References

[1] X. Liu, D. Liang, S. Yan, D. Chen, Y. Qiao, and J. Yan, "FOTS: Fast Oriented Text Spotting with a Unified Network," in *Proc. of CVPR*, 2018, pp. 5676–5685.

[2] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu, F. Shafait, S. Uchida, and E. Valveny, "Robust Reading Competition," in *Proc. of ICDAR*, 2015, pp. 1156–1160.

[3] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i. Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazàn, and L. P. de las Heras, "ICDAR 2013 Robust Reading Competition," in *Proc. of ICDAR*, 2013, pp. 1484–1493.

[4] A. Veit, T. Matera, L. Neumann, J. Matas, and S. Belongie, "COCO-Text: Dataset and Benchmark for Text Detection and Recognition in Natural Images," in *preprint arXiv:1601.07140*, 2016, pp. 1–8.

[5] Kai Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in *Proc. in ICCV*, 2011, pp. 1457–1464.

[6] C. K. Ch'ng and C. S. Chan, "Total-Text: A Comprehensive Dataset for Scene Text Detection and Recognition," in *Proc. of ICDAR*, 2017, pp. 935–942.

[7] Y. Liu, L. Jin, S. Zhang, and S. Zhang, "Detecting Curve Text in the Wild: New Dataset and New Solution," *ArXiv*, 2017, vol. abs/1712.02170, pp. 1–9.

[8] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic Data for Text Localisation in Natural Images," in *Proc. of CVPR*, 2016, pp. 2315–2324.

[9] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. of CVPR*, 2009, pp. 248–255.

[10] C. Fellbaum, *WordNet: An Electronic Lexical Database.* Bradford Books, 1998.

[11] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," in *Proc. of CVPR*, 2012, pp. 1083–1090.

[12] N. Nayef, F. Yin, I. Bizid, H. Choi, Y. Feng, D. Karatzas, Z. Luo, U. Pal, C. Rigaud, J. Chazalon, W. Khlif, M. M. Luqman, J. Burie, C. Liu, and J. Ogier, "ICDAR2017 Robust Reading Challenge on Multi-Lingual Scene Text Detection and Script Identification - RRC-MLT," in *Proc. in ICDAR*, vol. 01, 2017, pp. 1454–1459.

[13] T. Yuan, Z. Zhu, K. Xu, C. Li, T. Mu, and S. Hu, "A Large Chinese Text Dataset in the Wild," *Journal of Computer Science and Technology*, 2019, vol. 34, no. 3, pp. 509–521.

[14] B. Shi, C. Yao, M. Liao, M. Yang, P. Xu, L. Cui, S. Belongie, S. Lu, and X. Bai, "ICDAR2017 Competition on Reading Chinese Text in the Wild (RCTW-17)," in *Proc. of ICDAR*, vol. 01, 2017, pp. 1429–1434.

[15] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "EAST: An Efficient and Accurate Scene Text Detector," in *Proc. of CVPR*, 2017, pp. 2642–2651.

[16] W. He, X. Zhang, F. Yin, and C. Liu, "Deep Direct Regression for Multi-oriented Scene Text Detection," in *Proc. of ICCV*, 2017, pp. 745–753.

[17] B. Shi, X. Bai, and S. Belongie, "Detecting Oriented Text in Natural Images by Linking Segments," in *Proc. of CVPR*, 2017, pp. 3482–3490.

[18] Y. Liu and L. Jin, "Deep Matching Prior Network: Toward Tighter Multi-oriented Text Detection," in *Proc. of CVPR*, 2017, pp. 3454–3461.

[19] C. Xue, S. Lu, and F. Zhan, "Accurate Scene Text Detection Through Border Semantics Awareness and Bootstrapping," in *Proc. of ECCV*, 2018, pp. 370–387.

[20] S. Ruan, J. Lu, F. Xie, and Z. Jin, "A novel method for fast arbitrary-oriented scene text detection," in *Proc. of CCDC*, 2018, pp. 1652–1657.

[21] P. Xie, J. Xiao, Y. Cao, J. Zhu, and A. Khan, "RefineText: Refining Multi-oriented Scene Text Detection with a Feature Refinement Module," in *Proc. of ICME*, 2019, pp. 1756–1761.

