

# 基于卷积神经网络的图像验证码识别<sup>①</sup>

秦波, 顾乃杰, 张孝慈, 林传文

(中国科学技术大学 计算机科学与技术学院 网络计算与高效算法实验室, 合肥 230027)  
(安徽省计算与通信软件重点实验室, 合肥 230027)  
(中国科学技术大学 先进技术研究院, 合肥 230027)  
通讯作者: 顾乃杰, E-mail: [gunj@ustc.edu.cn](mailto:gunj@ustc.edu.cn)

**摘要:** 验证码作为一种安全手段, 被广泛应用于互联网领域. 本文提出了一种基于卷积神经网络的图像验证码识别方法, 通过卷积层级联、残差学习、全局池化、分组卷积等技术手段, 在保证识别准确率不受影响的前提下, 大大降低了网络的参数量. 本文以铁路购票网站验证码和正方教务系统验证码为例来测试模型性能. 对于铁路购票网站验证码, 实验结果显示本文提出的识别方法参数量最少, 对图形和中文词组验证码的识别准确率分别达到 98.76% 和 99.14%; 对于正方教务系统验证码, 本文方法参数量最少且识别准确率为 87.30%.

**关键词:** 图像验证码识别; 卷积神经网络; 残差学习; 可视化

引用格式: 秦波, 顾乃杰, 张孝慈, 林传文. 基于卷积神经网络的图像验证码识别. 计算机系统应用, 2018, 27(11): 142-148. <http://www.c-s-a.org.cn/1003-3254/6648.html>

## Image CAPTCHA Recognition Based on Convolutional Neural Network

QIN Bo, GU Nai-Jie, ZHANG Xiao-Ci, LIN Chuan-Wen

(Laboratory of Network Computing and High Efficient Algorithm, School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)  
(Anhui Provincial Key Laboratory of Computing and Communication Software, Hefei 230027, China)  
(Institute of Advanced Technology, University of Science and Technology of China, Hefei 230027, China)

**Abstract:** As a security measure, CAPTCHA is widely used in Internet. This study proposes a CAPTCHA identification method based on convolutional neural network. Through convolutional layer concatenation, residual learning, global pool, and other technical means, under the premise of ensuring the recognition accuracy rate is not affected, it greatly reduces the amount of network parameters. This study uses the CAPTCHA in the railway ticket website and the CAPTCHA in the educational system as examples to test the performance of the model. For the CAPTCHA in railway ticket website, the experimental results show that this method has the least amount of parameters, and the recognition accuracy of this method is 98.76% for image and the recognition accuracy of the Chinese phrases is 99.14%. For the CAPTCHA in educational system, it has the least amount of parameters and the accuracy is 87.30%.

**Key words:** image CAPTCHA recognition; convolutional neural network; residual learning; visualization

## 1 引言

验证码 (Completely Automated Public Turing test to tell Computers and Humans Apart, CAPTCHA) 是一种用于区分人类行为与机器行为的算法. 作为互联网

行业通用的基础安全策略, 验证码担负着保护网络账号安全体系的第一道防御, 用以验证和防范机器程序的恶意行为, 如何提升验证码的安全性并防止自动破解程序, 已经成为一个至关重要的问题.

<sup>①</sup> 收稿时间: 2018-04-02; 修改时间: 2018-05-11; 采用时间: 2018-05-16; csa 在线出版时间: 2018-10-24

作为深度神经网络的一种,卷积神经网络(Convolutional Neural Networks, CNN)在图像识别领域内表现出优异的性能<sup>[1]</sup>,且大大优于传统的机器学习方法.较传统方法而言,CNN的主要优势在于卷积层提取的图像特征具有很强的表达能力,避免了传统识别技术存在的数据预处理、人工设计特征等问题.在具备足够的标注样本的情况下,只需要定义一个合适的网络模型,就能够取得远高于传统方法所取得的识别效果.

本文设计了一种基于卷积神经网络的图像验证码识别方法,对传统的卷积神经网络训练中存在的模型复杂、输出层参数冗余等问题进行了改进.本文通过卷积层级联、残差学习、全局池化和分组卷积等技术手段,在保证网络识别精度的前提下,压缩了模型规模.

