

# 基于多特征卷积神经网络的手写公式符号识别

方定邦, 冯桂\*, 曹海燕, 杨恒杰, 韩雪, 易银城

华侨大学信息科学与工程学院厦门市移动多媒体通信重点实验室, 福建 厦门 361021

**摘要** 提出了基于多特征稠密卷积神经网络的模型框架(DenseNet-SE)。与传统方法相比,DenseNet-SE 采用数据驱动的方法,无需手工提取特征。该框架包含了稠密残差块的结构,能够获取深度特征。通过跳跃连接的方式,从浅层获取细粒度特征来辅助深度特征。同时,融合特征有助于网络结构获取更多全局信息,更好地表示公式符号的类别。利用在线手写数学表达式识别的竞赛组织(CROHME)提供的标准数学公式符号库来验证所提算法,结果表明,CROHME2014 和 CROHME2016 的识别率分别达到 93.38%和 92.93%,高于目前已有算法的识别率。

**关键词** 光计算; 稠密卷积神经网络; 手写公式符号; 稠密残差块; 深度特征; 细粒度特征

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.072001

## Handwritten Formula Symbol Recognition Based on Multi-Feature Convolutional Neural Network

Fang Dingbang, Feng Gui\*, Cao Haiyan, Yang Hengjie, Han Xue, Yi Yincheng

*Xiamen Key Laboratory of Mobile Multimedia Communications, College of Information Science and Engineering, Huaqiao University, Xiamen, Fujian 361021, China*

**Abstract** A model framework called DenseNet-SE is proposed based on a multi-featured dense convolutional neural network. Compared with the conventional methods, the DenseNet-SE adopts the data-driven approach and the manual extraction of features is not necessary. It contains the dense residual blocks so that the deep features can be acquired. In the jump-joining way, the fine-grained features are obtained from the shallow layers to assist the deep features. The fused features can help the network structure obtain more global information and better represent the categories of formula symbols. The standard mathematical formula symbol library provided by the competition organization on recognition of online handwritten mathematical expression (CROHME) is used to verify the proposed algorithm, results show that the recognition rates of CROHME2014 and CROHME2016 are 93.38% and 92.93%, respectively, higher than those of the existing algorithms.

**Key words** optics in computing; dense convolutional neural network; handwritten formula symbols; dense residual blocks; deep features; fine-grained features

**OCIS codes** 200.4560; 100.4996; 100.4995; 100.4999

## 1 引言

手写文字识别是图像处理与模式识别重要的研究领域,随着电子设备手写输入方式的广泛应用,手写识别技术受到了学术界和工业界的关注<sup>[1-2]</sup>。手写识别分为在线和离线两种形式。对于在线方式,图像数据中含有大量的轨迹信息,包括笔划的顺序以及曲率;而离线方式通过扫描仪和摄像头采集手

写图像。随着触摸屏和电子笔等移动终端设备的出现,大量的在线手写体涌现,在线数据通常与手写轨迹的坐标有关。手写数学公式符号识别是手写识别研究的重要部分,在一定程度上促进了手写识别的发展。在线手写数学表达式识别的竞赛组织(CROHME)提供的数据集以在线数据方式存在,本文将在线数据转化为离线的图像,使用离线的方法对手写公式符号进行分类。孤立的手写公式符号

收稿日期: 2018-09-26; 修回日期: 2018-10-12; 录用日期: 2018-10-31

基金项目: 福建省自然科学基金(2016J01306)、华侨大学研究生科研创新能力培育计划(17014082020)

\* E-mail: fengg@hqu.edu.cn

识别具有以下特点:1) 典型的 Mnist 数据集, 仅有 10 种类别, 而数学公式符号的种类多达上百种, 包括英文字符、数字, 以及一些特殊字符如“ $\Sigma$ ”、“ $\Pi$ ”、“ $\rightarrow$ ”; 2) 存在大量的手写风格, 导致同一类的符号差异很大; 3) 数据集中含有大量相似的符号, 如“COMMA”和“PERIOD”、大写字母“S”和小写字母“s”等。因此, 手写数学公式符号识别的研究存在很大的挑战性。

