

基于优化卷积神经网络的表面缺陷检测^①

姚明海^② 袁 惠

(浙江工业大学信息工程学院 杭州 310023)

摘 要 卷积神经网络(CNN)具有权值数量少,训练速度快等优点,在图像识别、机器视觉等领域得到广泛应用。本文提出了一种卷积神经网络的自适应加权池化算法,算法通过生成合并通道,并在学习掩模的引导下汇集特征,优化了子采样模型的特征提取,有效改善了网络的识别准确性和快速性。利用该算法对磁片表面缺陷进行检测实验,实验结果表明,本文提出的池化模型使卷积神经网络对特征的提取更加精确,同时提高了收敛速度和鲁棒性,并且可以应用于各种深度神经网络体系结构中。

关键词 卷积神经网络(CNN), 自适应加权池化模型, 缺陷检测, 子采样

0 引 言

表面缺陷的自动检测是机器视觉在现代工业自动化生产中的重要应用。传统的表面缺陷检测主要是通过人工来识别缺陷,为了解决这个问题,Win 等人^[1]提出对比度调整的 Otsu 方法和基于中值的 Otsu 方法,用于检测表面缺陷。Sayed^[2]提出使用熵滤波和最小误差阈值的表面缺陷检测算法来识别缺陷。Yang 等人^[3]提出了一种基于平稳小波变换的磁瓦图像缺陷提取方法。随着信号处理理论和计算机技术的发展,基于图像传感器和图像处理技术的机器视觉自动检测系统得到了广泛的应用。与传统方法相比,机器视觉检测技术具有非接触、高速度、高精度、低成本等优点。由于磁片表面对比度低、磨痕纹理干扰和缺陷块小且亮度变化大等难点,传统的模式识别方法想要分割出各种情况下的缺陷是十分困难的。大多数方法都是针对特定缺陷或者特定光源场景条件下进行检测,缺乏一定的通用性。近年来,深度学习的代表卷积神经网络(convolutional neural network, CNN)具有权值数量少、训练速度快

等优点,在图像分类与识别^[4]、自然语言识别、语音的识别与合成^[5,6]、信号处理和人脸识别^[7]等方面有着广泛的应用。

随着对网络性能的要求越来越高,国内外学者做了很多研究,Huang 等人^[8]通过构建局部卷积受限玻尔兹曼机模型来提升网络学习的性能。周飞燕等人^[9]通过增加 CNN 的深度,从而使特征面数增加,网络能够表示的特征空间增大,网络的学习能力也相应增强。Howard 等人^[10]通过深度可分离卷积的方式来代替传统的卷积方式,这种卷积结构可以极大地降低卷积神经网络的复杂度,减少模型的参数数量和计算量。Liu 等人^[11]建议使用卷积层激活的子阵列作为局部特征,并使用连续卷积层的特征图作为汇集信道,然后,将提取的特征与这些汇集通道合并,以生成更稳健的图像表示。

传统的池化操作都有其自身的局限性和缺点。关于最大池化,它仅考虑最大元素并忽略池化区域中的其他元素。如果汇集区域中的大多数元素具有高幅度,则在最大池化之后区别特征消失。关于平均池化,它计算合并区域内所有元素的平均值。该算子将考虑所有低幅度,并且将减少汇集后新特征

① 国家自然科学基金(61871350)资助项目。

② 男,1963年生,博士,博士生导师,教授;研究方向:模式识别和图像识别;联系人,E-mail: ymh@zjut.edu.cn
(收稿日期:2018-09-20)

图的对比度。如果有许多零元素,特征图的特征将大大减少。为了改善卷积神经网络对图像识别能力的有效性和鲁棒性,本文在深度卷积神经网络中融入预定义数量的自适应加权掩模和汇集通道,并结合经典池化模型,提出一种自适应加权池化(adaptive weighted pooling, AWP)模型,以提高模型的准确率及网络的学习能力。本文所提出的方法可作为标准池化操作的直接替代,因此它们可以与其他性能提升技术或替代架构^[12,13]一起使用。基于该模型的卷积神经网络应用在磁片表面缺陷检测上,不仅能提高检测的准确率,而且能满足工业生产流水线上实时检测的需求。