[22] C. Du, C. Wang, Y. Wang, Z. Feng, and J. Zhang, "TextEdge: Multi-oriented Scene Text Detection via Region Segmentation and Edge Classification," in *Proc. of ICDAR*, 2019, pp. 375–380.

[23] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals," *IEEE Trans. on Multimedia*, 2018, vol. 20, no. 11, pp. 3111–3122.

[24] Z. Tian, M. Shu, P. Lyu, R. Li, C. Zhou, X. Shen, and J. Jia, "Learning Shape-Aware Embedding for Scene Text Detection," in *Proc. of CVPR*, 2019, pp. 4229–4238.

[25] J. Duan, Y. Xu, Z. Kuang, X. Yue, H. Sun, Y. Guan, and W. Zhang, "Geometry Normalization Networks for Accurate Scene Text Detection," in *Proc. of ICCV*, 2019, pp. 9136–9145.

[26] Y. Liu, L. Jin, Z. Xie, C. Luo, S. Zhang, and L. Xie, "Tightness-Aware Evaluation Protocol for Scene Text Detection," in *Proc. of CVPR*, 2019, pp. 9604–9612.

[27] C. Wang, H. Fu, L. Yang, and X. Cao, "Text Co-Detection in Multi-View Scene," *IEEE Trans. on Image Processing*, 2020, vol. 29, pp. 4627–4642.

[28] Z. Zhong, L. Sun, and Q. Huo, "A Teacher-Student Learning Based Born-Again Training Approach to Improving Scene Text Detection Accuracy," in *Proc. of ICDAR*, 2019, pp. 281–286.

[29] P. Yang, G. Yang, X. Gong, P. Wu, X. Han, J. Wu, and C. Chen, "Instance Segmentation Network With Self-Distillation for Scene Text Detection," *IEEE Access*, 2020, vol. 8, pp. 45 825–45 836.

[30] M. Cao, Y. Zou, D. Yang, and C. Liu, "GISCA: Gradient-Inductive Segmentation Network With Contextual Attention for Scene Text Detection," *IEEE Access*, 2019, vol. 7, pp. 62 805–62 816.

[31] Q. Yang, M. Cheng, W. Zhou, Y. Chen, M. Qiu, and W. Lin, "IncepText: A New Inception-Text Module with Deformable PSROI Pooling for Multi-Oriented Scene Text Detection," in *Proc. of IJCAI*, 2018, pp. 1071–1077.

[32] S. Long, J. Ruan, W. Zhang, X. He, W. Wu, and C. Yao, "TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes," in *Proc. of ECCV*, 2018, pp. 19–35.

[33] Y. Xiao, M. Xue, T. Lu, Y. Wu, and S. Palaiahnakote, "A Text-Context-Aware CNN Network for Multi-oriented and Multi-language Scene Text Detection," in *Proc. of ICDAR*, 2019, pp. 695–700.

[34] F. Sheng, Z. Chen, T. Mei, and B. Xu, "A Single-Shot Oriented Scene Text Detector with Learnable Anchors," in *Proc. of ICME*, 2019, pp. 1516–1521.

[35] P. Lyu, C. Yao, W. Wu, S. Yan, and X. Bai, "Multi-oriented Scene Text Detection via Corner Localization and Region Segmentation," in *Proc. of CVPR*, 2018, pp. 7553–7563.

[36] M. Liao, Z. Zhu, B. Shi, G. Xia, and X. Bai, "Rotation-Sensitive Regression for Oriented Scene Text Detection," in *Proc. of CVPR*, 2018, pp. 5909–5918.

[37] H. Hu, C. Zhang, Y. Luo, Y. Wang, J. Han, and E. Ding, "WordSup: Exploiting Word Annotations for Character Based Text Detection," in *Proc. of ICCV*, 2017, pp. 4950–4959.

[38] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character Region Awareness for Text Detection," in *Proc. of CVPR*, 2019, pp. 9365–9374.

[39] S. Qin and R. Manduchi, "Cascaded Segmentation-Detection Networks for Word-Level Text Spotting," in *Proc. of ICDAR*, vol. 01, 2017, pp. 1275–1282.

[40] S. Mohanty, T. Dutta, and H. P. Gupta, "Recurrent Global Convolutional Network for Scene Text Detection," in *Proc. of ICIP*, 2018, pp. 2750–2754.

