

加权到先前的特征上,令原始特征得以重标定.注意力机制的算法结构如图2所示.

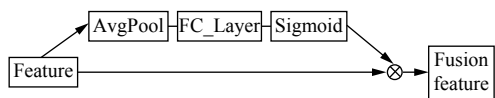


图2 注意力机制的算法结构

3 本文方法

本文提出的基于轻量级 SSD 的电力设备锈蚀目标检测算法在标准 SSD 的基础上结合 MobileNet 的深度可分离卷积进行轻量化操作,同时为保证检测精度不受影响,利用注意力机制和 concat 方式对多尺度特

征进行融合.完整算法结构如图3所示.

3.1 基于深度可分离卷积的特征提取网络

标准 SSD 采用扩展的 VGG-16 网络作为特征提取网络, VGG-16 采用的是密集连接的标准卷积方式.而深度可分离卷积是 MobileNet 的基本组成块,利用了分解卷积的方法对过参数化的标准卷积进行压缩.首先采用 1×1 大小的卷积核先对每一个通道进行卷积操作,再使用 3×3 大小的卷积核进行通道间的信息交流.通过将标准卷积中的乘法分解为加法的方式在不损失精度的情况下有效的减少大量参数.同时将激活函数由 ReLU 更换为性能更加优越的 h-swish 函数.图4为标准卷积和深度可分离卷积的结构对比图.

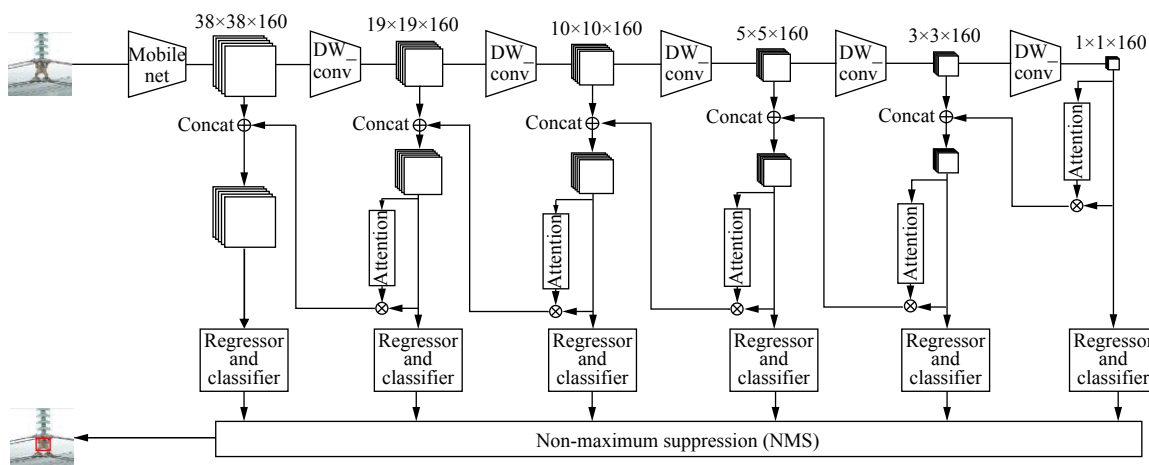


图3 基于轻量级 SSD 的电力设备锈蚀目标检测算法结构

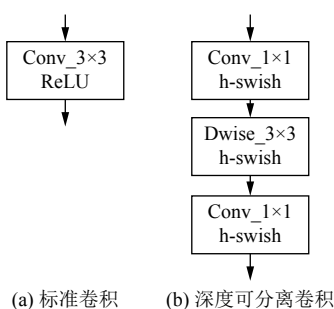


图4 标准卷积和深度可分离卷积的结构对比

3.2 基于注意力机制的上采样策略

标准 SSD 目标检测算法为了提升小目标的检测效果使用了多尺度的特征进行同时检测,但是由于没有做到特征的融合,导致小目标检测中并不能达到很好的预期效果.同时,虽然使用深度可分离卷积可以大量的减少参数量,但是由于失去了大部分可调参数,在一

定程度上牺牲了检测精度.因此本文采用了多尺度融合思想,将标准 SSD 中单独检测的多尺度特征进行上采样融合.在融合策略上,由于 FPN^[15]使用的 Elementwise add 特征融合方式对单通道特征图的特征相似度要求较高,但是经过上采样的高层特征图未必可以做到内容匹配.而 concat 特征组合方式则更加注重不同通道内的特征信息,同时再利用一个注意力模型对组合特征进行相关性标定,以此可以选择更具有价值的特征信息.但是由于采用 concat 的方式,所以通道数增加了一倍,因此采用一个卷积层将其通道降为原先的一半,同时起到特征融合的作用.基于注意力机制的上采样策略结构如图5所示.

图6(a)图为待检测的原图,其锈蚀目标部分为的防火门控制箱的外壳,图6(b)图为获得的原始特征图,图6(c)图为采用 Elementwise add 方式进行融合后的

特征图, 图 6(d) 图为本文特征融合方法进行融合后的特征图, 可以看到本文方法可以有效的抑制低能特征, 提取高能特征, 从而提高检测效果.