本文对最近几年手写数学公式符号识别算法进行了调研。手写公式符号识别的关键在于特征提取, Álvaro 等<sup>[3]</sup> 从在线符号的轨迹中获取 7 个特征, 结合了离线图像的剑桥大学模式识别和人类语言组提出的技术 (PRHLT)、提取几何特征的技术 (FKI) 以及 polar 特征, 利用这些特征并通过双向循环神经网络 (RNN-BLSTM) 对符号进行分类, 识别率为 91.24%。从在线数据中, Davila 等<sup>[4]</sup> 提取轨迹方向和梯度特征, 以及离线图像的特征, 包括水平的交叉特征、二维模糊直方图和方向模糊直方图特征, 并将这些特征进行组合, 利用传统的机器学习算法, 如随机森林和支持向量机 (SVM) 对公式符号进行分类, 识别率为 89.87%。Nguyen 等<sup>[5]</sup> 通过在线和离线两种分类器线性组合的方式来识别公式符号, 在线分类器采取 RNN-BLSTM, 而离线的分类器采用 CCN 对离线的公式图像进行分类, 将线性组合的分类器按照各占 1/2 的权重比例来分配, 获得最终的准确率为 91.08%。CROHME2016 的获胜者 Mouchère 等<sup>[6]</sup>, 通过在线的 inkml 文件获取特征信息, 这种特征信息包括曲率、轨迹的方向和直方图特征, 将这些特征送入多层感知机 (MLP) 对公式符号进行分类, 获得目前最高的符号识别率为 92.81%。

目前, 少数方法采用离线的特征对手写公式符号进行分类, 如 Dong 等<sup>[7]</sup> 提出 HMS-VGGN-et, 采用小尺寸  $3 \times 3$  和  $1 \times 1$  的卷积内核, 特点在于  $1 \times 1$  的卷积核不仅减少了特征图的数量, 还能保持特征图的大小不变。使用全局均值池化来代替模型中的全连接层, 可以减少网络中参数量过大的问题。在 CROHME2014 数据集中获得 top-1 的识别率为 91.82%; 在 CROHME2016 数据集中, 识别率仅次于 Mouchère 等<sup>[6]</sup> 提出的算法, 达到 91.77%。Ramadhan 等<sup>[8]</sup> 将在线的数据集转化为离线的数据集, 采用常规的 CNN 模型获取离线的图像特征, 馈送到 Softmax 层中对图像进行分类, 识别率为 87.72%, 明显低于结合在线特征的方法。通过以上调研, 大多数研究者采用了在线和离线特征的方法

对公式图像进行分类, 少数研究者使用离线的方法对手写公式符号进行分类。因此本文更加关注离线特征对公式符号进行分类的性能。

卷积神经网络框架可以运用到非常复杂的计算机视觉领域中, 如图像检测<sup>[9]</sup>、图像语义分割<sup>[10]</sup> 和图像超清分辨率的重建<sup>[11]</sup>。其中卷积神经网络是深度学习中的重要组成部分, 早期 LeCun 等<sup>[12]</sup> 提出 LeNet-5 结构的卷积神经网络, 用于识别单个手写数字, 并且这种网络结构广泛地应用于图像分类任务中, 取得了很好的效果。为了追求更高的识别精度, 通常需要更深的网络结构, 但网络深度的增加会导致梯度消失等问题。为了解决这种问题, Hinton 等<sup>[13]</sup> 将正则化项加在损失函数上, 并引入 dropout 层, 这种惩罚机制弱化了梯度消失的问题, 但并不能达到更好的效果。近些年来, He 等<sup>[14]</sup> 和 Huang 等<sup>[15]</sup> 分别提出了残差卷积神经网络 (ResNet) 和稠密卷积神经网络 (DenseNet), 这两种网络多达 100 层, 优势在于训练的过程中随机丢弃一些冗余的神经元, 从而达到更深网络拓扑结构, 而 DenseNet 网络结构更加紧凑, 所需的网络参数更少, 提高了网络运行的效率。本文在 DenseNet 基础上进行了扩展, 提出了一种新型的网络构架, 称作 DenseNet-SE。该网络能够获得深度特征和细粒度特征, 将两种特征进行融合, 能够更好地描述公式符号的每一个类别。

