

REFERENCES

- [1] Fatma Ben Abdallah, Ghada Feki, Mohamed Ezzarka, Anis Ben Ammar, and Chokri Ben Amar. 2018. Regim Lab Team at ImageCLEF Lifelog Moment Retrieval Task 2018. In *CLEF (Working Notes)*.
- [2] Fabian Caba Heilbron, Victor Escorcia, Bernard Ghanem, and Juan Carlos Nieves. 2015. Activitynet: A large-scale video benchmark for human activity understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 961–970.
- [3] Guangchun Cheng, Yiwen Wan, Abdullah N. Saudagar, Kamesh Namuduri, and Bill P. Buckles. 2015. Advances in human action recognition: A survey. *arXiv preprint arXiv:1501.05964*.
- [4] Berkan Demirel, Ramazan Gokberk Cinbis, and Nazli Ikizler-Cinbis. 2017. Attributes2classname: A discriminative model for attribute-based unsupervised zero-shot learning. In *Proceedings of the IEEE International Conference on Computer Vision*, 1232–1241.
- [5] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, Dang Nguyen, and Duc Tien. 2017. Overview of NTCIR-13 Lifelog-2 task. In *NTCIR*.
- [6] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Van-Tu Ninh, Tu-Khiem Le, Rami Albatal, Duc Tien Dang Nguyen and Graham Healy. 2019. Overview of NTCIR-14 Lifelog-3 task. In *NTCIR*.
- [7] Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. 2015. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, 1681–1691.
- [8] Mihir Jain, Jan C van Gemert, Thomas Mensink, and Cees GM Snoek. 2015. Objects2action: Classifying and localizing actions without any video example. In *Proceedings of the IEEE international conference on computer vision*, 4588–4596.
- [9] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 1725–1732.
- [10] Jamie Kiros, William Chan, and Geoffrey Hinton. 2018. Illustrative Language Understanding: Large-Scale Visual Grounding with Image Search. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 922–933, Melbourne, Australia, July. Association for Computational Linguistics.
- [11] Jie Lin, Ana Garcia del Molino, Qianli Xu, Fen Fang, Vigneshwaran Subbaraju, Joo-Hwee Lim, Liyuan Li, and Vijay Chandrasekhar. 2017. VCI2R at the NTCIR-13 Lifelog-2 Lifelog Semantic Access Task.
- [12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.
- [13] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.
- [14] Hamed Pirsiavash and Deva Ramanan. 2012. Detecting activities of daily living in first-person camera views. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2847–2854. IEEE.
- [15] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [16] Suriya Singh, Chetan Arora, and C. V. Jawahar. 2016. First person action recognition using deep learned descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2620–2628.
- [17] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. 2012. UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.
- [18] Robert Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [19] Tsun-Hsien Tang, Min-Huan Fu, Hen-Hsen Huang, Kuan-Ta Chen, and Hsin-Hsi Chen. 2018. Visual Concept Selection with Textual Knowledge for Understanding Activities of Daily Living and Life Moment Retrieval. In *CLEF (Working Notes)*.
- [20] Rowan Zellers and Yejin Choi. 2017. Zero-shot activity recognition with verb attribute induction. *arXiv preprint arXiv:1707.09468*. Conference Name: ACM Woodstock conference