1 池化模型改进的可行性研究

特征提取的有效程度往往决定卷积神经网络的学习性能,所以对特征提取过程(卷积和池化)进行优化与改进来提高算法的收敛速度和准确率等特性具有可行性。池化算法的选择决定了子采样提取到特征的有效性,从而对模型的准确率产生影响,因此池化操作在提高网络性能方面发挥着核心作用。

近年来,学者们的研究集中在复杂的网络结构、卷积层以及非线性激活函数上。实际上,很多理论和实验研究表明,在池化函数以及池化参数方面的改进,能在很大程度上提高网络的学习性能,会大幅提升图像分类效果。池化操作的参数有池化域的大小、池化步长等,这些参数也会在不同程度上影响池化操作的结果。国内外很多学者在池化方面做了研究,Giusti 等人^[14]通过使用最大池化模型的卷积神经网络实现了图像的快速扫描。Gong 等人^[15]使用多尺度无序池化解决了卷积神经网络全局激活缺乏几何不变性的问题。He 等人^[16]提出了空间金字塔池化(spatial pyramid pooling, SPP-NET)方法,SPP-NET 将任意大小的图像池化生成固定长度的图像表示,将区域卷积神经网络(region with CNN, RCNN)检测的速度提升 24 ~ 102 倍。Hu 等人^[17]提出的空间权重池化模型,大大增强了深度卷积神经网络对于细粒度分类的鲁棒性和有效性。Lee 等人^[18]引入树池的思想,这是一种学习不同池化过滤

器的方法,并且还响应性地组合这些过滤器。

2 自适应加权池化模型

卷积操作后,得到了一张张有着不同信息的特征图谱,数据量比原图少了很多,但深度学习的训练图片数量比较庞大,模型的计算过于复杂。而池化操作可以对输入的特征图进行压缩。它一方面使特征图变小,简化网络计算复杂度;另一方面提取主要特征。池化的两个主要作用是:(1)不变性,包括平移,旋转,缩放;这种不变性体现在光照或位置等外界扰动的鲁棒性和特征表示的紧凑性上。通常,汇集层总结了相同内核映射中相邻神经元的输出。在池化层中,通过在前一层的特征图上汇集局部邻域来减少特征图的分辨率,使得在输入失真的情况下也能保证不变性。(2)汇集主要的特征同时减少参数达到降维的目的,效果类似主成分分析^[19](principal component analysis, PCA),可以防止过拟合,提高模型泛化能力。

经典的池化操作有 2 种,即平均池化和最大池化。在池化的时候,如果取区域均值,往往能保留整体数据的特征,能突出背景的信息;而如果取区域最大值,则能更好保留(主要的)纹理上的特征^[20]。在从深度卷积神经网络的卷积层提取卷积特征图(convolutional feature map, CFM)之后,可以直接执行传统的最大池或平均池以获得图像表示。本文提出了一种新的汇集策略——自适应加权池,可以显著提高磁片分类的性能。该策略结合最大池和平均池的优点,同时将自适应加权掩模融入汇集操作。该模型将联合特征表示转换为更有用的表示,保留重要信息,同时丢弃不相关的细节。

与传统的池化方法相比,自适应加权池包含预定义数量的加权掩模,并在学习掩模的引导下汇集从前一层提取的特征。此外,它可以在深度卷积神经网络的端到端训练过程中学习 AWP 层的参数。自适应加权池将加权掩模视为多组池化通道,这些掩模可以指示卷积神经网络用于识别磁片的缺陷区域。从这些判别区域提取的特征对汇集特征有显著的效果,然后通过学习到的加权掩模来确定最大池

化和平均池化的“响应”混合比例;这种方法可以使池化混合比例自动根据当前池化域中的特征而变化,即自适应的池化。

自适应加权池函数如下所示:

$$P_{AWP}(X) = S(f_{AW})P_{\max}(X) + (1 - S(f_{AW}))P_{ave}(X) \quad (1)$$

其中, P_{AWP} 表示自适应加权池层的输出, $S(\cdot)$ 表示 sigmoid 函数, 它的输出用作最大值和平均值之间的混合比例; f_{AW} 为判别区域提取的特征; 最大池化 $P_{\max}(X) = \max(X_i)$ 选取区域中最大的值作为特征值; X_i 表示局部区域的像素值; 平均池化 $P_{ave} = \frac{1}{N} \sum_{i=1}^N X_i$ 指计算区域中的平均值作为特征值; N 为局部区域像素的个数。