[41] S. Mohanty, T. Dutta, and H. Prabhat Gupta, "Robust Scene Text Detection with Deep Feature Pyramid Network and CNN based NMS Model," in *Proc. of ICPR*, 2018, pp. 3741–3746.

[42] K. S. Raghunandan, P. Shivakumara, S. Roy, G. H. Kumar, U. Pal, and T. Lu, "Multi-Script-Oriented Text Detection and Recognition in Video/Scene/Born Digital Images," *IEEE Trans. on Circuits and Systems for Video Technology*, 2019, vol. 29, no. 4, pp. 1145–1162.

[43] S. Dey, P. Shivakumara, K. Raghunandan, U. Pal, T. Lu, G. H. Kumar, and C. S. Chan, "Script independent approach for multi-oriented text detection in scene image," *Neurocomputing*, 2017, vol. 242, pp. 96–112.

[44] V. Khare, P. Shivakumara, and P. Raveendran, "A new Histogram Oriented Moments descriptor for multi-oriented moving text detection in video," *Expert Systems with Applications*, 2015, vol. 42, no. 21, pp. 7627–7640.

[45] P. Cheng, W. Wang, and Y. Cai, "Multi-scale Scene Text Detection via Resolution Transform," in *Proc. of ICME*, 2019, pp. 988–993.

[46] D. He, X. Yang, W. Huang, Z. Zhou, D. Kifer, and C. L. Giles, "Aggregating Local Context for Accurate Scene Text Detection," in *Proc. of ACCV*, 2017, pp. 280–296.

[47] Y. Tang and X. Wu, "Scene Text Detection Using Superpixel-Based Stroke Feature Transform and Deep Learning Based Region Classification," *IEEE Trans. on Multimedia*, 2018, vol. 20, no. 9, pp. 2276–2288.

[48] ——, "Scene Text Detection and Segmentation Based on Cascaded Convolution Neural Networks," *IEEE Trans. on Image Processing*, 2017, vol. 26, no. 3, pp. 1509–1520.

[49] Y. Xu, Y. Wang, W. Zhou, Y. Wang, Z. Yang, and X. Bai, "TextField: Learning a Deep Direction Field for Irregular Scene Text Detection," *IEEE Trans. on Image Processing*, 2019, vol. 28, no. 11, pp. 5566–5579.

[50] D. He, X. Yang, D. Kifer, and L. Giles, "TextContourNet: A Flexible and Effective Framework for Improving Scene Text Detection Architecture With a Multi-Task Cascade," in *Proc. of WACV*, 2019, pp. 676–685.

[51] Z. Huang, Z. Zhong, L. Sun, and Q. Huo, "Mask R-CNN with Pyramid Attention Network for Scene Text Detection," in *Proc. of WACV*, 2019, pp. 764–772.

[52] X. Guo, J. Li, B. Chen, and G. Lu, "Mask-Most Net: Mask Approximation Based Multi-oriented Scene Text Detection Network," in *Proc. of ICME)*, 2019, pp. 206–211.

[53] X. Wang, Y. Jiang, Z. Luo, C.-L. Liu, H. Choi, and S. Kim, "Arbitrary Shape Scene Text Detection With Adaptive Text Region Representation," in *Proc. of CVPR*, 2019, pp. 6449–6458.

[54] Y. Liu, L. Jin, and C. Fang, "Arbitrarily Shaped Scene Text Detection With a Mask Tightness Text Detector," *IEEE Trans. on Image Processing*, 2020, vol. 29, pp. 2918–2930.

[55] P. Dai, H. Zhang, and X. Cao, "Deep Multi-Scale Context Aware Feature Aggregation for Curved Scene Text Detection," *IEEE Trans. on Multimedia*, 2020, vol. 22, no. 8, pp. 1969–1984.

[56] S. Zhang, Y. Liu, L. Jin, Z. Wei, and C. Shen, "OPMP: An Omni-directional Pyramid Mask Proposal Network for Arbitrary-shape Scene Text Detection," *IEEE Trans. on Multimedia*, 2020, pp. 1–14.

[57] C. Yao, X. Bai, N. Sang, X. Zhou, S. Zhou, and Z. Cao, "Scene Text Detection via Holistic, Multi-Channel Prediction," *CoRR*, 2016, vol. abs/1606.09002, pp. 1–10.

