Automatic Construction of Large Scale Image Data Set from Web Using Ontology and Deep Learning Model

Takahiro Yoshimura Yuki Mizoguchi Yuji Iwahori

Department of Computer Science
Chubu University
Kasugai 487-8501 Japan
Email: {yoshimura|mizoguchi}@cvl.cs.chubu.ac.jp
Email: iwahori@cs.chubu.ac.jp

Ryosuke Yamanishi

College of Information Science and Engineering
Ritsumeikan University
Kusatsu 525-8500, Japan
Email: ryama@media.ritsumei.ac.jp

***Abstract*—General object recognition requires a large scale image data set and recognition accuracy depends on the image data set used for the learning. This paper proposes a method to collect only the target images using Ontology and features obtained from obtained from Convolutional Neural Network (CNN). The proposed approach constructs large scale data set automatically by expanding the range of collected images. It is shown that the proposed approach is effective by collecting the image data set and a mean accuracy 88.6% was obtained from the subjective evaluation in experiments.**

Keywords–Image data set; Web image mining; Image processing; Web intelligence.

I. INTRODUCTION

General object recognition is the recognition of unconstrained image existing in the real world using computer system and this is one of the representative tasks in the computer vision. Since Convolutional Neural Network (CNN) [1] was proposed in 2012, the recognition ratio was improved dramatically. However, recognition requires a large scale image data set and the recognition accuracy depends on the image data set. Construction of large scale image data set by human requires a lot of time and human costs. Automatic or semi-automati construction of image data set using Web image mining searching images on the Web has been reported recently [2][3] to avoid the construction by human manually. Web image mining makes it possible to obtain large scale images taken under usual conditions by various humans with low cost since searching operation is available via posting service of a large scale images such as Flicker, Bing Image Search or Google Image Search and so on.

Only the Meta information added to the image, such as title, explanation sentence or tag is stil difficult to collect the target image data set. Image data collected automatically from the Web includes non-target images (noise images) and it is stil difficult to apply these approach directly to the general object recognition. This paper uses the low level concept of ontology and expands to increase the number of image data set and to exclude the noise images simultaneously, then how to perform the automatic construction of a large scale image data set is proposed.

II. PROPOSED APPROACH

The proposed approach constructs a large scale image data set as follows.

Step 1 Collecting image data from Web
Step 2 Recollecting image data using the low level concept of ontology
Step 3 Integrating image data set by excluding the collected duplication images
Step 4 Removeing the colleted noise images

The detail is shown below.

A. Collecting Images from Web

In this section, the Step 1) will be shown. Flickr is used to collect images as a collection method. Flickr is available with keyword for searching images. Related tag is added to the uploaded image by user. Tag consists of words related to the contents of image. The proposed approach searches the related images for the added tag by using the label of constructed data set as the searching query.

B. Recorrecting Images by Low Level Concept of Ontology

In this section, the Step 2) will be shown.

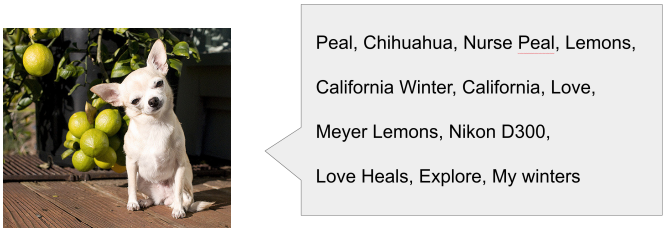

Figure 1. Tags added to Image

Image search by Flickr is applied according to the tags addef by user. When ”dog” is given as the label of image data set to construct, there are sometimes no hits even if ”dog” is the target label as shown in Figure 1. However, ”Chihuahua” is sometimes the tag added to the image, which is the low level concept of ”dog”. Recorrecting images using this low level concept of the label is used for this purpose to expand the range of correcting images. This paper uses DBpedia[4] as ontology and perform this strategy.

C. Excluding Duplication Images and Integrating Image set

In this section, the Step 3) will be shown. Integrating the corrected images is done by Step 1 and Step2. Corrected images sometimes include the case that both of the target label

 ISBN: 978-1-61208-612-5

and the target low level concept label are added to the image. In this case, images are corrected with duplication and processing to exclude the duplication images is necessary.

D. Removing Noise Images

In this section, the Step 4) will be shown. There is the case that noise images are included in the image corrected from Web. Noise images are inappropriate for the search query and removing noise images are necessary for the general object recognition.

