Deep learning for visual understanding

Guo, Y.; Guo Y.

Bibliography

[1] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110

[2] Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Proceedings of European Conference on Computer Vision. (2006) 404–417

[3] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2005) 886–893

[4] Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: Proceedings of the IEEE International Conference on Computer Vision. (2003) 1470–1477

[5] Jegou, H., Perronnin, F., Douze, M., Sánchez, J., Perez, P., Schmid, C.: Aggregating local image descriptors into compact codes. IEEE Transactions on Pattern Analysis and Machine Intelligence **34** (2012) 1704–1716

[6] Perronnin, F., Dance, C.: Fisher kernels on visual vocabularies for image categorization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2007) 1–8

[7] Sargano, A.B., Angelov, P., Habib, Z.: A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition. Applied Sciences **7** (2017) 110

[8] Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural computation **18** (2006) 1527–1554

[9] He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1026–1034

[10] Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence **39** (2017) 652–663

[11] Bordes, A., Glorot, X., Weston, J., Bengio, Y.: Joint learning of words and meaning representations for open-text semantic parsing. In: International Conference on Artificial Intelligence and Statistics. (2012) 127–135

[12] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems. (2013) 3111–3119

[13] Ciregan, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2012) 3642–3649

[14] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. (2012) 1097–1105

[15] Deng, L.: A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing **3** (2014) e2

[16] Bengio, Y., et al.: Learning deep architectures for ai. Foundations and trends® in Machine Learning **2** (2009) 1–127

[17] Schmidhuber, J.: Deep learning in neural networks: An overview. Neural networks **61** (2015) 85–117

[18] Bengio, Y.: Deep learning of representations: Looking forward. In: International Conference on Statistical Language and Speech Processing. (2013) 1–37

[19] Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 1798–1828

[20] LeCun, Y.: Learning invariant feature hierarchies. In: Proceedings of European Conference on Computer Vision. (2012) 496–505

[21] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision **115** (2015) 211–252

[22] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Proceedings of European Conference on Computer Vision. (2014) 818–833

[23] He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Proceedings of European Conference on Computer Vision. (2014) 346–361

[24] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations. (2015)

[25] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 1–9

[26] Salakhutdinov, R., Hinton, G.: Deep boltzmann machines. In: International Conference on Artificial Intelligence and Statistics. (2009) 448–455

[27] Ngiam, J., Chen, Z., Koh, P.W., Ng, A.Y.: Learning deep energy models. In: International Conference on Machine Learning. (2011) 1105–1112

[28] Poultney, C., Chopra, S., Cun, Y.L., et al.: Efficient learning of sparse representations with an energy-based model. In: Advances in Neural Information Processing Systems. (2007) 1137–1144

[29] Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: International Conference on Machine Learning. (2008) 1096–1103

[30] Rifai, S., Vincent, P., Muller, X., Glorot, X., Bengio, Y.: Contractive auto-encoders: Explicit invariance during feature extraction. In: International Conference on Machine Learning. (2011) 833–840

[31] Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2009) 1794–1801

[32] Gao, S., Tsang, I.W.H., Chia, L.T., Zhao, P.: Local features are not lonely–laplacian sparse coding for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2010) 3555–3561

[33] Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In: Advances in Neural Information Processing Systems. (2009) 2223–2231

[34] Zhou, X., Yu, K., Zhang, T., Huang, T.S.: Image classification using super-vector coding of local image descriptors. In: Proceedings of European Conference on Computer Vision. (2010) 141–154

[35] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86** (1998) 2278–2324

[36] Zeiler, M.D.: Hierarchical convolutional deep learning in computer vision. PhD thesis, NEW YORK UNIVERSITY (2013)

[37] Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Is object localization for free?-weakly-supervised learning with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 685–694

[38] Lin, M., Chen, Q., Yan, S.: Network in network. In: International Conference on Learning Representations. (2014)

[39] Boureau, Y.L., Ponce, J., LeCun, Y.: A theoretical analysis of feature pooling in visual recognition. In: International Conference on Machine Learning. (2010) 111–118

[40] Scherer, D., Müller, A., Behnke, S.: Evaluation of pooling operations in convolutional architectures for object recognition. In: International Conference on Artificial Neural Networks. (2010) 92–101

[41] Cireşan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J.: High-performance neural networks for visual object classification. arXiv preprint arXiv:1102.0183 (2011)

[42] Zeiler, M.D., Fergus, R.: Stochastic pooling for regularization of deep convolutional neural networks. In: International Conference on Learning Representations. (2013)

[43] Ouyang, W., Luo, P., Zeng, X., Qiu, S., Tian, Y., Li, H., Yang, S., Wang, Z., Xiong, Y., Qian, C., et al.: Deepid-net: multi-stage and deformable deep convolutional neural networks for object detection. arXiv preprint arXiv:1409.3505 (2014)

[44] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 580–587

[45] Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 1717–1724

[46] Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)

[47] Baldi, P., Sadowski, P.J.: Understanding dropout. In: Advances in Neural Information Processing Systems. (2013) 2814–2822

[48] Ba, J., Frey, B.: Adaptive dropout for training deep neural networks. In: Advances in Neural Information Processing Systems. (2013) 3084–3092

[49] McAllester, D.: A pac-bayesian tutorial with a dropout bound. arXiv preprint arXiv:1307.2118 (2013)

[50] Wager, S., Wang, S., Liang, P.S.: Dropout training as adaptive regularization. In: Advances in Neural Information Processing Systems. (2013) 351–359

