

针对复杂电器零件的轻量化分类算法研究

田金玉, 曾志强*, 甄德鑫

五邑大学智能制造学部, 广东 江门

收稿日期: 2024年1月23日; 录用日期: 2024年2月22日; 发布日期: 2024年2月29日

摘要

本文针对复杂电器零件的自动化检测问题, 在实际工业场景下, 采集并构建了具有复杂特性的(不同形状、厚度、颜色及透明度等)的多种电器零件样本数据集, 并提出了轻量化的实时高精度分类模型。在模型构建中, 通过引入轻量化的卷积残差块, 多尺度的金字塔特征提取模块, 以及显著减小模型计算量的Skip-Attention结构, 使模型具有较低检测延时性的同时, 并保证了较高的检测准确度。实验结果证明, 本文所提出算法的实时检测效果优于较多数成熟的实时检测模型, 具备应用于工业零件实时检测的可行性。

关键词

电器零件分类, 机器视觉, 实时检测

Research on Lightweight Classification Algorithm for Complex Electrical Parts

Jinyu Tian, Zhiqiang Zeng*, Dexin Zhen

Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen Guangdong

Received: Jan. 23rd, 2024; accepted: Feb. 22nd, 2024; published: Feb. 29th, 2024

Abstract

This article is based on the automation detecting of complex electrical parts. In actual industrial scenarios, varieties of electrical parts datasets with complex characteristics (different shapes, thickness, color and transparency) are collected and constructed real-time classification model. In the model construction, by introducing lightweight convolutional residual blocks, multi-scale pyramid characteristics extraction modules, and Skip-Attention structures with significantly reduced model computing, it has achieved lower detection delayed models while ensuring higher

*通讯作者。

文章引用: 田金玉, 曾志强, 甄德鑫. 针对复杂电器零件的轻量化分类算法研究[J]. 计算机科学与应用, 2024, 14(2): 317-324. DOI: 10.12677/csa.2024.142032

detection accuracy. The experimental results prove that the real-time detection effect of the method proposed in this article is better than the most mature real-time detection model, and it has the feasibility of real-time detection in industrial parts.

Keywords

Electrical Part Classification, Machine Vision, Real-Time Detection

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在现代工业生产中,需要对各种零件进行快速的分拣。大多数电器零件种类繁多,体积微小,而一部分零件生产厂家现阶段仍是采用传统的人工分拣方法,使零件分拣效率低、成本高。随着工业生产对机器自动化和智能化需求的不断增长,基于机器视觉的检测[1]、测量[2]和定位[3]等技术被越来越多地应用到实际生产中。然而,生产过程中种类繁多、外形复杂且相似度高的电器零件,其视觉识别和处理工作仍然存在极大的困难和挑战。因此,利用机器视觉技术,对实际生产环境中的电器零件实现自动化的高效准确的分类变得尤为重要。

基于机器视觉的图像分类方法,主要包括特征提取和深度学习两种策略。采用特征提取的图像分类方法,侧重于通过人工设计的规则,将图像映射成一个独特的特征向量。随后,利用机器学习分类器,如支持向量机[4]等,对特征向量进行有效的分类。经典的特征提取方法有:局部二值模式[5]、尺度不变特征变换[6]、方向梯度直方图[7]等。尽管在形成最终的特征描述向量过程中,通过特征编码和特征整合等手段,在一定程度上增强了特征描述的鲁棒性,但在应对各种光照变化、细粒度分类等方面仍然存在一些误差。

深度学习作为人工智能领域的一项关键技术,其强大的特征学习能力和模式识别能力为解决电器零件的分类问题提供了全新的途径。目前,针对图像分类的深度学习算法,有两种主流架构:一种是以 ResNet [8], AlexNet [9]等为代表性的卷积神经网络的特征提取算法,具有较强的局部感知能力,且网络结构相对轻量化,对资源有限的端侧设备较友好。然而,对于图像全局特征的识别有一定局限性。另一种是以 Transformer [10]为代表性的主干网络的提取算法,具备较强的全局关联能力,能更好地捕捉到图像中的全局特征。然而一般参数量较大,不易部署与推理。比如 Swin Transformer [11]、Vision Transformer [12]等。而在工业检测中,我们期望模型能够达到较强的实时检测能力,并且保证获得较准确的检测结果。因此,本文通过结合 CNN [13]与 Transformer 两种算法的优势,提出了一种针对复杂电器零件的轻量化、高精度分类算法,以实现电器零件的实时检测。