[58] C. Xue, S. Lu, and W. Zhang, "MSR: Multi-Scale Shape Regression for Scene Text Detection," in *Proc. of IJCAI*, 2019, pp. 989–995.

[59] W. Wang, E. Xie, X. Li, W. Hou, T. Lu, G. Yu, and S. Shao, "Shape Robust Text Detection With Progressive Scale Expansion Network," in *Proc. of CVPR*, 2019, pp. 9336–9345.

[60] W. Wang, E. Xie, X. Song, Y. Zang, W. Wang, T. Lu, G. Yu, and C. Shen, "Efficient and Accurate Arbitrary-Shaped Text Detection With Pixel Aggregation Network," in *Proc. of ICCV*, 2019, pp. 8439–8448.

[61] Q. Wang, J. Gao, and Y. Yuan, "Embedding Structured Contour and Location Prior in Siamesed Fully Convolutional Networks for Road Detection," *IEEE Trans. Intell. Transp. Syst.*, 2018, vol. 19, no. 1, pp. 230–241.

[62] J. Gao, Q. Wang, and Y. Yuan, "Convolutional Regression Network for Multi-Oriented Text Detection," *IEEE Access*, 2019, vol. 7, pp. 96 424–96 433.

[63] E. Xie, Y. Zang, S. Shao, G. Yu, C. Yao, and G. Li, "Scene Text Detection with Supervised Pyramid Context Network," in *Proc. of AAAI*, 2019, pp. 9038–9045.

[64] Y. Wang, H. Xie, Z. Zha, Y. Tian, Z. Fu, and Y. Zhang, "R-Net: A Relationship Network for Efficient and Accurate Scene Text Detection," *IEEE Trans. on Multimedia*, 2020, pp. 1–14.

[65] Y. Wang, H. Xie, Z. Fu, and Y. Zhang, "DSRN: A Deep Scale Relationship Network for Scene Text Detection," in *Proc. of IJCAI-19*, 2019, pp. 947–953.

[66] J. Hou, X. Zhu, C. Liu, K. Sheng, L. Wu, H. Wang, and X. Yin, "HAM: Hidden Anchor Mechanism for Scene Text Detection," *IEEE Trans. on Image Processing*, 2020, vol. 29, pp. 7904–7916.

[67] J. Hou, X. Zhu, C. Liu, C. Yang, L. Wu, H. Wang, and X. Yin, "Detecting Text in Scene and Traffic Guide Panels With Attention Anchor Mechanism," *IEEE Trans. on Intelligent Transportation Systems*, 2020, pp. 1–10.

[68] C. Zhang, B. Liang, Z. Huang, M. En, J. Han, E. Ding, and X. Ding, "Look More Than Once: An Accurate Detector for Text of Arbitrary Shapes," in *Proc. of CVPR*, 2019, pp. 10 552–10 561.

[69] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-time Scene Text Detection with Differentiable Binarization," in *Proc. of AAAI*, 2020, pp. 11 475–11 481.

[70] Y. Wang, H. Xie, Z.-J. Zha, M. Xing, Z. Fu, and Y. Zhang, "ContourNet: Taking a Further Step Toward Accurate Arbitrary-Shaped Scene Text Detection," in *Proc. of CVPR*, 2020, pp. 11 753–111 762.

[71] F. Cong, W. Hu, Q. Huo, and L. Guo, "A Comparative Study of Attention-Based Encoder-Decoder Approaches to Natural Scene Text Recognition," in *Proc. of ICDAR*, 2019, pp. 916–921.

[72] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, vol. 39, no. 11, pp. 2298–2304.

[73] B. Shi, X. Wang, P. Lyu, C. Yao, and X. Bai, "Robust Scene Text Recognition with Automatic Rectification," in *Proc. of CVPR*, 2016, pp. 4168–4176.

[74] C. Y. Lee and S. Osindero, "Recursive Recurrent Nets with Attention Modeling for OCR in the Wild," in *Proc. of CVPR*, 2016, pp. 2231–2239.

[75] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu, and S. Zhou, "Focusing Attention: Towards Accurate Text Recognition in Natural Images," in *Proc. of ICCV*, 2017, pp. 5086–5094.