## 2 相关工作

鉴于验证码在互联网的广泛应用,国内外学者对验证码的设计和识别进行了广泛的研究.Zhang和Wang<sup>[2]</sup>利用传统图像处理方法并结合KNN算法进行验证码识别研究.李兴国等人<sup>[3]</sup>提出一种滴水算法分割验证码.Lu等人<sup>[4]</sup>分别提出了不同的字符分割算法并结合SVM分类算法及BP神经网络进行字符识别.Yan等人<sup>[5]</sup>分析并研究了现有微软验证码的缺陷,并设计了虚拟问答及基于情感方法对验证码进行识别.Mori和Malik<sup>[6]</sup>利用形状上下文方法对验证码进行了识别研究.

深度神经网络在科学研究中被广泛使用,Yann LeCun等人<sup>[7]</sup>提出了基于CNN的文字识别系统LeNet-5.Hinton等人<sup>[8]</sup>提出的深度置信网络(Deep Belief Network, DBN)可以更快地训练深度网络.在图像应用中,CNN采用随机梯度下降(Stochastic Gradient Decent, SGD)和GPU(Graphics Processing Unit)加快了深度网络的训练速率.IDSIA实验室<sup>[9]</sup>针对手写中文汉字识别提出一种多列CNN模型,通过训练多个CNN网络并对结果进行简单的平均集成实现并行训练.Zhong等人<sup>[10]</sup>提出一种HCCR-GoogLeNet模型,利用GoogLeNet的网络结构并引入手写汉字的多尺度多方向特征,完成脱机手写中文汉字识别.范望等人<sup>[11]</sup>构造出卷积神经网络用于汉字验证码识别.针对不分割的验证码识别,Google研究人员<sup>[12]</sup>采用概率模型和卷积神经网络开发了一种无需字符分割的街道门牌识别系统,对门

牌准确率高达90%.Yann LeCun等人<sup>[13]</sup>采用空间位移神经网络(SDNN)和隐马尔科夫模型(HMM)对带有粘连的手写字符进行识别.Shi等人<sup>[14]</sup>利用CNN和RNN结合思想,提出一种卷积循环神经网络(CRNN)完成验证码的整体识别.

## 3 网络模型设计

本文提出的卷积神经网络模型主要由以下模块组成:级联卷积层、残差层以及分类池化层,如图1所示,下面对这些结构进行具体介绍.

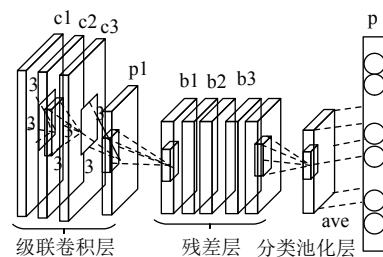


图1 本文网络结构

### 3.1 级联卷积层

传统卷积神经网络的卷积层通常由 $7 \times 7$ 的卷积核构成,用于从前一层的特征图中提取更高级的特征.如图2所示,本文采用的级联卷积层由若干个 $3 \times 3$ 卷积核的卷积层和一个池化层组成,用于替代传统网络中的卷积层.这种级联卷积层的优势主要在于以下几点.

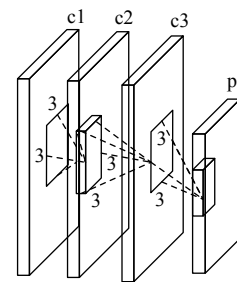


图2 级联卷积层

#### (1) 参数量

假设图2中所有数据有 $C$ 个通道,下面将比较3个 $3 \times 3$ 卷积核和1个 $7 \times 7$ 卷积核的参数量.

1个 $7 \times 7$ 卷积核的参数量为:

$$KP_{7 \times 7} = 7^2 C^2 = 49C^2 \quad (1)$$

3个 $3 \times 3$ 卷积核的参数量为:

$$3KP_{3 \times 3} = 3(3^2 C^2) = 27C^2 \quad (2)$$

由公式 (1) 和 (2) 可知, 两者的参数量倍数为:

$$RT = \frac{KP_{7 \times 7}}{3KP_{3 \times 3}} \approx 1.81 \quad (3)$$

(2) 感受野

感受野 (Receptive Field) 计算方法如下:

$$RF_i = \begin{cases} (RF_{i-1} - 1) \times stride + ksize, & i > 0 \\ 1, & i = 0 \end{cases} \quad (4)$$

其中,  $RF_i$  表示第  $i$  层感受野;  $stride$  表示该层卷积步长;  $ksize$  表示卷积核大小. 由公式 (4) 可知, 当  $stride = 1$  时, 3 个  $3 \times 3$  的卷积核感受野与 1 个  $7 \times 7$  卷积核相同.

