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



# 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

## "Onomatopoeia" images 5 Exp.(1) : Automatic building of dataset

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

Onomatopoeia words in Japanese  $\rightarrow$  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.

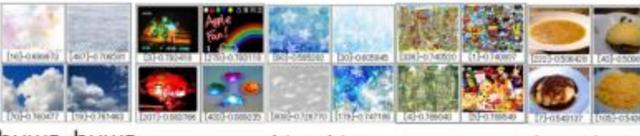


-Collected Web images regarding twenty kinds of onomatopoeia words -Achieved the average precision(AP) on the top 50 imgs were <u>89.6%</u> (layer 6)

## DNN features outperformed IFV clearly at image retrieval



toge-toge zara-zara gotu-gotu siwa-siwa



huwa-huwa kira-kira toro-toro pika-pika gotya-gotya

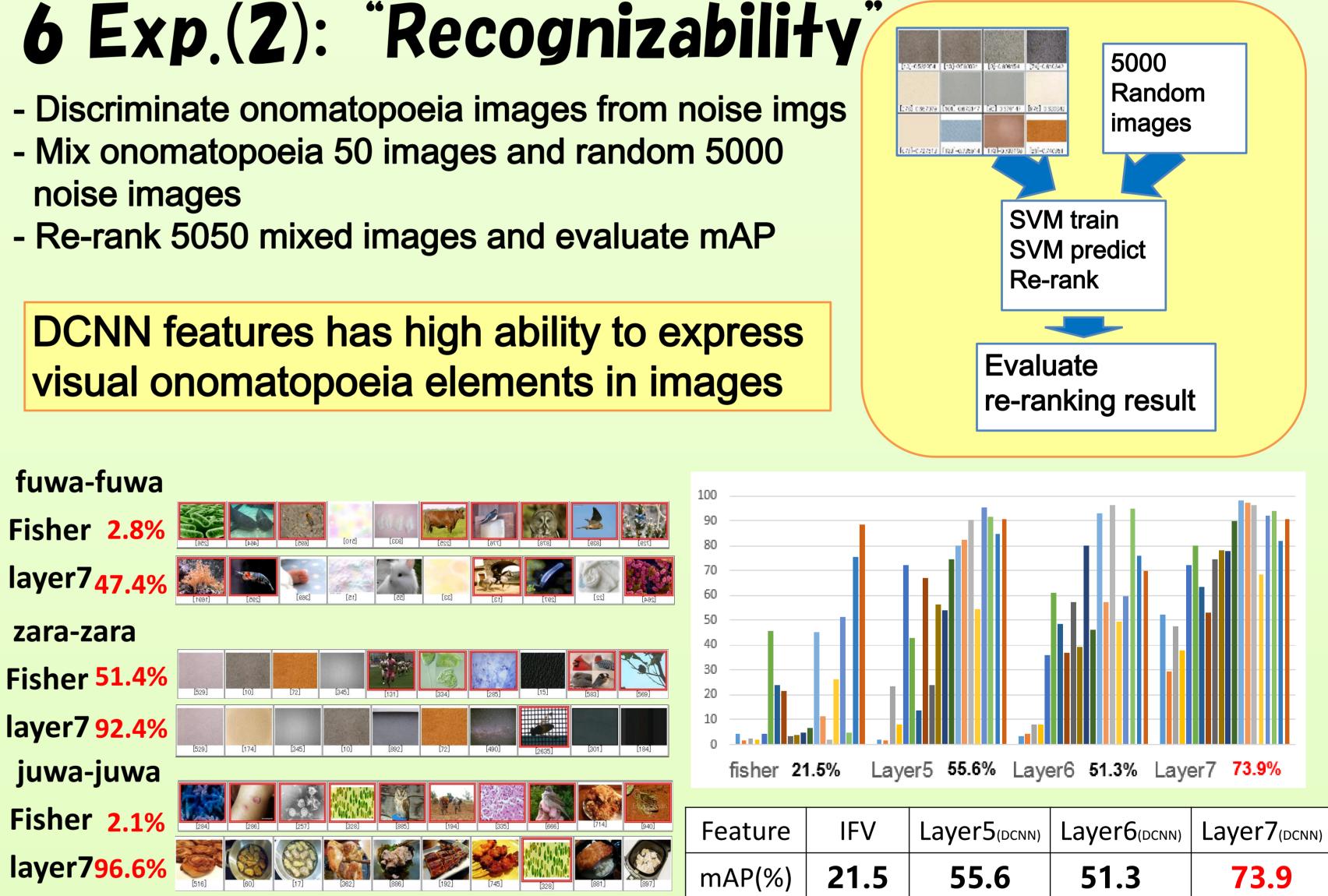
		_			
onomatopoeia	meaning	onor	natopoeia	meaning	g
pika-pika	shining gold	mo	ofu-mofu	softly	
basha-basha	splashing water	mo	ck-mock	mountain	ous smoke or clouds
fuwa-fuwa	softly; spongy	ka	ira-kara	hanging	g many metals
nyoki-nyoki	shooting up one after another	b b	ou-bou	overgro	wn
kira-kira	shining stars	fuy	wa-fuwa	well-roa	asted
gune-gune	winding	shiv	wa-shiwa	wrinkle	d; crumpled
toge-toge	thorny; prickly	za	ira-zara	sandy; g	gritty
butsu-butsu	a rash	ka	ari-kari	crispy; o	crunch
puru-puru	fresh and juicy	gu	ru-guru	whirling	g
gotsu-gotsu	rugged; angular; hard; stiff	gi	za-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