[51] Wang, S.I., Manning, C.D.: Fast dropout training. In: International Conference on Machine Learning. (2013) 118–126

[52] Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research **15** (2014) 1929–1958

[53] Warde-Farley, D., Goodfellow, I.J., Courville, A., Bengio, Y.: An empirical analysis of dropout in piecewise linear networks. In: International Conference on Learning Representations. (2014)

[54] Wan, L., Zeiler, M., Zhang, S., Cun, Y.L., Fergus, R.: Regularization of neural networks using dropconnect. In: International Conference on Machine Learning. (2013) 1058–1066

[55] Howard, A.G.: Some improvements on deep convolutional neural network based image classification. arXiv preprint arXiv:1312.5402 (2013)

[56] Dosovitskiy, A., Springenberg, J.T., Brox, T.: Unsupervised feature learning by augmenting single images. arXiv preprint arXiv:1312.5242 (2013)

[57] Wu, R., Yan, S., Shan, Y., Dang, Q., Sun, G.: Deep image: Scaling up image recognition. arXiv preprint arXiv:1501.02876 (2015)

[58] Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P., Bengio, S.: Why does unsupervised pre-training help deep learning? Journal of Machine Learning Research **11** (2010) 625–660

[59] He, K., Sun, J.: Convolutional neural networks at constrained time cost. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 5353–5360

[60] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 3431–3440

[61] Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y., Li, J.: Deep learning for content-based image retrieval: A comprehensive study. In: Proceedings of the ACM International Conference on Multimedia. (2014) 157–166

[62] Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems. (2014) 3320–3328

[63] Ngiam, J., Chen, Z., Chia, D., Koh, P.W., Le, Q.V., Ng, A.Y.: Tiled convolutional neural networks. In: Advances in Neural Information Processing Systems. (2010) 1279–1287

[64] Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1440–1448

[65] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. (2015) 91–99

[66] Zhu, Y., Urtasun, R., Salakhutdinov, R., Fidler, S.: segdeepm: Exploiting segmentation and context in deep neural networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 4703–4711

[67] Gidaris, S., Komodakis, N.: Object detection via a multi-region and semantic segmentation-aware cnn model. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1134–1142

[68] Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Simultaneous detection and segmentation. In: Proceedings of European Conference on Computer Vision. (2014) 297–312

[69] Zhang, Y., Sohn, K., Villegas, R., Pan, G., Lee, H.: Improving object detection with deep convolutional networks via bayesian optimization and structured prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 249–258

[70] Yoo, D., Park, S., Lee, J.Y., So Kweon, I.: Multi-scale pyramid pooling for deep convolutional representation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop. (2015) 71–80

[71] Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1395–1403

[72] Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 3476–3483

[73] Wang, X., Zhang, L., Lin, L., Liang, Z., Zuo, W.: Deep joint task learning for generic object extraction. In: Advances in Neural Information Processing Systems. (2014) 523–531

[74] Zeng, X., Ouyang, W., Wang, X.: Multi-stage contextual deep learning for pedestrian detection. In: Proceedings of the IEEE International Conference on Computer Vision. (2013) 121–128

[75] Miclut, B.: Committees of deep feedforward networks trained with few data. In: German Conference on Pattern Recognition. (2014) 736–742

[76] Weston, J., Ratle, F., Mobahi, H., Collobert, R.: Deep learning via semi-supervised embedding. In: Neural Networks: Tricks of the Trade. (2012) 639–655

[77] Simonyan, K., Vedaldi, A., Zisserman, A.: Deep fisher networks for large-scale image classification. In: Advances in Neural Information Processing Systems. (2013) 163–171

[78] Chen, Q., Song, Z., Dong, J., Huang, Z., Hua, Y., Yan, S.: Contextualizing object detection and classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **37** (2015) 13–27

[79] Hinton, G.E., Sejnowski, T.J.: Learning and releaming in boltzmann machines. Parallel Distrilmted Processing **1** (1986)

[80] Carreira-Perpinan, M.A., Hinton, G.E.: On contrastive divergence learning. In: International Conference on Artificial Intelligence and Statistics. (2005) 33–40

[81] Hinton, G.E.: A practical guide to training restricted boltzmann machines. In: Neural networks: Tricks of the trade. (2012) 599–619

[82] Cho, K., Raiko, T., Ihler, A.T.: Enhanced gradient and adaptive learning rate for training restricted boltzmann machines. In: International Conference on Machine Learning. (2011) 105–112

[83] Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: International Conference on Machine Learning. (2010) 807–814

[84] Arel, I., Rose, D.C., Karnowski, T.P.: Deep machine learning-a new frontier in artificial intelligence research [research frontier]. IEEE Computational Intelligence Magazine **5** (2010) 13–18

[85] Lee, H., Ekanadham, C., Ng, A.Y.: Sparse deep belief net model for visual area v2. In: Advances in Neural Information Processing Systems. (2008) 873–880

[86] Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: International Conference on Machine Learning. (2009) 609–616

[87] Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Unsupervised learning of hierarchical representations with convolutional deep belief networks. Communications of the ACM **54** (2011) 95–103

[88] Nair, V., Hinton, G.E.: 3d object recognition with deep belief nets. In: Advances in Neural Information Processing Systems. (2009) 1339–1347

[89] Tang, Y., Eliasmith, C.: Deep networks for robust visual recognition. In: International Conference on Machine Learning. (2010) 1055–1062

