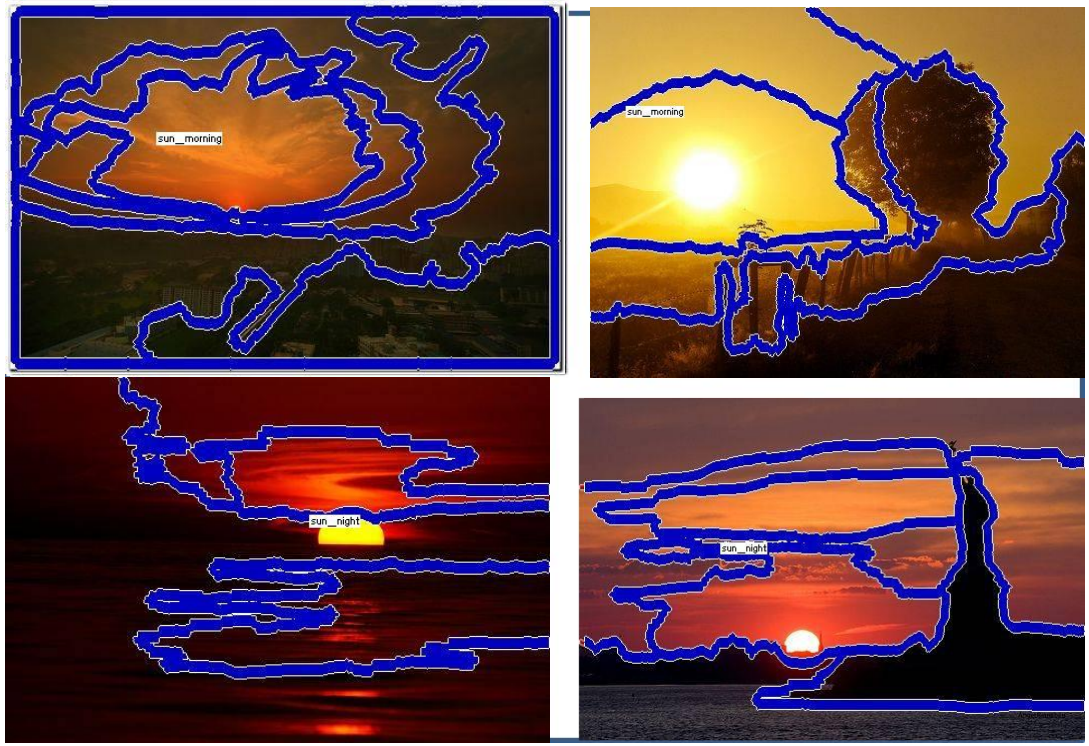
An example of high visual relationships (1)

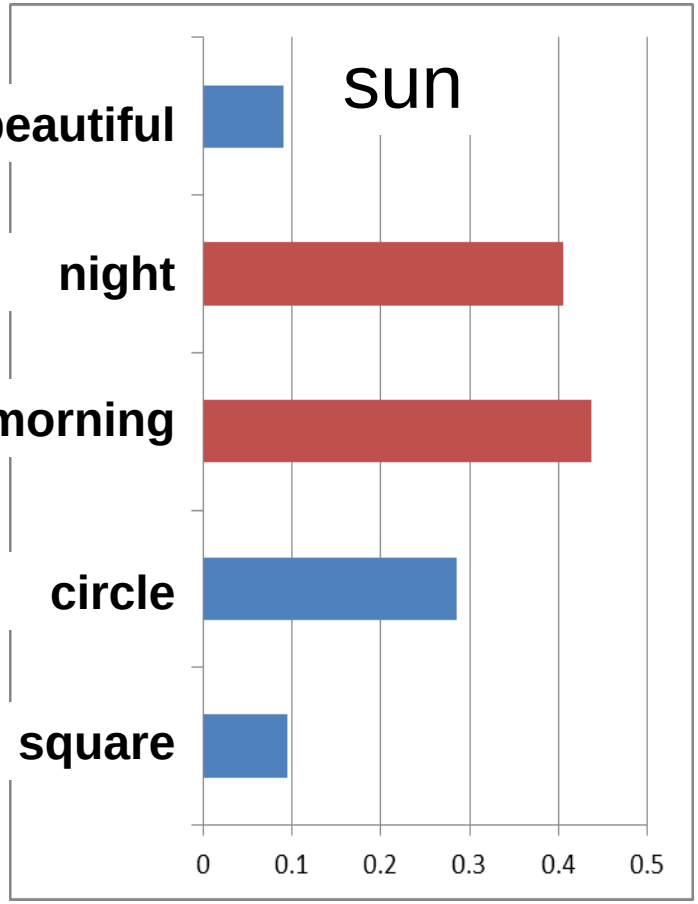
- Color adjective and object noun
  - "blue + car"



