

DOI:10.11992/tis.201509018

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20160105.1532.004.html>

# 一种卷积神经网络的图像矩正则化策略

殷瑞<sup>1,2</sup>, 苏松志<sup>1,2</sup>, 李绍滋<sup>1,2</sup>

(1. 厦门大学 信息科学与技术学院, 福建 厦门 361005; 2. 厦门大学 福建省仿脑智能系统重点实验室, 福建 厦门 361005)

**摘要:**卷积神经网络的池化策略包含极大池化和平均池化,极大池化选择池化区域中的最大值,极易出现过拟合现象;平均池化对池化区域中所有元素赋予相同权重,降低了高频分量的权重。本文提出将矩池化作为卷积神经网络的正则化策略,矩池化将几何矩概念引入到卷积神经网络的池化过程中,首先计算池化区域的中心矩,然后根据类插值法依概率随机地从中心矩的 4 个邻域中选择响应值。在数据集 MNIST、CIFAR10、CIFAR100 上的实验结果表明随着训练迭代次数的增加,矩池化的训练误差和测试误差最低,矩池化的高差别性和强鲁棒性使其获得了比极大池化和平均池化更好的泛化能力。

**关键词:**中心矩;随机选择;池化;卷积神经网络;过拟合

**中图分类号:**TP391.4 **文献标志码:**A **文章编号:**1673-4785(2016)01-0043-06

中文引用格式:殷瑞,苏松志,李绍滋.一种卷积神经网络的图像矩正则化策略[J].智能系统学报,2016,11(1):43-48.

英文引用格式:YIN Rui, SU Songzhi, LI Shaozi. Convolutional neural network's image moment regularizing strategy[J]. CAAI Transactions on Intelligent Systems, 2016, 11(1): 43-48.

## Convolutional neural network's image moment regularizing strategy

YIN Rui<sup>1,2</sup>, SU Songzhi<sup>1,2</sup>, LI Shaozi<sup>1,2</sup>

(1. School of Information Science and Technology, Xiamen University, Xiamen 361005, China; 2. Fujian Key Laboratory of the Brain-Like Intelligent System, Xiamen University, Xiamen 361005, China)

**Abstract:** There are two kinds of pooling strategies for convolutional neural network (CNN) as follows: max pooling and average pooling. Max pooling simply chooses the maximum element, which makes this strategy extremely prone to overfitting. Average pooling endows all elements with the same weight, which lowers the weight of the high-frequency components. In this study, we propose moment pooling as a regularization strategy for CNN. First, we introduce the geometric moment to CNN pooling and calculate the central moment of the pooling region. Then, we randomly select the response values based on the probability-like interpolation method from the four neighbors of the moment as per their probability. Experiments on the MNIST, CIFAR10, and CIFAR100 datasets show that moment pooling obtains the fewest training and test errors with training iteration increments. This strategy's robustness and strong discrimination capability yield better generalization results than those from the max and average pooling methods.

**Keywords:** central moment; random selection; pooling; convolutional neural network; overfitting

卷积神经网络模型训练过程中存在一个共同的问题,即当模型复杂度比较高时,由于训练数据相对不足而产生过拟合现象,导致模型泛化能力比较差。一系列正则化技术被用来解决该问题,如  $L_1$  正则化,  $L_2$  正则化,即对损失函数增加  $L_1$  范数或  $L_2$  范

数,根据图像变换扩增训练数据等。虽然上述方法可以在一定程度上抑制过拟合,训练出比未正则化模型泛化性更强的模型,但是由于以上方法是基于参数的方法,在解决过拟合问题的同时却增加了模型训练复杂度,在一定程度上弱化了其效果。

Hinton 等<sup>[2]</sup>提出的 dropout 是另外一种正则化方法,它在训练过程中以 0.5 的概率随机将网络中神经元的响应值置为 0。该方法在计算机视觉的多个问题中都取得了不错的成绩,然而目前仍不能完

收稿日期:2015-09-16. 网络出版日期:2016-01-05.

基金项目:国家自然科学基金资助项目(61202143, 61572409);福建省自然科学基金资助项目(2013J05100).

