ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

Article Annotation Using Image and Text Processing

Prof. Nilam Shaikh[1], Arsha Shyamsundar[2], Kajal Kadulkar[3], Satish Kasar[4], Heramb Limaye[5], Prof. Pratibha Rane[6]

U.G. Students, Dept. of Computer Engineering, S.S.P.M's College of Engineering, Kankavli, Maharashtra, India[2,3,4,5]

Assistant Professor, Dept. of Computer Engineering, S.S.P.M'sCollege of Engineering, Kankavli, Maharashtra, India[1,6]

ABSTRACT: In earlier days tasks in language processing like image captioning and automatic image retrieval were never seen before approaches. These language processing task rely on large training sets of image associated with human annotations that specifically describe the visual content. In this paper we propose to explore more complex cases of language processing were textual descriptions are loosely related to images. We focus on various domains of articles in which the textual contents often expresses loose relation that are not directly inferred from images .We introduce architectures and structures for multiple tasks including article annotation, source detection, text annotation, caption generation and image retrieval.We show this processing to be appropriate to explore earlier problems,for which we can provide baseline performance using various learning architectures and different representation of the textual and visual features. We deliver very appreciable results and overcome several limitations of current approaches of this kind of domain, which we hope it will help to develop more progress in this field.

KEYWORDS: Image captioning; source detection; caption generation; image retrieval; visual and textual features.

I. INTRODUCTION

From last few years, there has been a lot of interest in exploring relation between natural language and images. At the same time there was growing advancement in the field like natural language processing (NLP) which led to appealing results in learning both text-to-image and image-to-text relations. Techniques like automatic image captioning, image generation or image retrieval from sentences has shown never seen before results.

Due to popularity of crowd sourcing tools the production of images based on visual and language contents has been easier. Moving towards annotation, it is usually short and correct sentences which describe the visual contents of the image or action taking place in image. The difference in this approach is that previously the complicated types of documents have been hardly explored. We believe in present success in the field of NLP and image captioning is mature enough for more difficult objectives than those posed by existing approaches.

In this paper we propose several different schemes, especially for source detection, text annotation and image retrieval. In order to evaluate these approaches, these tasks should be combined and performed to get appropriate results. For this we collect different articles with large data and the same is provided with one to four images with their captions. The articles are collected from various sources through internet. The overall results are very appreciable but there can be still few improvements, and we hope it will be our future research in this field.

II. LITERATURE SURVEY

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei of Dept. of Computer Science, Princeton University, USA."Imagenet: A large-scale hierarchical image database"[1], illustrated the usefulness of ImageNet through three simple methods in object recognition, image classification and automatic object clustering.In

DOI: 10.15680/IJIRCCE.2018.0603074

which they show that ImageNet is much larger in scale and diversity and much more accurate than the current image datasets.Constructing such a large-scale database is a challenging task.The main disadvantage of this method is aligning the cognitive hierarchy with the “visual” hierarchy also remains an unexplored area.

J. Donahue, L. Hendricks, S. Guadarrama, M. Rohrbach, S. Venu- gopalan, . Saenko, and T.Darrell. "Long-term recurrent convolutional networks for visual recognition and description", In CVPR, 2015[2].In which they described a class of recurrent convolutional architectures which is end-to-end trainable and suitable for large-scale visual understanding tasks, and demonstrate the value of these models for activity recognition, image captioning, and video description.They need to improve upon methods which learn a deep hierarchy of para fixed visual representation of the input and only learn the dynamics of the output sequence.Tasks with static input and predictions, deep sequence modeling tools like LRCN vision systems for problems with sequential structure were barely explored.

Chang, M. Savva, and C. Manning. "Interactive learning of spatial knowledge for text to 3d scene generation". Sponsor: Idibon, page 14, 2014[3].They presented an interactive text to 3D scene generation system that learns the expected spatial layout of objects from data. A user provides input natural language text from which we extract explicit constraints on the objects that should appear in the scene.User interaction is essential for text to scene generation since the process is fundamentally under-constrained. Most natural textual descriptions of scenes will not mention many visual aspects of a physical scene. However, it is still possible to automatically generate a plausible starting scene for refinement. They need to improve the system by incorporating more feedback mechanisms for the user, and the learning algorithm.