## 2 离线数据处理模型

### 2.1 离线图像的生成

CROHME 库中提供了大量 inkml 格式的文件, inkml 属于一种标记语言<sup>[16]</sup>, 这种标记语言通过电子笔或触摸屏在终端设备留下用户手写轨迹信息。普遍用于绘制公式、流程图、音乐符号以及公式符号, 兼容于任何平台。inkml 的基本元素包含了符号所有的轨迹, 这些轨迹是在交互界面留下的  $x$  和  $y$  坐标构成。假设符号由一系列的轨迹  $t_i (i=1, 2, \dots, n)$  构成, 即  $T=[t_1, t_2, t_3, \dots, t_n]$ , 共有  $n$  组轨迹, 轨迹中包含大量的坐标点  $t_i=[f_{i1}, f_{i2}, f_{i3}, \dots, f_{iU}]$ , 第  $i$  组轨迹由  $U$  组坐标点构成, 即  $j=1, 2, 3, \dots, U, j$  为其中的任意一对坐标, 边界框的大小为  $S$ 。连接相邻两点的坐标, 直到所有轨迹的坐标点完成, 形成符号图像。简要描述在线数据转离线图像算法如下:

1) 确定轨迹的坐标点中垂直坐标和水平坐标的最大值和最小值, 即  $X_{\min}, Y_{\min}, X_{\max}, Y_{\max}$ 。

2) 确定容纳所有轨迹点的维度,即长( $h$ )和宽( $w$ ),轨迹的长宽比  $T_r$ ,三者表达式分别为  $h = Y_{\max} - Y_{\min}$ ,  $w = X_{\max} - X_{\min}$ ,  $T_r = w/h$ 。

3) 对于轨迹  $T$  的所有的坐标点  $f_{ij} = (x_{ij}, y_{ij})$ ,通过  $x'_{ij} = x_{ij} - x_{\min}$ ,  $y'_{ij} = y_{ij} - y_{\min}$  得到一组新的坐标点  $f'_{ij} = (x'_{ij}, y'_{ij})$ ,其中  $i = 1, 2, \dots, n, j = 1, 2, 3, \dots, U$ 。

4) 确定尺度因子  $F$ ,如果  $T_r < S$ ,则  $F = (S - 1)/w$ ,否则,  $F = (S - 1)/h$ 。

5) 调整所有的坐标点。通过尺度因子  $F$  使得所有的点在指定的边界框内,获得一组新的轨迹  $t'_i$  ( $i = 1, 2, \dots, n$ ),即  $T' = [t'_1, t'_2, t'_3, \dots, t'_n]$ 。

6) 遍历所有的新的轨迹  $T'$ ,连接相邻两点的坐标形成线段,直到所有的坐标点完成,形成手写符号图像。

由于移动终端的尺寸大小不一致,导致图片的规格不同,需要统一每一张图片的规格。在实验中,规定每一张图片尺寸大小为  $48 \times 48$ ,并对每一张图片的灰度值进行归一化处理,最后进行高斯滤波去噪。

### 2.2 数据扩增

符号的种类繁多,而每一类符号的数量又极其不均匀,造成了公式符号整体识别率难以提升。图1为101类符号数量分布图,其中符号“一”的数量为7940;而一些特殊的字符,如“日”的数量仅为4;另外,大部分符号种类的数量较少,导致符号的特征很难在深度网络模型中学习。为此,对数据集进行扩增,保持所有种类的符号数量处于均衡的状态,减小类别间符号数量的差异性。