通信作者:李绍滋. E-mail: szlig@xmu.edu.cn.

全解释它的有效性。Dropout 并不能在神经网络的每一层都提高模型的泛化能力,甚至会起反作用,因此 dropout 在网络中的位置选择非常重要,然而目前并没有特定的标准,依赖于个人经验。

池化方法是一种无参的正则化方法。极大池化(max pooling)简单地从池化区域中选择最大值作为最终响应值,虽然保留了高频分量,如图像的边缘等信息,但它对噪声信息非常敏感,并且在训练中极易出现过拟合。平均池化(average pooling)对池化区域的所有值取平均,该方法虽然考虑了区域中所有信息,有效地降低了噪声信息的影响,然而它平滑了图像,从而导致非常重要的高频信息丢失。

为确保模型的判别性及鲁棒性,池化策略必须保留高频分量以提高判别性,同时通过一定的随机性以增强鲁棒性。本文从全新的角度提出了一项卷积神经网络池化策略:矩池化。首先,计算池化区域的中心矩(即图像灰度重心),一般情况下中心矩是一个浮点值,并不指向图像中一个确定的像素位置,本文根据概率随机地从中心矩的 4 个邻域中选择一个作为最终的响应值。该策略应用中心矩方法在保留高响应值的同时弱化了噪声信息的影响,随机性避免了训练过程中的过拟合现象。

## 1 卷积神经网络

人类的视觉系统对外界的认识是从局部到全局的过程,因而卷积神经网络(convolutional neural network CNN)认为图像的局部像素联系比较远距离的像素联系更紧密。因此卷积神经网络在高分辨率提取局部特征,然后在低分辨率将局部特征组合形成更加复杂的特征。CNN 通过增加较高层特征图的数量来补偿空间信息的丢失。CNN 的基础框架由卷积层和池化层组成,卷积层类似于简单细胞,池化层使图像特征具有平移不变性。

### 1.1 卷积层

卷积层是实现图像局部联系的一种途径,它通过参数共享学习适用于所有数据的底层特征。图 1 是第一个卷积层学习到通用的边缘特征,更高的卷积层可以提取更加具体的特征如轮廓。卷积层将整个图片与卷积核卷积,然后计算卷积核在图片每个位置的内积,将输出作为特征图  $c$ ,  $c$  可以用来评价图像的每一个部分与卷积核的匹配度。

然后将非线性变换函数应用于特征图  $c$ , 即:  $a=f(c)$ 。非线性变换函数  $f$  有很多种选择,最常用的是 tanh 和 logistic 函数。本文选择受限线性单元(ReLU)  $f(c) = \max(0, c)$ , V. Nair<sup>[3]</sup> 证明 ReLU 比

tanh 和 logistic 泛化能力更强。

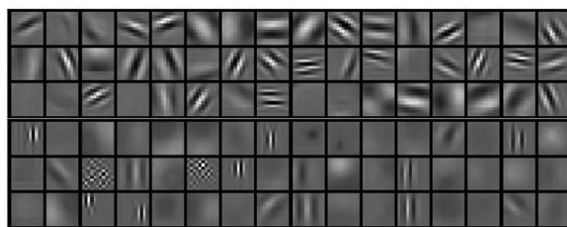


图 1 CNN 网络第一卷积层学习到的部分边缘特征

Fig.1 The part of edge features learned by the first convolution layer of CNN

### 1.2 池化层

池化层通过池化函数  $\text{pool}(\cdot)$  将图像中一个较小的子区域  $R$  的信息集合起来,其作用是通过降低分辨率获得模型对图像平移及旋转变换的不变性。

$$s_i = \text{pool}(f(c_j)) \quad \forall j \in R_i \quad (1)$$

式中:  $R_i$  表示特征图  $c$  中池化区域,  $i, j$  是该区域每个元素的索引。

Pool 有 2 种常用函数:  $\max$ <sup>[4]</sup> 和  $\text{ave}$ <sup>[5]</sup>。前者取池化区域的最大值:

$$s_i = \max_{j \in R_i} a_j \quad (2)$$

后者取池化区域的平均值:

$$s_i = \frac{1}{|R_i|} \sum_{j \in R_i} a_j \quad (3)$$

在深度卷积神经网络中,上述 2 种池化方法都有其不可避免的缺点。由于  $\max$  仅简单地选择最大值,因此对噪声非常敏感,并且实验发现该方法极易过拟合。 $\text{ave}$  考虑了池化区域中所有元素,避免了噪声的影响,然而在 CNN 中我们期望池化层尽可能保留强响应值,但  $\text{ave}$  考虑了池化区域中的所有元素,并且赋予了相同权重。当选择 ReLU 函数作为非线性变换函数时,  $\text{ave}$  考虑了大量的 0 值,间接地降低了强响应值的权重。假如非线性变换函数选择 tanh,  $\text{ave}$  会抵消掉正负强响应值(梯度方向相反),返回一个完全不具代表性的响应值。

### 1.3 全连接层

全连接层一般在网络的最后几层,与卷积层不同,它感知全局信息,将卷积层学习到的局部特征聚集起来形成全局特征用于特定的图像处理任务,如分类、检测、识别等。

## 2 矩池化

矩不变性是图像处理中一个经典问题,自 1962 年 H. Ming et.al<sup>[6]</sup> 将矩特征应用于模式识别中后,该方法被广泛应用到计算机视觉中的多个邻域。由于

矩特征具有平移不变性、相似变换不变性、旋转不变性以及正交变换不变性,因此它被成功地应用于图像分类中。P.Rosin<sup>[7]</sup>认为局部区域的灰度值分布相对于其中心具有偏移性,因此他利用图像中心到中心矩的向量来表示局部区域的主方向。E. Rublee<sup>[8]</sup>在提取二值特征时,将矩特征应用于估计主方向以提高特征的鲁棒性。本文将中心矩与随机选择结合起来,将其应用于 CNN 的池化操作中,为了不增加 CNN 模型的计算复杂性,本文采用一阶矩。

灰度矩定义如下:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \quad (4)$$

式中: $I(x,y)$  表示图像位置  $(x,y)$  处的灰度值, $p,q$  取值为 1。

中心矩定义如下:

$$c = \begin{bmatrix} \frac{m_{10}}{m_{00}} & \frac{m_{01}}{m_{00}} \end{bmatrix} \quad (5)$$

根据式(4)、(5)计算池化区域的中心矩  $c(x,y)$ ,一般情况下  $c$  为浮点值,不指向任何一个确定的离散值,如图 2 所示,其上下边界具有 4 个确定位置( $Q_{11}, Q_{12}, Q_{22}, Q_{21}$ ),称其为  $c$  的四邻域。在图像放大中采用插值法计算插入点的像素值,本文借鉴该思想通过中心矩的四邻域计算其响应值。

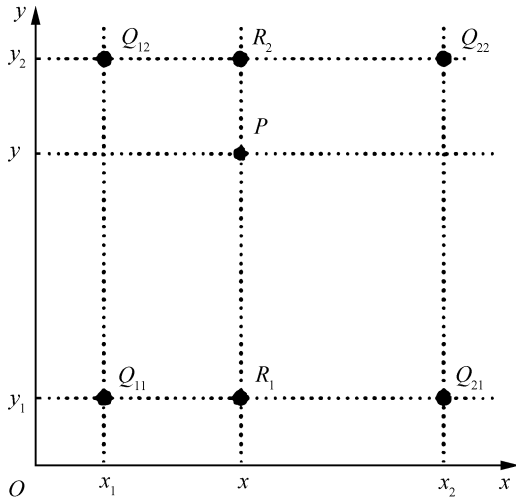


图 2 矩池化选择方法

Fig.2 The moment pooling's selection strategy

$$\begin{aligned} x_1 &= (\text{int})x & y_1 &= (\text{int})y \\ x_2 &= x_1 + 1 & y_2 &= y_1 + 1 \\ Q_1 &= (x_1, y_1) \\ Q_2 &= (x_2, y_1) \\ Q_3 &= (x_2, y_2) \\ Q_4 &= (x_1, y_2) \end{aligned} \quad (6)$$