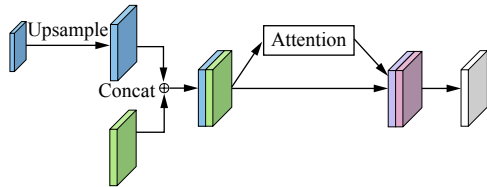


图 5 基于注意力机制的上采样策略

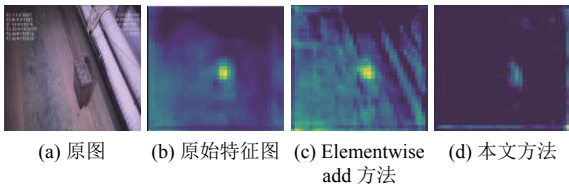


图 6 Elementwise add 与本文方法产生的特征图

4 实验分析

4.1 实验数据集

由于利用目标检测技术对电力设备进行故障检测仍处于发展阶段, 目前暂无公开的数据集用于电力设备锈蚀的目标检测, 因此本文就已有的电力设备锈蚀图像提出 RustDetection 数据集, 该数据集图片由架空输电线路、电缆隧道、电表电箱等多种电力设备锈蚀故障图片组成, 通过实地采集、网络获取等多种方式, 结合数据增加方法最终采用了 600 张锈蚀图片作为训练集, 200 张作为测试集, 并通过 LabelImg 标注软件对锈蚀区域进行标定, 按照 VOC2012 数据集的格式进行处理, 图 7 为标签标定的效果.



图 7 RustDetection 数据集

由于本文提出的锈蚀检测的数据集样本量不大, 若直接使用该数据集进行训练会导致网络不能很快的收敛, 最终的检测效果也不好. 因此本文先利用包含共 21 类的 17 125 张图片的 VOC2012 通用公共数据集上进行预训练, 再采用迁移学习的方法, 对本文提出的 RustDetection 数据集进行微调训练.

4.2 实验过程

在训练阶段, 输入的图像首先经过缩放将输入大小变为 $300 \times 300 \times 3$ 的 RGB 图像, 并对其作归一化处理后进行训练, 训练阶段在 NVIDIA GTX 1080Ti GPU 上进行.

网络训练采用迁移学习, 首先将搭建好的网络模型在 VOC2012 数据集上进行 300 轮训练, 并已经在该数据集上拥有较好的检测效果后移除多分类子网络部分结构, 添加二分类子网络, 该二分类子网络由 6 个卷积层构成, 分别用于 6 个尺度下的目标预测, 对该部分的参数采用 Kaiming 初始化方法进行初始化, 相比较随机初始化而言, 该种参数初始化方法可以有效的避免激活函数的输出值趋向于 0, 从而保证网络的训练可以顺利进行. 随后将完整的算法模型在本文提出的 RustDetection 数据集上同样采用阶段性学习率的方式进行训练, 即在训练初始阶段采用大学习率, 在后期调整为小学习率, 这样可以加速模型收敛, 加快训练速度.

在测试和验证阶段, 目标图像首先通过已经训练好的算法模型, 整个模型最后输出 11 620 个候选框信息, 每个候选框包括 2 个分类值 (锈蚀、背景) 和 4 个坐标值 (候选框的中心点坐标和长宽), 随后过滤掉所有被识别为背景的候选框, 对剩下的目标候选框进行非极大值抑制, 最后选出 IOU 最大的候选框作为目标框, 完成锈蚀目标的识别.

4.3 实验结果分析

具体的检测结果如图 8 所示, 其中图 8(a) 为电缆隧道内的防火门控制箱锈蚀目标检测结果, 图 8(b) 为检修电源箱的锈蚀目标检测结果, 图 8(c) 为输电线路的锈蚀目标检测结果, 图 8(d) 为电表箱的锈蚀目标检测结果. 锈蚀区域由彩色框进行标注, 左上角为该区域的分类标签.

为了进一步的验证本文提出算法在模型体量、检测速度和检测精度上的优势, 本文将分别采用 VGG-16 和 ResNet-50 为主干网络的标准 SSD 模型和本文提出基于注意力上采样策略的轻量级 SSD 模型进行对比. 本文的判别标准主要由准确率 (Precision), 召回率 (Recall) 和 AP 值 (Average Precision) 组成, 其中准确率和召回率的计算如式 (1), 式 (2) 所示:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

式中, TP 表示正样本判定正确的个数, FP 数值表示正样本判定错误的个数, FN 数值表示负样本判定错误的个数. 而 AP 值的计算采用 VOC2007 的 11-Point 方法进行计算, 11-Point 方法是结合 $Recall$ 为 [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], 分别找到这 11 个点

下的最大精度值 x , 随后按照式 (3) 计算平均精度:

$$AP_{11\text{-point}} = \frac{1}{11} \left(\sum_{x \in \text{MaxPrecision}} x \right) \quad (3)$$

表 1 为不同的算法模型在本文 RustDetection 数据集下的检测效果对比.

