**Content-Based Image Selection for Automatic Report Generation**

Jeongwook Choi, Jeongin Seo, Hyeyoung Park

Dept. of Computer Science and Engineering, Kyoungpook National University

Daehak-ro 80, Buk-gu

Daegu, Korea

cho2oi@knu.ac.kr ,lain@knu.ac.kr, hypark@knu.ac.kr

**Abstract**

**Recently, with the rapid development of internet environment and smart devices, a lot of images and information are being produced and shared quickly. The images are uploaded with several keyword tags, and people can get a necessary image through search with text query. However, acquiring images through such a conventional search engine simply depends on the given tag information, and thus they cannot always provide images suitable for the intention of users. In this paper, we propose a method to recommend appropriate images for a given report subject by evaluating the relevance of the images searched with the keyword text and the suitability of the images for the report document. As a basic experiment to see the applicability of the proposed method, we show that the proposed method can provide better image selection results than the naïve method of using conventional search engine.**

Keywords- Image recommendation; Automatic report generation; Image search and reranking; Auto-tagging; Deep learning

**I. Introduction**

According to development of internet and big data technologies, the amount of information that people can easily obtain has rapdly increased, and the studies on extracting and summarizing meaningful information from the huge amount of data have been been actively conducted [1]. Moreoever, as a more challenging topic, the automatic report generation [2], which is about automatically generates a report format document containing text, images and video on a given subject, is also getting attention in the field of big data and artificial intelligence.

In this paper, we propose a method to recommend appropriate images for a given report subject by evaluating the relevance of the images searched with the keyword text as well as the suitability of the images for the report document. In order to achieve the goal, we exploit conventional image retrieval techniques and improve the search result trhough analyzing image contents by using image auto-tagging method based on deep learning technologies.

Image retrieval, which is to find queried images from an image database, can be categoried two classes according to the types of querys: the text-based search and the content-based retrieval. In the text-based search, a query is given as text keywords, and the search engine tries to match the given keywords with the tag information that are provided along with the stored images. The well-known image search services such as google image search takes this approach. Although this method can retrieve queried images very efficiently, from huge database its result completely depends on the tag made by the image provider, and thus it does not consider the whole contents in the images as well as other characteristics and qualities of the images.

On the other hand, in the content based image retrieval, a query is given as an image, and it is required to search other images that have similar contents to the query [3]. The main difficulty of this method is to find a good similarity measure between the query image and another one in database, which need to consider the contents of the images. Recently, deep learning techniques have provided successful solutions to find good image features an similarity through learning. However, since the calculation of the similarity needs to be conducted on each images in the database, it is practically implementable only for a limited size of database therefore this approach is not suitable for image recommendation for automatic report generation.

Based on the considerations, in order to recommend suitable images for the given report keyword, this paper proposes a hybrid method which combines the text-based search and image content analysis by deep learning. Details of the proposed methods are described in Chapter 2.

**II. Proposed System**

*A. Overall Structure*

The overall structure of the proposed image recommendation system is shown in Figure 2. The system first gets a specific keyword related to the report subject as an input, and apply it to the conventional text-based image search engine so as to get a set of candidate images. By using the popular search engine such as Google, we can easily obtain various candidate images from a plenty of internet resources. However, as we mentioned earlier, the search research is mainly depends on the subjective tags made by image provider, and thus we need further analysis to evaluate the appropriateness of the image for the report to be generated. The proposed image re-ranking module gets the set of image candidates as an input, analyzes the image contents by using deep neural networks, and evaluate the images to make modified rank of the candidates. The detail description of the CBE (content-based evaluation) module, which is the core of this paper, will be given in the next section.

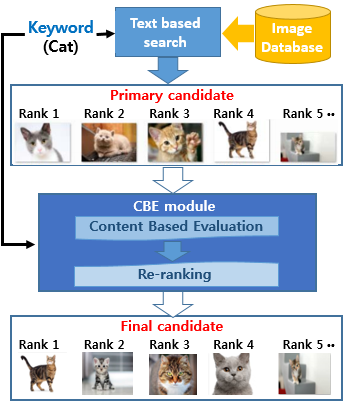


Fig. 1 Overall structure of the proposed method

*B. Content Based Image Evaluation*

Fig. 2 shows the structure of the proposed CBE module, which analyzes the contents of each candidate image and calculate evaluation score for re-ranking the candidate list. In order to analyze the contents of the images, we utilize a deep neural network model that is designed and trained for recognizing various objects in the images [4]. By applying each image to the deep neural network, we can get the recognition results, which are the recognized object tags and corresponding confidence value.

