

Recopilación de referencias

September 8, 2020

1 Papers individuales

1.1 Show, Attend & Tell (Xu et al., 2015)

1.2 Automatic Generation of Medical Imaging Reports (Jing, Xie, and Xing, 2018)

- Image captioning with deep learning: (Vinyals et al., 2014)(Fang et al., 2014)(Karpathy and Fei-Fei, 2014)(Xu et al., 2015)(Krause et al., 2016)(You et al., 2016)(Liang et al., 2017)
- Hierarchical LSTM: (Krause et al., 2016; Liang et al., 2017)
- Co-attention: (Ba, Mnih, and Kavukcuoglu, 2014)(Xu et al., 2015)
- Datasets: (D. Demner-Fushman et al., 2016)
- Metrics: BLEU(Papineni et al., 2002), METEOR (Denkowski and Lavie, 2014), ROUGE (Lin, 2004), CIDEr (Vedantam, Zitnick, and Parikh, 2014)
- LSTM cannot effectively model long sequences: (Liu et al., 2015)(Martin and Cundy, 2018)

1.3 Feature Difference Makes Sense (Park et al., 2020)

- Image captioning with deep learning: (Vinyals et al., 2014; Xu et al., 2015; Karpathy and Fei-Fei, 2014; You et al., 2016)(L. Zhou et al., 2017)(Anderson et al., 2017)
- X-ray datasets: (Kougia, Pavlopoulos, and Androutsopoulos, 2019)(Johnson et al., 2019)
- Feature difference (Jhamtani and Berg-Kirkpatrick, 2018)(Tan et al., 2019)(Forbes et al., 2019)
- Feature vectors from lower convolutional layers(Darlow and Storkey, 2019)(Bau et al., 2017)(B. Zhou et al., 2017)

1.4 Visual Question Generation from Radiology Images (Sarrouiti, Ben Abacha, and Dina Demner-Fushman, 2020)

1.5 A Survey on Biomedical Image Captioning (Kougia, Pavlopoulos, and Androutsopoulos, 2019)

- Datasets:
 - * IU X-ray (images, tags & reports): <https://openi.nlm.nih.gov/faq>(D. Demner-Fushman et al., 2016)
 - * PEIR-GROSS: <https://peir.path.uab.edu/library/index.php?/category/2>
 - * ICLEF-CAPTION (images, tags & reports): (Eickhoff et al., 2017)