2. 数据集与预处理

如图 1 所示,本文基于实际生产工业生产环境下,采集并制作了 6 种代表性的电器零件图像数据集。并采用了多种的数据增强方式,获得了 3000 多张高分辨率的零件数据集。数据集由多种复杂零件图像构成,涵盖实际作业场景下的多种拍摄问题(光照不均、零件表面存在阴影、反光、摆放角度多变等)。将采集到的零件数据集,首先按照 6:2:2 比例划分为训练集、验证集、测试集,之后单独对训练集进行多种线

下数据增强(随机旋转、色域变换、均衡化等),从而获得最终的复杂电器零件的工业数据集。

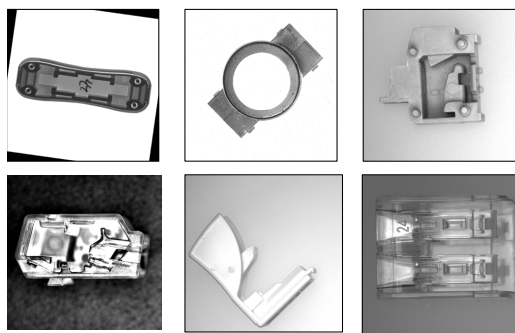


Figure 1. Sample diagram of electrical parts
图 1. 电器零件样本图

3. 轻量化模型

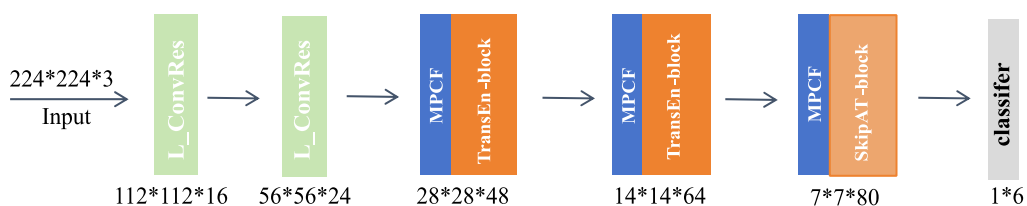


Figure 2. Model structure diagram
图 2. 模型结构图

模型结构如图 2 所示。输入图像被统一压缩成 $224 \times 224 \times 3$ (高*宽*通道数)大小,而后进行特征提取。模型结构共六层,在前五层中,每一层都对特征图进行 2 倍的下采样,以提高网络感受野,捕获更多的全局特征。最后一层,通过将特征图展平为 1 维向量,从而得到最后的分类得分。模型前两层采用提出的轻量化卷积残差块结构,第三、四层采用提出的多尺度的金字塔特征提取结构,并结合 MobileVit [14] 中的多头注意力结构(TransEn-block),进行多尺度的全局特征提取。在模型第五层,结合提出的 SkipAT-block 结构,实现复用第四层的注意力计算结果,实现轻量化的注意力块计算。

在面对实际工业应用场景,本文重点考虑模型的轻量化和实时推理速度。具体地说,本文在 MobileVit 基础上,做了以下三种改进:

1) 引入了 Skip-Attention [15]对原模型的 TransEn-block 结构进行改进,即 SkipAT-block。通过利用前面层的自注意力计算来近似一个或多个后续层的注意力,在维持模型整体性能的同时,减少了计算和存储开销。这一改进不仅有利于模型在有限硬件资源下的高效部署,还加速了推理速度,使其更适用于实时应用场景。

2) 为了进一步降低模型的复杂度,引入深度卷积代替原网络中的常规卷积,并利用通道打乱以及残差连接的结构,增加多个通道特征信息的有效交互,从而构建了 L-ConvRes 模块。

3) 针对本文零件特征,我们设计了多尺度的金字塔特征提取模块(MPCF),该模块通过引入多尺度的卷积核,能够灵活捕捉零件在不同尺度上的特征信息,从而更全面地理解零件的结构和形态。

3.1. SkipAT-block

为了实现模型的轻量化,我们借鉴了先前研究中提出的 Skip-Attention 概念。在视觉 Transformer 的

网络结构中,不同层之间存在着显著的特征图相关性。基于这一观察,通过采用 Skip-Attention (SKIPAT),在前面层的多头注意力(MSA)计算的输出上引入简单的函数控制,实现了重用自注意力块的目的。这一设计旨在保持模型基准性能的同时,提高计算速度,使模型更适应轻量化需求。详细实现如下:

$$\varphi(Z_{l-1}^{MSA}) = FC_2(DwC(FC_1(Z_{l-1}^{MSA}))) \tag{1}$$

$$Z_l = \varphi(Z_{l-1}^{MSA}) + Z_{l-1} \tag{2}$$

$$Z_l = MLP(Z_l) + Z_l \tag{3}$$

SKIPAT 参数函数,如公式(1)所示。将第 L-1 层的 MSA 块(Z_{l-1}^{MSA})通过两个线性层(FC_1 、 FC_2)和中间的深度卷积[16] (DwC),因此将特征映射到第 L 层,作为 L 层的注意力计算结果,从而省略中间层的注意力计算,以实现层注意力复用。引入参数函数的目的在于确保直接特征重用不会影响这些注意力块中的平移不变性和等价性。此外,该参数函数还充当了一个强大的正则化器,有助于提升模型的泛化性能。SkipAT-block 的实现框架,如公式(2)、(3)所示。它将 L-1 级的整体输出(Z_{l-1})与 L-1 的多头注意力经参数函数后的输出($\varphi(Z_{l-1}^{MSA})$),同时作为第 L 级的输入,而后第 L 级只通过多层感知机(MLP),而后与原输入进行残差连接,从而得到 L 级最终的输出。相比于原先的多头注意力结构(TransEn-block),L 级省略了过多的注意力计算,节省了模型整体的耗时。

3.2. 轻量化残差卷积块(L-ConvRes)和多尺度的金字塔特征提取模块(MPCF)

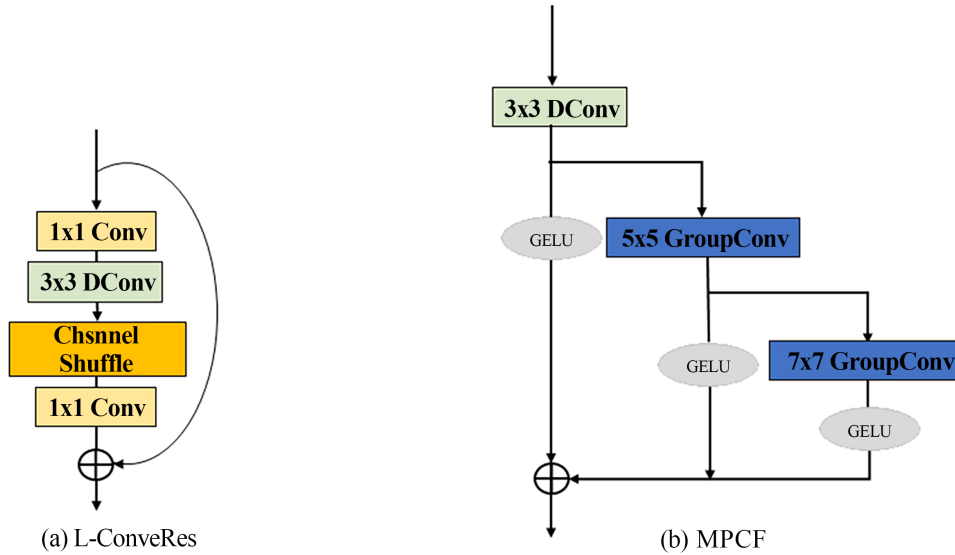


Figure 3. Lightweight residual convolution block (L-ConvRes) and multi-scale pyramid feature extraction module (MPCF)

图 3. 轻量化残差卷积块和多尺度的金字塔特征提取模块

轻量化卷积残差模块如图 3(a)所示。该结构采用两个 1*1 的常规卷积,分别进行通道的降维和升维,一方面,能够有效地降低模型的复杂度。另一方面,允许模型学习不同通道之间的非线性关系。采用 3*3 的深度卷积在增加网络感受野的同时,保证了模型的轻量化。并且,为了保持不同组之间特征的信息交互,以及更好的捕捉多尺度特征,进行了通道打乱,在不引入额外参数的同时,有效地提高了模型的鲁棒性。

多尺度的金字塔特征提取模块如图 3(b)所示。模块采用了大核卷积,金字塔式的特征提取和融合结构。

首先将输入特征图先通过一个 3×3 的标准卷积进行通道压缩，以减少计算量。此时，输出特征信息将分成两条分支，一支进行非线性激活作为该层的单独输出特征，另一支是将该层特征继续通过一个 5×5 的组卷积(分 8 组)，进一步地进行多尺度特征提取。同理，此时输出继续作为单独的一支进行输出，而另一支继续通过一个 7×7 的组卷积(分 16 组)来提取有效特征。最后，将提取到的特征图相加，实现进一步地多尺度特征的融合。经上一个卷积输出的特征每次通过下一个卷积，都会进一步扩大感受野，提取到更多的特征信息。且每个卷积的输出特征信息又被单独保存输出，不同尺度的特征得到高效地利用。