假设提取的特征图可以表示为大小为 $H \times W \times D$ 的张量, 其中 H, W 表示每个特征图的高度和宽度, D 表示特征图的数量。可以将自适应加权池层的加权掩模公式化为 $H \times W$ 大小的 K 个合并通道。这些掩模的高度和宽度与要输入的特征图的大小相同。第 k 个学习掩模的汇集特征, 表示为 f_{AW}^k , 可以通过下式计算:

$$f_{AW}^k = \sum_i^{H \times W} b_i^k x_i^t \quad (2)$$

其中, x_i^t 表示输入第 t 张特征图的第 i 个局部特征, b_i^k 是第 k 个学习掩模中的对应系数。

为了训练具有自适应加权池层的深度卷积神经网络 (deep convolutional neural network, DCNN), DCNN 需要从损失函数的导数反向传播到输入图像。这意味着可以通过反向传播损失的梯度来训练加权掩模的参数。假设 M 是 AWP 层中的学习掩模的集合, 并且具有 $H \times W \times K$ 的大小, K 表示掩模的数量。 X 是 AWP 层的输入特征图, 其尺寸为 $H \times W \times D$ 。由以下公式, 可以获得掩模的参数:

$$\frac{\partial l}{\partial M} = \frac{\partial l}{\partial p} \cdot \frac{\partial p}{\partial M} = \Delta \cdot S(f_{AW}) \cdot [1 - S(f_{AW})] \cdot X \cdot (P_{\max} - P_{ave}) \quad (3)$$

$$\frac{\partial l}{\partial X} = \frac{\partial l}{\partial p} \cdot \frac{\partial p}{\partial X} = \Delta [S(f_{AW}) \cdot (1 - S(f_{AW})) \cdot M \cdot (P_{\max} - P_{ave}) + S(f_{AW}) \cdot 1 [P_{\max}] + (1 - S(f_{AW})) \cdot \frac{1}{N}] \quad (4)$$

其中, l 为损失函数, $\Delta = \frac{\partial l}{\partial p}$ 表示高层反向传播到底层的误差, $1[\cdot]$ 为 0/1 指数函数。

3 实验与结果分析

本文采用深度卷积神经网络 AlexNet^[21] 和 ResNets^[22] 进行实验, 应用在铁氧体磁片表面进行缺陷检测。池化层分别使用最大池化模型、平均池化模型以及本文提出的自适应加权池化模型进行实验, 并比较了它们的性能。

3.1 图像预处理

铁氧体磁片原图来自浙江省某磁片厂, 采用 basler 工业相机拍摄了 1 050 张磁片图片, 分辨率为 659×494 , 表面缺陷主要表现为边缘磕碰导致的凹坑, 部分样本如图 1 所示。图像预处理主要分为背景分割、感兴趣区域的图像增强和图像分块 3 步骤。首先进行背景分割, 将圆磁片和背景进行分离。本文采用的是结合边缘轮廓检测的模板匹配方法, 由于磁片在流水线上运动时, 光源与磁片存在各个角度, 磁片周围存在一个较浅的黑色光晕。本文根据圆磁片表面轮廓与背景像素差距较大的特点, 通过轮廓提取方法, 找到轮廓并加黑像素, 加深这种像素偏差, 生成新的模板, 大大提高了模板匹配的准确度, 感兴趣区域提取结果如图 2(a) 所示。