References

- Anderson, Peter, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang (2017). *Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering*. arXiv: 1707.07998 [cs.CV].
- Ba, Jimmy, Volodymyr Mnih, and Koray Kavukcuoglu (2014). *Multiple Object Recognition with Visual Attention*. arXiv: 1412.7755 [cs.LG].
- Bau, David, Bolei Zhou, Aditya Khosla, Aude Oliva, and Antonio Torralba (2017). *Network Dissection: Quantifying Interpretability of Deep Visual Representations*. arXiv: 1704.05796 [cs.CV].
- Darlow, Luke Nicholas and Amos Storkey (2019). *What Information Does a ResNet Compress?* URL: <https://openreview.net/forum?id=Hk1bTjRcKX>.
- Demner-Fushman, D., M. D. Kohli, M. B. Rosenman, S. E. Shooshan, L. Rodriguez, S. Antani, G. R. Thoma, and C. J. McDonald (Mar. 2016). "Preparing a collection of radiology examinations for distribution and retrieval". In: *J Am Med Inform Assoc* 23.2. [PubMed Central:PMC5009925] [DOI:10.1093/jamia/ocv080] [PubMed:24108713], pp. 304–310.
- Denkowski, Michael and Alon Lavie (June 2014). "Meteor Universal: Language Specific Translation Evaluation for Any Target Language". In: *Proceedings of the Ninth Workshop on Statistical Machine Translation*. Baltimore, Maryland, USA: Association for Computational Linguistics, pp. 376–380. DOI: 10.3115/v1/W14-3348. URL: <https://www.aclweb.org/anthology/W14-3348>.
- Eickhoff, Carsten, Immanuel Schwall, Alba García Seco de Herrera, and Henning Müller (Sept. 2017). "Overview of Image-CLEFcaption 2017 - the Image Caption Prediction and Concept Extraction Tasks to Understand Biomedical Images". In: *CLEF2017 Working Notes*. CEUR Workshop Proceedings. Dublin, Ireland: CEUR-WS.org <<http://ceur-ws.org>>.
- Fang, Hao, Saurabh Gupta, Forrest Iandola, Rupesh Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C. Platt, C. Lawrence Zitnick, and Geoffrey Zweig (2014). *From Captions to Visual Concepts and Back*. arXiv: 1411.4952 [cs.CV].
- Forbes, Maxwell, Christine Kaeser-Chen, Piyush Sharma, and Serge Belongie (2019). *Neural Naturalist: Generating Fine-Grained Image Comparisons*. arXiv: 1909.04101 [cs.CL].
- Jhamtani, Harsh and Taylor Berg-Kirkpatrick (2018). *Learning to Describe Differences Between Pairs of Similar Images*. arXiv: 1808.10584 [cs.CL].
- Jing, Baoyu, Pengtao Xie, and Eric Xing (July 2018). "On the Automatic Generation of Medical Imaging Reports". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, pp. 2577–2586. DOI: 10.18653/v1/P18-1240. URL: <https://www.aclweb.org/anthology/P18-1240>.
- Johnson, Alistair E. W., Tom J. Pollard, Nathaniel R. Greenbaum, Matthew P. Lungren, Chih-ying Deng, Yifan Peng, Zhiyong Lu, Roger G. Mark, Seth J. Berkowitz, and Steven Horng (2019). *MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs*. arXiv: 1901.07042 [cs.CV].
- Karpathy, Andrej and Li Fei-Fei (2014). *Deep Visual-Semantic Alignments for Generating Image Descriptions*. arXiv: 1412.2306 [cs.CV].
- Kougia, Vasiliki, John Pavlopoulos, and Ion Androutsopoulos (2019). *A Survey on Biomedical Image Captioning*. arXiv: 1905.13302 [cs.CV].
- Krause, Jonathan, Justin Johnson, Ranjay Krishna, and Li Fei-Fei (2016). *A Hierarchical Approach for Generating Descriptive Image Paragraphs*. arXiv: 1611.06607 [cs.CV].
- Liang, Xiaodan, Zhiting Hu, Hao Zhang, Chuang Gan, and Eric P. Xing (2017). *Recurrent Topic-Transition GAN for Visual Paragraph Generation*. arXiv: 1703.07022 [cs.CV].
- Lin, Chin-Yew (July 2004). "ROUGE: A Package for Automatic Evaluation of Summaries". In: *Text Summarization Branches Out*. Barcelona, Spain: Association for Computational Linguistics, pp. 74–81. URL: <https://www.aclweb.org/anthology/W04-1013>.
- Liu, Pengfei, Xipeng Qiu, Xinchu Chen, Shiyu Wu, and Xuanjing Huang (Sept. 2015). "Multi-Timescale Long Short-Term Memory Neural Network for Modelling Sentences and Documents". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, pp. 2326–2335. DOI: 10.18653/v1/D15-1280. URL: <https://www.aclweb.org/anthology/D15-1280>.
- Martin, Eric and Chris Cundy (2018). "Parallelizing Linear Recurrent Neural Nets Over Sequence Length". In: *International Conference on Learning Representations*. URL: <https://openreview.net/forum?id=HyUNwulC->.
- Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu (July 2002). "Bleu: a Method for Automatic Evaluation of Machine Translation". In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, pp. 311–318. DOI: 10.3115/1073083.1073135. URL: <https://www.aclweb.org/anthology/P02-1040>.
- Park, Hyeryun, Kyungmo Kim, Jooyoung Yoon, Seongkeun Park, and Jinwook Choi (July 2020). "Feature Difference Makes Sense: A medical image captioning model exploiting feature difference and tag information". In: *Proceedings of the*

- 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. Online: Association for Computational Linguistics, pp. 95–102. DOI: 10.18653/v1/2020.acl-srw.14. URL: <https://www.aclweb.org/anthology/2020.acl-srw.14>.
- Sarrouti, Mourad, Asma Ben Abacha, and Dina Demner-Fushman (July 2020). “Visual Question Generation from Radiology Images”. In: *Proceedings of the First Workshop on Advances in Language and Vision Research*. Online: Association for Computational Linguistics, pp. 12–18. DOI: 10.18653/v1/2020.alvr-1.3. URL: <https://www.aclweb.org/anthology/2020.alvr-1.3>.
- Tan, Hao, Franck Dernoncourt, Zhe Lin, Trung Bui, and Mohit Bansal (2019). *Expressing Visual Relationships via Language*. arXiv: 1906.07689 [cs.CL].
- Vedantam, Ramakrishna, C. Lawrence Zitnick, and Devi Parikh (2014). *CIDEr: Consensus-based Image Description Evaluation*. arXiv: 1411.5726 [cs.CV].
- Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan (2014). *Show and Tell: A Neural Image Caption Generator*. arXiv: 1411.4555 [cs.CV].
- Xu, Kelvin, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio (2015). *Show, Attend and Tell: Neural Image Caption Generation with Visual Attention*. arXiv: 1502.03044 [cs.LG].
- You, Quanzeng, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo (2016). *Image Captioning with Semantic Attention*. arXiv: 1603.03925 [cs.CV].
- Zhou, Bolei, David Bau, Aude Oliva, and Antonio Torralba (2017). *Interpreting Deep Visual Representations via Network Dissection*. arXiv: 1711.05611 [cs.CV].
- Zhou, Luowei, Chenliang Xu, Parker Koch, and Jason J. Corso (2017). “Watch What You Just Said: Image Captioning with Text-Conditional Attention”. In: *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*. Thematic Workshops ’17. Mountain View, California, USA: Association for Computing Machinery, pp. 305–313. ISBN: 9781450354165. DOI: 10.1145/3126686.3126717. URL: <https://doi.org/10.1145/3126686.3126717>.