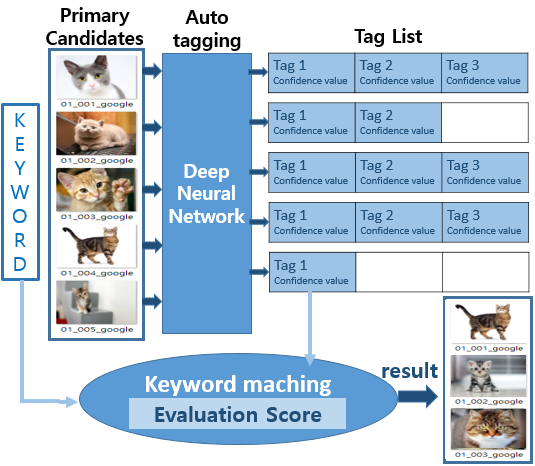


Fig. 2 Structure of CBE module

In order to evaluate the appropriateness of the image based on the tags that are automatically made by deep neural networks, we need to calculate the similarity between the object tag and the given keyword. To obtain the similarity, we exploit word hierarchy and related simiarlity measure provided by WordNet [6-7]. The evaluation score for each tag is calculated by the multiplication of the tag-keyword similty and the tag confidence value from deep network, and The total evaluation score for each image is then obtained by summation of the evaluation scores of all the tags made for the image. Finally, the candidate images are re-ranked based on the evaluation scores.

**III. Experimental Results**

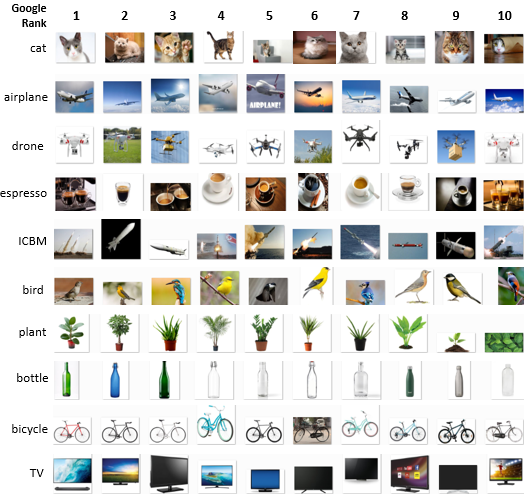


Fig. 3 10 candidate images searched for 10 keywords (by Google)

In order to check the practical applicability of the proposed method, we conducted a preliminary experiments for 10 different keywords. For each keyword, we first apply Google image search and obtained top 10 images as the candidate image set. Fig 3 shows the 10 keywords and corresponding 10 images obtained through Google search.

The proposed CBE module are then applied to re-rank the candidate list. The trained deep network model used for auto-tagging of the object labels was the convolution neural network (CNN) – based model called ResNet-152 [4-5], which is introduced in the ImageNet Competition 2015. The network is composed of 152 layers and can classify 1000 object classes. For each given input, ResNet-152 returns one or more object tag and its correpoding confidence values. (Fig 4 shows the first tag with highest confidence value of each image.). For the calculation of tag-keyword matching similarity, we used the Wu & Palmer similarity measure (wup function) [6] provided by WordNet 3.0[7]. The final evaluation score of each image is then obtained by adding all the tag evaluation scores for each image, which are calculated as the multiplication of tag-keyword similarity and tag-confidence. By re-ranking the 10 candidate images based on the evaluation score, we choose top three images as the recommendation of the proposed system.

In Fig. 4, we compared the original google search results and the recommendation results of the proposed methods. From the results, we can say two promising points. First, as shown from the results for “espresso”, “airplane”, and “bird”, the proposed system gives higher scores to the images with bigger object and/or clearer background. Also, from the case of “cat”, we can also say that the proposed evaluation system prefers more typical shape of the objects. Second promising thing is that the proposed system can give plausible recommendation even when the given keyword is not exactly matched with the object tag of the deep neural networks. For example, it can operates on the new vocabulary such ICBM by utilizing hierarchy in the WordNet. From the results, we can say that the proposed method can provide better image recommendation for report generations, compared to the naïve method using conventional search engine.

Fig. 4 Experimental Results on each keywords

**IV. Conclusion**

In this paper, we propose a content based image selection method to recommend appropriate image candidates for a given subject in order for automatic report generation. From a prepliminary experiment, we could get promising results that the proposed CBE module gives high evaluation score on the images with appropriate size and position of the objects and thus improve the recommendation quality. This research is still in the starting stage, and will be improved through further works such as retraining deep network models for larger number of object classes.

**Acknowledgment**

This work was partially supported by ICT R&D program of MSIP/IITP. [ R7124-16-0004, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding ]. This work was partially supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.2016-0-00145,Smart Summary Report Generation from Big Data Related to a Topic)

**References**

[1] Zhong, S. H., Liu, Y., Li, B., & Long, J. (2015). Query-oriented unsupervised multi-document summarization via deep learning model. Expert Systems with Applications, 42(21), 8146-8155.

1. Zhang, J., Li, X., Nie, W., & Su, Y. (2017). Automatic report generation based on multi-modal information. Multimedia Tools and Applications, 76(9), 12005-12015.
2. Liu, P., Guo, J. M., Wu, C. Y., & Cai, D. (2017). Fusion of Deep Learning and Compressed Domain Features for Content-Based Image Retrieval. IEEE Transactions on Image Processing, 26(12), 5706-5717.
3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
4. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 675-678). ACM.
5. Wu, Z., & Palmer, M. (1994, June). Verbs semantics and lexical selection. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics (pp. 133-138). Association for Computational Linguistics.
6. Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004, May). WordNet:: Similarity: measuring the relatedness of concepts. In Demonstration papers at HLT-NAACL 2004 (pp. 38-41). Association for Computational Linguistics.