4. 实验及结构分析

4.1. 实验环境与相关参数

训练过程中，训练集被压缩成 224×224 大小，之后进行随机旋转、裁剪等方法进行线上增强。同时，验证集也压缩到 224×224 进行效果评估。考虑到零件数据集与公开数据集的差异性，我们未使用任何预训练权重，从头开始训练。我们的网络训练和测试在基于 Pytorch1.10.0+cuda11.5+Nvidia GeForce RTX3090 平台上进行。实验相关参数的具体设置如表 1 所示。

Table 1. Experimental related parameter settings

表 1. 实验相关参数设置

参数名称	参数值
优化器	AdamW
初始学习率	0.01
参数衰减(Weight-Decay)	$1e-4$
训练批大小(Batchsize)	8
训练总轮数(Epoch)	100

4.2. 评价指标

面对真实的工业应用场景，本文主要从检测速度和检测精度两个方面来评估针对复杂电器零件的检测算法的性能。因此，本文从检测准确率、参数量、实时推理速度三个指标，进行模型整体性能的评估。

1) 准确率(Accuracy)

$$\text{Accuracy} = \frac{TP + TN}{N} \quad (4)$$

表示正确分类的样本个数占总样本个数的比率，其中 N 表示样本总数， TP 表示该零件准确分类时的样本数， TN 表示将该类零件判断为非该类的样本数。准确率越高，表明模型的检测精度越高。

2) 参数量(Params)

是指网络模型中需要训练的参数总数。模型参数量较多会增加在推理过程中的计算资源需求，包括 CPU、GPU 或其他专用硬件。在工业场景中，部署硬件通常是受限的，因此较大的模型可能导致计算资源短缺，降低推理速度。越小的参数，对部署设备更加友好，推理速度一般更快。

3) 实时推理速度(FPS)

FPS (Frame Per Second)即每秒可以处理图片的数量。不同的检测平台其性能各不相同，因此评估 FPS 参数时须在同一设备上进行测试。本文在单位时间内处理图片数据，通过 1 秒内处理图片的数量，数量越多，表示实时检测速度越快。

4.3. 实验细节

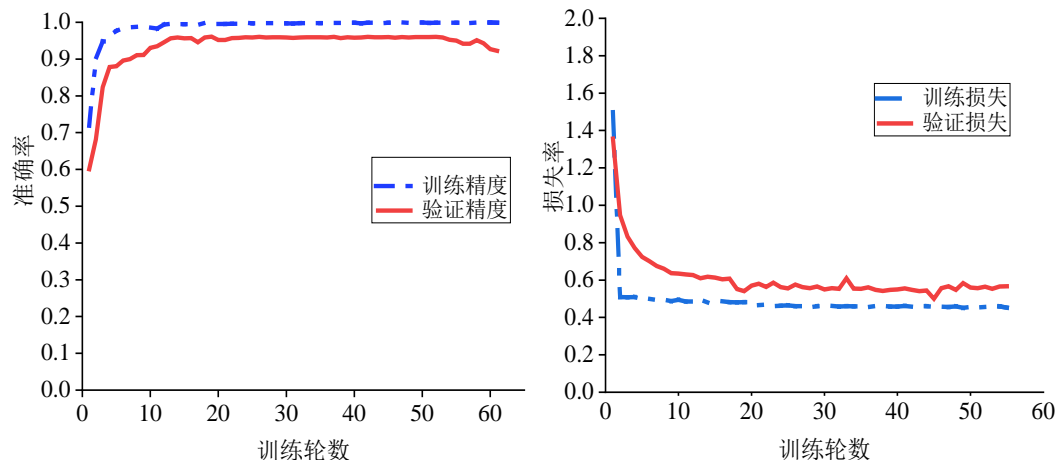


Figure 4. Accuracy curves (left panel) and loss curves (right panel) on training and validation sets
图 4. 训练集和验证集上的精度曲线(左图)和损失曲线(右图)

本文模型训练过程中，训练集以及验证集上的精度和损失如图 4 所示。从左图可以看到，在模型训练 50 轮左右时，模型逐渐收敛，训练集与验证集上的精度曲线变化平缓，训练集上精度接近 99%，验证集上的准确度可达 97.8%。同时，从右图可以看到，两 s 数据集上的损失在 45 轮左右也趋于平稳，训练集上的损失可稳定在 0.5，验证集上的损失可稳定在 0.6。由此看出，本文模型在达到较高检测准确性的同时，具备良好的鲁棒性。

5. 实验结果与分析

Table 2. Test results of each model on the electrical parts dataset
表 2. 各模型在电器零件数据集上的测试结果