[90] Huang, G.B., Lee, H., Learned-Miller, E.: Learning hierarchical representations for face verification with convolutional deep belief networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2012) 2518–2525

[91] Younes, L.: On the convergence of markovian stochastic algorithms with rapidly decreasing ergodicity rates. Stochastics: An International Journal of Probability and Stochastic Processes **65** (1999) 177–228

[92] Salakhutdinov, R., Larochelle, H.: Efficient learning of deep boltzmann machines. In: International Conference on Artificial Intelligence and Statistics. (2010) 693–700

[93] Salakhutdinov, R.R., Hinton, G.E.: An efficient learning procedure for deep boltzmann machines. Neural computation **24** (2012) 1967–2006

[94] Hinton, G.E., Salakhutdinov, R.R.: A better way to pretrain deep boltzmann machines. In: Advances in Neural Information Processing Systems. (2012) 2447–2455

[95] Cho, K., Raiko, T., Ilin, A., Karhunen, J.: A two-stage pretraining algorithm for deep boltzmann machines. In: International Conference on Artificial Neural Networks, Springer (2013) 106–113

[96] Montavon, G., Müller, K.R.: Deep boltzmann machines and the centering trick. In: Neural Networks: Tricks of the Trade. (2012) 621–637

[97] Goodfellow, I.J., Courville, A., Bengio, Y.: Joint training deep boltzmann machines for classification. arXiv preprint arXiv:1301.3568 (2013)

[98] Goodfellow, I., Mirza, M., Courville, A., Bengio, Y.: Multi-prediction deep boltzmann machines. In: Advances in Neural Information Processing Systems. (2013) 548–556

[99] Elfwing, S., Uchibe, E., Doya, K.: Expected energy-based restricted boltzmann machine for classification. Neural Networks **64** (2015) 29–38

[100] Eslami, S.A., Heess, N., Williams, C.K., Winn, J.: The shape boltzmann machine: a strong model of object shape. International Journal of Computer Vision **107** (2014) 155–176

[101] Kae, A., Sohn, K., Lee, H., Learned-Miller, E.: Augmenting crfs with boltzmann machine shape priors for image labeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 2019–2026

[102] Dahl, G., Mohamed, A.r., Hinton, G.E., et al.: Phone recognition with the mean-covariance restricted boltzmann machine. In: Advances in Neural Information Processing Systems. (2010) 469–477

[103] Liou, C.Y., Cheng, W.C., Liou, J.W., Liou, D.R.: Autoencoder for words. Neurocomputing **139** (2014) 84–96

[104] Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. science **313** (2006) 504–507

[105] Zhang, J., Shan, S., Kan, M., Chen, X.: Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In: Proceedings of European Conference on Computer Vision. (2014) 1–16

[106] Jiang, X., Zhang, Y., Zhang, W., Xiao, X.: A novel sparse auto-encoder for deep unsupervised learning. In: International Conference on Advanced Computational Intelligence. (2013) 256–261

[107] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research **11** (2010) 3371–3408

[108] Goroshin, R., LeCun, Y.: Saturating auto-encoders. arXiv preprint arXiv:1301.3577 (2013)

[109] Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: International Conference on Artificial Neural Networks. (2011) 52–59

[110] Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., Baskurt, A.: Spatio-temporal convolutional sparse auto-encoder for sequence classification. In: British Machine Vision Conference. (2012)

[111] Leng, B., Guo, S., Zhang, X., Xiong, Z.: 3d object retrieval with stacked local convolutional autoencoder. Signal Processing **112** (2015) 119–128

[112] Konda, K., Memisevic, R., Krueger, D.: Zero-bias autoencoders and the benefits of co-adapting features. In: International Conference on Learning Representations. (2015)

[113] Goodfellow, I., Lee, H., Le, Q.V., Saxe, A., Ng, A.Y.: Measuring invariances in deep networks. In: Advances in Neural Information Processing Systems. (2009) 646–654

[114] Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., Le, Q.V., Ng, A.Y.: On optimization methods for deep learning. In: International Conference on Machine Learning. (2011) 265–272

[115] Zou, W.Y., Ng, A.Y., Yu, K.: Unsupervised learning of visual invariance with temporal coherence. In: Advances in Neural Information Processing Systems Workshop. (2011)

[116] Simoncelli, E.P.: Statistical modeling of photographic images. In: Handbook of image and video processing. (2005)

[117] Le, Q.V.: Building high-level features using large scale unsupervised learning. In: IEEE International Conference on Acoustics, Speech and Signal Processing. (2013) 8595–8598

[118] Alain, G., Bengio, Y.: What regularized auto-encoders learn from the data-generating distribution. Journal of Machine Learning Research **15** (2014) 3563–3593

[119] Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I.J., Lavoie, E., Muller, X., Desjardins, G., Warde-Farley, D., et al.: Unsupervised and transfer learning challenge: a deep learning approach. In: International Conference on Machine Learning Workshop. (2012) 97–110

[120] Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set: A strategy employed by v1? Vision research **37** (1997) 3311–3325

[121] Raina, R., Battle, A., Lee, H., Packer, B., Ng, A.Y.: Self-taught learning: transfer learning from unlabeled data. In: International Conference on Machine Learning. (2007) 759–766

[122] Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2010) 3360–3367

[123] Donoho, D.L.: For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. Communications on pure and applied mathematics **59** (2006) 797–829

[124] Censor, Y., Zenios, S.A.: Parallel optimization: Theory, algorithms, and applications. Oxford University Press on Demand (1997)

