(a) 原始网络实验结果 (b) 特征融合网络实验结果

图 7 原始网络和特征融合网络实验结果对比

相邻三层特征融合网络与相邻两层特征融合网络相比较, 在准确率和召回率上均有所下降. 此外, 在训练过程中, 多层特征进行融合存在计算量大、消耗内存的情况, 因此本文没有采用三层以三层以上的特征融合网络.

本文所提出的三种特征融合网络中, 最高层特征融合网络的性能最好. 由于最高层的语义信息比较强, 高层的语义特征融合至其他层后, 使网络在各个层级上都具有丰富的语义, 性能上取得显著的提升, 并且不牺牲速度和内存. 因此, 之后的实验中, 本文使用最高层特征融合网络作为最佳的特征融合网络, 与常用的自然场景文本检测方法进行比较.

表 2 和表 3 分别展示了最高层特征融合网络与其他方法在 ICDAR2011 和 ICDAR2013 数据集上的实验结果. 从表中可以看出, 本文的方法在 ICDAR2011 和 ICDAR2013 数据集上, F 值都达到 0.83, 比原始网络 (Fast TextBoxes) 的 F 值的提高了 3%, 比之前最好的方法提高了 2%. 本文方法最大的优势在于召回率得到显著的提升, 在 ICDAR2011 数据集上, 本文方法比之前最好的方法 Text Flow 在召回率上提升了 4%; 在 ICDAR2013 数据集上, 本文方法比之前最好的方法 FCN 在召回率上提高了 5%, 这主要因为小尺度文本检测的召回率得到提升. 综上所述, 本文的方法相比于之前的方法, 能有效地检测出小尺度文本, 文本检测的整体性能有显著的改善.

由上述实验结果可知, 本文方法在自然场景文本检测上能够有效地检测出文字的位置. 图 8 展示了使用本文的最高层特征融合网络检测文本成功和失败的图例. 检测成功的图例 (图 8(a)) 显示出本文方法具有较高的定位准确性和鲁棒性, 能有效地从复杂背景中检测出不同大小和不同形状的文字. 对于检测失败的图例 (图 8(b)), 图像中的文字极其模糊或者文字与背景具有较低的对比度, 即使人眼也很难识别出图像中的文字区域.

表 2 在 ICDAR2011 数据集上的实验结果

方法	准确率	召回率	F 值
SFT-TCD[17]	0.82	0.75	0.73
Yin et al.[18]	0.86	0.68	0.76
MSERs-CNN[10]	0.88	0.71	0.78
Zhang et al.[19]	0.84	0.76	0.80
Fast TextBoxes[15]	0.86	0.74	0.80
Text Flow[20]	0.86	0.76	0.81
最高层特征融合网络	0.86	0.80	0.83

表 3 在 ICDAR2013 数据集上的实验结果

方法	准确率	召回率	F 值
Text Spotter[21]	0.88	0.65	0.75
Iwrr2014[22]	0.86	0.68	0.76
Text Flow[20]	0.88	0.71	0.78
Zhang et al.[19]	0.84	0.76	0.80
Fast TextBoxes[15]	0.86	0.74	0.80
FCN[23]	0.86	0.76	0.81
最高层特征融合网络	0.86	0.81	0.83

(a) 检测成功示例图

(b) 检测失败示例图

图 8 本文方法检测文本示例图

5 结论与展望

本文提出了一种基于特征融合的深度神经网络, 该网络将高层特征与低层特征相融合, 利用网络高层的强语义特征增强低层输出层的语义信息, 使整个网络的输出层都具有较强的表达能力. 特征融合后的网络能在不同的输出层上预测不同尺度以及不同形状的

文字. 本文在两个公开的数据集上验证了特征融合网络的性能, 实验结果表明本文提出的特征融合网络对小尺度的文字, 定位效果显著. 其中, 本文提出的最高层特征融合网络能取得最佳的检测效果, 具有较高的定位准确性和鲁棒性, 并优于常用的自然场景文本检测方法, F 值在 ICDAR2011 和 ICDAR2013 两个数据集上均达到了 0.83. 本文的特征融合网络只支持单尺度的图像输入, 在一定程度上限制算法性能的提升. 因此, 下一步的工作, 我们将尝试把改进后的网络改为多尺度输入的网络. 网络将会从以下两方面进行修改, 一方面是改变网络中卷积层的卷积核大小, 建立输出层中不同大小的特征图之间的整体关联性, 使网络能支持多尺度图像输入. 另一方面, 使用其他方式放大高层的特征图, 例如, 反池化操作, 即记录池化过程中最大激活值所在的坐标位置, 然后上采样得到放大的特征图, 使网络中融合的特征图能自适应进行变化而不依赖于固定计算. 接下来的工作, 我们将尝试用这两种方法, 进一步提高网络的性能.