综上所述, 3 个  $3 \times 3$  的卷积核与 1 个  $7 \times 7$  的卷积核有同样大小的感受野, 而参数量仅是  $7 \times 7$  卷积核的 0.55 倍. 这种级联卷积层的另一个优势在于, 其内部每个卷积层后均会加入一个激活层, 进行一次非线性变换. 与传统卷积层相比, 这种结构增加了网络的非线性函数拟合能力, 对特征的学习能力更强<sup>[15]</sup>.

### 3.2 残差层

随着网络深度的增加, 网络出现无法收敛、性能退化等问题<sup>[16]</sup>. 主要原因在于经典卷积神经网络参数梯度值等于所有前层参数梯度的乘积, 当浅层的参数梯度值过小就容易产生梯度消失的问题, 从而影响到网络的表现性能. 研究显示, 残差结构<sup>[17]</sup>不仅能够解决网络深度深而性能退化的问题, 而且也解决了梯度消失的问题.

下面将对残差层进行分析. 为了不失一般性, 从第  $l$  层到第  $L$  层的前向过程可表示为:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i) \quad (5)$$

其中,  $x_l$  表示第  $l$  个残差结构的输入特征图,  $F(\cdot)$  为残差结构的特征学习函数. 假设网络损失函数为  $E$ , 第  $L$  层的梯度值为  $\partial E / \partial x_L$ , 那么第  $l$  层梯度值为:

$$\frac{\partial E}{\partial x_l} = \frac{\partial E}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial E}{\partial x_L} \left( 1 + \frac{\partial \sum_{i=1}^{L-1} F(x_i)}{\partial x_l} \right) \quad (6)$$

可以看出, 网络参数学习过程实际上等价于对前层面的参数梯度的求和过程, 即从  $\partial E / \partial x_L$  到  $\partial E / \partial x_l$  是线性叠加的过程, 因而允许网络深度增加, 而不会出现梯度消失的问题.

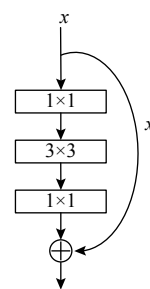


图3 残差模块

传统神经网络学习目标函数是  $H(x)$ , 由于残差模块引入了恒等映射, 使得目标函数转变为残差函数  $F(x) = H(x) - x$ . 如果  $F(x) = 0$ , 那么就转变为恒等映射, 而且恒等映射并未引入额外参数. 如果优化目标是逼近恒等映射, 那么学习扰动特征比重学习输入数据特征要容易. 残差层中有大量的  $1 \times 1$  卷积, 这部分的计算量占残差层将近一半的计算量. 本文对  $1 \times 1$  卷积核进行分组 (group) 卷积, 并对 group 设置在网络识别准确率上的影响进行了实验. 卷积分组方法如图 4 所示.

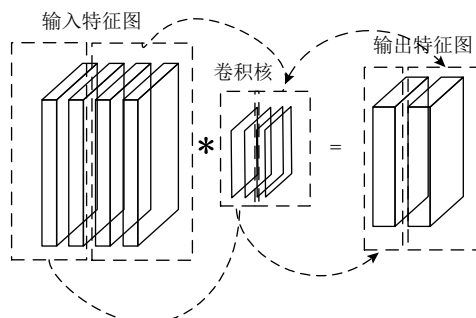


图4 卷积分组

假如输入特征图个数为  $N$ , 该层的卷积核 (kernel) 个数为  $M$ , 计算下一层特征图是利用  $M$  个 kernel 和  $N$  个输入特征图. 引入 group 后, 计算下一层特征图则利用  $M/group$  个 kernel 和  $N/group$  个输入特征图.