Feature extration is applied to the image i which was corrected from the Web by the search query q. Here, let the number of the corrected images by the search query q be NI. CNN is used as the image feature extractor, and VGG16 [5] is used as the CNN model. 4096 dimensional feature vector obtained from fc2 layer of VGG16 was used as image features $G(i)$. Euclidian distance $FD^q(i)$ was obtained between each image feature vector $G(i)$ and those centroid vector M^q of all images.

$$\boldsymbol{VI}^q = \begin{cases} \text{Appropriate Image} & FD^q(i) <= VT^q \\ \text{Noise Image} & FD^q(i) > VT^q \end{cases} \quad (1)$$

$$VT^q = \frac{1}{NI}\sum_{i=1}^{NI} FD^q(i) \quad (2)$$

where threshold value VT^q is given by the mean value of $FD^q(i)$ of each image feature vector and entroid vector.

III. Experiment

Subjective evaluation was applied for whether the proposed approach adds the appropriate label for the corrected image data set or not. Image data set was constructed by correcting from Flickr and labels used for the evaluation were "cat", "crab", "elephant", "fox", "giraffe", and "lion" as the general words. 100 images are randomly extracted from the constructed image data set and 6 evaluators evaluated whether images in the constructed data set are matched for the object labels or not.

Constructed image data set is shown in Figure 2 and evaluation is shown in Table I.

TABLE I. RESULT BY SUBJECTIVE EVALUATION [%]

Search Keyword	Precision
cat	93.3
crab	94.0
elephant	88.3
fox	79.6
giraffe	79.5
lion	96.7
AVG	88.6

Table I suggests that "cat", "crab", and "lion" gave more than 90 % accuracy. "crab" included some cuisine images but noise images were removed correctly. "fox" corrected some inappropriate images including building or humans. "giraffe" was judged as zebra or kangaroo and this decreased accuracy.

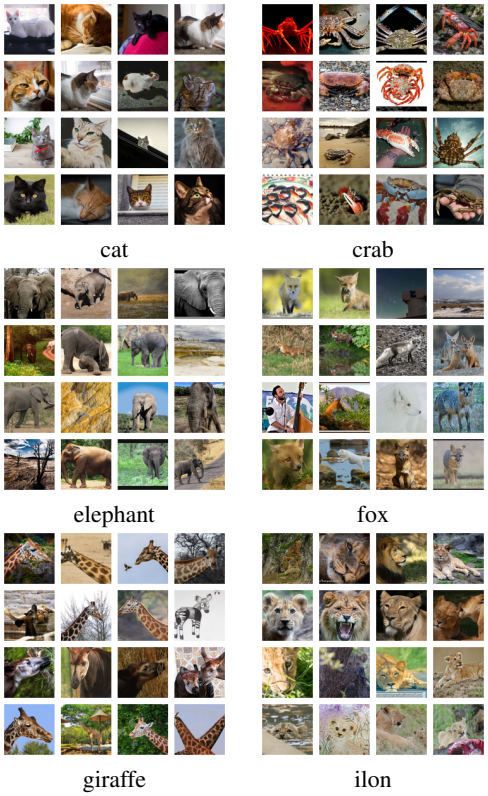

Figure 2. Example of Constructed Image Data Set

IV. Conclusion and Future Work

This paper proposed an automatic construction of Web image dataset by removing noise images using ontology and CNN features. Low level concept of ontorogy made it possible to recorrecting images and expand the range of correcting images. Removing noise images was also applied using the image features obtained by CNN for the corrected images. It was confirmed that image data set constructed by the proposed approach is available to the general object recognition. Evaluation using this generated dataset for the general object recognition is our future task.

Acknowledgment

Iwahori's research is supported by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid Scientific Research(C)(#17K00252) and Chubu University Grant.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, pp. 1097-1105, 2012.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009.

[3] S.Otani, R.Yamanishi, Y.Iwahori, "Generation of Web Image Database Based on Hybrid Noise Removal Method of Visual and Semantic Features", Journal of JSAI (Japanese Society for Artificial Intelligence), vol. 32, no. 1, WII-N 1-10, 2017.

[4] J. Lehmann, et al. DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web Journal, 2014.

[5] K.Simonyan,A.Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv 2014.

 ISBN: 978-1-61208-612-5