[125] Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Cognitive modeling **5** (1988) 1

[126] Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Advances in Neural Information Processing Systems. (2007)

[127] Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: International Conference on Machine Learning. (2009) 689–696

[128] Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research **11** (2010) 19–60

[129] Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., et al.: Pathwise coordinate optimization. The Annals of Applied Statistics **1** (2007) 302–332

[130] Gregor, K., LeCun, Y.: Learning fast approximations of sparse coding. In: International Conference on Machine Learning. (2010) 399–406

[131] Chambolle, A., De Vore, R.A., Lee, N.Y., Lucier, B.J.: Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. IEEE Transactions on Image Processing **7** (1998) 319–335

[132] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In: IEEE International Conference on Acoustics, Speech and Signal Processing. (2009) 693–696

[133] Kavukcuoglu, K., Ranzato, M., LeCun, Y.: Fast inference in sparse coding algorithms with applications to object recognition. arXiv preprint arXiv:1010.3467 (2010)

[134] Balasubramanian, K., Yu, K., Lebanon, G.: Smooth sparse coding via marginal regression for learning sparse representations. In: International Conference on Machine Learning. (2013) 289–297

[135] Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2006) 2169–2178

[136] Coates, A., Ng, A.Y.: The importance of encoding versus training with sparse coding and vector quantization. In: International Conference on Machine Learning. (2011) 921–928

[137] Gao, S., Tsang, I.W.H., Chia, L.T.: Laplacian sparse coding, hypergraph laplacian sparse coding, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 92–104

[138] Yu, K., Lin, Y., Lafferty, J.: Learning image representations from the pixel level via hierarchical sparse coding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2011) 1713–1720

[139] Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R.: Deconvolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2010) 2528–2535

[140] Zeiler, M.D., Taylor, G.W., Fergus, R.: Adaptive deconvolutional networks for mid and high level feature learning. In: Proceedings of the IEEE International Conference on Computer Vision. (2011) 2018–2025

[141] Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., Cao, L., Huang, T.: Large-scale image classification: fast feature extraction and svm training. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2011) 1689–1696

[142] He, Y., Kavukcuoglu, K., Wang, Y., Szlam, A., Qi, Y.: Unsupervised feature learning by deep sparse coding. In: Proceedings of the 2014 SIAM International Conference on Data Mining. (2014) 902–910

[143] Master, S.: Large scale object detection. PhD thesis, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University (2014)

[144] Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Proceedings of European Conference on Computer Vision Workshop. (2004)

[145] Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the 5th Annual Workshop on Computational Learning Theory. (1992) 144–152

[146] Wang, X., Han, T.X., Yan, S.: An hog-lbp human detector with partial occlusion handling. In: Proceedings of the IEEE International Conference on Computer Vision. (2009) 32–39

[147] Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Proceedings of European Conference on Computer Vision. (2010) 143–156

[148] Jaakkola, T.S., Haussler, D., et al.: Exploiting generative models in discriminative classifiers. In: Advances in Neural Information Processing Systems. (1999) 487–493

[149] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2009) 248–255

[150] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. In: International Conference on Learning Representations. (2014)

[151] Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)

[152] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence **32** (2010) 1627–1645

[153] Szegedy, C., Toshev, A., Erhan, D.: Deep neural networks for object detection. In: Advances in Neural Information Processing Systems. (2013) 2553–2561

[154] Erhan, D., Szegedy, C., Toshev, A., Anguelov, D.: Scalable object detection using deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 2147–2154

[155] Ren, S., He, K., Girshick, R., Zhang, X., Sun, J.: Object detection networks on convolutional feature maps. IEEE Transactions on Pattern Analysis and Machine Intelligence (2016)

[156] Uijlings, J.R., van de Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. International Journal of Computer Vision **104** (2013) 154–171

[157] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 779–788

[158] Alexe, B., Deselaers, T., Ferrari, V.: Measuring the objectness of image windows. IEEE Transactions on Pattern Analysis and Machine Intelligence **34** (2012) 2189–2202

[159] Endres, I., Hoiem, D.: Category independent object proposals. In: Proceedings of European Conference on Computer Vision. (2010) 575–588

[160] Cheng, M.M., Zhang, Z., Lin, W.Y., Torr, P.: Bing: Binarized normed gradients for objectness estimation at 300fps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 3286–3293

[161] Zitnick, C.L., Dollár, P.: Edge boxes: Locating object proposals from edges. In: Proceedings of European Conference on Computer Vision. (2014) 391–405

[162] Hosang, J., Benenson, R., Schiele, B.: How good are detection proposals, really? In: British Machine Vision Conference. (2014)

[163] Dai, Q., Hoiem, D.: Learning to localize detected objects. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2012) 3322–3329

[164] Hoiem, D., Chodpathumwan, Y., Dai, Q.: Diagnosing error in object detectors. In: Proceedings of European Conference on Computer Vision. (2012) 340–353

[165] Dong, J., Chen, Q., Yan, S., Yuille, A.: Towards unified object detection and semantic segmentation. In: Proceedings of European Conference on Computer Vision. (2014) 299–314

[166] Hoffman, J., Guadarrama, S., Tzeng, E.S., Hu, R., Donahue, J., Girshick, R., Darrell, T., Saenko, K.: Lsda: Large scale detection through adaptation. In: Advances in Neural Information Processing Systems. (2014) 3536–3544

[167] Hoffman, J., Guadarrama, S., Tzeng, E., Donahue, J., Girshick, R., Darrell, T., Saenko, K.: From large-scale object classifiers to large-scale object detectors: An adaptation approach. In: Advances in Neural Information Processing Systems. (2014)