[76] Z. Cheng, Y. Xu, F. Bai, Y. Niu, S. Pu, and S. Zhou, "AON: Towards Arbitrarily-Oriented Text Recognition," in *Proc. of CVPR*, 2018, pp. 5571–5579.

[77] F. Bai, Z. Cheng, Y. Niu, S. Pu, and S. Zhou, "Edit Probability for Scene Text Recognition," in *Proc. of CVPR*, 2018, pp. 1508–1516.

[78] L. Xing, Z. Tian, W. Huang, and M. R. Scott, "Convolutional Character Networks," in *Proc. of ICCV*, 2019, pp. 9126–9136.

[79] F. Zhan and S. Lu, "ESIR: End-To-End Scene Text Recognition via Iterative Image Rectification," in *Proc. of CVPR*, 2019, pp. 2054–2063.

[80] S. Wang, Y. Wang, X. Qin, Q. Zhao, and Z. Tang, "Scene Text Recognition via Gated Cascade Attention," in *Proc. of ICME*, 2019, pp. 1018–1023.

[81] M. Yang, Y. Guan, M. Liao, X. He, K. Bian, S. Bai, C. Yao, and X. Bai, "Symmetry-Constrained Rectification Network for Scene Text Recognition," in *Proc. of ICCV*, 2019, pp. 9146–9155.

[82] F. Sheng, Z. Chen, and B. Xu, "NRTR: A No-Recurrence Sequence-to-Sequence Model for Scene Text Recognition," in *Proc. of ICDAR*, 2019, pp. 781–786.

[83] Z. Zhang, P. Chen, X. Shi, and L. Yang, "Text-guided Neural Network Training for Image Recognition in Natural Scenes and Medicine," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2019, pp. 1–13.

[84] M. Liao, J. Zhang, Z. Wan, F. Xie, J. Liang, P. Lyu, C. Yao, and X. Bai, "Scene Text Recognition from Two-Dimensional Perspective," in *Proc. of AAAI*, 2019, pp. 8714–8721.

[85] H. Li, P. Wang, C. Shen, and G. Zhang, "Show, attend and read: a simple and strong baseline for recognising irregular text," in *Proc. of AAAI*, 2019, pp. 8610–8617.

[86] S. Long and C. Yao, "UnrealText: Synthesizing Realistic Scene Text Images from the Unreal World," *Proc. of CVPR*, 2020, pp. 1–17.

[87] D. Yu, X. Li, C. Zhang, T. Liu, J. Han, J. Liu, and E. Ding, "Towards Accurate Scene Text Recognition With Semantic Reasoning Networks," in *Proc. pf CVPR*, 2020, pp. 12 113–12 122.

[88] X. Xu, J. Chen, J. Xiao, L. Gao, F. Shen, and H. T. Shen, "What Machines See Is Not What They Get: Fooling Scene Text Recognition Models With Adversarial Text Images," in *Proc. of CVPR*, 2020, pp. 12 304–12 314.

[89] Z. Qiao, Y. Zhou, D. Yang, Y. Zhou, and W. Wang, "SEED: Semantics Enhanced Encoder-Decoder Framework for Scene Text Recognition," in *Proc. of CVPR*, 2020, pp. 13 528–13 537.

[90] Y. Liu, H. Chen, C. Shen, T. He, L. Jin, and L. Wang, "ABCNet: Real-Time Scene Text Spotting With Adaptive Bezier-Curve Network," in *Proc. of CVPR*, 2020, pp. 9809–9818.

[91] R. Litman, O. Anschel, S. Tsiper, R. Litman, S. Mazor, and R. Manmatha, "SCATTER: Selective Context Attentional Scene Text Recognizer," in *Proc. of CVPR*, 2020, pp. 11 962–11 972.

[92] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Deep Features for Text Spotting," in *Proc. of ECCV*, 2014, pp. 512–528.

[93] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading Text in the Wild with Convolutional Neural Networks," *Int J Comput Vis*, 2016, vol. 116, pp. 1–20.

[94] M. Bušta, L. Neumann, and J. Matas, "Deep TextSpotter: An End-to-End Trainable Scene Text Localization and Recognition Framework," in *Proc. of ICCV*, 2017, pp. 2223–2231.