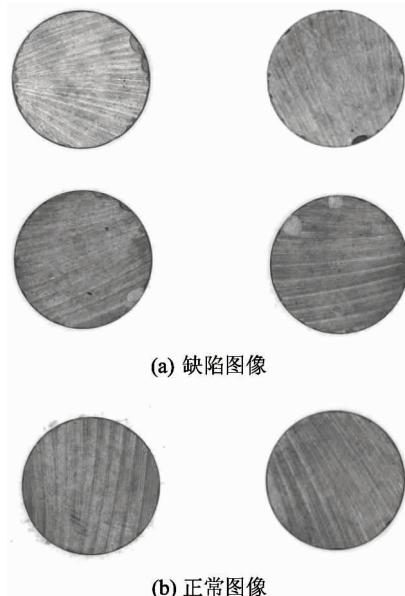


图 1 原始瓷片图

然后进行图像分块,这是一种数据扩充的方法,同时由于缺陷面积小,这样也能突出缺陷特征。经过实验分析,为了使缺陷特征都能包含在一个窗口中,本文以圆心处 18° 扇形区域的外接矩形作为一个分块单元,每次以圆心到外接矩形的中心为轴顺时针旋转 18° ,确定出新矩形的位置,如图 2(b) 所示。这样每个分割窗口的重叠率为 15.4%,还保留了轮廓信息,这样每个磁片图像能得到 20 个分块图像数据,所有的最终图像都经过均值化和归一化处理。图像分块处理不仅能大大扩大数据集,而且能完整地突出缺陷特征,减小后续 CNN 的网络参数。

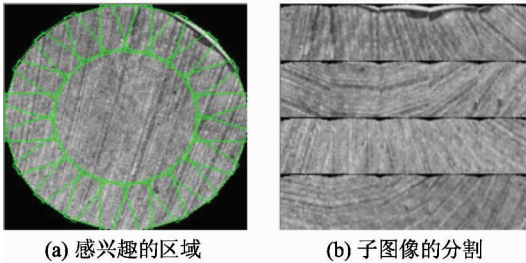


图 2 图像预处理结果

3.2 实验过程

本文总共采集到 1 050 张磁片样本,切割完经筛选后共得到 18 000 张子图像,将 14 000 张作为训练集,2 000 张作为验证数据,2 000 张作为测试数据。在磁片数据集上,本文主要针对磁片凹坑缺陷进行检测。

本文实验设备是一台显卡为 TeslaK40C 的服务器,使用 TensorFlow 框架。在本文的实验中,使用预训练的 DCNN、AlexNet、ResNets 作为基础网络模型,评估自适应加权池化方法引起的性能改进。训练时使用 Xavier 初始化网络参数,将 AdamOptimizer 作为优化器的优化算法,softmax 交叉熵损失函数作为损失函数。网络训练的超参数如下:批处理大小(batch size)为 128,通过从高斯分布绘制的具有固定标准偏差 0.005 的随机权重来初始化加权掩模层和全连接层。对于卷积层,初始学习率设置为 0.001,对于加权掩模层和全连接层,初始学习率设置为 0.01。在每 40 个训练时期之后,学习率除以 10,这些模型需要经过多达 90 个时期的训练。本文使用 0.0005 的权重和 0.9 的动量衰减。

对于 AlexNet,本文通过在最后一个最大池和第一个完全连接层之间插入自适应加权池层。对于 ResNets,本文不是在平均池层和全连接层之间插入自适应加权池层,而是分别用自适应加权池和最大池替换平均池,并在自适应加权池层和最后一个全连接层之间插入一个额外的全连接层和批量标准层,以保持快速训练。掩码数量会影响自适应加权池方法的质量,实验从 {1,4,9,16,25} 中设置掩码 K 的数量。

3.3 结果分析

本文提出的自适应加权池化模型代替原来的最大池与平均池,在 AlexNet and ResNets 结构上进行实验,表 1 为几种模型的卷积神经网络对磁片检测的精度。由表 1 可以看出,在 AlexNet 中应用自适应加权池模型,检测精度由原来的 68% 提升到 86%;ResNets 结合的自适应加权池模型,在精确度方面,也明显优于传统的池化方法。表 2 为随着 K 值的变化,自适应加权池的性能变化,结果表明,对

