

# 一种结合 Dropblock 和 Dropout 的正则化策略

胡辉,司凤洋,曾琛,舒文璐

(华东交通大学 信息工程学院,南昌 330013)

**摘要:**为了能够全面且高效加快卷积分类网络的收敛速度和提升稳定性,提出了一种新的正则化策略,将 Dropblock 算法和 Dropout 算法相结合,从而实现对整个卷积分类网络的浅层、中层和深层网络进行正则化.其中,Dropblock 通过隐藏部分特征图实现卷积层正则化,Dropout 通过隐藏部分权重参数实现全连接层正则化,从而实现对整个卷积分类网络进行全面正则化.通过 Kaggle 猫狗分类大赛提供的数据集进行训练和测试实验表明,提出的新的正则化策略可有效加快分类网络的收敛速度和提升稳定性,此外,能有效提高深度卷积分类网络的分类准确率.

**关键词:**正则化;Dropout;Dropblock;收敛速度;稳定性

**中图分类号:**TN951

**文献标志码:**A

卷积神经网络(Convolutional Neural Network, CNN)<sup>[1]</sup>已经广泛应用,并统领了图像分类<sup>[2]</sup>领域,但是,如今越来越深的卷积神经网络使得网络模型参数数量爆炸性增长,导致网络收敛<sup>[3]</sup>困难,且整个训练过程极不稳定,最终降低卷积神经网络的分类准确率.通常采用 L1 和 L2 范数正则化和软权重共享加快卷积神经网络的收敛速度和提升稳定性<sup>[4]</sup>,但是该方案无法降低网络模型的复杂度,且仅作用于局部卷积神经网络,只能提高局部网络的稳定性.为此, Hinton 等人<sup>[5]</sup>提出 Dropout 正则化,通过在训练过程中以某概率隐藏部分神经元<sup>[6]</sup>,使得网络在训练过程参数数量大幅降低,且网络模型不会太依赖某些局部特征<sup>[7]</sup>,有效加快卷积神经网络的收敛速度和提高区域神经网络的稳定性.然而一个完整的分类网络可以划分为浅层、中层和深层网络<sup>[8]</sup>,其中,浅层和中层网络由卷积层组成,深层网络由全连接层(Fully Connected Layer, FC)组成. Dropout 是仅仅能有效作用于全连接层区域,从而无法对整个分类网络进行全局正则化. 2018 年底, Golnaz Ghiasi 等人<sup>[9]</sup>提出 Dropblock 正则化,与 Dropout 不同, Dropblock 正则化是以某概率隐藏某块特征图,然后对未被隐藏的特征图进行归一化,从而实现了对卷积层的正则化.于是基于 Dropblock 正则化和 Dropout 正则化,提出首次一种新的正则化策略,将 Dropblock 应用于分类网络的浅层和中层网络实现这 2 个区域的正则化,将 Dropout 应用于分类网络的深层网络从而实现该区域的正则化,最终全面正则化分类网络,以此进行图像分类应用,通过 Kaggle 猫狗分类比赛提供的数据集,实验验证了该正则化策略能有效加快分类网络的收敛速度、提升收敛性且能有效提高深度分类网络的分类准确率.

## 1 基于 Dropblock 和 Dropout 的正则化组合策略的分类算法方案

将基于结合 Dropblock 和 Dropout 的正则化策略应用于分类网络中,完成猫狗图像分类任务.算法共分为 3 个步骤,预处理输入图像、特征提取和特征分类,算法框架如图 1 所示.

### 1.1 数据预处理

将待训练的图片随机旋转和随机裁剪,进行数据增强处理,在一定程度上扩充了数据的数量,增加了数据

收稿日期:2019-01-18;修回日期:2019-09-12.

基金项目:国家自然科学基金(61761019);江西省自然科学基金(20142BAB207001).

作者简介:胡辉(1970-),男,江西南昌人,华东交通大学教授,硕士生导师,主要研究方向为计算机视觉, DSP&FPGA, GNSS.

通信作者:司凤洋(1992-),男,安徽安庆人,硕士研究生,主要研究方向为深度学习、计算机视觉, E-mail:2448351454@qq.com.

集的多样性,让分类网络学习到更多有效的特征信息,发掘网络模型的分类潜力.