式中  $x_1, x_2$  分别表示  $x, y$  向下取整。

最近邻插值法<sup>[9]</sup>选择与插入点最近的像素作为中心矩的像素值,虽然该方法插值速度快,但它容易出现棋盘格效应。且在池化操作中,中心矩与最大值像素位置非常接近,最近邻插值法几乎等价于极大池化法。

双线性插值法<sup>[10]</sup>用待插入点的  $2 \times 2$  区域的 4 个邻近像素的值通过加权平均计算得到。其核心思想是在  $x, y$  方向上分别进行一次线性插值,距离待插入点越近,权值越高。在池化操作中双线性内插算法不会出现像素值不连续的情况,然而此算法和平均池化类似,具有低通滤波器的性质,使高频分量受损,所以可能会使轮廓信息在一定程度上变得模糊。除此之外,双线性插值法要分别在 2 个方向上插值,增加了模型的计算复杂度。

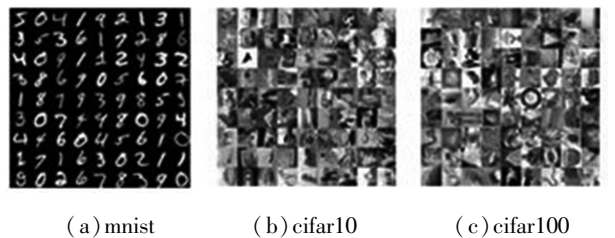
本文在 2 个方向上以概率  $p_x, p_y$  随机选择  $x, y$ , 式(7)、(8)所示距离中心矩越近,被选择的可能性越大。最后根据选择的  $(x, y)$  坐标从四邻域中选择池化区域的响应值。矩池化的随机性使每次选择都不相同,因此可以有效地达到预防过拟合的效果。在计算复杂度上,该策略只需要随机生成 2 个概率,远小于双线性插值法,与最近邻插值法接近。

$$\begin{aligned} p_1 &= (x_2 - x) \\ p_2 &= (y_2 - y) \end{aligned} \quad (7)$$

$$\begin{aligned} x &= \begin{cases} x_1 & p_x \leq p_1 \\ x_2 & p_x > p_1 \end{cases} \\ y &= \begin{cases} y_1 & p_y \leq p_1 \\ y_2 & p_y > p_1 \end{cases} \end{aligned} \quad (8)$$

### 3 实验

实验部分将矩池化、极大池化、平均池化分别在 MNIST, CIFAR-10, CIFAR-100 数据集上进行比较,图 3 代表 3 个数据集的部分数据。实验采用深度学习框架 caffe<sup>[11]</sup>,caffe 是一个清晰而高效的深度学习框架,它具有上手快、速度快、模块化、开放性、社区好的特点,并且支持命令行、python 和 matlab 接口,可以在 CPU 和 GPU 间无缝切换,大大提高了模型的学习效率。



(a) mnist (b) cifar10 (c) cifar100

图 3 实验中用到的部分数据

Fig.3 The part experiments data



模型优化采用随机梯度下降法 (stochastic gradient descent), 通过损失函数的负梯度  $\nabla L(W_t)$  与  $t$  时刻权重更新值  $v_t$  的线性组合来更新权重。学习率  $\alpha$  代表负梯度的权重, 势  $\mu$  代表  $v_t$  的权重。形式上, 在给定  $t$  时刻的  $v_t, w_t$  时, 通过式 (9) 更新  $t+1$  时刻的权重。这两个参数需要通过调整来得到最好的结果<sup>[12]</sup>, 一般经过  $\text{stepsize}$  个训练回合, 将  $\alpha$  更新为原来的  $\gamma_{th}$ 。

训练中需要优化的参数有训练的回合数 (epoch), 学习率  $\alpha$ , 势  $\mu$ , 权重衰减  $\lambda$ ,  $\alpha$  的变化步长  $\text{stepsize}$  和变化率  $\gamma$ , 本文中  $\mu=0.9$ ,  $\gamma=0.01$ , 其他参数设置与具体任务有关。

$$\begin{aligned} v_{t+1} &= \mu v_t - \alpha \nabla L(w_t) \\ w_{t+1} &= w_t + v_{t+1} \end{aligned} \quad (9)$$