表 1 几种模型的卷积神经网络对磁片检测精度

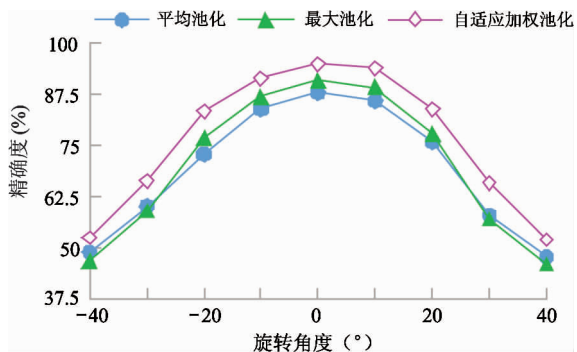
池化方法	分类精确度(%)
AlexNet(最大池)	68.4
AlexNet + 自适应加权池	86.9
ResNets(平均池)	88.7
ResNets + 最大池	91.6
ResNets + 自适应加权池	95.3

表 2 不同参数设置的 DCNN 的分类精度

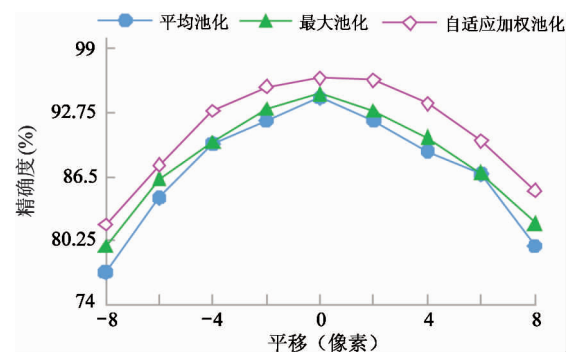
加权掩模 K 值(个)	AlexNet + 自适应 加权池(%)	ResNets + 自适应 加权池(%)
1	80.9	92.1
4	82.8	94.4
9	86.7	95.2
16	84.4	94.8
25	82.6	93.6

于 AlexNet + AWP,由表可知随着掩模数量增加到 9,精度通常会提高。但是,将掩模数量设置得太大(例如, $K=25$)可能会影响其性能;对于 ResNets + AWP,性能对学习掩模的数量不是很敏感。图 3 为几种池化方法对旋转、平移和缩放 3 种变换的鲁棒

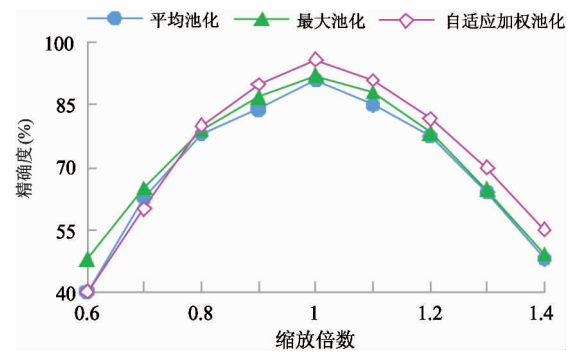
性的实验结果,从图3可以得到结论,本文提出自适应加权池化模型的鲁棒性效果远好于单纯的最大池化和平均池化模型。



(a) 旋转变换



(b) 平移变换



(c) 缩放变换

图3 3种不同类型变换对应不同池化模型的精度

4 结论

本文提出了一种称为自适应加权池的新方法,它指示 DCNN 对对象分类过程中,从输入图像的判别区域提取更有用的信息,增强图像的表达。并且可以在卷积神经网络的端到端训练过程中学习这些

池化通道。

本文采用卷积神经网络对磁片表面缺陷进行检测分析,研究了经典的池化方法以及自适应加权池化模型,对比不同模型对特征提取的优化程度。实验表明,本文提出的自适应加权池化模型在检测准确率、鲁棒性等性能方面远优于经典池化模型,可以显著地提高分类效果。此外,这种模型也可以进一步应用于其他与卷积神经网络相关的深度学习算法。但是目前本文只是针对“凹坑”缺陷,由于还缺乏其他缺陷的数据集,笔者正在努力采集更多数据,下一步可以尝试增加缺陷种类,使此网络模型针对不同类别的缺陷对象更加具有应用价值。