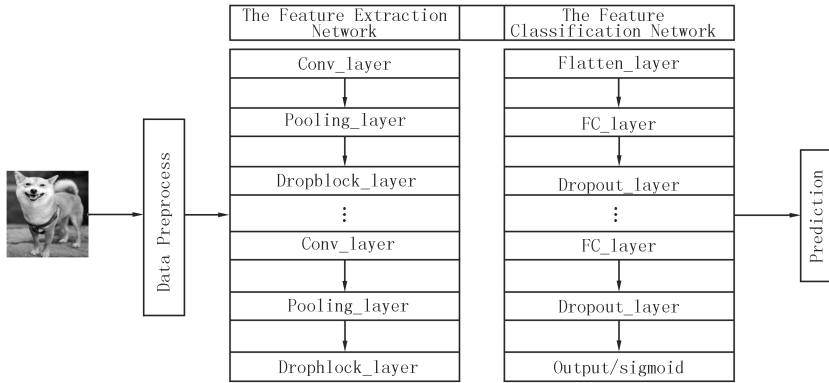


图1 算法流程概图

Fig.1 Structure diagram of the algorithm

## 1.2 共享特征提取

通过基础网络提取输入图片的共享特征见图2.根据像素亮度可判断激活程度.它包含了用于图像分类的高语义信息.语义信息的优劣直接影响网络的分类性能.本文采用权值共享的卷积层,减少了网络模型参数和运算时间,此外,应用Dropblock对浅层网络和中层网络中的卷积层进行正则化,激励分类网络学习到更加有效的特征、加快分类网络收敛速度和提升稳定性.

### 1.2.1 权值共享的卷积层

卷积网络中第 $l$ 层输出的特征图由第 $l-1$ 层和第 $l$ 层卷积核决定,第 $l$ 层特征图单个像素计算公式为

$$y_{p,q}^l = \sum_i \sum_j \omega_{i,j}^l x_{p+i,q+j}^{l-1}, \quad (1)$$

其中 $i, j$ 表示第 $l$ 层卷积核权值位置索引, $p, q$ 表示第 $l$ 层特征图像素位置索引, $\omega_{i,j}^l$ 表示第 $l$ 层 $i, j$ 位置卷积核权值,同理 $x_{p+i,q+j}^{l-1}$ 表示第 $l-1$ 层 $p+i, q+j$ 位置特征图像素值.

由于权重共享,每个 $\omega_{i,j}$ 的变化都会影响到下一层特征输出,因此在权重更新时,对其进行全微分

$$\frac{\partial E}{\partial \omega_{i,j}^l} = \sum_p \sum_q \frac{\partial E}{\partial y_{p,q}^l} \frac{\partial y_{p,q}^l}{\partial \omega_{i+j,p+q}^l}, \quad (2)$$

其中 $E$ 表示卷积核元素值的矩阵.

### 1.2.2 Dropblock 正则化

Dropblock有2个重要的参数 $s$ 和 $\gamma$ , $s$ 用于控制被丢弃块的大小,用于控制被丢弃激活单元数量的

$$\gamma = \frac{1-p}{s^2} \frac{f_s^2}{(f_s - s + 1)^2}, \quad (3)$$

其中, $p$ 表示使某个单元保持活性的概率, $f_s$ 表示该处特征图的大小,本文设置 $p$ 为0.8.

## 1.3 特征分类

根据上节提取的共享特征图,首先将共享特征图拉平为512维向量,随后经过一个4 096维FC层,将拉平后的512维向量转变为 $1 \times 4 096$ 向量.在该全连接层之后,应用Dropout正则化,从而实现分类网络深层网络区域正则化,采用二值交叉熵损失函数和相应的梯度下降法进行参数更新.通过训练迭代获得模型参数.基于测试集,通过分类准确率及损失评估模型的性能.

