REFERENCES

[1] S. Abu-El-Haija, N. Kothari, J. Lee, P. Natsev, G. Toderici, B. Varadarajan, and S. Vijayanarasimhan. 2016. Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675* (2016).

[2] J. Dong, X. Li, W. Lan, Y. Huo, and C. G. M. Snoek. 2016. Early Embedding and Late Reranking for Video Captioning. In *MM*.

[3] J. Dong, X. Li, and C. G.M. Snoek. 2018. Predicting visual features from text for image and video caption retrieval. *T-MM* (2018). https://doi.org/10.1109/TMM.2018.2832602

[4] F. Faghri, D. J Fleet, J. R. Kiros, and S. Fidler. 2017. VSE++: improved visual-semantic embeddings. *arXiv preprint arXiv:1707.05612* (2017).

[5] A. Frome, G. S Corrado, J. Shlens, S. Bengio, J. Dean, T. Mikolov, et al. 2013. Devise: A deep visual-semantic embedding model. In *NIPS*.

[6] R. Hadsell, S. Chopra, and Y. LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *CVPR*.

[7] A. Karpathy and L. Fei-Fei. 2015. Deep visual-semantic alignments for generating image descriptions. In *CVPR*. 3128–3137.

[8] D. P Kingma and J. Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[9] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4 (2014), 2065–2073.

[10] M. Liu, X. Xie, and H. Zhou. 2018. Content-based Video Relevance Prediction Challenge: Data, Protocol, and Baseline. *arXiv preprint arXiv:1806.00737* (2018).

[11] M. Mazloom, X. Li, and C. G.M. Snoek. 2016. TagBook: A Semantic Video Representation Without Supervision for Event Detection. *T-MM* 18, 7 (2016), 1378–1388.

[12] Y. Pan, T. Mei, T. Yao, H. Li, and Y. Rui. 2016. Jointly Modeling Embedding and Translation to Bridge Video and Language. In *CVPR*.

[13] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. 2015. Learning spatiotemporal features with 3d convolutional networks. In *ICCV*.

[14] L. van de Maaten and G. Hinton. 2008. Visualizing Data using T-SNE. *JMLR* 9 (2008), 2579–2605.