### 3.1 训练集 CIFAR-10

CIFAR-10<sup>[13]</sup> 包含 6 万个  $32 \times 32$  的 RGB 图片, 共分为 10 种类型。训练数据 50 000 张图片, 测试数据 10 000 张图片。实验采用 Hinton 等提出的 dropout 模型 2, 该模型包含 3 个卷积层, 每个卷积层由 64 个  $5 \times 5$  的卷积核构成, 卷积层将非线性变换函数 ReLU 的结果作为它的输出。池化层选择大小为 3 的池化区域, 以步长为 2 实施池化操作, 然后通过局部响应归一化层 (LRN: local response normalization) 对池化结果进行局部归一化以抑止非常大的 ReLU 的输出值。最后一层是全连接层, 它的 softmax 输出是整个网络的分类预测结果。

实验中  $\alpha$  的调整步长  $\text{stepsize}$  的大小非常重要, 其太小将会导致模型迟迟不能跳出局部最优; 反之由于学习率太大, 模型一直在全局最优附近徘徊, 前者会降低模型泛化能力, 后者延缓了模型的训练速度。图 4 是  $\text{stepsize}=20, 100$  时矩池化的误差曲线, 第一次学习率调整后, 误差曲线都会加速下降。后续调整仅降低训练误差, 对测试误差影响甚微, 因此最大迭代次数不超过  $2 \times \text{stepsize}$ 。由图 4 可看出选 1.2~1.5 倍即可。当  $\text{stepsize}=20$  时, 曲线仍保持比较大的下降趋势, 即模型还未达到全局最优附近, 此时降低学习率, 使模型在后面的迭代过程中极易陷入局部最优, 测试误差为 20.57%; 当  $\text{stepsize}=100$  时, 曲线变化已趋平缓, 此时降低学习率可使模型趋于全局最优, 测试误差降低至 17.24%, 后续实验参数设置如表 1 所示。

表 1 CIFAR-10 数据集上的参数设置

Table 1 The parameter settings on dataset CIFAR-10

epoch	$\alpha$	stepsize	$\gamma$	$\lambda$
120	0.001	100	0.01	0.004

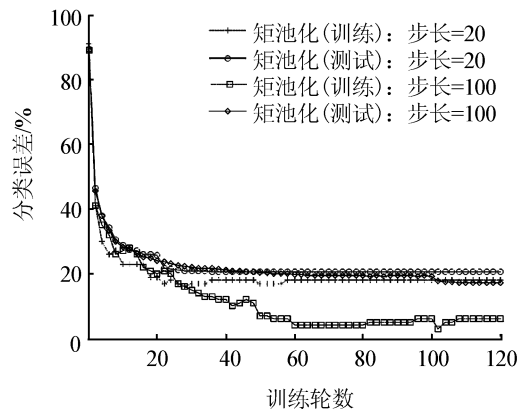


图 4 学习率  $\alpha$  的迭代步长对实验结果的影响

Fig.4 The effect of learning rate to  $\alpha$ 's iteration stepsize

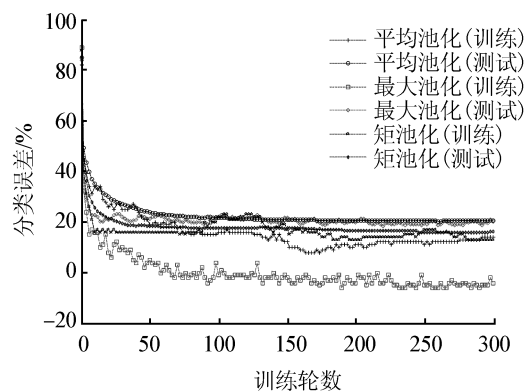


图 5 数据集 CIFAR-10 上的实验结果

Fig.5 Experiments result on dataset CIFAR-10

使用上述 CNN 模型, 在池化层分别采用极大池化, 平均池化以及本文提出的矩池化方法, 并比较它们的分类结果。如图 5 所示, 最大池化曲线训练过程中快速下降到 0, 平均池化的训练误差及测试误差都比较高。矩池化在有效避免过抑合的同时, 又能保证最低的预测误差, 这与矩特征的 2 个特性息息相关。表 2 比较 3 种池化方法在 CIFAR-10 数据集上的训练误差及预测误差, 矩池化预测误差最低。

