Video Analysis using Color spaces and RL

Jude Law

1. INTRODUCTION

Color spaces have shown promising results in computer vision [1][2][3][4]. However, the use of color spaces in videos has not been analyzed. Video applications such as action recognition and video segmentation have been gaining a lot of attention in recent years. Using deep learning, action recognition results have reached a saturation point on some datasets [5][6][7]. In order to overcome these saturations, large scale datasets have been introduced such as the Sports1M [8] and Youtube8M[9]. Along with these datasets, the problem of limited data- namely zero-shot learning [10] has been introduced. ZSL has shown to be a very challenging problem because of introduction of unseen classes at test time. This is essentially a very practical problem and hence is of real value to the computer vision industry. To help optimize the learning with limited data, RL has shown great promise [11][12]. Along these lines, this work proposes the use of RL to determine what color spaces can help in video understanding namely action recognition and localization in the context of ZSL.

2. REFERENCES

- [1] Gowda, S.N. and Yuan, C., 2018, December. ColorNet: Investigating the importance of color spaces for image classification. In *Asian Conference on Computer Vision* (pp. 581-596). Springer, Cham.
- [2] Gowda, S.N. and Yuan, C., 2020. StegColNet: Steganalysis based on an ensemble colorspace approach. *arXiv preprint arXiv:2002.02413*.
- [3] Chai, D. and Bouzerdoum, A., 2000. A Bayesian approach to skin color classification in YCbCr color space. In *TENCON 2000. Proceedings* (pp. 421-424). IEEE.
- [4] Vandenbroucke, N., Macaire, L. and Postaire, J.G., 1998, October. Color pixels classification in an hybrid color space. In *Image Processing, 1998. ICIIP 98. Proceedings. 1998 International Conference on* (Vol. 1, pp. 176-180). IEEE.
- [5] Carreira, J. and Zisserman, A., 2017. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6299-6308).
- [6] Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y. and Paluri, M., 2018. A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 6450-6459).
- [7] Gowda, S.N., 2017. Human activity recognition using combinatorial Deep Belief Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1-6).
- [8] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1725-1732).
- [9] Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B. and Vijayanarasimhan, S., 2016. Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*.
- [10] Socher, R., Ganjoo, M., Manning, C.D. and Ng, A., 2013. Zero-shot learning through cross-modal transfer. In *Advances in neural information processing systems* (pp. 935-943).
- [11] Gowda, S.N., Eustratiadis, P., Hospedales, T. and Sevilla-Lara, L., 2020. ALBA: Reinforcement Learning for Video Object Segmentation. *arXiv preprint arXiv:2005.13039*.
- [12] Ye, Z., Geng, Y., Chen, J., Chen, J., Xu, X., Zheng, S., Wang, F., Zhang, J. and Chen, H., 2020, July. Zero-shot Text Classification via Reinforced Self-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3014-3024).