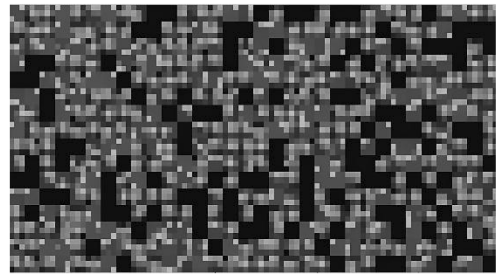


图2 可视化共享特征图(基于VGG-16)

Fig.2 Visualize shared feature maps(based on VGG-16)

### 1.3.1 Dropout 正则化

对于标准的卷积神经网络全连接层,第  $l$  层卷积网络的输出

$$z_i^l = w_i^l y^{l-1} + b_i^l, \tag{4}$$

其中  $y^{l-1}$  为第  $l-1$  层网络输出,  $w$  和  $b$  分别表示第  $l$  层卷积网络的权重和偏置,最后  $y_i^l = f(z_i^l)$  激活函数对第  $l$  层卷积输出进行非线性映射. 引入 Dropout 正则化,首先使用一个独立的伯努利随机函数  $r_j^{l-1} \sim \text{Bernoulli}(p)$  随机失活 FC 层部分神经元

$$\bar{y}^{l-1} = r^{l-1} y^{l-1}, \tag{5}$$

其中,  $y^{l-1}$  为第  $l-1$  层网络输出,  $r^{l-1}$  为随机失活函数.通过第  $l$  层卷积  $z_i^l = w_i^l \cdot \bar{y}^{l-1} + b_i^l$  处理第  $l-1$  层网络输出  $y^{l-1}$ ,最后使用  $y_i^l = f(z_i^l)$  激活函数对  $z_i^l$  进行非线性映射.

### 1.3.2 损失函数

为了训练该分类网络,采用二值交叉熵损失函数

$$l_y = -\frac{1}{N} \sum_{i=0}^N (y_i \ln(y'_i) + (1 - y_i) \ln(1 - y'_i)), \tag{6}$$

其中,  $N$  表示样本数量,取样本点  $y_i$  的概率为  $y'_i$ . 由(6)式可知,对于整个模型来说,损失函数是所有样本点损失的平均值,最终损失值非负,取值在  $(0, 1)$  之间.

## 2 实验安排

实验数据来自于 Kaggle 比赛中猫狗分类所用的图片数据集<sup>[10]</sup>. 该数据集共有 17 000 张图片,其中训练集 10 000 张,验证集 4 000 张,测试集 3 000 张,猫和狗的图片数量各占一半. 实验所用硬件: Intel i7-7700K CPU, 32 GB RAM, 520 GB SSD ROM, NVIDIA GTX 1070Ti GPU; 软件: CUDA 9.0, CUDNN 7.0, Python 2.7.8, TensorFlow, Ubuntu 16.04.

实验分为 2 部分. 第 1 部分实验基于 AlexNet<sup>[11]</sup> 和 ResNet-34<sup>[12]</sup>, 从 0 开始训练分类网络, 分为 2 阶段训练策略: 第 1 阶段, 迭代前 100 epoch, 采用 AdaDelta<sup>[13]</sup> 优化函数, 该阶段学习率为 1.0; 第 2 阶段, 剩下的 500 epoch, 采用 AdaMax 优化函数, 此阶段学习率为 0.002. 第 2 部分实验基于 VGG-16<sup>[14]</sup>, 使用预训练模型参数初始化网络, 随后通过 fine-tuning 策略将用于特征提取的卷积层的解冻部分和新添加的网络层共同参与分类网络训练, 在此部分实验的中, 迭代 100 epoch, 采用 RMSprop 优化函数, 学习率为 0.000 1. 基础网络参数初始化都通过 ImageNet 图像数据集<sup>[15]</sup> 预训练获得. 基于上述 3 种基础网络的分类网络模型见图 3.