An example of high visual relationships (2)

- Time adjective and noun related sky
  - “sun+morning” , “sun+night”





Experiments (2)

Dataset:64,600 images for 320 tagged pairs

1. Evaluate mutual information of tag pairs

2. Compare visual feature based relation with text based relation

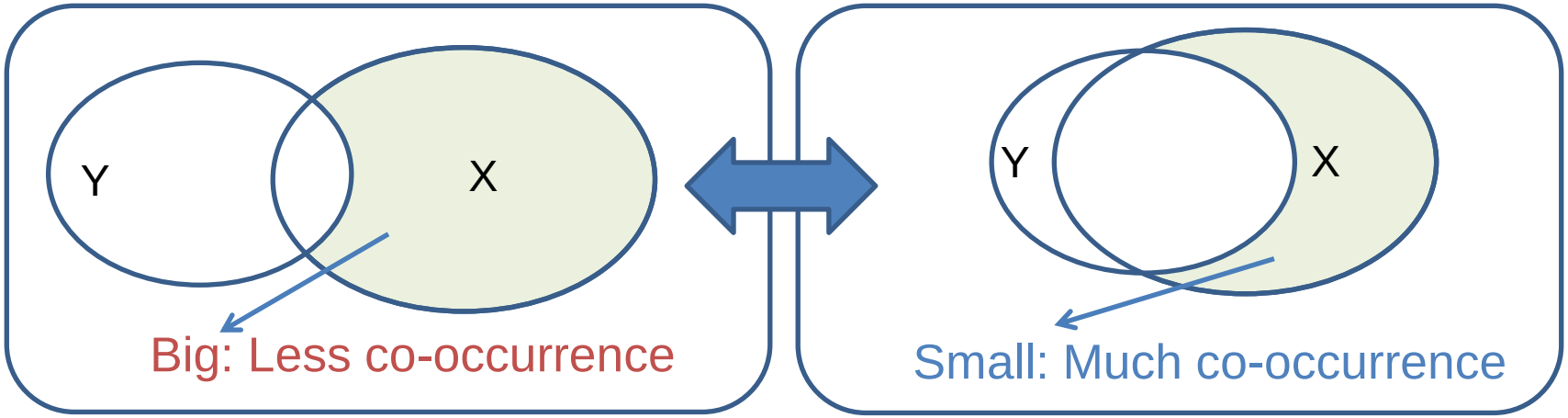
Similarity by textual relationships

2. Compare visual relation and textual relation

The Normalized Google Distance (NGD)

– Estimates similarity from number of search results

$$NGD = \frac{\max\{\log|X|, \log|Y|\} - \log|X \cap Y|}{\log N - \min\{\log|X|, \log|Y|\}}$$