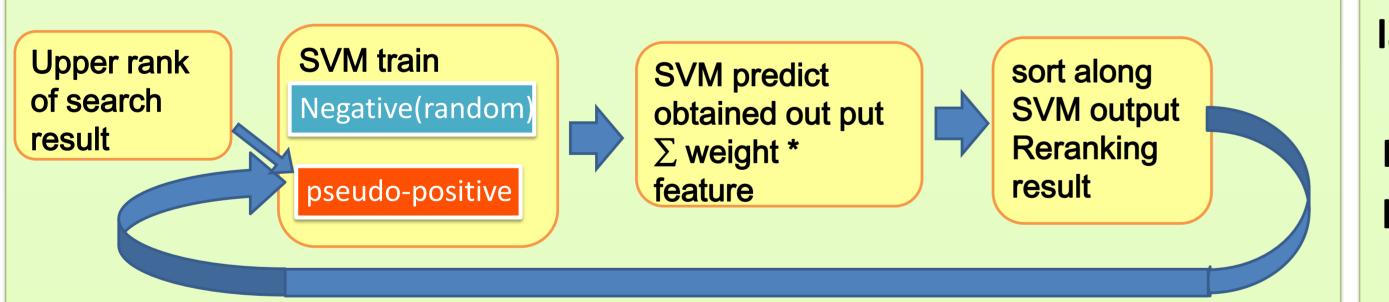
## 3 Re-ranking method



bing

-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 Fisher 51.4%



## **Re-ranking flow**



# 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"

confusion matrix

# 4 Image features

- Improved fisher vector (IFV) SURF 64 dim(128 to 64 by PCA), GMM = 256Feature 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)

Input 3x231 x231	•	Layer1 <u>conv</u> 96x24 x24	Layer2 conv 256x 12x12	Layer3 conv 512x 12x12	Layer4 conv 1024x 12x12	Layer5 <u>conv</u> 1024x 6x6	Layer6 <u>conv</u> 3072x 1x1	Layer7 full 4096x 1x1	Layer8 full 1000x 1x1	•	output 1000x 1x1
x231		x24	12x12	12 <b>x</b> 12	12x12	6x6	1x1	1 <b>x</b> 1	1x1		1x1

- Pair words are selected based on text	
co-occurrence counts in Bing Img Searc	;h

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

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

'fuwa-fuwa'' flower Image: Section of the section										
"saku-saku" cake Image: Saku set of the sake set	"goro-goro" cake	2 🖉 🧶	31	3	10	2	3	1	0	62.0
"fuwa-fuwa" cake Image: Simooth cake <td>"pasa-pasa" cake</td> <td>A 🔍 🔍</td> <td>3</td> <td>26</td> <td>3</td> <td>7</td> <td>5</td> <td>6</td> <td>0</td> <td>52.0</td>	"pasa-pasa" cake	A 🔍 🔍	3	26	3	7	5	6	0	52.0
smooth cake Image: Simple	"saku-saku" cake		8	4	24	8	2	3	1	48.0
deep cake Image: Second se	"fuwa-fuwa" cake	in 💭 💓	1	2	2	42	3	0	0	84.0
light cake Image: Second s	smooth cake	0 9	3	2	3	3	37	2	0	74.0
66.0 70.3 54.5 66.7 69.8 79.3 97.9 72.3   rpon-pon flower 33 35 3 2 2 4 0 4 70.0   fuwa-fuwa flower 34 70 36 2 3 2 4 0 4 70.0   fuwa-fuwa flower 36 7 36 2 3 2 0 0 4 70.0   resh flower 36 7 36 2 3 2 0 0 72.0   main flower 36 7 36 2 3 2 4 0 4 70.0   main flower 36 7 38 7 1 0 1 76.0   wellow flower 36 0 2 3 44 1 0 0 88.0   wellow flower 36 0 0 0 0 3 0 44 88.0   wellow flower 36 0 0 3 0	deep cake		1	0	2	1	0	46	0	92.0
Yoon-pon flower   Single Si	light cake		0	0	0	0	3	0	47	94.0
'pon-pon'' flower image: second s			66.0	70.3	54.5	66.7	69.8	79.3	97.9	72.3
'fuwa-fuwa'' flower Image: Sector			confusion matrix							
investigation investigation<	′pon-pon″ flower	Se Se Se C	35	3	2	2	4	0	4	70.0
nain flower Image: Market in the second	'fuwa-fuwa" flower	R 💽 🎗	7	36	2	3	2	0	0	72.0
Image: Second	resh flower		2	1	38	7	1	0	1	76.0
vellow flower Image: Second secon	nain flower	<b>新林</b> 德	0	2	3	44	1	0	0	88.0
red flower 3 0 0 0 3 0 44 88.0	lue flower		1	0	0	0	49	0	0	98.0
		100 100 100 to								
72.9 85.7 84.4 78.6 81.7 100.0 89.8 84.6	vellow flower			0	0	0	0	50	0	100.0
			0		-					