Amir R. Zamir,Tilman Wekel,Jitendra Malik,Pulkit Agrawal,Silvio Savarese,Colin Wei of Stanford University,University of California, Berkeley."Generic 3D Representationvia Pose Estimation and Matching"[4]. They learned a generic 3D representation through solving a set of foundational proxy 3D tasks: object-centric camera pose estimation and wide baseline feature matching.Their method is based upon the premise that provide supervision over a set of carefully selected foundational tasks, generalization to novel tasks and abstraction capabilities,they developed independent semantic and 3D representations, but investigating concrete techniques for integrating them (beyond simplistic late fusion or ConvNet fine-tuning) is a worthwhile future direction for research.

III. PROPOSED SYSTEM

This paper is concerned with the task of automatically annotating the article, generating captions for images and retrieval of appropriate images. Here we promise to explore more complex types of articles. Our approach leverages the vast resource of pictures available on web and the fact that many of them are loosely related to article. Annotation model that suggest keywords for an image search and retrieval of meaningful images from set of large collection of images.

In most of the cases, the textual descriptions are loosely related to images where we will focus on the different domains of articles where textual content often expresses connotative and ambiguous relations between given textual query and images associated with it.

Our proposed system reduces the task of searching the particular information about the article. Using the article annotation algorithm the user can know the keywords for searching and visualizing the contents of articles. Any type of the user can use our system efficiently as the task of article annotation, image captioning and image retrieval is done automatically ones the article is entered or browsed.

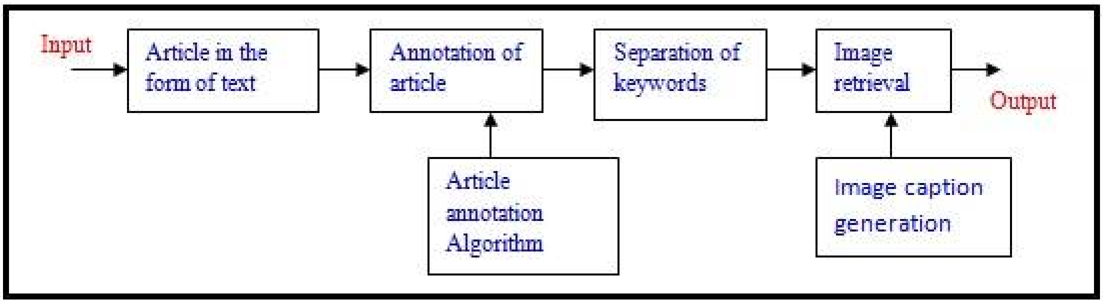

Fig1: Block diagram for article annotation and image retrieval.

IV. TASK DESCRIPTION

Next we describe the tasks that we will be dealing in this paper.

A. *Source Detection*

Here in source detection the task is to analyze details of the articles and to find in which agency it was formerly issued. Each article need to be carefully analyzed in order to extract the minute details. The problem can be addressed by doing some kind of refinements.

For now we are not aware of any other technique to detect the source of articles, but there exists a huge amount of related works, which is mostly motivated by the detection of plagiarism or by tackling the authorship identification of the problem.

B. *Text Annotation*

Basically annotation is summarization or evaluation of any kind of story, article, survey reports or opinion piece. Here in this paper we will be annotating the articles so that we can retrieve the appropriate images associated with that article. The annoted articles may contain important keywords like, names of people, location information, incident or event occurred etc. Keywords may vary depending on the types of articles.

Outcome of annoted article is short test description, now the image captioning strategies and retrieval of image could be carried out.

Algorithm for article annotation:

```
1. Start
2. Read an article.
3. If article has more tokens then
        a. Read next string.
        b. Check if it is token.
        c. Replace out it with blank space.
        d. Update an article.
        e. Go to step no.2.
4. Display the result.
```

C. *Caption Generation*

Along with the problem of tackling text and image relation, automatic caption generation is the one receiving most focus. Image captioning has made significant progress from last few years. Previously the results of automatic caption generation was outstanding, but they were all focused only on describing the visual content of the image which doesn't generate impact needed as it should have a caption that will illustrate the contents of article related to image.

The image is captioned with short description, which differs from existing image captioning aspects, as the images needs to be mapped with the article. There need to be focus on both the images and the articles. The corresponding caption may be sometimes loosely related to the image, this is one of the main challenge posed by our paper.

D. *Image Retrieval*

The annoted article and the caption accompanying image are mapped, based on the relation between the annoted article and caption of image, one to four images is retrieved from set of large collection of images.