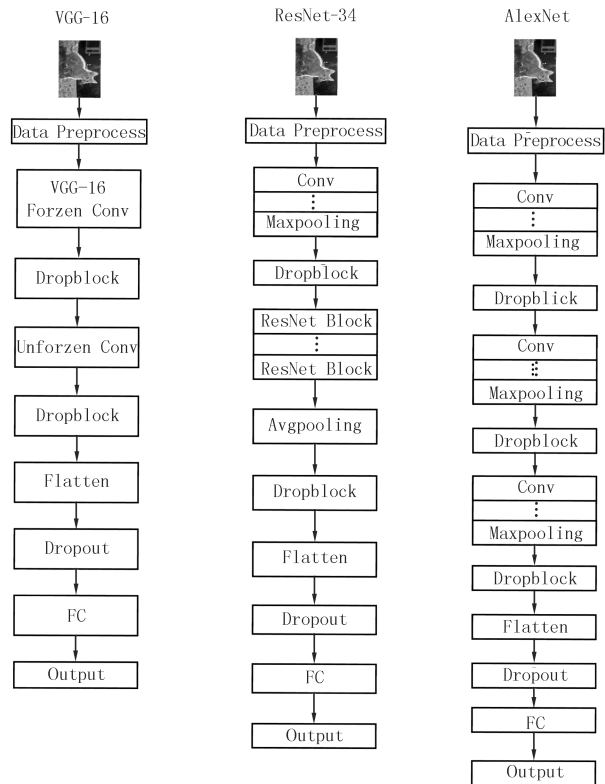


图 3 基于 VGG-16, ResNet-34 和 AlexNet 的分类网络模型

Fig. 3 Classification network model based on VGG-16, Resnet-34 and AlexNet

### 3 实验结果

通过实验(结果见图 4)可知,基于 Dropblock 和 Dropout 的正则化策略使得深度卷积分类网络获得最高的分类准确率.基于 VGG-16 基础网络的实验表明,不添加任何正则化策略(None)的网络模型分类准确率为 95.3%,添加 Dropout 使其分类准确率提高了 0.9%,添加了 Dropout&Dropblock 正则化策略使其分类准确率提高了 2.5%;基于 ResNet-34 基础网络的实验表明,不添加任何正则化策略(None)的网络模型分类准确率为 94.5%,添加 Dropout 使其分类准确率提高了 1.2%,添加了 Dropout&Dropblock 正则化策略使其分类准确率提高了 2.5%;但基于 AlexNet 基础网络的实验表明,添加了 Dropout&Dropblock 正则化策略比 Dropout 正则化的网络分类准确率下降了 0.3%,由于 Alexnet 网络较浅,当浅层网络卷积层中有较多的神经元失活后,一定程度上会影响网络提取输入数据的特征,导致分类准确率下降,由此表明本文提出的复合型正则化策略能有效提升深度卷积神经网络分类准确率,但在一定程度上降低了浅层卷积神经网络分类准确率.

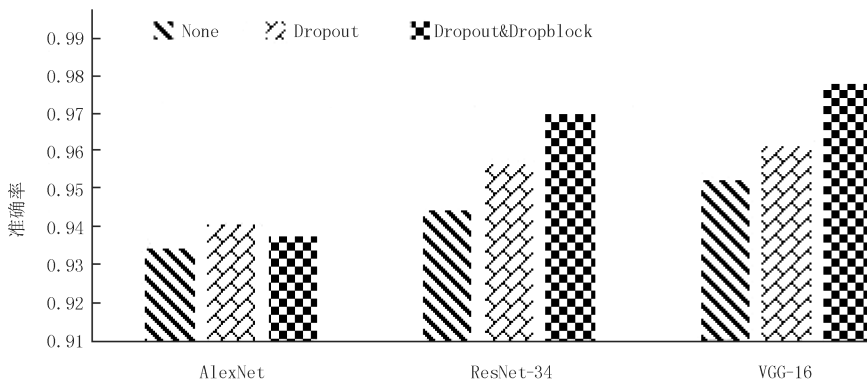


图 4 网络模型分类结果

Fig.4 Classification results of network models

如图 5 所示,基于 Dropblock 和 Dropout 正则化策略应用到分类网络中使得网络在训练过程中更加稳定,且收敛速度更快.基于 ResNet-34,应用 Dropout&Dropblock 正则化策略的网络训练至第 10 epoch 左右时,验证集损失急剧下跌,随后缓慢下降直至平稳;而训练集损失则先快速下降,直至平稳.对比 None 和 Dropout,Dropblock &Dropout 正则化策略使整个训练过程损失波动性更小,下降速度快且最终的损失值更小,充分说明该正则化策略提高了网络训练的稳定性、加快了网络收敛速度.同样地,基于 VGG-16 和 ResNet-34 的实验结果一致.此外,基于 AlexNet 的实验发现,虽然应用 Dropout&Dropblock 正则化策略比 Dropout 的训练集损失稍大,但应用 Dropout&Dropblock 正则化策略使得训练过程损失曲线更加稳定且收敛到最小损失的速度更快.由实验结果表明本文提出的正则化策略能有效提高分类网络的稳定性且加快训练分类网络的收敛速度,特别地,对于深度分类网络,还能提高其分类准确率.