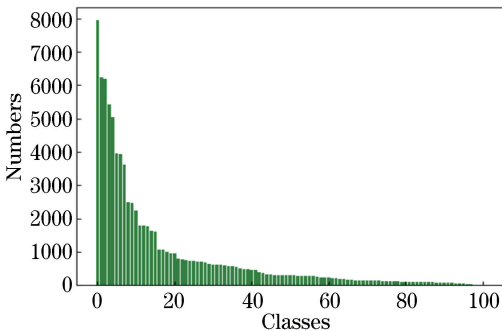


图1 101种符号类别分布图

Fig. 1 Class distribution of 101 symbols

灵活的失真模型<sup>[17]</sup>与图像增强中的仿射变化有关,引入了一种类似于控制手部肌肉弹性不变的因子  $\alpha$ ,文献[4]对在线数据进行扩增,而本文对离线的图像进行扩增。这种模型的优点是能够保留原

始图像的位移不变性。 $\nabla_x(x, y) = (-1, 1)$ ,  $\nabla_y(x, y) = (-1, 1)$ 分别表示像素的坐标点  $(x, y)$  经过失真模型后,在水平和垂直方向随机移动的位置,并采用双线性插值方法对其进行校正。其中  $\nabla_x$  和  $\nabla_y$  表示图像中所有的像素经过标准差为  $\sigma$  的高斯核卷积后的结果。经过实验检验,参数的设置为  $\sigma = 8$ 。如果标准差  $\sigma > 8$ ,经过位移场变化后的图像会趋于严重扭曲,失去了图像原有的特性,若设置过小,则产生的图像没有任何区分性,弹性因子  $\alpha$  主要控制位移后变化的强度,一般设置为  $36.0 \sim 38.0$ 。从图1可以看出,大部分符号类型的数量小于3000,因此,将符号类别的数量均衡在3000,对于大于3000类别的符号,保持原来的数量不变,经过弹性失真模型后的图像如图2所示。

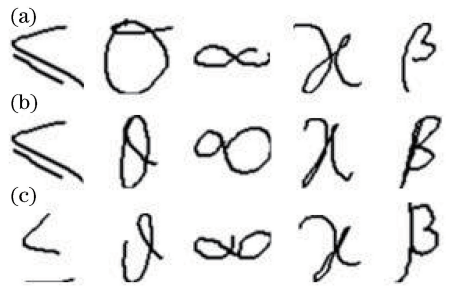


图2 原始图像通过弹性失真模型后随机生成的图像。

(a)原始图像;(b)第一次随机生成的图像;(c)第二次随机生成的图像

Fig. 2 Images randomly generated after original images passing through elastic distortion model. (a) Original images; (b) images randomly generated for first time; (c) images randomly generated for second time

### 3 网络模型结构

DenseNet-SE 网络主要由 DenseNet 中的密集块和 SeNet 网络模块所构成,这两个模块构建出的新模块,即稠密残差块,通过稠密残差块重新组建新的网络框架。DenseNet-SE 网络框架如图3所示。

#### 3.1 DenseNet 的 Dense 模块

DenseNet<sup>[15]</sup>采用级联的连接方式,对于一般的  $L$  层卷积神经网络,共有  $L$  层连接,而 DenseNet 有  $L(L + 1)/2$  层连接。这种稠密连接的方式,充分利用了每一层的网络参数,实现了参数共享,提高了参数的利用率,促进了梯度的传播。将所有层的特征输出作为当前层的输入,即  $x_1, x_2, x_3, \dots, x_{L-1}$  为第  $1 \sim L - 1$  层的特征图,以级联的方式经过第  $L$  层,通过复合函数  $H_L(*)$  得到输出  $x_L$ ,公式为

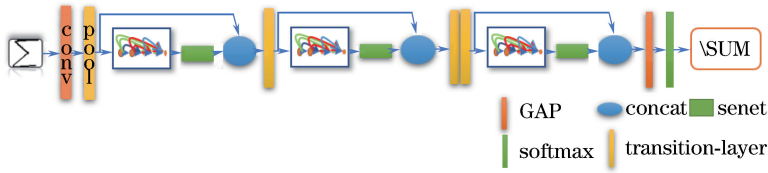


图3 DenseNet-SE网络结构图

Fig. 3 Structural diagram of DenseNet-SE network

$$x_L = H_L(x_1, x_2, x_3, \dots, x_{L-1}), \quad (1)$$

式中  $H_L(*)$  定义为依次经过  $3 \times 3$  的卷积、池化、批量归一化(BN)以及非线性激活函数(ReLu) 4种运算的复合函数,将这样的网络构建称为 Dense 块。