[95] H. Li, P. Wang, and C. Shen, "Towards End-to-End Text Spotting with Convolutional Recurrent Neural Networks," in *Proc. of ICCV*, 2017, pp. 5248–5256.

[96] T. He, Z. Tian, W. Huang, C. Shen, Y. Qiao, and C. Sun, "An End-to-End TextSpotter with Explicit Alignment and Attention," in *Proc. of CVPR*, 2018, pp. 5020–5029.

[97] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu, "TextBoxes: A Fast Text Detector with a Single Deep Neural Network," in *Proc. of AAAI*, 2016, pp. 4161–4167.

[98] M. Liao, B. Shi, and X. Bai, "TextBoxes++: A Single-Shot Oriented Scene Text Detector," *IEEE Trans. on Image Processing*, 2018, vol. 27, no. 8, pp. 3676–3690.

[99] J. Li, Z. Zhou, Z. Su, S. Huang, and L. Jin, "A New Parallel Detection-Recognition Approach for End-to-End Scene Text Extraction," in *Proc. of ICDAR*, 2019, pp. 1358–1365.

[100] Z. Hong, Y. Petillot, D. Lane, Y. Miao, and S. Wang, "TextPlace: Visual Place Recognition and Topological Localization Through Reading Scene Texts," in *Proc. of ICCV*, 2019, pp. 2861–2870.

[101] Y. Zhou, S. Fang, H. Xie, Z. Zha, and Y. Zhang, "MLTS: A Multi-Language Scene Text Spotter," in *Proc. of ICME*, 2019, pp. 163–168.

[102] M. Bušta, Y. Patel, and J. Matas, "E2E-MLT - An Unconstrained End-to-End Method for Multi-language Scene Text," in *Proc. of ACCV Workshop*, 2019, pp. 127–143.

[103] A. Lundgren, D. Castro, E. Lima, and B. Bezerra, "OctShuffleMLT: A Compact Octave Based Neural Network for End-to-End Multilingual Text Detection and Recognition," in *Proc. of ICDARW*, vol. 4, 2019, pp. 37–42.

[104] M. Liao, P. Lyu, M. He, C. Yao, W. Wu, and X. Bai, "Mask TextSpotter: An End-to-End Trainable Neural Network for Spotting Text with Arbitrary Shapes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2019, pp. 1–17.

[105] W. Feng, W. He, F. Yin, X. Zhang, and C. Liu, "TextDragon: An End-to-End Framework for Arbitrary Shaped Text Spotting," in *Proc. of ICCV*, 2019, pp. 9075–9084.

[106] Y. Gao, Z. Huang, Y. Dai, K. Chen, J. Guo, and W. Qiu, "Wacnet: Word Segmentation Guided Characters Aggregation Net for Scene Text Spotting With Arbitrary Shapes," in *Proc. of ICIP*, 2019, pp. 3382–3386.

[107] J. Liu, Z. Chen, B. Du, and D. Tao, "ASTS: A Unified Framework for Arbitrary Shape Text Spotting," *IEEE Trans. on Image Processing*, 2020, vol. 29, pp. 5924–5936.

[108] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. of ICCV*, 2017, pp. 2980–2988.

[109] Z. Liu, G. Lin, S. Yang, F. Liu, W. Lin, and W. L. Goh, "Towards Robust Curve Text Detection With Conditional Spatial Expansion," in *Proc. of CVPR*, 2019, pp. 7261–7270.

[110] H. Wang, P. Lu, H. Zhang, M. Yang, X. Bai, Y. Xu, M. He, Y. Wang, and W. Liu, "All You Need Is Boundary: Toward Arbitrary-Shaped Text Spotting," in *Proc. of AAAI*, 2020, pp. 12 160–12 167.

[111] R. Bagi, T. Dutta, and H. P. Gupta, "Cluttered TextSpotter: An End-to-End Trainable Light-Weight Scene Text Spotter for Cluttered Environment," *IEEE Access*, 2020, vol. 8, pp. 111 433–111 447.

[112] L. Qiao, S. Tang, Z. Cheng, Y. Xu, Y. Niu, S. Pu, and F. Wu, "Text Perceptron: Towards End-to-End Arbitrary-Shaped Text Spotting," *Proc. of AAAI*, 2020.

