A VISUAL ANALYSIS ON RECOGNIZABILITY AND DISCRIMINABILITY OF ONOMATOPOEIA WORDS WITH DCNN FEATURES

The University of Electro-Communications, Tokyo, Japan

Wataru Shimoda and Keiji Yanai

UEC TOKYO

100th ANNIVERSARY SINCE 1918

Summary

Objective: Examine the relation between onomatopoeia words and images using a large number of tagged images.

Approach: [1] Evaluate "recognizability" of onomatopoeia concepts by noise separation

[2] Evaluate "discriminablity" of onomatopoeia concepts within the same nouns

1 "Onomatopoeia" images

Onomatopoeia words in English
→ Source of the sound that it describes such as "tic tac" and "quack"

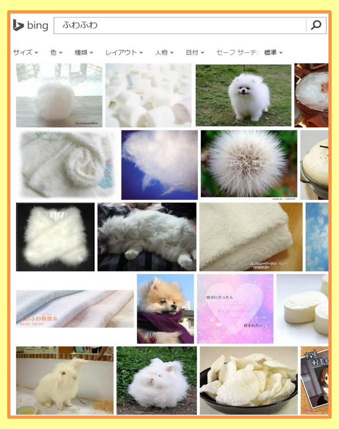
Onomatopoeia words in Japanese
→ Expressing feeling of visual appearance or touch of objects or materials.

Fuwa-fuwa means being very softy like very soft cotton

2 Collecting images from Web

- Use Bing Image Search API
- Upper-ranked images can be regarded corresponding to the given onomatopoeia.
- There are some noisy images.

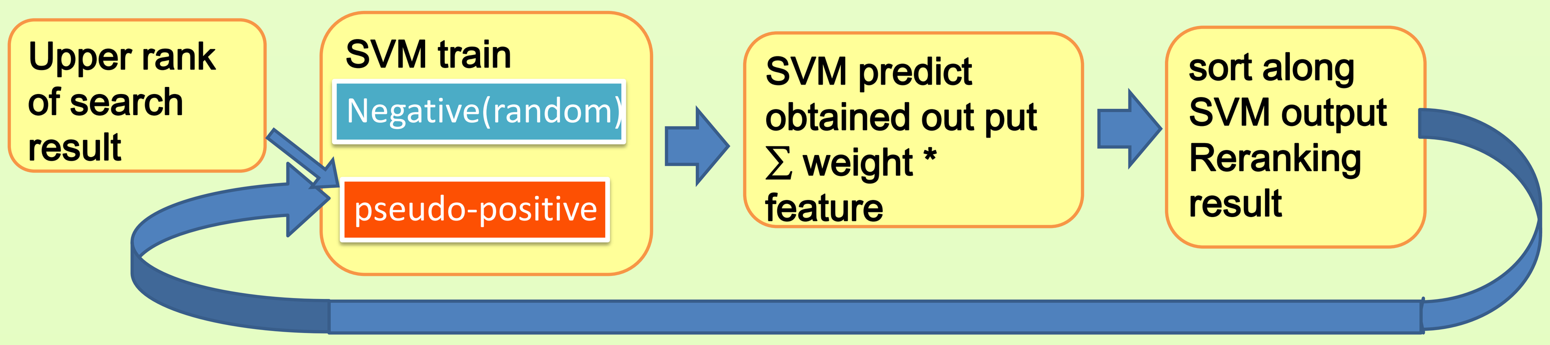
bing API



3 Re-ranking method

-Semantic annotation is a hard task due to ambiguous definition.
-Mine similar images to upper ranked images by recognition.

- Train SVM to the images in the original search results
- Sort images in the descending order of the SVM output values
- If needed, we can improve result by repeating re-ranking loop

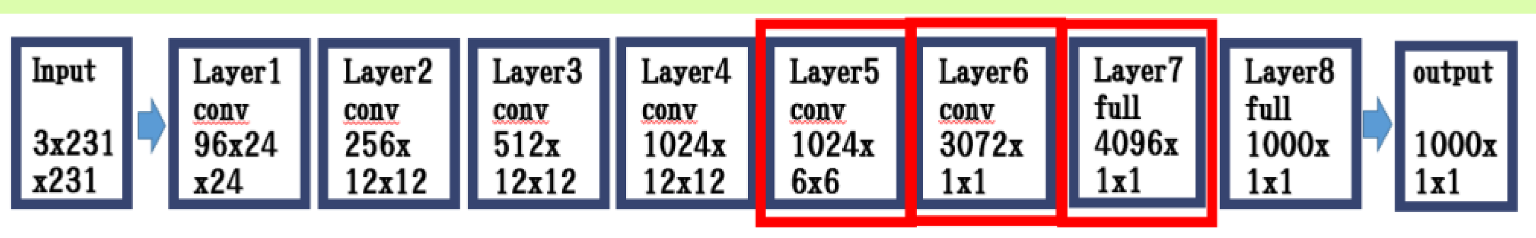


Re-ranking flow

Iterative re-ranking

4 Image features

- Improved fisher vector (IFV)
  SURF 64 dim(128 to 64 by PCA), GMM = 256
  Feature 2*64*256 dim
- Deep Convolutional Neural Network (DCNN) (Over feat)
  Pretrained with ImageNet of 1000 class (1,000,000 images)
  Use activation from layer 5,6,7 (36864, 3072 and 4096 dim)