### 3.3 分类池化层

传统的卷积神经网络中, 卷积层和池化层等用于特征提取, 随后利用一到多个全连接层将学习到的特征空间隐式地映射到样本标记空间, 最后在输出层由 softmax 激活函数完成目标分类任务. 全连接层的弊端在于参数过多, 容易出现过拟合等情况. 本文使用分类池化层来替代全连接层<sup>[18]</sup> (图 5). 这里分类池化层实际上是全局平均池化层 (图 6), 其没有需要优化的参数, 从而可以避免出现过拟合情况. 同时其对空间信息求和, 因而对空间变换更加健壮.

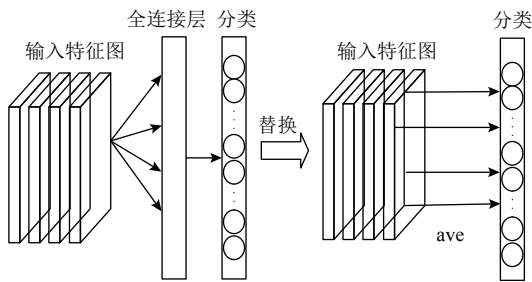


图5 分类池化层

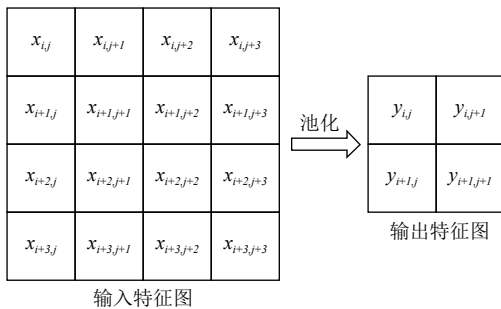


图6 全局平均池化层

图6显示是全局平均池化层,输出特征图的计算方法如下:

$$y_{i,j} = \frac{\sum_{k=0}^M \sum_{t=0}^N x_{i+k,j+t}}{MN} \quad (7)$$

### 3.4 网络结构

本文结构的具体网络参数见表1.其中Conv1\_x和Pool1是级联卷积层,Conv2\_x至Conv5\_x是残差层,Pool2是分类池化层.

## 4 实验与分析

### 4.1 平台介绍

本文在深度学习框架Caffe的基础上,实现了基于卷积神经网络的图像验证码识别方法.

实验平台使用Intel XeonE5-2620 v2 2.1 GHZ CPU, 126 GB 内存, 以及NVIDIA Tesla K40m GPU. 其中Tesla K40m GPU 单精度峰值4.29 Tflops, 显存为12 GB GDDR5, 显存带宽为288 Gbytes/sec.

### 4.2 数据集介绍

本文实验数据以铁路购票网站验证码和正方教务系统验证码为例.

#### (1) 购票网站验证码

购票网站验证码共收集109 900张,包含图形

302类,训练图片数112 599张;包含中文词组302类,训练图片数39 931张.对于验证码识别之前,首先需要对验证码进行预处理操作,分割成单独子图和中文字符或词组.具体购票网站验证码如图7所示.

表1 网络参数

层名	卷积核大小 (stride, padding, group)	输出
Conv1_x	[3×3,64]×3 (1,0,1)	224×224
Pool1	[3×3,64],max (2,2,1)	112×112
Conv2_x	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 128 \end{bmatrix} \times 3$ $\begin{pmatrix} 1, 0, 2 \\ 1, 1, 1 \\ 1, 0, 2 \end{pmatrix} \times 2$ $\begin{pmatrix} 1, 0, 2 \\ 1, 1, 1 \\ 2, 0, 2 \end{pmatrix}$	56×56
Conv3_x	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 256 \end{bmatrix} \times 4$ $\begin{pmatrix} 2, 0, 2 \\ 1, 1, 1 \\ 1, 0, 2 \end{pmatrix} \times 3$ $\begin{pmatrix} 1, 0, 2 \\ 1, 1, 1 \\ 1, 0, 2 \end{pmatrix}$	28×28
Conv4_x	$\begin{bmatrix} 1 \times 1, 192 \\ 3 \times 3, 192 \\ 1 \times 1, 512 \end{bmatrix} \times 6$ $\begin{pmatrix} 2, 0, 2 \\ 1, 1, 1 \\ 1, 0, 2 \end{pmatrix} \times 5$ $\begin{pmatrix} 1, 0, 2 \\ 1, 1, 1 \\ 1, 0, 2 \end{pmatrix}$	14×14
Conv5_x	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 3$ $\begin{pmatrix} 2, 0, 2 \\ 1, 1, 1 \\ 1, 0, 2 \end{pmatrix} \times 2$ $\begin{pmatrix} 1, 0, 2 \\ 1, 1, 1 \\ 1, 0, 2 \end{pmatrix}$	7×7
Conv6	[1×1,cls] (1,0,1)	7×7
Pool2	[7×7,cls],ave (7,0,1)	1×1