V. REPRESENTATION

A. *Text Representation*

This representation encodes every word in a real-valued vector that preserves semantic similarity, e.g. "king" minus "man" plus "woman" will be close to "queen" in the embedding space. Words are modeled based on their context defined as a window that spans both past and future words. Method have been proposed to learn their representation: article annotation algorithm, where the objective is predicting a word given in its context, continuous skip-grammar model.

B. *Image Representation*

Our image representation is based on annotated articles and keywords separated from the articles. Images are represented in the form that the articles could be easily mapped to it. Images are shown to user to elaborate the article in more detail form. According to the article, one to four images are retrieved.

VI. CONCLUSION

This paper introduces annotation of articles using image and text processing. The objective is to map the article to appropriate images for better understanding and to visualize the contents of articles. In order to combine several tasks in the domain of articles, source detection and automatic caption generation are used for the illustration of articles. The automatic caption generation task, however, is clearly more sensitive to loosely related text and images. Designing larger directory of articles and images able to handle this situation is part of our future work.

ACKNOWLEDGEMENT

We would like express deep sense of gratitude to Prof. Nilam. N. Shaikh and Prof. P. S. Rane for their valuable guidance during the preparation of our project work without which completion of this task would be a difficult task. We are highly indebted to them for constantly encouraging us by giving their critics on our work.

We are also grateful to all our faculty member of Computer Department especially our head of department Prof. D. P. Mhapasekar and our respected principal Dr. S. R. Sange who give us idea of significant cooperation and confidence during completion of this work.

We are also immensely thankful to our parents for inspiring and supporting us always.

REFERENCES

1. J. Deng, W. Dong, R. Socher, K. Li, K. Li, and L. Fei-Fei, “ Imagenet:A large-scale hierarchical image database”, In CVPR, pages 248–255, 2009.
2. J. Donahue, L. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan,. Saenko, and T.Darrell, “ Long-term recurrent convolutional networks for visual recognition and description”, In CVPR,2015.
3. A. Chang, M. Savva, and C. Manning, “ Interactive learning of spatial knowledge for text to 3d scene generation”, Sponsor: Idibon,page 14, 2014.
4. Amir R. Zamir,Tilman Wekel,Jitendra Malik,Pulkit Agrawal,Silvio Savarese,Colin Wei, "Generic 3D Representation via Pose Estimation and Matching", University of California, Berkeley, (June):831–839, 2010.
5. Arnau Ramisa, Fei Yan, Francesc Moreno-Noguer, and Krystian Mikolajczyk, ” Breaking News: Article Annotation by Image and Text Processing.” The journal of pattern analysis and machine intelligence, IEEE, 2017.
6. K. Barnard, P. Duygulu, D. Forsyth, N. De Freitas, D. Blei, and M. Jordan, ” Matching words and pictures,” The Journal of Machine Learning Research, 3:1107–1135, 2003.
7. K. Barnard and D. Forsyth.” Learning the semantics of words and pictures”, In ICCV, volume 2, pages 408–415. IEEE, 2001.
8. X. Chen and C. Zitnick. “Mind’s eye: A recurrent visual representation for image caption generation”, In CVPR, 2015.
9. B. Coyne and R. Sproat.” Wordseye: an automatic text-to-scene conversion system.”, In Conference on Computer Graphics and Interactive Techniques, pages 487–496. ACM, 2001.
10. M. Douze, A. Ramisa, and C. Schmid. “Combining attributes and fisher vectors for efficient image retrieval”, In CVPR, pages 745–752, 2011.
11. H. Fang, S. Gupta, F. Iandola, R. Srivastava, L. Deng, P. Dollar, J. Gao, X. He, M. Mitchell, J. Platt, C. Zitnick, and G. Zweig. From “captions to visual concepts and back”, In CVPR, 2015.
12. A. Farhadi, M. Hejrati, M. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth.” Every picture tells a story: Generating sentences from images,” In ECCV, pages 15–29. Springer, 2010.
13. G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. Berg, and T. Berg, “Baby talk: Understanding and generating image descriptions.” In CVPR. Citeseer, 2011.
14. Y. Feng and M. Lapata. “Topic Models for Image Annotation and Text Illustration”, Conference of the North American Chapter of the ACL: Human Language Technologies, (June):831–839, 2010.
15. Feng and M. Lapata. Automatic caption generation for news images,PAMI, 35(4):797–812, 2013.

DOI: 10.15680/IJIRCCE.2018.0603074