[168] Zhou, B., Jagadeesh, V., Piramuthu, R.: Conceptlearner: Discovering visual concepts from weakly labeled image collections. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 1492–1500

[169] Liang, X., Liu, S., Wei, Y., Liu, L., Lin, L., Yan, S.: Towards computational baby learning: A weakly-supervised approach for object detection. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 999–1007

[170] Gong, Y., Wang, L., Guo, R., Lazebnik, S.: Multi-scale orderless pooling of deep convolutional activation features. In: Proceedings of European Conference on Computer Vision. (2014) 392–407

[171] Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop. (2014) 806–813

[172] Sun, S., Zhou, W., Li, H., Tian, Q.: Search by detection: Object-level feature for image retrieval. In: Proceedings of International Conference on Internet Multimedia Computing and Service. (2014) 46

[173] Babenko, A., Slesarev, A., Chigorin, A., Lempitsky, V.: Neural codes for image retrieval. In: Proceedings of European Conference on Computer Vision. (2014) 584–599

[174] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Semantic image segmentation with deep convolutional nets and fully connected crfs. In: International Conference on Learning Representations. (2015)

[175] Dai, J., He, K., Sun, J.: Convolutional feature masking for joint object and stuff segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 3992–4000

[176] Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 447–456

[177] Lin, G., Shen, C., van den Hengel, A., Reid, I.: Efficient piecewise training of deep structured models for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 3194–3203

[178] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1529–1537

[179] Papandreou, G., Chen, L.C., Murphy, K., Yuille, A.L.: Weakly-and semi-supervised learning of a dcnn for semantic image segmentation. arXiv preprint arXiv:1502.02734 (2015)

[180] Dai, J., He, K., Sun, J.: Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1635–1643

[181] Jain, A., Tompson, J., LeCun, Y., Bregler, C.: Modeep: A deep learning framework using motion features for human pose estimation. In: Asian Conference on Computer Vision. (2014) 302–315

[182] Pfister, T., Simonyan, K., Charles, J., Zisserman, A.: Deep convolutional neural networks for efficient pose estimation in gesture videos. In: Asian Conference on Computer Vision. (2014) 538–552

[183] Pfister, T., Charles, J., Zisserman, A.: Flowing convnets for human pose estimation in videos. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1913–1921

[184] Yu, J., Guo, Y., Tao, D., Wan, J.: Human pose recovery by supervised spectral embedding. Neurocomputing **166** (2015) 301–308

[185] Chen, X., Yuille, A.L.: Articulated pose estimation by a graphical model with image dependent pairwise relations. In: Advances in Neural Information Processing Systems. (2014) 1736–1744

[186] Jain, A., Tompson, J., Andriluka, M., Taylor, G.W., Bregler, C.: Learning human pose estimation features with convolutional networks. In: International Conference on Learning Representations. (2014)

[187] Tompson, J.J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: Advances in Neural Information Processing Systems. (2014) 1799–1807

[188] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C.: Efficient object localization using convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 648–656

[189] Ouyang, W., Chu, X., Wang, X.: Multi-source deep learning for human pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 2329–2336

[190] Fan, X., Zheng, K., Lin, Y., Wang, S.: Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 1347–1355

[191] Carreira, J., Agrawal, P., Fragkiadaki, K., Malik, J.: Human pose estimation with iterative error feedback. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 4733–4742

[192] Huang, C.H., Boyer, E., Ilic, S.: Robust human body shape and pose tracking. In: 3DTV-Conference, 2013 International Conference on. (2013) 287–294

[193] Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. International Journal of Computer Vision **61** (2005) 55–79

[194] Tian, Y., Zitnick, C.L., Narasimhan, S.G.: Exploring the spatial hierarchy of mixture models for human pose estimation. In: Proceedings of European Conference on Computer Vision. (2012) 256–269

[195] Wang, F., Li, Y.: Beyond physical connections: Tree models in human pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 596–603

[196] Pishchulin, L., Andriluka, M., Gehler, P., Schiele, B.: Poselet conditioned pictorial structures. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 588–595

[197] Dantone, M., Gall, J., Leistner, C., Van Gool, L.: Human pose estimation using body parts dependent joint regressors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 3041–3048

[198] Sapp, B., Taskar, B.: Modec: Multimodal decomposable models for human pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 3674–3681

[199] Johnson, S., Everingham, M.: Clustered pose and nonlinear appearance models for human pose estimation. In: British Machine Vision Conference. (2010)

[200] Eichner, M., Marin-Jimenez, M., Zisserman, A., Ferrari, V.: 2d articulated human pose estimation and retrieval in (almost) unconstrained still images. International Journal of Computer Vision **99** (2012) 190–214

[201] Toshev, A., Szegedy, C.: Deeppose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 1653–1660

[202] Chu, J.L., Krzyżak, A.: Analysis of feature maps selection in supervised learning using convolutional neural networks. In: Canadian Conference on Artificial Intelligence. (2014) 59–70

[203] Yu, W., Yang, K., Bai, Y., Yao, H., Rui, Y.: Visualizing and comparing convolutional neural networks. arXiv preprint arXiv:1412.6631 (2014)

[204] Agrawal, P., Girshick, R., Malik, J.: Analyzing the performance of multilayer neural networks for object recognition. In: Proceedings of European Conference on Computer Vision. (2014) 329–344