图7 购票网站验证码

#### (2) 正方教务系统验证码

正方教务系统验证码由字母和数字构成,共包含33类,训练验证码数5950张.对于验证码识别之前,需要对验证码进行预处理,包含去噪以及分割等步骤.该验证码如图8所示.



图8 正方教务系统验证码

### 4.3 铁路购票验证码实验

铁路购票网站验证码包含图形验证码和中文验证码.

#### (1) 图形验证码



对于图形分割部分, 由于验证码中八张子图尺寸与位置均固定, 故考虑直接设置分割点实现分割. 同时, 对于图片数较少的类别, 采取数据增强方式来扩充训练数据集, 主要包括图像缩放、旋转和颜色变换等.

对于图形识别部分, 图形验证码共 302 类, 训练集图片 82 599 张, 测试集图片 30 000 张. 实验参数如下: 训练 mini-batch 大小设置为 32, 训练最大迭代为 250 000 次. 训练过程中前 100 000 次迭代的学习率为 0.01, 后面每隔 40 000 次学习率衰减 1/10, 梯度更新方法为 Momentum, 值设置为 0.9. 本文测试传统卷积神经网络模型, 包含 AlexNet<sup>[1]</sup>, GoogLeNet<sup>[19]</sup>以及 ResNet50<sup>[17]</sup>, 并与本文方法进行比较, 实验结果如表 2.

表 2 模型在图形验证码上的准确率和模型大小

方法	模型大小 (MB)	top1 准确率 (%)	top5 准确率 (%)
AlexNet	221.66	97.52	99.12
GoogLeNet	48.33	98.32	99.48
ResNet50	92.26	98.70	99.58
本文方法	24.96	98.76	99.55

### (2) 中文验证码

对于字符分割部分, 本文利用垂直投影算法分割出单个词组, 而并非是中文字符. 该方法好处是尽可能避免切分单字带来的误差. 下面是中文词组分割的算法.

#### 算法 1 基于中文词组分割算法

- 1) 从验证码中截取出中文汉字部分 *ChiWords*;
- 2) 对 *ChiWords* 进行灰度化、二值化处理, 得到 *BChiWords*;
- 3) 从 *BChiWords* 中依次统计出每列黑色像素点个数 *BNP*;
- 4) 设置阈值  $T_1, T_2$ . 如果 *BNP* 小于阈值  $T_1$ , 则说明此位置有可能是分割点, 记连续分割点的起止位置  $s$  和  $t$ . 如果  $t - s \geq T_2$  时, 保存  $s$  和  $t$ .
- 5) 根据结果切分中文词组, 并获取下一张验证码, 返回第 1) 步.
- 6) 直至终止条件满足, 结束.

对于字符识别部分, 中文验证码包含中文词组 302 类, 训练集图片数 27 982 张, 测试集图片 11 949 张. 基于字符分割识别实验, 本文利用开源 Tesseract 软件实验和基于 AlexNet 模型的中文识别实验, 实验记为 SplitAlexNet; 基于词组分割识别实验, 主要考虑 AlexNet, GoogLeNet, ResNet50 和本文方法. 实验结果见表 3.

表 3 模型在中文验证码上的准确率和模型大小

方法	模型大小 (MB)	top1 准确率 (%)	top5 准确率 (%)	
基于字符分割识别	Tesseract SplitAlexNet	- 230.1	60.22 75.20	- -
基于词组分割识别	AlexNet GoogLeNet ResNet50 本文方法	221.66 48.33 92.26 21.31	98.92 99.10 99.13 99.14	99.39 99.46 99.42 99.42

### (3) 整体验证码

实验数据为购票网站验证码, 测试验证码 988 张, 与上述实验使用的数据集不同.

整体验证码识别过程: 首先对验证码预处理, 然后对中文词组和子图识别. 中文词组识别结果与词组标注结果进行比对, 如果词组识别无误, 将符合词组的子图识别结果下标与标签文件进行比较 (图 9). 实验结果见表 4.