|X|, |Y| : The number of search results obtained from Flickr

Results of similarity by textual relation

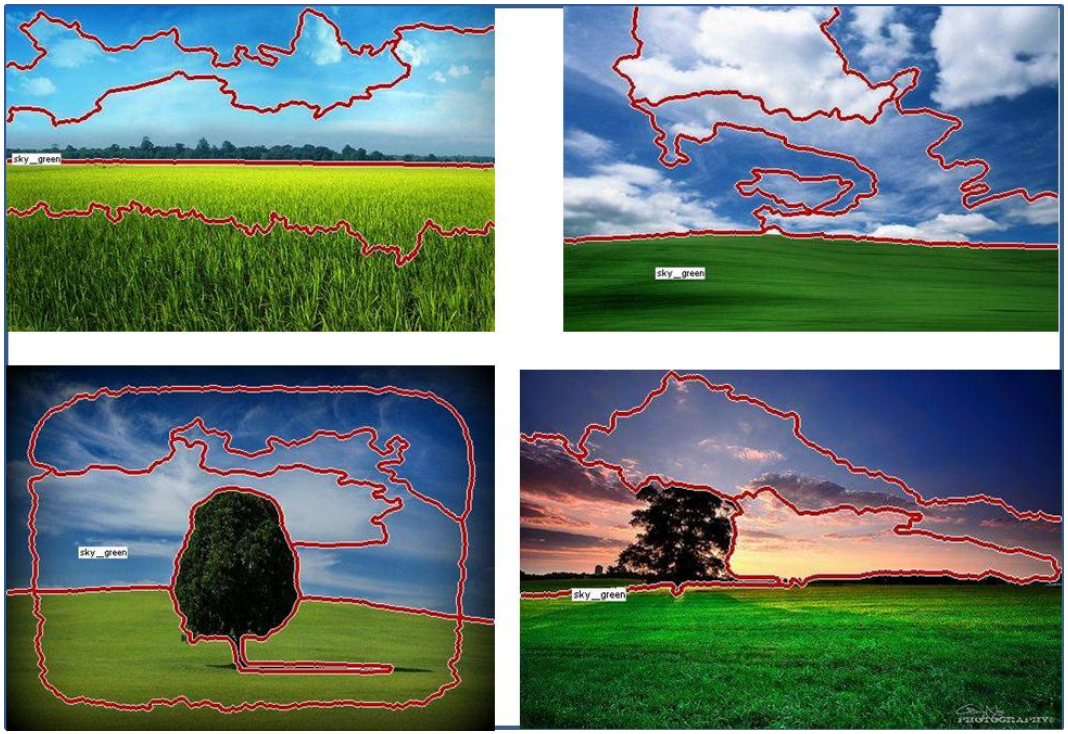
名詞/形容詞	red	blue	green	black	white	circle	square	morning	night	winter	summer	new	old	beautiful	cool
beach	0.678	0.492	0.650	0.646	0.630	0.802	0.983	0.620	0.639	0.669	0.445	0.669	0.715	0.539	0.717
bird	0.587	0.518	0.566	0.550	0.563	0.798	0.960	0.666	0.776	0.602	0.708	0.758	0.757	0.616	0.726
boat	0.618	0.516	1.556	0.683	0.647	0.676	0.954	0.606	0.629	0.725	0.575	0.690	0.593	0.618	0.710
bridge	0.646	0.579	0.616	0.623	0.619	0.665	0.798	0.600	0.487	0.613	0.680	0.584	0.567	0.640	0.728
car	0.508	0.556	0.613	0.523	0.573	0.766	0.918	0.747	0.573	0.704	0.683	0.623	0.425	0.666	0.542
cat	0.666	0.624	0.630	0.462	0.518	0.884	0.934	0.761	0.735	0.754	0.774	0.794	0.759	0.660	0.714
cloud	0.579	0.422	0.552	0.588	0.532	0.666	0.859	0.462	0.616	0.640	0.630	0.731	0.651	0.548	0.617
cup	0.659	0.721	0.720	0.711	0.679	0.671	0.943	0.628	0.853	0.853	0.858	0.700	0.734	0.831	0.770
dog	0.638	0.621	0.646	0.477	0.528	0.828	0.925	0.744	0.785	0.617	0.684	0.746	0.720	0.697	0.708
flower	0.405	0.480	0.379	0.579	0.408	0.724	0.878	0.666	0.730	0.709	0.523	0.739	0.765	0.517	0.707
fruit	0.508	0.687	0.534	0.694	0.663	0.667	0.890	0.699	0.809	0.779	0.647	0.812	0.735	0.707	0.671
house	0.594	0.583	0.555	0.604	0.543	0.722	0.895	0.689	0.597	0.618	0.649	0.521	0.434	0.623	0.657
people	0.597	0.589	0.600	0.525	0.524	0.763	0.788	0.700	0.506	0.640	0.527	0.625	0.576	0.474	0.604
sea	0.541	0.394	0.571	0.579	0.560	0.788	0.917	0.588	0.600	0.614	0.472	0.700	0.615	0.525	0.699
sky	0.463	0.226	0.415	0.535	0.444	0.696	0.806	0.489	0.450	0.504	0.480	0.645	0.599	0.498	0.635
snow	0.633	0.560	0.669	0.644	0.435	0.732	0.922	0.603	0.567	0.157	0.763	0.667	0.717	0.653	0.717
sun	0.495	0.408	0.468	0.530	0.496	0.673	0.871	0.420	0.606	0.515	0.416	0.664	0.582	0.405	0.585
tower	0.679	0.573	0.663	0.666	0.629	0.722	0.728	0.649	0.522	0.683	0.725	0.668	0.557	0.670	0.713
train	0.697	0.694	0.731	0.664	0.681	0.748	0.910	0.690	0.649	0.684	0.759	0.692	0.571	0.742	0.711
tree	0.483	0.447	0.376	0.536	0.483	0.709	0.826	0.541	0.565	0.447	0.601	0.691	0.574	0.558	0.654

