



















## REFERENCES

- [1] Grigory Antipov, Sid-Ahmed Berrani, Natacha Ruchaud, and Jean-Luc Dugelay. 2015. Learned vs. hand-crafted features for pedestrian gender recognition. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 1263–1266.
- [2] Michael M Bronstein, Alexander M Bronstein, Fabrice Michel, and Nikos Paragios. 2010. Data fusion through cross-modality metric learning using similarity-sensitive hashing. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 3594–3601.
- [3] Bokai Cao, Hucheng Zhou, Guoqiang Li, and Philip S Yu. 2016. Multi-view machines. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 427–436.
- [4] Changxing Ding and Dacheng Tao. 2015. Robust face recognition via multimodal deep face representation. *IEEE Transactions on Multimedia* 17, 11 (2015), 2049–2058.
- [5] Guiguang Ding, Yuchen Guo, and Jile Zhou. 2014. Collective matrix factorization hashing for multimodal data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2075–2082.
- [6] Yunhao Gong, Qifa Ke, Michael Isard, and Svetlana Lazebnik. 2014. A multi-view embedding space for modeling internet images, tags, and their semantics. *International journal of computer vision* 106, 2 (2014), 210–233.
- [7] Sung Ju Hwang and Kristen Grauman. 2012. Reading between the lines: Object localization using implicit cues from image tags. *IEEE transactions on pattern analysis and machine intelligence* 34, 6 (2012), 1145–1158.
- [8] Qing-Yuan Jiang and Wu-Jun Li. 2017. Deep Cross-Modal Hashing. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 3270–3278.
- [9] Lu Jin, Shenghua Gao, Zechao Li, and Jinhui Tang. 2014. Hand-crafted features or machine learnt features? together they improve RGB-D object recognition. In *Multimedia (ISM), 2014 IEEE International Symposium on*. IEEE, 311–319.
- [10] Tamara G Kolda and Brett W Bader. 2009. Tensor decompositions and applications. *SIAM review* 51, 3 (2009), 455–500.
- [11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
- [12] Shaishav Kumar and Raghavendra Udupa. 2011. Learning hash functions for cross-view similarity search. In *IJCAI proceedings-international joint conference on artificial intelligence*, Vol. 22. 1360.
- [13] Kai Li, Guo-Jun Qi, Jun Ye, and Kien A Hua. 2017. Linear subspace ranking hashing for cross-modal retrieval. *IEEE transactions on pattern analysis and machine intelligence* 39, 9 (2017), 1825–1838.
- [14] Zijia Lin, Guiguang Ding, Jungong Han, and Jianmin Wang. 2017. Cross-view retrieval via probability-based semantics-preserving hashing. *IEEE transactions on cybernetics* 47, 12 (2017), 4342–4355.
- [15] Hong Liu, Rongrong Ji, Yongjian Wu, and Gang Hua. 2016. Supervised matrix factorization for cross-modality hashing. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press, 1767–1773.
- [16] Sean Moran and Victor Lavrenko. 2015. Regularised cross-modal hashing. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 907–910.
- [17] Morten Mørup, Lars Kai Hansen, and Sidse M Arnfred. 2008. Algorithms for sparse nonnegative Tucker decompositions. *Neural computation* 20, 8 (2008), 2112–2131.
- [18] Yuxin Peng, Xin Huang, and Jinwei Qi. 2016. Cross-Media Shared Representation by Hierarchical Learning with Multiple Deep Networks. In *IJCAI*. 3846–3853.
- [19] Yuxin Peng, Jinwei Qi, Xin Huang, and Yuxin Yuan. 2018. CCL: Cross-modal Correlation Learning With Multigrained Fusion by Hierarchical Network. *IEEE Transactions on Multimedia* 20, 2 (2018), 405–420.
- [20] Nikhil Rasiwasia, Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Gert RG Lanckriet, Roger Levy, and Nuno Vasconcelos. 2010. A new approach to cross-modal multimedia retrieval. In *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 251–260.
- [21] Steffen Rendle. 2010. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 995–1000.
- [22] Abhishek Sharma, Abhishek Kumar, Hal Daume, and David W Jacobs. 2012. Generalized multiview analysis: A discriminative latent space. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2160–2167.
- [23] Xiaobo Shen, Fumin Shen, Quan-Sen Sun, Yang Yang, Yun-Hao Yuan, and Heng Tao Shen. 2017. Semi-paired discrete hashing: Learning latent hash codes for semi-paired cross-view retrieval. *IEEE transactions on cybernetics* 47, 12 (2017), 4275–4288.
- [24] Xiaobo Shen, Fumin Shen, Quan-Sen Sun, and Yun-Hao Yuan. 2015. Multi-view latent hashing for efficient multimedia search. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 831–834.
- [25] Jingkuan Song, Yang Yang, Yi Yang, Zi Huang, and Heng Tao Shen. 2013. Inter-media hashing for large-scale retrieval from heterogeneous data sources. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. ACM, 785–796.
- [26] Jinhui Tang, Lu Jin, Zechao Li, and Shenghua Gao. 2015. RGB-D object recognition via incorporating latent data structure and prior knowledge. *IEEE Transactions on Multimedia* 17, 11 (2015), 1899–1908.
- [27] Jun Tang, Ke Wang, and Ling Shao. 2016. Supervised matrix factorization hashing for cross-modal retrieval. *IEEE Transactions on Image Processing* 25, 7 (2016), 3157–3166.
- [28] Jingdong Wang, Heng Tao Shen, Jingkuan Song, and Jianqiu Ji. 2014. Hashing for similarity search: A survey. *arXiv preprint arXiv:1408.2927* (2014).
- [29] Kaiye Wang, Qiyue Yin, Wei Wang, Shu Wu, and Liang Wang. 2016. A Comprehensive Survey on Cross-modal Retrieval. *arXiv preprint arXiv:1607.06215* (2016).
- [30] Xing Xu, Fumin Shen, Yang Yang, Heng Tao Shen, and Xuelong Li. 2017. Learning discriminative binary codes for large-scale cross-modal retrieval. *IEEE Transactions on Image Processing* 26, 5 (2017), 2494–2507.
- [31] Tao Yao, Xiangwei Kong, Haiyan Fu, and Qi Tian. 2016. Semantic consistency hashing for cross-modal retrieval. *Neurocomputing* 193 (2016), 250–259.
- [32] Dongqing Zhang and Wu-Jun Li. 2014. Large-Scale Supervised Multimodal Hashing with Semantic Correlation Maximization. In *AAAI*, Vol. 1. 7.
- [33] Jian Zhang and Yuxin Peng. 2017. SSDH: semi-supervised deep hashing for large scale image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology* (2017).
- [34] Jian Zhang, Yuxin Peng, and Mingkuan Yuan. 2018. Unsupervised Generative Adversarial Cross-modal Hashing. (2018).
- [35] Yi Zhen and Dit-Yan Yeung. 2012. Co-regularized hashing for multimodal data. In *Advances in neural information processing systems*. 1376–1384.
- [36] Yi Zhen and Dit-Yan Yeung. 2012. A probabilistic model for multimodal hash function learning. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 940–948.
- [37] Jile Zhou, Guiguang Ding, and Yuchen Guo. 2014. Latent semantic sparse hashing for cross-modal similarity search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 415–424.
- [38] Xiaofeng Zhu, Zi Huang, Heng Tao Shen, and Xin Zhao. 2013. Linear cross-modal hashing for efficient multimedia search. In *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 143–152.