[205] Cadieu, C.F., Hong, H., Yamins, D.L., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., DiCarlo, J.J.: Deep neural networks rival the representation of primate it cortex for core visual object recognition. PLoS Comput Biol **10** (2014) e1003963

[206] Nguyen, A., Yosinski, J., Clune, J.: Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 427–436

[207] Firat, O., Aksan, E., Oztekin, I., Vural, F.T.Y.: Learning deep temporal representations for brain decoding. arXiv preprint arXiv:1412.7522 (2014)

[208] Chen, X., Shrivastava, A., Gupta, A.: Neil: Extracting visual knowledge from web data. In: Proceedings of the IEEE International Conference on Computer Vision. (2013) 1409–1416

[209] Divvala, S.K., Farhadi, A., Guestrin, C.: Learning everything about anything: Webly-supervised visual concept learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 3270–3277

[210] Song, H.O., Lee, Y.J., Jegelka, S., Darrell, T.: Weakly-supervised discovery of visual pattern configurations. In: Advances in Neural Information Processing Systems. (2014) 1637–1645

[211] Li, H., Zhao, R., Wang, X.: Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification. arXiv preprint arXiv:1412.4526 (2014)

[212] Ren, J.S., Xu, L.: On vectorization of deep convolutional neural networks for vision tasks. In: AAAI Conference on Artificial Intelligence. (2015)

[213] Liu, Y., Guo, Y., Wu, S., Lew, M.S.: Deepindex for accurate and efficient image retrieval. In: Proceedings of the ACM International Conference on Multimedia Retrieval. (2015) 43–50

[214] Zheng, L., Wang, S., He, F., Tian, Q.: Seeing the big picture: Deep embedding with contextual evidences. arXiv preprint arXiv:1406.0132 (2014)

[215] Yan, Z., Jagadeesh, V., DeCoste, D., Di, W., Piramuthu, R.: Hd-cnn: Hierarchical deep convolutional neural network for image classification. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 2740–2748

[216] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems. (2014) 487–495

[217] Koskela, M., Laaksonen, J.: Convolutional network features for scene recognition. In: Proceedings of the ACM International Conference on Multimedia. (2014) 1169–1172

[218] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the ACM International Conference on Multimedia. (2014) 675–678

[219] Jolliffe, I.: Principal component analysis. Wiley Online Library (2002)

[220] Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. Computer Vision and Image Understanding **106** (2007) 59–70

[221] Quattoni, A., Torralba, A.: Recognizing indoor scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2009) 413–420

[222] Jegou, H., Douze, M., Schmid, C.: Hamming embedding and weak geometric consistency for large scale image search. In: Proceedings of European Conference on Computer Vision. (2008) 304–317

[223] Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology **2** (2011) 27

[224] Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2007) 1–8

[225] Seber, G.A.: Multivariate observations. Volume 252. John Wiley & Sons (2009)

[226] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: A review. Neurocomputing **187** (2016) 27–48

[227] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 770–778

[228] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: International Conference on Machine Learning. (2014) 647–655

[229] Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of

the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 806–813

[230] Xiu-Shen, W., Bin-Bin, G., Jianxin, W.: Deep spatial pyramid ensemble for cultural event recognition. In: Proceedings of the IEEE International Conference on Computer Vision Workshop. (2015)

[231] Liu, L., Shen, C., Hengel, A.v.d.: The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 4749–4757

[232] Babenko, A., Lempitsky, V.: Aggregating local deep features for image retrieval. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1269–1277

[233] Cimpoi, M., Maji, S., Vedaldi, A.: Deep filter banks for texture recognition and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 3828–3836

[234] Liu, L., Shen, C., van den Hengel, A.: Cross-convolutional-layer pooling for image recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2016)

[235] Ng, J., Yang, F., Davis, L.: Exploiting local features from deep networks for image retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop. (2015) 53–61

[236] Cimpoi, M., Maji, S., Kokkinos, I., Vedaldi, A.: Deep filter banks for texture recognition, description, and segmentation. International Journal of Computer Vision **118** (2016) 65–94

[237] Xiao, J., Hays, J., Ehinger, K., Oliva, A., Torralba, A., et al.: Sun database: Large-scale scene recognition from abbey to zoo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2010) 3485–3492

[238] Parizi, S.N., Vedaldi, A., Zisserman, A., Felzenszwalb, P.: Automatic discovery and optimization of parts for image classification. In: International Conference on Learning Representations. (2015)

[239] Singh, S., Gupta, A., Efros, A.: Unsupervised discovery of mid-level discriminative patches. In: Proceedings of European Conference on Computer Vision. (2012) 73–86

[240] Juneja, M., Vedaldi, A., Jawahar, C., Zisserman, A.: Blocks that shout: Distinctive parts for scene classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2013) 923–930

[241] Kulkarni, P., Jurie, F., Zepeda, J., Pérez, P., Chevallier, L.: Spleap: Soft pooling of learned parts for image classification. In: Proceedings of European Conference on Computer Vision. (2016) 329–345

[242] Wu, J., Gao, B.B., Liu, G.: Representing sets of instances for visual recognition. In: AAAI Conference on Artificial Intelligence. (2016)

[243] Liu, L., Wang, L., Liu, X.: In defense of soft-assignment coding. In: Proceedings of the IEEE International Conference on Computer Vision. (2011) 2486–2493

[244] Fernando, B., Fromont, E., Muselet, D., Sebban, M.: Discriminative feature fusion for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2012) 3434–3441