### 4 结论

本文提出了一种结合 Dropblock 和 Dropout 新的正则化策略针对图像分类进行了研究.采用 Dropblock 实现分类网络浅层和中间层卷积区域的正则化,采用 Dropout 对分类网络深层全连接区域进行正则化,从而实现对整个分类网络全面正则化,且有效降低模型复杂度.最后,计算每个 epoch 的验证集损失和测试集损失评估分类网络的收敛速度和稳定性,及计算测试集分类准确率.通过实验结果表明,该正则化策略能有效加快分类网络的收敛速度和提升其稳定性,且能有效提升深度卷积分类网络的分类准确率.

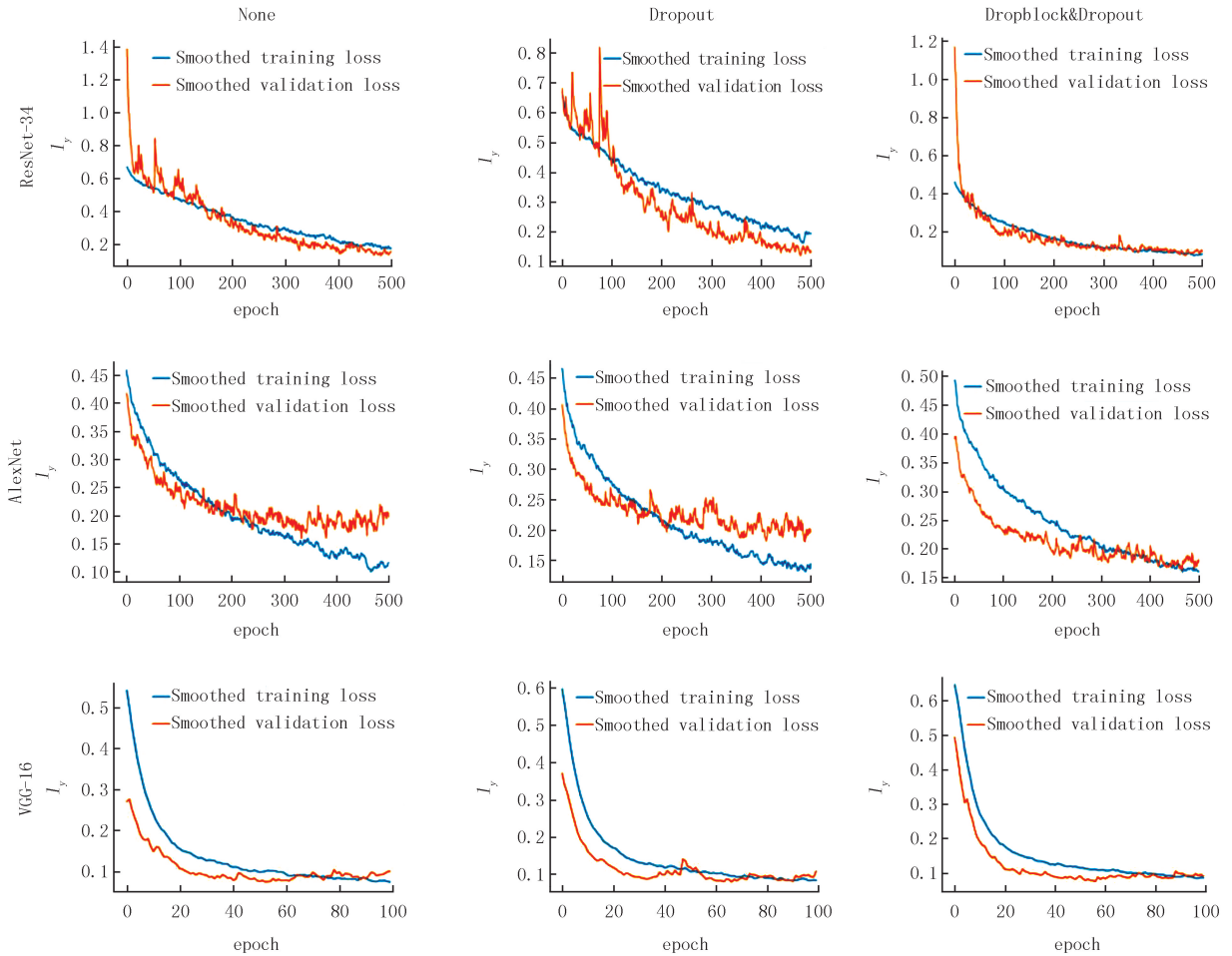


图 5 分类网络模型的训练集损失和验证集损失

Fig. 5 Training set loss and validation set loss of classification network models

## 参 考 文 献

- [1] 陈伟,何家欢,裴喜平.基于相空间重构和卷积神经网络的电能质量扰动分类[J].电力系统保护与控制,2018,46(14):87-93.
- [2] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//2015 Conference on Computer Vision and Pattern Recognition. Boston: IEEE Press, 2017: 1-12.
- [3] Luo D, Ding C, Huang H. Linear discriminant analysis. New formulations and overfit analysis[C]//2011 American Association for Artificial Intelligence. San Francisco: AAAI Press, 2011: 417-422.
- [4] Goryainov V B, Goryainova E R. Comparative analysis of robust and classical methods for estimating the parameters of a threshold autoregression equation[J]. Automation and Remote Control, 2019, 80(4): 666-675.
- [5] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout, a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [6] Huang Juiting, Li Jinyu, Yu dong, et al. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers[C]//2013 International Conference on Acoustics, Speech and Signal Processing. Vancouver: IEEE Press, 2013: 7304-7308.
- [7] Acharya U R, Meiburger K M, Koh J E W, et al. A novel algorithm for breast lesion detection using textons and local configuration pattern features with ultrasound imagery[J]. IEEE Access, 2019(7): 22829-22842.
- [8] Kong Tao, Yao Anbang, Chen Yurong, et al. HyperNet. Towards accurate region proposal generation and joint object detection[C]//2016 Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press, 2017: 845-853.
- [9] Ghiasi G, Lin T Y, Le Q V. DropBlock. A regularization method for convolutional networks[C]//2018 Neural Information Processing Systems. Montréal: IEEE Press, 2018: 10750-10760.