表 2 不同池化方法在 CIFAR-10 数据集上的训练误差与测试误差

Table 2 Pooling methods' training errors and test errors on dataset CIFAR-10

方法	训练误差/%	预测误差/%
平均池化	10.40	19.11
最大池化	0.00	19.38
矩池化	6.00	17.24

3.2 训练集 MNIST

MNIST<sup>[14]</sup>由大小为 28×28 的手写体 0~9 组成,数据集中包含 60 000 张训练图片和 10 000 张测试数据,实验中预处理时将图片归一化到[0,1]。

Lecun Y<sup>[15]</sup>提出的 LeNet-5 模型在数字分类任务中取得了非常好的效果,本文采用的模型与 LeNet-5 略有不同,将原来的非线性变换函数 sigmoid 替换为 ReLU。实验中参数设置如表 3 所示,stepsize=fixed 表示学习率在学习过程中不变,实验中为 20。

表 3 MNIST 数据集上的参数设置  
Table 3 The parameter settings on dataset MNIST

epoch	$\alpha$	stepsize	$\gamma$	$\lambda$
100	0.01	fixed	null	0.000 5

训练过程中,3 种池化方法几乎都完全过度抑合训练数据,权重衰减虽然可以预防过抑合,但在该数据集上效果甚微,表 4 比较各个池化方法在 MNIST 数据集上的结果。矩池化方法只是一种正则化方法,因此可以与其他方法结合提高模型泛化能力。

表 4 不同池化方法在 MNIST 数据集上的训练误差与测试误差

Table 4 Pooling methods' training errors and test errors on dataset MNIST %		
方法	训练误差	预测误差
平均池化	0.05	1.51
最大池化	0.02	1.07
矩池化	0.10	0.93

3.3 训练集 CIFAR-100

CIFAR-100<sup>[8]</sup>数据集与 CIFAR-10 类似,不同之处在于它增加了图片类别数到 100,包含 50 000 张训练数据(每类 500 张)及 10 000 张测试数据。CIFAR-100 的训练模型以及参数设置如表 5 所示,与 CIFAR-100 的基本一致,为提高准确率将第 3 个卷积层数增加到 128。相较于 CIFAR-10, CIFAR-100 的训练数据相当有限,而模型却更加复杂。如表 6 所示有限的训练数据下传统的池化方法并没有很好的泛化能力,而矩池化方法有效地降低了分类误差。

数据集 CIFAR-100 上的实验结果如图 6 所示,可以看出,第 1 次调整学习率前,曲线已趋平缓,调整后曲线迅速下降,与 CIFAR-10 实验结果一致。

表 5 CIFAR-100 数据集上的参数设置  
Table 5 The parameter settings on dataset CIFAR-100

epoch	$\alpha$	stepsize	$\gamma$	$\lambda$
300	0.001	100	0.01	0.004

表 6 不同池化方法在 CIFAR-100 数据集上的训练误差与测试误差

Table 6 Pooling methods' training errors and test errors on dataset CIFAR-100 %		
方法	训练误差	预测误差
平均池化	27.00	47.03
最大池化	1.00	49.64
矩池化	14.00	45.38