### 3.2 SeNet 网络模块

SeNet 是一种轻量级的模块<sup>[18]</sup>,增强了符号特征的表示能力,通过对通道和门控制单元共同调节资源的分配,提升有用的特征同时抑制与形状无关的特征信息。SeNet 网络由挤压层和抽取层组成。

由于下采样层的操作,只能获取局部感受野,而无法获取全局信息,同时破坏了像素点之间的位置关系,并降低了特征图的分辨率。为了解决这种问题,在挤压层使用全局均值池化(GAP),相关运算步骤描述如下:输入的图像经过若干层卷积之后,将所有特征层堆叠在一起,记为  $R = \{r_1, r_2, r_3, \dots, r_Z\}$ ,其中  $r_s \in \mathbf{R}^{a \times c}$ ,表示  $a \times c$  大小的特征图,其中  $s$  表示其中任意一个特征图, $s = 1, 2, \dots, Z$ ,所有的特征层经过全局均值池化后,记为  $z = \{z_1, z_2, z_3, \dots, z_b\}$ ,其中  $z_l \in \mathbf{R}$ , $l = 1, 2, \dots, b$ , $l$  为池化后的任意特征值, $H$  和  $W$  分别表示经过卷积后特征图的高度和宽度, $u, v$  为依次遍历图像的每一个位置。可以将均值池化看作是对所有的局部信息统计求和,公式为

$$r_s = \frac{1}{H \times W} \sum_{u=1}^H \sum_{v=1}^W x_s(u, v). \quad (2)$$

抽取层主要对局部特征层之间的相互关联度和通道之间的独立性产生作用。两层带有非线性激活函数的全连接层(FC),第一层采用非线性激活函数,激活单元的数量为  $C/r$ ,其中  $r$  表示衰减因子( $r = 2.5$ ),其目的是减少计算的复杂性;第二层采用 Sigmoid 非线性激活函数,激活单元数量为  $C$ ,输出的激活响应值归一化到 0 和 1 之间。将挤压层和抽取层通过堆叠的方式构建成一个残差块,经过若干卷积输出的特征层作用到残差块中得到相应的响应值  $\hat{X}$ ,整个运算过程可以表示为

$$\hat{X} = \text{Sigmoid}\{\text{FC}_1\{\text{Relu}\{\text{FC}_2[\text{GAP}(X)]\}\}\}, \quad (3)$$

式中  $\text{FC}_1$  和  $\text{FC}_2$  分别表示第一层、第二层全连接层。为了方便表示,可以描述为  $\hat{X} = \lambda X$ , $\lambda$  为压缩因子,通过网络结构的反向传播算法,自适应调整  $\lambda$  来适合整个框架。

### 3.3 DenseNet-SE 网络结构

受到 Dense 和 SeNet 模块结构的启发,结合这两种网络各自的特点,本文提出了新型结构模块,称作稠密残差块,为了使这两模块更好地衔接,在中间引入  $1 \times 1$  的卷积和批量归一化层。采用线性激活函数对特征层进行处理,保持原始特征层的特性。其优点在于, $1 \times 1$  的卷积减少了通道的数量,解决了网络瓶颈的问题;通过批量归一化层<sup>[19]</sup>,将上一层输出的所有激活值均匀地分布在每一层网络中,促进模型的训练。每层卷积中有  $d$  个维度的激活值,记作  $M = (m^{(1)}, m^{(2)}, \dots, m^{(d)})$ ,并对激活层中的每一个维度进行归一化处理,公式描述为

$$\hat{m}^{(k)} = \frac{m^{(k)} - E[m^{(k)}]}{\sqrt{V[m^{(k)}] + \epsilon}}, \quad (4)$$