- [10] Goldbloom A, Hamner B, Moser J, et al. The Kaggle dogs vs cats competition[EB/OL]. [2019-06-16]. <https://www.Kaggle.com/c/dogs-vs-cats>

vs-cats-redux-kernels-edition.

- [11] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [12] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Press, 2017: 770-778.
- [13] Qu Zhijian, Yuan Shengao, Chi Rui, et al. Genetic optimization method of pantograph and catenary comprehensive monitor status prediction model based on adadelta deep neural network[J]. IEEE Access, 2019(7): 23210-23221.
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//2015 International Conference on Learning Representations. San Diego: Computer Science, 2015: 1-14.
- [15] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2014, 115(3): 211-252.

## A regularization strategy combining Dropblock and Dropout

Hu Hui, Si Fengyang, Zeng Chen, Shu Wenlu

(School of Information Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** In order to enhance the convergence rate and robustness of classification networks, a new regularization strategy is proposed. The Dropblock algorithm and Dropout algorithm are combined to regularize outputs of the shallow, middle and deep layers in a convolutional classification network. The Dropblock can implement regularization by hiding part of feature maps of convolution layers, the Dropout can implement regularization by hiding part of weight parameters of fully connected layers. Therefore, a whole convolutional classification network can be regularized. Training and testing experiments were conducted through the dataset provided by the Kaggle dogs vs cats classification competition. It shows that the new regularization strategy can effectively enhance the convergence rate and robustness of classification networks. In addition, the regularization strategy can effectively improve the classification accuracy of deep convolutional classification network.

**Keywords:** regularization; Dropout; Dropblock; convergence speed; stability

[责任编辑 杨浦 王凤产]