5 Exp.(1) : Automatic building of dataset

-Collected Web images regarding twenty kinds of onomatopoeia words
-Achieved the average precision(AP) on the top 50 imgs were 89.6% (layer 6)

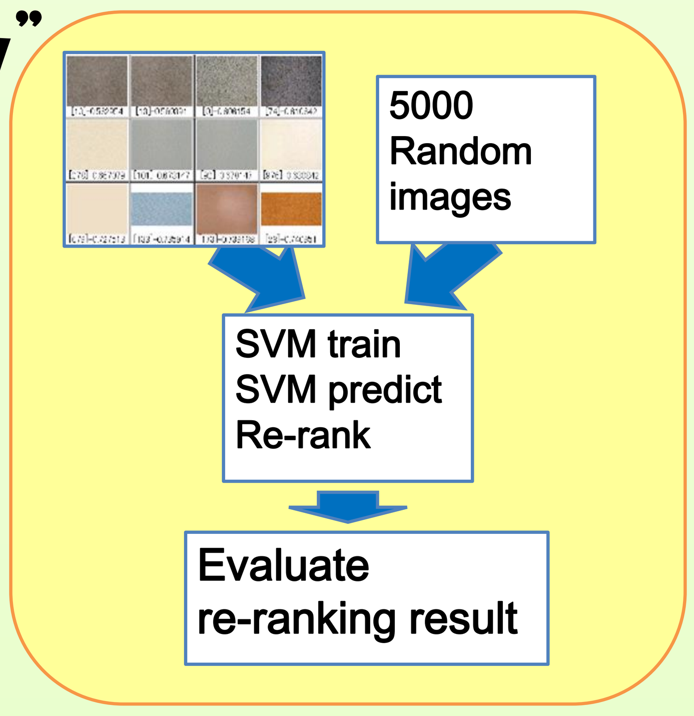
DNN features outperformed IFV clearly at image retrieval

onomatopoeia	meaning	onomatopoeia	meaning
pika-pika	shining gold	mofu-mofu	softly
basha-basha	splashing water	mock-mock	mountainous smoke or clouds
fuwa-fuwa	softly; spongy	kara-kara	hanging many metals
nyoki-nyoki	shooting up one after another	bou-bou	overgrown
kira-kira	shining stars	fuwa-fuwa	well-roasted
gune-gune	winding	shiwa-shiwa	wrinkled; crumpled
toge-toge	thorny; prickly	zara-zara	sandy; gritty
butsu-butsu	a rash	kari-kari	crispy; crunch
puru-puru	fresh and juicy	guru-guru	whirling
gotsu-gotsu	rugged; angular; hard; stiff	giza-giza	notched; corrugated

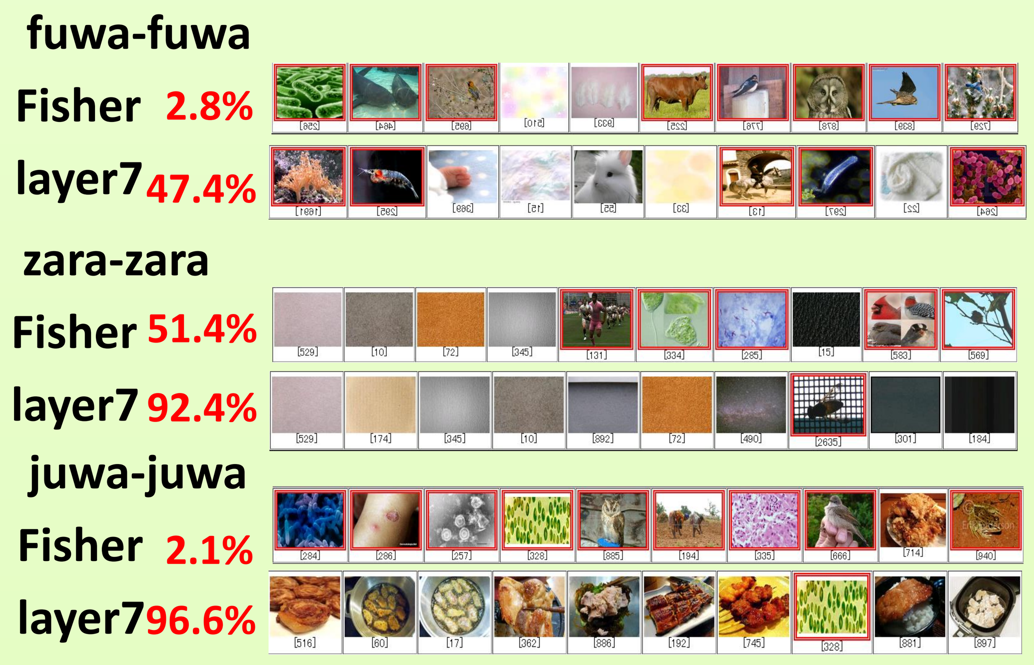
Feature	IFV	Layer5(DCNN)	Layer6(DCNN)	Layer7(DCNN)
mAP(%)	60.4	85.2	89.6	85.1