[245] Minka, T.P.: A comparison of numerical optimizers for logistic regression. Technical report (2003)

[246] Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9** (2008) 2579–2605

[247] Nilsback, M.E., Zisserman, A.: Automated flower classification over a large number of classes. In: Indian Conference on Computer Vision, Graphics & Image Processing. (2008) 722–729

[248] Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. Journal of machine learning research **9** (2008) 1871–1874

[249] Swain, M.J., Ballard, D.H.: Color indexing. International Journal of Computer Vision **7** (1991) 11–32

[250] Zhou, B., Khosla, A., Lapedriza, A., Torralba, A., Oliva, A.: Places: An image database for deep scene understanding. arXiv preprint arXiv:1610.02055 (2016)

[251] Xie, L., Tian, Q., Hong, R., Zhang, B.: Image classification and retrieval are one. In: Proceedings of the ACM International Conference on Multimedia Retrieval. (2015) 3–10

[252] Yang, S., Ramanan, D.: Multi-scale recognition with dag-cnns. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1215–1223

[253] Sharma, G., Schiele, B.: Scalable nonlinear embeddings for semantic category-based image retrieval. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 1296–1304

[254] Li, Y., Liu, L., Shen, C., van den Hengel, A.: Mining mid-level visual patterns with deep cnn activations. International Journal of Computer Vision (2016) 1–21

[255] Herranz, L., Jiang, S., Li, X.: Scene recognition with cnns: objects, scales and dataset bias. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 571–579

[256] Kim, Y.D., Jang, T., Han, B., Choi, S.: Learning to select pre-trained deep representations with bayesian evidence framework. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 5318–5326

[257] Wei, X.S., Luo, J.H., Wu, J., Zhou, Z.H.: Selective convolutional descriptor aggregation for fine-grained image retrieval. IEEE Transactions on Image Processing (2017)

[258] Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2008) 1–8

[259] Guo, Y., Lew, M.S.: Bag of surrogate parts: one inherent feature of deep cnns. In: British Machine Vision Conference. (2016)

[260] Shirahama, K., Grzegorzek, M.: Towards large-scale multimedia retrieval enriched by knowledge about human interpretation. Multimedia Tools and Applications **75** (2016) 297–331

[261] Cao, L., Gao, L., Song, J., Shen, F., Wang, Y.: Multiple hierarchical deep hashing for large scale image retrieval. Multimedia Tools and Applications (2017) 1–14

[262] Zagoruyko, S., Komodakis, N.: Wide residual networks. In: British Machine Vision Conference. (2016)

[263] Liu, Y., Guo, Y., Lew, M.S.: On the exploration of convolutional fusion networks for visual recognition. In: International Conference on Multimedia Modeling. (2017) 277–289

[264] Deng, J., Ding, N., Jia, Y., Frome, A., Murphy, K., Bengio, S., Li, Y., Neven, H., Adam, H.: Large-scale object classification using label relation graphs. In: Proceedings of European Conference on Computer Vision. (2014) 48–64

[265] Ristin, M., Gall, J., Guillaumin, M., Van Gool, L.: From categories to subcategories: large-scale image classification with partial class label refinement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 231–239

[266] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T.: Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 2625–2634

[267] Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: A neural image caption generator. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 3156–3164

[268] You, Q., Jin, H., Wang, Z., Fang, C., Luo, J.: Image captioning with semantic attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 4651–4659

[269] Zuo, Z., Shuai, B., Wang, G., Liu, X., Wang, X., Wang, B., Chen, Y.: Convolutional recurrent neural networks: Learning spatial dependencies for image representation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop. (2015) 18–26

[270] Liang, M., Hu, X.: Recurrent convolutional neural network for object recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 3367–3375

[271] Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., Xu, W.: Cnn-rnn: A unified framework for multi-label image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 2285–2294

[272] Visin, F., Kastner, K., Cho, K., Matteucci, M., Courville, A., Bengio, Y.: Renet: A recurrent neural network based alternative to convolutional networks. arXiv preprint arXiv:1505.00393 (2015)

[273] Yan, G., Wang, Y., Liao, Z.: Lstm for image annotation with relative visual importance. In: British Machine Vision Conference. (2016)

[274] Salakhutdinov, R., Torralba, A., Tenenbaum, J.: Learning to share visual appearance for multiclass object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2011) 1481–1488

[275] Murdock, C., Li, Z., Zhou, H., Duerig, T.: Blockout: Dynamic model selection for hierarchical deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 2583–2591

[276] Elman, J.L.: Finding structure in time. Cognitive science **14** (1990) 179–211

[277] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9** (1997) 1735–1780

[278] Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images. Master's thesis, Department of Computer Science, University of Toronto (2009)

[279] He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Proceedings of European Conference on Computer Vision. (2016) 630–645

[280] Romero, A., Ballas, N., Kahou, S.E., Chassang, A., Gatta, C., Bengio, Y.: Fitnets: Hints for thin deep nets. In: International Conference on Learning Representations. (2015)

[281] Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: International Conference on Artificial Intelligence and Statistics. (2015) 562–570

[282] Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. In: International Conference on Learning Representations Workshop. (2015)

[283] Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks. In: International Conference on Learning Representations Workshop. (2015)

[284] Agostinelli, F., Hoffman, M., Sadowski, P., Baldi, P.: Learning activation functions to improve deep neural networks. In: International Conference on Learning Representations Workshop. (2015)