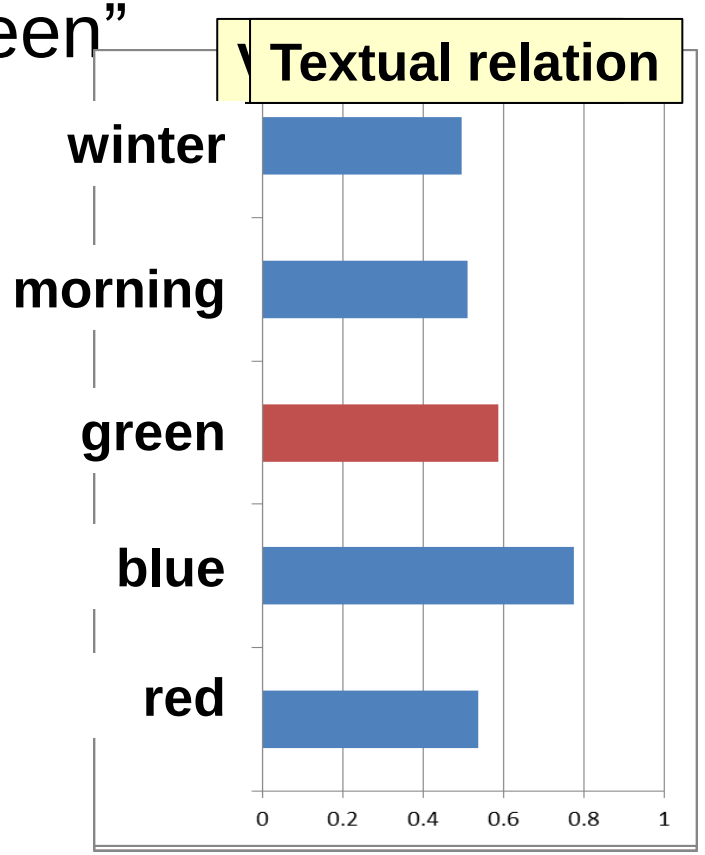
2. Compare visual relation and textual relation

- Low visual relationship

  ⇔ high similarity by textual relation

  – “beach+summer “, “sky+green”





Summary

- Analyze visual similarity between the tag pairs
- Calculate mutual information from the images of 360 tag pairs
  - High visual relationship

    The pair of color adjectives and object nouns

    The pair of time adjectives and related sky nouns
- Compare visual relation and textual relation
  - Exist the tag pair of Low visual relationship but High textual relationship

Future works

- Large scale experiments
  - e.g. 1000 nouns × 1000 adjectives
- Simultaneous recognition of the nouns and adjectives

Future works

dog

bag

red

cute

Summary

- Analyze visual similarity between the tag pairs
- Calculate mutual information from the images of 360 tag pairs
  - High visual relationship

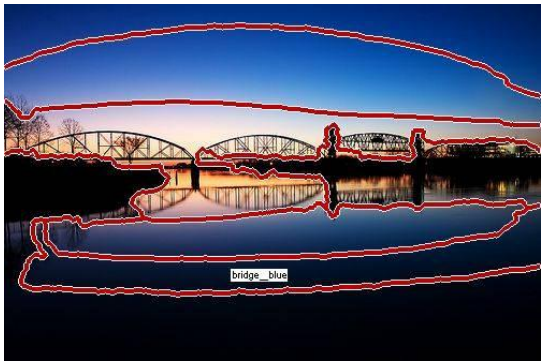
    The pair of color adjectives and object nouns

    The pair of time adjectives and related sky nouns
- Compare visual relation and textual relation
  - Exist the tag pair of Low visual relationship but High textual relationship

Thank you for your attention

Failed result

- blue + bridge



High visual relation

But

Most of bridge iare not blue

- An adjective modifies some nouns.

