IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 7.301

Volume 3, Issue 2, March 2023

Multi Task Learning for Captioning Images with Novel Words

Sreekantha B[1], Saniya Sultana[2], Shabreen Taj[3], Shikhar[4], Tasmiya Khanum[5]

Associate Professor, Department of Information Science and Engineering[1]
Students, Department of Information Science and Engineering[2,3,4,5]
HKBK College of Engineering, Bangalore, Karnataka, India
shreekantha.is@hkbk.edu.in, saniya544505@gmail.com, shabreenshabu240@gmail.com,
shikhar0055@gmail.com, khntasmiya@gmail.com

Abstract: *In this article, we introduce a Multi-task Learning Approach for Image Captioning (MLAIC), which is inspired by the fact that people can readily finish this job given their proficiency in a variety of fields. There are three crucial components that make up MLAIC in particular:(i)A multi-object categorization model that uses a CNN image decoder to learn intricate category-aware picture representations (ii) A model for creating image captions that uses an LSTM-based decoder that is grammar conscious and shares its CNN encoder and LSTM decoder with an object categorization job to create text summaries of pictures. The additional object categorization and grammatical skills are particularly relevant to the job of creation. (ii) A syntactic generation model that enhances LSTM-based decoders that are syntax cognizant. An effective grammar creation model for the image labeling model is (iii).Our model beats other strong competitors in terms of efficiency, according to testing results on the MS-COCO dataset.*

Keywords: Multi-task Learning Approach for Image Captioning

REFERENCES

[1]. [Anderson et al., 2017] Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. vqa and image captioning require both bottom-up and top-down focus. ArXiv preprint 1707.07998, 2017

[2]. [Bernardi et al., 2016] Raffaella Bernardi, RuketCakici, Desmond Elliott, AykutErdem, ErkutErdem, NazliIkizler-Cinbis, Frank Keller, Adrian Muscat, and Barbara Plank. A review of models, datasets, and assessment metrics for automatic description creation from photos. JAIR,\s55:409–442, 2016.

[3]. [Caruana, 1998] Multitask learning, by Rich Caruana, is discussed on pages 95 to 133 in Learning to Learn. Springer, 1998.

[4]. [Chen et al., 2017] Long Chen, Hanwang Zhang, Jun Xiao, LiqiangNie, Jian Shao, Wei Liu, and Tat- Seng Chua. Scacnn: Spatial and channel-wise attention in convolutional networks for picture captioning. In CVPR, 2017.

[5]. [Gu et al., 2017a] Jiuxiang Gu, Jianfei Cai, Gang Wang, and Tsuhan Chen. Coarse-to-fine learning for captioning images is called stack captioning. ArXiv preprint 1709.03376, 2017.

[6]. [Gu et al., 2017b] Jiuxiang Gu, Gang Wang, Jianfei Cai, and Tsuhan Chen. a linguistic empirical research for captioning images on CNN. In ICCV, 2017.

[7]. [He et al., 2016] Kaiming He, Jian Sun, Xiangyu Zhang, and Shaoqing Ren. Image identification using deep residual learning. Pages 770–778 of CVPR, 2016.

[8]. [Karpathy and Fei-Fei, 2015] Li FeiFei with Andrej Karpathy. Deep visual-semantic alignments for producing picture descriptions. 2015 CVPR, pages 3128–3137

[9]. [Li et al., 2017] Yale Song, Jiebo Luo, and Yuncheng Li. improving pairwise ranking for multi- label image classification. ArXiv preprint 1704.03135, 2017.

DOI: 10.48175/IJARSCT-8928

[10]. [Lin et al., 2014] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. Microsoft Coco: "Common items in context." Pages 740–755 of ECCV,Springer, 2014