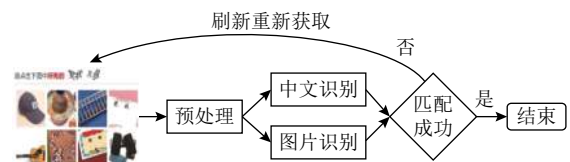


图 9 整体识别过程

表 4 模型在整体验证码上的准确率和模型大小

方法	模型大小 (MB)	准确率 (%)
AlexNet	443.32	66.09
GoogLeNet	96.66	73.07
ResNet50	184.52	77.42
本文方法	46.27	77.93

从表 2 可知, 本文方法参数量最少, 大小为 24.96 MB. 且在测试集的 top1 准确率最高, 比 AlexNet, GoogLeNet, ResNet50 分别高 1.24%, 0.44%, 0.06%.

表 3 的结果显示, 基于字符分割识别的方法, 词组识别准确率较低, 其中较好的模型是 SplitAlexNet, 其 top1 准确率达到 75.20%; 基于中文词组识别的方法, 本文方法准确率达到 99.14%, 高于传统的 AlexNet, GoogLeNet, ResNet50 神经网络模型准确率. 与此同时本文方法参数量最少, 分别比传统神经网络参数量降低 9.40 倍, 23.25 倍, 1.26 倍和 3.32 倍.

从表 4 可知, 本文提出的整体验证码识别准确率要比其它网络模型高, 且模型更小.

### 4.4 正方教务系统验证码实验

该验证码训练集图片 4630 张, 测试集图片 1320 张.

本文测试传统的卷积神经网络模型, 包含 AlexNet, GoogLeNet, ResNet50, 并与本文方法进行对比。

实验参数如下: 训练 mini-batch 大小设置为 32, 训练最大迭代 30 000 次. 训练过程中前 8 000 次学习率为 0.001, 后面每隔 6 000 次学习率衰减 1/10, 梯度更新方法为 Momentum, 值设置为 0.9.

通过表 5 和表 6 可知, 准确率方面, 本文方法的准确率不低于传统神经网络模型; 模型大小方面, 本文方法参数量比 AlexNet, GoogLeNet 和 ResNet50 要少。

表 5 模型在字符上的准确率及模型大小

方法	模型大小 (MB)	top1 准确率 (%)	top5 准确率 (%)
AlexNet	217.45	91.36	99.62
GoogLeNet	39.71	93.50	99.62
ResNet50	90.16	94.31	99.69
本文方法	23.70	95.07	99.69

表 6 模型在正方验证码上的准确率及模型大小

方法	模型大小 (MB)	准确率 (%)
AlexNet	217.45	75.40
GoogLeNet	39.71	80.10
ResNet50	90.16	84.30
本文方法	23.70	87.30

#### 4.5 实验分析

本文重点考虑在不影响模型准确率的前提下, 进一步减少模型的参数量, 即模型规模. 通过上述实验可以看到, 本文网络结构是基准模型中参数量最少的, 同时验证码的识别效果比其它网络要好. 接下来将具体分析原因.

一是模型参数量缩减方面. 本文利用级联卷积层减少了网络参数, 增加了网络的非线性. 同时在残差层调整卷积分组个数, 利用全连接层替换. 这几个方面的改进能够很明显地看出网络规模变小.

本文调整卷积分组的个数, 一方面使得卷积层计算加快, 模型参数量减少, 同时也影响识别的准确率, 下面将实验说明卷积分组对网络结构准确率的影响.

##### (1) 卷积分组实验

卷积分组不仅能够降低网络的规模, 同时卷积分组个数也影响着网络的识别率. 本文将对分组个数进行实验, 考虑在不同的分组下, 本文方法的准确率变化情况.

为了说明卷积分组对实验准确率的影响, 将在图形验证码和中文验证码上进行实验分析. 通过表 7 和表 8 容易看出, 对于不同数据集最优的 group 选择也不同.

表 7 卷积分组对图形验证码准确率影响

Group	1	2	4	8
acc	98.45	98.76	98.23	98.17

表 8 卷积分组对中文验证码准确率影响

Group	1	2	4	8
acc	99.04	99.12	99.14	98.15

##### (2) 特征图可视化

通过特征图可视化能够更好地理解网络的特征学习过程, 下图是利用本文方法识别的最后层可视化结果.