[113] S.-X. Zhang, X. Zhu, J.-B. Hou, C. Liu, C. Yang, H. Wang, and X.-C. Yin, "Deep Relational Reasoning Graph Network for Arbitrary Shape Text Detection," in *Proc. of CVPR*, 2020, pp. 9699–9708.

[114] Y. Zhu, C. Zhao, J. Wang, X. Zhao, Y. Wu, and H. Lu, "CoupleNet: Coupling Global Structure with Local Parts for Object Detection," in *Proc. of ICCV*, 2017, pp. 4146–4154.

[115] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. of CVPR*, 2018, pp. 4510–4520.

[116] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2018, vol. 40, no. 4, pp. 834–848.

[117] Y. Luo, Y. Wong, M. S. Kankanhalli, and Q. Zhao, "G-softmax: Improving Intra-class Compactness and Inter-class Separability of Features," *IEEE Trans. on Neural Networks and Learning Systems*, 2019, vol. abs/1904.04317, pp. 1–15.

[118] P. Lyu, M. Liao, C. Yao, W. Wu, and X. Bai, "Mask TextSpotter: An End-to-End Trainable Neural Network for Spotting Text with Arbitrary Shapes," in *Proc. of ECCV*, 2018, pp. 71–88.

[119] S. Wang, Z. Li, C. Ding, B. Yuan, Q. Qiu, Y. Wang, and Y. Liang, "C-LSTM: Enabling Efficient LSTM using Structured Compression Techniques on FPGAs," in *Proc. of ACM/SIGDA FPGA*, 2018, pp. 11–20.

[120] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. of ICML*, 2009, pp. 41–48.

[121] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proc. of CVPR*, 2017, pp. 6517–6525.

[122] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. of CVPR*, 2016, pp. 779–788.

[123] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, vol. 39, no. 6, pp. 1137–1149.

[124] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," in *Proc. of ECCV*. Springer International Publishing, 2018, pp. 122–138.

[125] G. Xie, J. Wang, T. Zhang, J. Lai, R. Hong, and G. Qi, "Interleaved Structured Sparse Convolutional Neural Networks," in *Proc. of CVPR*, 2018, pp. 8847–8856.

[126] T. Dutta and H. P. Gupta, "Leveraging Smart Devices for Automatic Mood-Transferring in Real-Time Oil Painting," *IEEE Trans. on Industrial Electronics*, 2017, vol. 64, no. 2, pp. 1581–1588.

[127] R. Bagi, S. Mohanty, T. Dutta, and H. P. Gupta, "Leveraging Smart Devices for Scene Text Preserved Image Stylization: A Deep Gaming Approach," *IEEE MultiMedia*, 2020, pp. 1–10.

[128] Y. Sun, C. Zhang, Z. Huang, J. Liu, J. Han, and E. Ding, "TextNet: Irregular Text Reading from Images with an End-to-End Trainable Network," in *Proc. of ACCV*, 2019, pp. 83–99.

[129] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *preprint arXiv:1409.0473*, 2014, pp. 1–15.

[130] P. Cheng and W. Wang, "A Multi-Oriented Scene Text Detector with Position-Sensitive Segmentation," in *Proc. of ICMR*, 2018, pp. 152–159.

[131] L. Neumann and J. Matas, "Real-Time Lexicon-Free Scene Text Localization and Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2016, vol. 38, no. 9, pp. 1872–1885.

[132] O. Alsharif and J. Pineau, "End-to-End Text Recognition with Hybrid HMM Maxout Models," in *Proc. of ICLR*, 2014.

[133] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proc. of ICPR2012*, 2012, pp. 3304–3308.

[134] Z. Yu, C. Feng, M. Liu, and S. Ramalingam, "CASENet: Deep Category-Aware Semantic Edge Detection," in *Proc. of CVPR*, 2017, pp. 1761–1770.

[135] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. of CVPR*, 2018.

[136] B. Jiang, R. Luo, J. Mao, T. Xiao, and Y. Jiang, "Acquisition of Localization Confidence for Accurate Object Detection," in *Proc. of ECCV*, 2018, pp. 816–832.

[137] W. Postl, "Detection of linear oblique structures and skew scan in digitized documents," in *Proc. of ICPR*, 1986, pp. 1–4.