参考文献

1 陈利. 车牌识别系统设计与实现. 现代电子技术, 2012, 35(15): 142–144.

2 胡二雷, 冯瑞. 基于深度学习的图像检索系统. 计算机系统应用, 2017, 26(3): 8–19.

3 王琦, 陈临强, 梁旭. 视频中的字幕提取. 计算机工程与应用, 2012, 48(5): 177–178, 216.

4 Ozuysal M, Fua P, Lepetit V. Fast keypoint recognition in ten lines of code. Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA. 2007. 1–8. [doi: 10.1109/CVPR.2007.383123]

5 Lee JJ, Lee PH, Lee SW, *et al*. AdaBoost for text detection in natural scene. Proceedings of 2011 International Conference on Document Analysis and Recognition. Beijing, China. 2011. 429–434. [doi: 10.1109/ICDAR.2011.93]

6 Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA. 2010. 2963–2970. [doi: 10.1109/CVPR.2010.5540041]

7 Neumann L, Matas J. A method for text localization and recognition in real-world images. Proceedings of the 10th Asian Conference on Computer Vision. Queenstown, New Zealand. 2010. 770–783. [doi: 10.1007/978-3-642-19318-7_60]

8 易尧华, 申春辉, 刘菊华, 等. 结合 MSCRs 与 MSERs 的自然场景文本检测. 中国图象图形学报, 2017, 22(2): 154–160. [doi: 10.11834/jig.20170202]

9 Jaderberg M, Vedaldi A, Zisserman A. Deep features for text spotting. Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland. 2014. 512–528

10 Huang WL, Qiao Y, Tang XO. Robust scene text detection with convolution neural network induced MSER trees. Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland. 2014. 495–511. [doi: 10.1007/978-3-319-10593-2_33]

11 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587. [doi: 10.1109/CVPR.2014.81]

12 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448. [doi: 10.1109/ICCV.2015.169]

13 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

14 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands. 2016. 21–37. [doi: 10.1007/978-3-319-46448-0_2]

15 Liao MH, Shi BG, Bai X, *et al*. TextBoxes: A fast text detector with a single deep neural network. Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, CA, USA. 2017.

16 Fu CY, Liu W, Ranga A, *et al*. DSSD: Deconvolutional single shot detector. arXiv:1701.06659, 2017.

17 Huang WL, Lin Z, Yang JC, *et al*. Text localization in natural images using stroke feature transform and text covariance descriptors. Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, NSW, Australia. 2013. 1241–1248. [doi: 10.1109/ICCV.2013.157]

18 Yin XC, Yin XW, Huang KZ, *et al*. Robust text detection in natural scene images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(5): 970–983. [doi: 10.1109/TPAMI.2013.182]

19 Zhang Z, Shen W, Yao C, *et al*. Symmetry-based text line detection in natural scenes. Proceedings of 2015 IEEE

Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 2558–2567.

20 Tian SX, Pan YF, Huang C, *et al*. Text flow: A unified text detection system in natural scene images. Proceedings of 2015 International Conference on Computer Vision. Santiago, Chile. 2015. 4651–4659.

21 Neumann L, Matas J. Real-time scene text localization and recognition. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA. 2012. 3538–3545. [doi: 10.1109/CVPR.2012.6248097]

22 Zamberletti A, Noce L, Gallo I. Text localization based on fast feature pyramids and multi-resolution maximally stable extremal regions. In: Jawahar CV, Shan SG, eds. Computer Vision - ACCV 2014 Workshops. Cham: Springer, 2014. 91–105. [doi: 10.1007/978-3-319-16631-5_7]

23 Zhang Z, Zhang CQ, Shen W, *et al*. Multi-oriented text detection with fully convolutional networks. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 4159–4167. [doi: 10.1109/CVPR.2016.451]