图 10 显示的是本文方法在最后层特征图的激活程度. 可以看出无论是中文词组还是图形验证码, 相同类别的图片在最后层的激活程度相似, 而不同类别之间的激活程度却迥然不同.

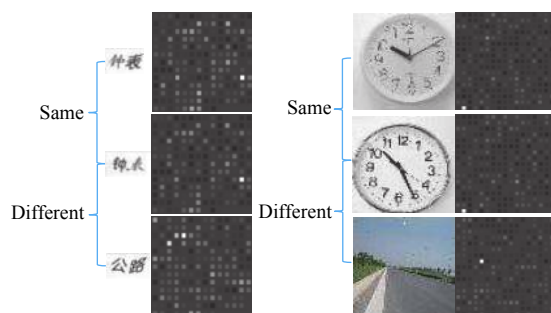


图 10 最后层特征图可视化结果

## 5 总结

本文利用卷积层级联、残差学习、全局池化、卷积分组等方法, 提出了一种基于卷积神经网络的验证码识别方法. 实验表明该方法较传统网络模型具有参数量少的特点, 而且有着更高的识别准确率. 此外, 本文还讨论了卷积分组对准确率的影响, 并通过实验选出性能最优的卷积分组. 接下来的工作将继续优化网络模型结构, 同时也将根据本文的实验结果分析如何提升验证码的安全性.

### 参考文献

- 1 Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- 2 Zhang JS, Wang XF. Breaking Internet banking CAPTCHA based on instance learning. Proceedings of 2010 IEEE International Symposium on Computational Intelligence and

- Design. Hangzhou, China. 2010. 39–43.
- 3 李兴国, 高炜. 基于滴水算法的验证码中粘连字符分割方法. 计算机工程与应用, 2014, 50(1): 163–166. [doi: [10.3778/j.issn.1002-8331.1208-0310](https://doi.org/10.3778/j.issn.1002-8331.1208-0310)]
  - 4 Lu PP, Shan L, Li J, *et al.* A new segmentation method for connected characters in CAPTCHA. Proceedings of 2015 International Conference on Automation and Information Sciences. Changshu, China. 2015. 128–131.
  - 5 Yan J, El Ahmad AS. A low-cost attack on a Microsoft CAPTCHA. Proceedings of the 15th ACM Conference on Computer and Communications Security. Alexandria, VA, USA. 2008. 543–554.
  - 6 Mori G, Malik J. Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA. Proceedings of 2003 IEEE Conference on Computer Vision and Pattern Recognition. Madison, WI, USA. 2003, 1. 134–141.
  - 7 LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791)]
  - 8 Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Computation, 2006, 18(7): 1527–1554. [doi: [10.1162/neco.2006.18.7.1527](https://doi.org/10.1162/neco.2006.18.7.1527)]
  - 9 Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA. 2012. 3642–3649.
  - 10 Zhong ZY, Jin LW, Xie ZC. High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps. Proceedings of 2015 13th International Conference on Document Analysis and Recognition. Tunis, Tunisia. 2015. 846–850.
  - 11 范望, 韩俊刚, 苟凡, 等. 卷积神经网络识别汉字验证码. 计算机工程与应用, 2018, 54(3): 160–165.
  - 12 Goodfellow IJ, Bulatov Y, Ibarz J, *et al.* Multi-digit number recognition from street view imagery using deep convolutional neural networks. arXiv: 1312.6082, 2013.
  - 13 Matan O, Burges CJC, LeCun Y, *et al.* Multi-digit recognition using a space displacement neural network. Moody JM, Hanson SJ, Lippman RP. Neural Information Processing Systems. San Mateo, CA: Morgan Kaufmann Publishers, 1992: 488–495.
  - 14 Shi BG, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2298–2304. [doi: [10.1109/TPAMI.2016.2646371](https://doi.org/10.1109/TPAMI.2016.2646371)]
  - 15 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
  - 16 He KM, Sun J. Convolutional neural networks at constrained time cost. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 5353–5360.
  - 17 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
  - 18 Lin M, Chen Q, Yan SC. Network in network. arXiv: 1312.4400, 2013.
  - 19 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.