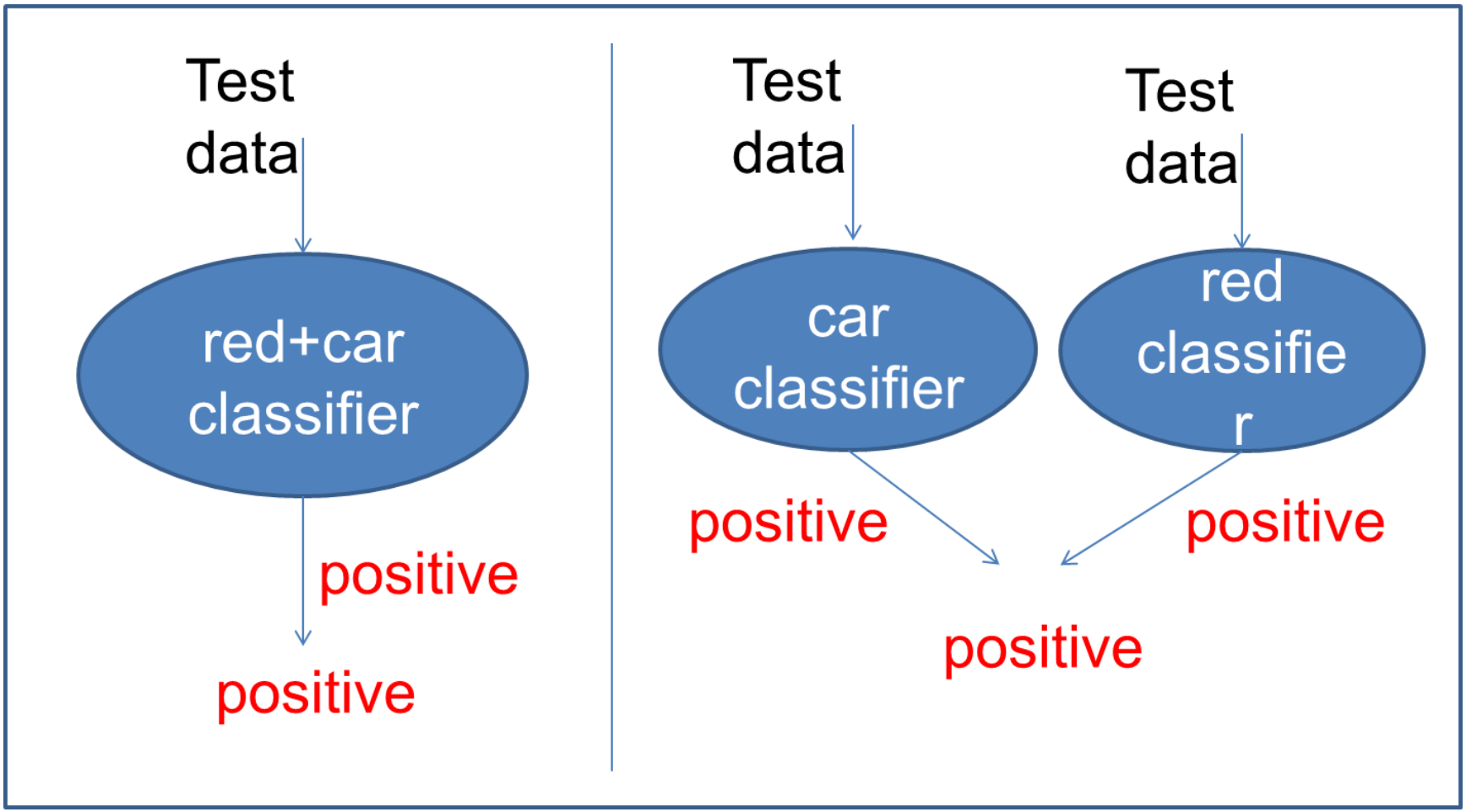
blue + car
and
blue + sky

blue + car
or
blue + sky

Future works

- There are two ways of simultaneous recognition.
  1. Prepare for each noun and adjective class
  2. Prepare the classes that combine nouns and adjectives
- Reduce the class number by the second method
- Use of mutual information as an index for it