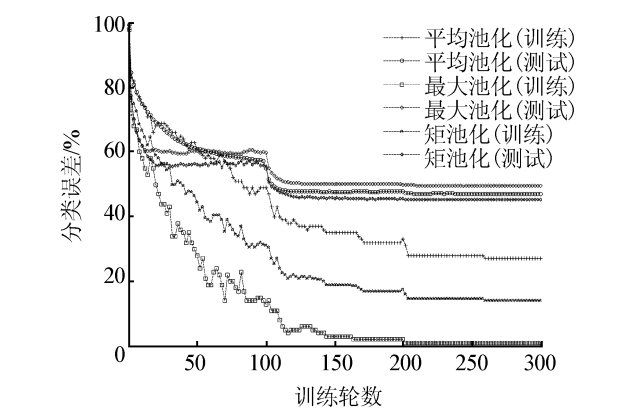


图 6 数据集 CIFAR-100 上的实验结果  
Fig.6 Experiments result on dataset CIFAR-100

4 结束语

本文在卷积神经网络的框架上提出了一种可以与其他正则化方法结合使用的池化策略,如 dropout、权重衰减、数据扩增等。矩池化将中心矩与随机选择应用于 CNN 的池化层中,中心矩选择池化区域的灰度重心而非最大值,在保持模型判别性的同时有效地消除噪声的影响;而选择的随机性,使每次选择具有一定的不确定性,从而更好地避免过抑合,提高模型鲁棒性。除此之外,矩池化是一种无参的正则化方法,不会影响模型复杂度,可以与任何已有的 CNN 模型相结合。实验表明该方法可以有效地预防过抑合,提高模型的泛化能力。

参考文献:

[1] MONTAVON G, ORR G, MÜLLER K R. Neural networks: tricks of the trade[M]. 2nd ed. Berlin Heidelberg: Springer, 2012.

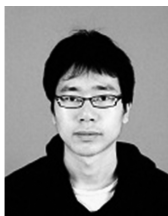
[2] HINTON G E, SRIVASTAVE N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors [EB/OL]. [2012-07-03]. <http://arxiv>.

- org/pdf/1207.0580.pdf.
- [3] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines [C]//Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel, 2010.
- [4] RANZATO M, BOUREAU Y L, LECUN Y. Sparse feature learning for deep belief networks [C]//Proceedings of Advances in Neural Information Processing Systems (NIPS). Cambridge, MA, 2007.
- [5] LECUN Y, BOSER B E, DENKER J S, et al. Handwritten digit Recognition with a back-propagation network [C]//Proceedings of Advances in Neural Information Processing Systems (NIPS). Cambridge, MA, 1989.
- [6] HU M K. Visual pattern recognition by moment invariants [J]. IRE Transactions on Information Theory, 1962, 8(2): 179-187.
- [7] ROSIN P L. Measuring corner properties [J]. Computer vision and image understanding, 1999, 73(2): 291-307.
- [8] RUBLEE E, RABAUDE V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF [C]//Proceedings of IEEE International Conference on Computer Vision (ICCV). Barcelona, 2011: 2564-2571.
- [9] EVANS O D, KIM Y. Efficient implementation of image warping on a multimedia processor [J]. Real-time imaging, 1998, 4(6): 417-428.
- [10] GONZALEZ R C, WOODS R E. Digital image processing [M]. 2nd ed. New Jersey: Prentice-Hall, 2002.
- [11] JIA Y, SHEHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding [C]//Proceedings of the ACM International conference on Multimedia. ACM, 2014: 625-678.
- [12] BOTTOU L. Stochastic gradient descent tricks [M]//MONTAVON G, ORR G B, MÜLLER K R. Neural Networks: Tricks of the Trade. 2nd ed. Berlin Heidelberg: Springer, 2012: 421-436.
- [13] KRIZHEVSKY A. The CIFAR-10, CIFAR-100 database [EB/OL]. <http://www.cs.toronto.edu/~kriz/cifar.html>. LECUN Y, CORTES C, BURGESS C J C. The MNIST database of handwritten digits [EB/OL]. <http://yann.lecun.com/exdb/mnist/>.
- [14] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

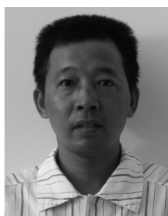
#### 作者简介:



殷瑞,女,1993年生,硕士研究生,主要研究方向为图像特征表示、计算机视觉、深度学习。



苏松志,男,1982年生,讲师,博士。主要研究方向为行人检测和人体行为分析。



李绍滋,男,1963年生,教授,博士生导师,福建省人工智能学会副理事长。主要研究方向为人工智能及其应用、计算机视觉与机器学习、运动目标检测与识别等。主持过多项国家、省市级项目研究,获得省科学技术三等奖2项,发表学术论文200余篇,其中SCI检索27篇、EI检索171篇。