参考文献

- [1] Win M, Bushroa A R, Hassan M A, et al. A contrast adjustment thresholding method for surface defect detection based on mesoscopy[J]. *IEEE Transactions on Industrial Informatics*, 2017, 11(3): 642-649
- [2] Sayed M S. Robust fabric defect detection algorithm using entropy filtering and minimum error thresholding[C]. In: *IEEE 59th International Midwest Symposium on Circuits and Systems*, Abu Dhabi, United Arab, 2017, 1-4
- [3] Yang C, Liu P, Yin G, et al. Defect detection in magnetic tile images based on stationary wavelet transform[J]. *NDT & E International*, 2016, 83: 78-87
- [4] Lee H, Kwon H. Going deeper with contextual CNN for hyperspectral image classification[J]. *IEEE Transactions on Image Processing*, 2017, 26(10): 4843-4855
- [5] Graves A, Mohamed A R, Hinton G. Speech recognition with deep recurrent neural networks[C]. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, 2013, 6646-6649
- [6] Saitoh T, Zhou Z H, Zhao G Y, et al. Concatenated frame image based CNN for visual speech recognition [C]. In: *Workshop on Multi-view Lip-reading Challenges*, Asian Conference on Computer Vision, Taipei, China, 2016. 277-289
- [7] He R, Wu X, Sun Z N, et al. Wasserstein cnn: Learning invariant features for nir-vis face recognition [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2019, 41(7):1761-1773
- [8] Huang G B, Lee H, Learned-Miller E. Learning hierarchical representations for face verification with convolutional deep belief networks[C]. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

- tion, Providence, USA, 2012. 2518-2525
- [9] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251
- [10] Howard A G, Zhu M, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobilevision applications[J]. *arXiv:1704.04861*, 2017
- [11] Liu L Q, Shen C H, Henge A V D. The treasure beneath convolutional layers: cross-convolutional-layer pooling for image classification[C]. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015. 4749-4757
- [12] Szegedy C, Liu W, Sermanet P, et al. Going deeper with convolutions[C]. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015. 1-9
- [13] Wang J D, Zhen W, Zhang T, et al. Deeply-fused nets [J]. *arXiv: 1605.07716*, 2016
- [14] Giusti A, Ciresan D C, Masci J, et al. Fast image scanning with deep max—pooling convolutional neural networks[C]. In: Proceedings of IEEE International Conference on Image Processing, Melbourne, Australia, 2013. 4034-4038
- [15] Gong Y C, Wang L W, Guo R Q, et al. Multi-scale orderless pooling of deep convolutional activation features [C]. In: Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 2014. 392-407
- [16] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37 (9) :1904-1916
- [17] Hu Q C, Wang H B, Li T, et al. Deep cnns with spatially weighted pooling for fine-grained car recognition[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(11) :3147-3156
- [18] Lee C Y, Gallagher P, Tu Z W. Generalizing pooling functions in CNNs: mixed, gated, and tree[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4) :863-875
- [19] Chan T H, Jia K, Gao S H, et al. PCANet: a simple deep learning baseline for image classification? [J]. *IEEE Transactions on Image Processing*, 2015, 24(12) : 5017-5032
- [20] Phan H, Koch P, Hertel L, et al. CNN-LTE: a class of 1-X pooling convolutional neural networks on label tree embeddings for audio scene classification[C]. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA, 2017.
- [21] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012. 1097-1105
- [22] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. In: IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016. 770-778

Surface defect detection based on optimized convolutional neural network

Yao Minghai, Yuan Hui

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

Abstract

The convolutional neural networks (CNN) have been proved to be effective in image recognition, machine vision and other fields with the advantages of small number of weights and fast training speed. An adaptive weighted pooling algorithm is proposed for convolutional neural networks. The proposed algorithm optimizes the feature extraction of sub-sampling models by generating merge channels and collecting features under the guidance of learning masks, and effectively improves the recognition accuracy and speed of the network. The experiments carried out on the surface defect detection of the magnetic disks show that the proposed pooling model can improve the accuracy of the features extraction and so can effectively detect the defects of surface in faster convergence and robustness with convolutional neural network. Also the pooling model proposed in this paper can be applied to various deep neural network architectures.

Key words: convolutional neural network (CNN), adaptive weighted pooling model, defect detection, sub-sampling