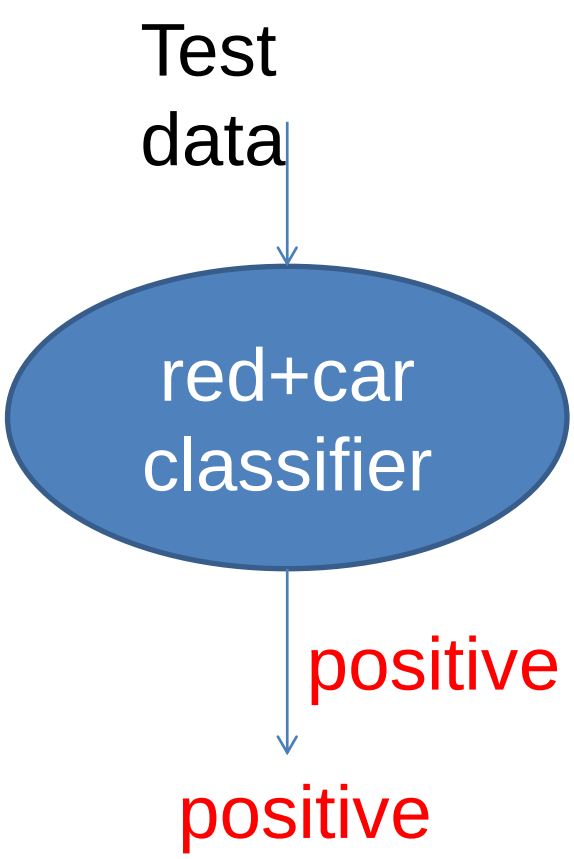
Future work

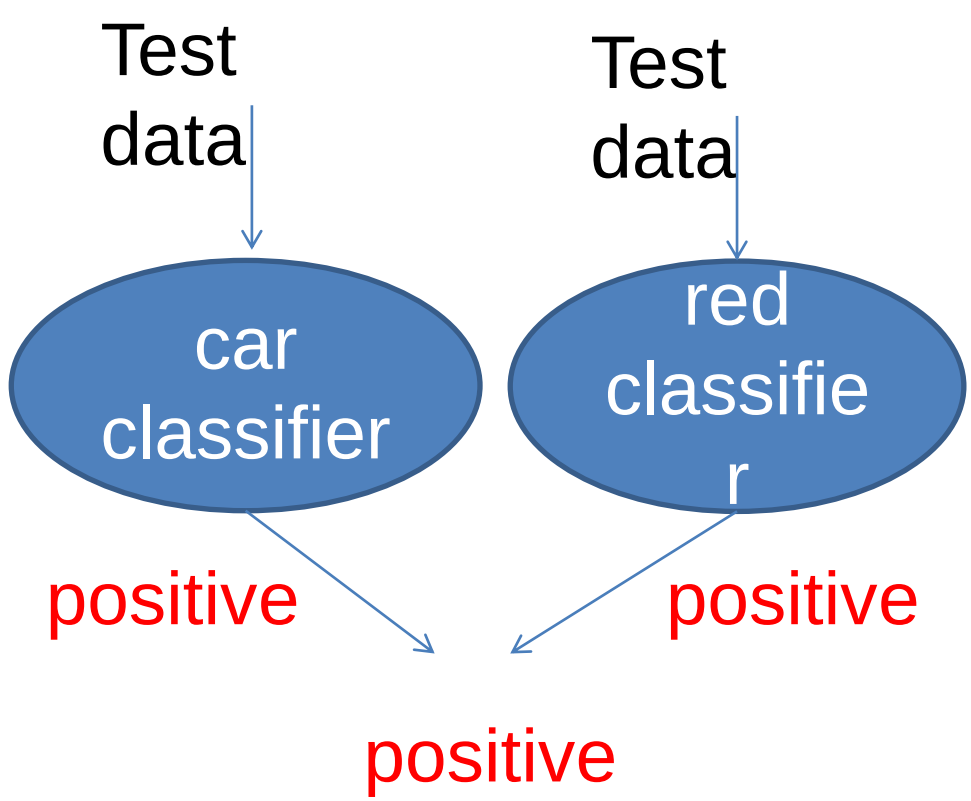


Minus of mutual information

- The noise included in the images, mutual information is in the negative.

Classification experiment





all regions

Results of Classification