[138] I. Goodfellow, Y. Bengio, A. C. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[139] Y. Fathullah, C. Zhang, and P. C. Woodland, "Improved Large-Margin Softmax Loss for Speaker Diarisation," in *Proc. of ICASSP*, 2020, pp. 7104–7108.

[140] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "ASTER: An Attentional Scene Text Recognizer with Flexible Rectification," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2019, vol. 41, no. 9, pp. 2035–2048.

[141] L. GÃmez and D. Karatzas", "TextProposals: A text-specific selective search algorithm for word spotting in the wild," *Pattern Recognition*, 2017, vol. 70, pp. 60 – 74.

[142] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in *Proc. of CVPR*, 2012, pp. 3538–3545.

[143] X. Yue, Z. Kuang, Z. Zhang, Z. Chen, P. He, Y. Qiao, and W. Zhang, "Boosting up Scene Text Detectors with Guided CNN," in *Proc. of BMVC*, 2018, pp. 1–14.

[144] L. Yang, Y. Song, and Y. Zhang, "Enhanced EAST: Improving Network's Feature Extraction Ability and Text Complete Shape Perception," in *Proc. of ICDAR*, 2019, pp. 769–774.

[145] F. Liu, D. Gu, and C. Chen, "IoU-Related Arbitrary Shape Text Scoring Detector," *IEEE Access*, 2019, vol. 7, pp. 180 428–180 437.

[146] S. Qin, A. Bissaco, M. Raptis, Y. Fujii, and Y. Xiao, "Towards Unconstrained End-to-End Text Spotting," in *Proc. of ICCV*, 2019, pp. 4703–4713.

[147] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial Transformer Networks," in *Proc. of NIPS*, 2015, pp. 2017–2025.

[148] A. Neubeck and L. Van Gool, "Efficient Non-Maximum Suppression," in *Proc. of ICPR*, vol. 3, 2006, pp. 850–855.

[149] J. H. Hosang, R. Benenson, and B. Schiele, "A Convnet for Non-maximum Suppression," in *Pattern Recognition*. Springer International Publishing, 2016, pp. 192–204.

[150] Y. Liu, L. Jin, S. Zhang, C. Luo, and S. Zhang, "Curved scene text detection via transverse and longitudinal sequence connection," *Pattern Recognition*, 2019, vol. 90, pp. 337–345.

[151] Z. Tian, W. Huang, T. He, P. He, and Y. Qiao, "Detecting Text in Natural Image with Connectionist Text Proposal Network," in *Proc. of ECCV*. Springer, 2016, pp. 56–72.

[152] W. Wu, J. Xing, and H. Zhou, "TextCohesion: Detecting Text for Arbitrary Shapes," *CoRR*, 2019, vol. abs/1904.12640, pp. 1–13.

LIST OF PUBLICATIONS

Refereed Journal Papers (Published/ Accepted)

1. "Cluttered Text Spotter: An End-to-End Trainable Light- weight Scene Text Spotter for Cluttered Environment", IEEE Access, 2020, Inpress.
2. "Leveraging Smart Devices for Scene Text Preserved Image Stylization: A Deep Gaming Approach", IEEE MultiMedia, 2020, Inpress.
3. "Cost-effective & Smart Text Sensing & Spotting in Blurry Scene Images using Deep Networks", IEEE Sensors Journal, 2020, Accepted.

Refereed Journal Papers (Revision/ Communicated)

1. "Multi-Attention Network for Partially Occluded Arbitrary-Shaped Scene Text Detection", IEEE Transactions on Cybernetics, 2^{nd} revision.
2. "Robust Text Detection in Blurry Scene Images", IEEE Signal Processing Letters, Communicated.
3. "Leveraging Smartpnones for End-to-end Spotting of Multilingual Oriented Scene Texts and Traffic Signs in Adverse Meteorological Conditions", IEEE Transactions on Intelligent Transportation Systems, Communicated.
4. "Multi-Attention Network for Arbitrary-shaped Text Detection in Scene Images with Faint Edges", IEEE Transactions on Circuits and Systems for Video Technology, Communicated.

Refereed Conference Papers

1. "Deep Learning Architectures for Computer Vision Applications: A Study", Advances in Data and Information Sciences, Springer, 2019, Inpress.