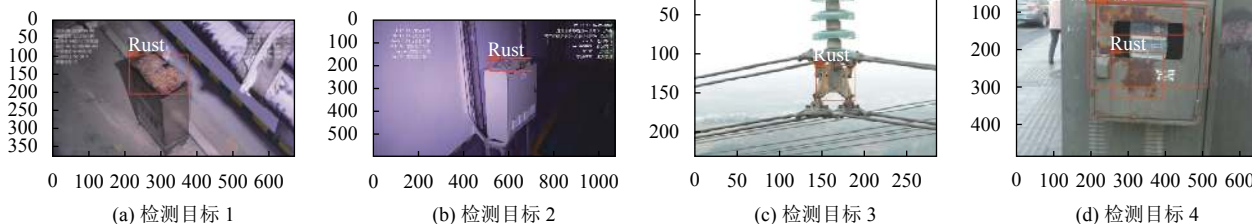


图 8 锈蚀目标检测结果

表 1 不同网络模型的检测效果对比 (%)

算法模型	Recall	Precision	AP
SSD (VGGbase)	78.04	86.49	65.36
SSD (ResNetbase)	75.61	93.94	70.92
SSD (MobileNetbase)	63.41	83.87	59.76
本文算法	78.05	96.96	71.35

由表 1 可见, 若只采用轻量级的 MobileNet 结构对 SSD 模型进行轻量化处理, 其检测效果会因为参数损失而变差, 而本文方法由于添加了上采样和特征融合模块, 可以做到有效的提升检测效果, 甚至超越了原标准 SSD 算法. 不同算法结构的模型参数量、权重大小和在 Intel Core i5-7200U CPU 上的检测时间对比如表 2 所示.

表 2 不同模型大小对比

算法模型	参数量	权重大小 (MB)	检测时间 (s)
SSD (VGGbase)	23 745 908	90.58	1.84
SSD (ResNetbase)	25 433 488	97.02	1.24
SSD (MobileNetbase)	4021 220	15.34	0.50
本文算法	8638 820	32.95	0.98

综上所述, 本文方法相较于只采用轻量级的 MobileNet 的 SSD 模型在上采样网络上扩张了网络结构, 增加了 53.4% 的参数量, 但是相比较拥有庞大参数量的以 VGG-16 为主干网络的标准 SSD 模型在参数量减少 63.6%, 速度提升 46.7% 的情况下提升 10.47% 的准确度和 5.99% 的平均精度, 相比较以 ResNet-50 为主干网络的标准 SSD 也可以做到在参数量减少 66% 的情况下, 提升 2.98% 的准确度和 0.43% 的平均精度.

5 结论与展望

本文提出了一种基于轻量级 SSD 目标检测模型的电力设备锈蚀目标检测方法, 该方法针对目标检测模型参数量巨大, 设备计算能力要求高等特点, 提出了一种轻量级的 SSD 目标检测模型, 并采用了一种基于注意力模型的上采样策略对轻量化后的网络结构进行优化, 弥补了由于减少参数带来的精度损失. 本文提出的模型可以做到在大幅削减参数量的同时保证 96.96% 的检测准确度和 71.35% 的平均精度, 同时检测时间仅为 980 ms, 若设备允许可以使用 GPU 加速, 检测时间仅需 240 ms, 可以满足电力系统安全监控的现实需求. 本文的进一步工作为将网络模型移植和加载进入终端设备, 做到工业现场的实时监测.

参考文献

- 涂洁, 冯智慧, 梁文勇, 等. 小型无人机在电力线路巡检中的应用分析. 电气时代, 2016, (11): 75-77.
- 韩正新, 乔耀华, 孙阳, 等. 基于图像识别的无人机输电线路绝缘子故障检测方法研究. 现代电子技术, 2017, 40(22): 179-181, 186.
- 宋伟, 左丹, 邓邦飞, 等. 高压输电线防震锤锈蚀缺陷检测. 仪器仪表学报, 2016, 37(S1): 113-117.
- 李辉, 钟平, 戴玉静, 等. 基于深度学习的输电线路锈蚀检测方法的研究. 电子测量技术, 2018, 41(22): 54-59.
- Nash W, Drummond T, Birbilis N. Quantity beats quality for semantic segmentation of corrosion in images. arXiv preprint arXiv: 1807.03138, 2018.
- 周自强, 纪扬, 苏焱, 等. 基于迁移学习卷积神经网络的电

- 缆隧道锈蚀识别算法. 中国电力, 2019, 52(4): 104–110.
- 7 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 21–37.
 - 8 Howard A, Sandler M, Chu G, *et al.* Searching for mobilenetv3. arXiv preprint arXiv: 1905.02244, 2019.
 - 9 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.
 - 10 Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448.
 - 11 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada. 2015. 91–99.
 - 12 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.
 - 13 Mnih V, Heess N, Graves A, *et al.* Recurrent models of visual attention. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada. 2014. 2204–2212.
 - 14 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7132–7141.
 - 15 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2117–2125.