[285] Jin, X., Xu, C., Feng, J., Wei, Y., Xiong, J., Yan, S.: Deep learning with s-shaped rectified linear activation units. In: AAAI Conference on Artificial Intelligence. (2016)

[286] Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M.M.A., Prabhat, M., Adams, R.P.: Scalable bayesian optimization using deep neural networks. In: International Conference on Machine Learning. (2015) 2171–2180

[287] Mishkin, D., Matas, J.: All you need is a good init. In: International Conference on Learning Representations. (2016)

[288] Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). In: International Conference on Learning Representations. (2016)

[289] Li, H., Ouyang, W., Wang, X.: Multi-bias non-linear activation in deep neural networks. In: International Conference on Machine Learning. (2016)

[290] Huang, G., Sun, Y., Liu, Z., Sedra, D., Weinberger, K.Q.: Deep networks with stochastic depth. In: Proceedings of European Conference on Computer Vision. (2016) 646–661

[291] Targ, S., Almeida, D., Lyman, K.: Resnet in resnet: generalizing residual architectures. In: International Conference on Learning Representations Workshop. (2016)

[292] Larsson, G., Maire, M., Shakhnarovich, G.: Fractalnet: Ultra-deep neural networks without residuals. In: International Conference on Learning Representations. (2017)

[293] Singh, S., Hoiem, D., Forsyth, D.: Swapout: Learning an ensemble of deep architectures. In: Advances in Neural Information Processing Systems. (2016) 28–36

[294] Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38** (1995) 39–41

[295] Mensink, T., Verbeek, J., Perronnin, F., Csurka, G.: Distance-based image classification: Generalizing to new classes at near-zero cost. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 2624–2637

[296] Akata, Z., Perronnin, F., Harchaoui, Z., Schmid, C.: Good practice in large-scale learning for image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **36** (2014) 507–520

[297] Ordonez, V., Kulkarni, G., Berg, T.L.: Im2text: Describing images using 1 million captioned photographs. In: Advances in Neural Information Processing Systems. (2011) 1143–1151

[298] Hodosh, M., Young, P., Hockenmaier, J.: Framing image description as a ranking task: Data, models and evaluation metrics. Journal of Artificial Intelligence Research **47** (2013) 853–899

[299] Kuznetsova, P., Ordonez, V., Berg, A.C., Berg, T.L., Choi, Y.: Collective generation of natural image descriptions. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. (2012) 359–368

[300] Mitchell, M., Han, X., Dodge, J., Mensch, A., Goyal, A., Berg, A., Yamaguchi, K., Berg, T., Stratos, K., Daumé III, H.: Midge: Generating image descriptions from computer vision detections. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. (2012) 747–756

[301] Kulkarni, G., Premraj, V., Dhar, S., Li, S., Choi, Y., Berg, A.C., Berg, T.L.: Baby talk: Understanding and generating image descriptions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2011)

[302] Karpathy, A., Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 3128–3137

[303] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y.: Show, attend and tell: Neural image caption generation

with visual attention. In: International Conference on Machine Learning. (2015) 2048–2057

[304] Mao, J., Xu, W., Yang, Y., Wang, J., Huang, Z., Yuille, A.: Deep captioning with multimodal recurrent neural networks (m-rnn). In: International Conference on Learning Representations. (2015)

[305] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Proceedings of European Conference on Computer Vision. (2014) 740–755

[306] Wang, C., Yang, H., Bartz, C., Meinel, C.: Image captioning with deep bidirectional lstms. In: Proceedings of the 2016 ACM on Multimedia Conference. (2016) 988–997

[307] Jia, X., Gavves, E., Fernando, B., Tuytelaars, T.: Guiding the long-short term memory model for image caption generation. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 2407–2415

[308] Jin, J., Fu, K., Cui, R., Sha, F., Zhang, C.: Aligning where to see and what to tell: image caption with region-based attention and scene factorization. arXiv preprint arXiv:1506.06272 (2015)

[309] Fang, H., Gupta, S., Iandola, F., Srivastava, R.K., Deng, L., Dollár, P., Gao, J., He, X., Mitchell, M., Platt, J.C., et al.: From captions to visual concepts and back. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 1473–1482

[310] Johnson, J., Karpathy, A., Fei-Fei, L.: Densecap: Fully convolutional localization networks for dense captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 4565–4574

[311] Yao, L., Ballas, N., Cho, K., Smith, J.R., Bengio, Y.: Oracle performance for visual captioning. In: British Machine Vision Conference. (2016)

[312] Wu, Q., Shen, C., Liu, L., Dick, A., van den Hengel, A.: What value do explicit high level concepts have in vision to language problems? In: Proceed-

ings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 203–212

[313] Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: The 40th annual meeting of the Association for Computational Linguistics. (2002) 311–318

[314] Lavie, M.D.A.: Meteor universal: Language specific translation evaluation for any target language. In: The 52nd Annual Meeting of the Association for Computational Linguistics. (2014)

[315] Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: The 42nd Annual Meeting of the Association for Computational Linguistics Workshop. (2004)

[316] Vedantam, R., Lawrence Zitnick, C., Parikh, D.: Cider: Consensus-based image description evaluation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 4566–4575

[317] Chen, X., Fang, H., Lin, T.Y., Vedantam, R., Gupta, S., Dollár, P., Zitnick, C.L.: Microsoft coco captions: Data collection and evaluation server. arXiv preprint arXiv:1504.00325 (2015)

[318] Liu, Y., Lew, M.S.: Learning relaxed deep supervision for better edge detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 231–240