模型	准确率	参数量	实时推理速度(img/s)
Mobilenext [17]	0.9841	8.63	166
Mobilenetv3 [18]	0.9621	5.81	250
Mobilenetv2 [19]	0.9570	8.51	270
Espnetv2 [20]	0.9748	2.47	113
MobileVit	0.9785	3.64	62
Ours	0.9907	2.98	125

多个轻量化模型在本文电器零件测试集上的实验结果，如表 2 所示。可以看出，本文模型针对零件数据集有较高的检测正确率，且参数量较低，实时推理速度较快，可达到每秒 125 张图像的检测速度。其中，MobileVit 也是将 CNN 与 Transformer 相联合的混合模型，本文模型在原有的 MobileVit 上，准确度提升了 1.2%，实时推理速度是其 2 倍。并且，与纯卷积结构的 Mobilenetv3、Mobilenetv2、Espnetv2 相比，本文模型实时检测精度各有 3%、3.3%、1.6% 的提升。其中，Mobilenetv2 在实时推理速度上有显著优势，但模型参数量较大。同时，本文模型与检测准确度相近的 Mobilenext 相比，本模型的参数量约是其 1/3，而提升了将近 1% 的检测准确度。总体上，本文模型在兼顾检测准确度和模型轻量化的同时，保证了较优的实时推理速度。

通过比较各模型的实时推理速度,可明显看出,纯卷积模型普遍具有较高的实时检测速度。将参数量相近的纯卷积结构的 Espnetv2 与注意力块较多的 MobileVit 相比,前者比后者的实时推理速度快了将近一倍。由此也证明,本文采用卷积块结合 Skip-Attention 方法构建较轻量化的模型,进行零件特征提取的合理性。一方面,利用少量注意力块有效解决 CNN 中无法捕获长距离依赖信息的局限性,另一方面,利用卷积结构在端侧检测中的推理优势,实现较高的实时检测效率。

6. 结论

本文创建了多种复杂电器零件图像数据集,采取多种数据增强方法处理,并构建了针对零件的轻量化分类模型,实现了对多种零件的有效分类。实验结果表明,本文模型相比其他较成熟的轻量化模型,具有较高的检测速度和较低的延迟,同时保持了较高的准确率,能够满足工业检测的实际需求。

参考文献

- [1] 吕广贤. 基于机器视觉的散热器钎焊缺陷检测系统研发[J]. 图像与信号处理, 2021, 10(3): 146.
- [2] 李超, 许杰. 电子元器件外形尺寸机器视觉测量系统设计[J]. 光电子, 2020, 10(3): 84.
- [3] 刘瑞欣, 严春雨, 李飞, 等. 基于改进 YOLOX 的茶叶嫩芽目标检测研究[J]. 软件工程与应用, 2022, 11(6): 1404-1414.
- [4] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/BF00994018>
- [5] Ojala, T., Pietikainen, M. and Maenpaa, T. (2002) Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 971-987. <https://doi.org/10.1109/TPAMI.2002.1017623>
- [6] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [7] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, 20-25 June 2005, 886-893.
- [8] Shafiq, M. and Gu, Z. (2022) Deep Residual Learning for Image Recognition: A Survey. *Applied Sciences*, **12**, Article 8972. <https://doi.org/10.3390/app12188972>
- [9] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**, 1-8.
- [10] Vaswani, A., et al. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 1-10.
- [11] Liu, Z., et al. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-1000. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [12] Dosovitskiy, A., et al. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [13] Lecun, Y., et al. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [14] Mehta, S. and Rastegari, M. (2021) Mobilevit: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer. arXiv: 2110.02178.
- [15] Venkataramanan, S., et al. (2023) Skip-Attention: Improving Vision Transformers by Paying Less Attention. arXiv: 2301.02240.
- [16] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 1800-1807. <https://doi.org/10.1109/CVPR.2017.195>
- [17] Zhou, D.Q., et al. (2020) Rethinking Bottleneck Structure for Efficient Mobile Network Design. *Computer Vision—ECCV 2020: 16th European Conference*, Glasgow, 23-28 August 2020, 680-697. https://doi.org/10.1007/978-3-030-58580-8_40
- [18] Howard, A., et al. (2019) Searching for Mobilenetv3. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October 2019 - 2 November 2019, 1314-1324. <https://doi.org/10.1109/ICCV.2019.00140>

- [19] Howard, A., *et al.* (2018) Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation.
- [20] Mehta, S., *et al.* (2019) ESPNetv2: A Light-Weight, Power Efficient, and General Purpose Convolutional Neural Network. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 9182-9192. <https://doi.org/10.1109/CVPR.2019.00941>