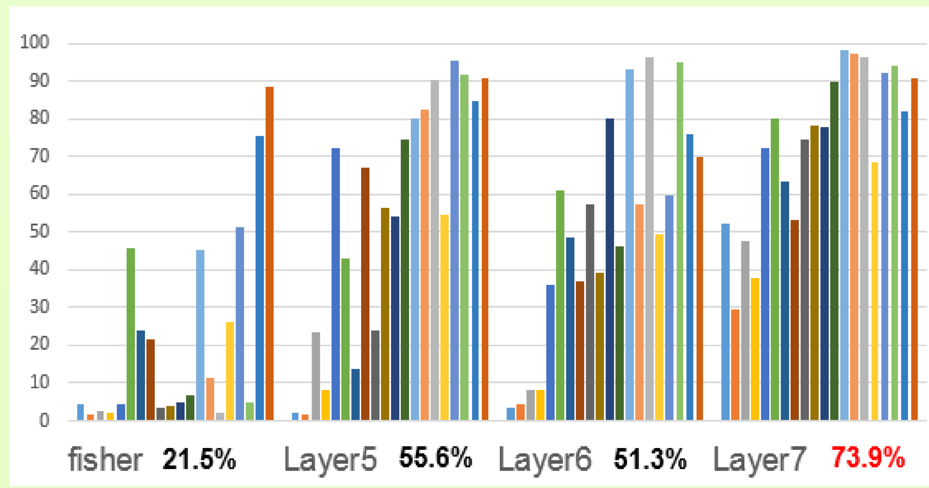
6 Exp.(2): "Recognizability"

- Discriminate onomatopoeia images from noise imgs
- Mix onomatopoeia 50 images and random 5000 noise images
- Re-rank 5050 mixed images and evaluate mAP



DCNN features has high ability to express visual onomatopoeia elements in images





Feature	IFV	Layer5(DCNN)	Layer6(DCNN)	Layer7(DCNN)
mAP(%)	21.5	55.6	51.3	73.9

7 Exp.(3): "Discriminability" : Fixed nouns and the different onomatopoeia words

- Some onomatopoeia words are strongly related to specific kinds of objects
- Examine visual discriminablity of onomatopoeia words with images associated with noun/onomatopoeia(adjective) pairs
- 4 nouns:"dog", "shoes", cake" and "flower"
- Pair words are selected based on text co-occurrence counts in Bing Img Search

Noun	#class	Rate(%)
dog	8	52.5
shoes	6	85.7
cake	7	72.3
flower	7	84.6

	confusion matrix							
"goro-goro" cake	31	3	10	2	3	1	0	62.0
"pasa-pasa" cake	3	26	3	7	5	6	0	52.0
"saku-saku" cake	8	4	24	8	2	3	1	48.0
"fuwa-fuwa" cake	1	2	2	42	3	0	0	84.0
smooth cake	3	2	3	3	37	2	0	74.0
deep cake	1	0	2	1	0	46	0	92.0
light cake	0	0	0	0	3	0	47	94.0
	66.0	70.3	54.5	66.7	69.8	79.3	97.9	72.3

	confusion matrix							
"pon-pon" flower	35	3	2	2	4	0	4	70.0
"fuwa-fuwa" flower	7	36	2	3	2	0	0	72.0
fresh flower	2	1	38	7	1	0	1	76.0
main flower	0	2	3	44	1	0	0	88.0
blue flower	1	0	0	0	49	0	0	98.0
yellow flower	0	0	0	0	0	50	0	100.0
red flower	3	0	0	0	3	0	44	88.0
	72.9	85.7	84.4	78.6	81.7	100.0	89.8	84.6

Onomatopoeia words have visual characteristics which can be discriminated even within the same object category