式中: $m^{(k)}$  为  $M$  中的第  $k$  个任意激活值; $E[m^{(k)}]$  和  $V[m^{(k)}]$  分别表示  $m^{(k)}$  的均值和方差; $\epsilon$  为确保数值稳定的恒量; $\hat{m}^{(k)}$  为归一化后的激活值。但是进行归一化处理后的激活层可能会影响前面一层的特征表示。因此,引入两个超参数  $\gamma$  和  $\vartheta$  来恢复原始的特征信息,得到新的响应值  $y^{(l)}$ ,其中  $l$  为任意的一层, $l = 1, 2, \dots, d$ 。响应值  $y^{(l)}$  的公式为

$$y^{(l)} = \gamma^{(l)} \hat{m}^{(k)} + \vartheta^{(l)}. \quad (5)$$

在稠密残差块的最后,将细粒度特征和经过自适应调整的深度特征通过 Concat 方式并接在一起,图 4 为稠密残差块的结构图。在稠密残差块之间加入过渡层,包括  $1 \times 1$  的 Conv 以及  $2 \times 2$  的均值池化。将若干个稠密残差密集块和过渡层之间共同组成 DenseNet-SE 网络。

在 DenseNet-SE 的最后一层 Softmax 之前,引入局部归一化层(LRN)<sup>[20]</sup>,即在神经元之间创建竞争机制,对于较大输出值产生更大的响应,抑制较小响应的输出,有效地提高了符号图像分类的精度,其

中 LRN 公式描述为

$$b_{x,y}^p = \frac{a_{x,y}^p}{\left\{ d + \alpha \sum_{z=\max(0,p-q/2)}^{\min(N-1,p+q/2)} [a_{x,y}^z]^2 \right\}^\beta}, \quad (6)$$

式中  $a_{x,y}^p$  表示坐标  $(x,y)$  位置上的神经元  $p$  的激活值, 参数设置为  $d=2, \alpha=10^{-1}, \beta=0.75, N=128, q=5$ 。在网络的最后采用全局均值池化来代替全连接层, 对公式符号进行分类。在一定程度上降低了模型参数的计算量。

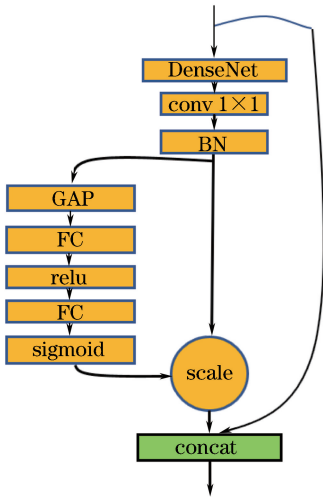


图 4 稠密残差块模块构造图

Fig. 4 Structural diagram of residual-dense block module

从网络的构建可以看出, DenseNet-SE 具备这两种网络各自的优势。在连接方式上, 包含了跳跃连接的方式和更深的网络结构, 跳跃连接的方式能够捕获浅层的细粒度特征, 而深层的网络获取深度

特征, 细粒度特征辅助深度特征产生融合的特征对公式符号进行分类。在 DenseNet-SE 的配置中, 首先, 输入的图像经过 32 个大小为  $7 \times 7$ 、步长为 2 的 Conv 操作, 再经过  $3 \times 3$  的最大池化, 然后反馈到三个稠密残差块中, 其网络层数分别为 12、16、32。在第一个和第二个稠密残差块之间引入过渡层, 在最后一个稠密残差块中引入两层过渡层, 以减少深层网络中特征的冗余性。其中三个 SeNet 模块滤波器的数量分别设置为 32、64、128, 采用递进的数量。原因在于, 如果网络的浅层滤波器的数量设置过多, 会造成浅层网络的参数量过大, 训练的效果不佳。在模型判别过程中采用 LRN 和 GAP 来获取更高的判别分数, 提高网络的稳健性。

## 4 实验结果及分析

### 4.1 公式库简介

CROHME 提供了标准数学公式库, CROHME 是目前最大公共的手写数学公式库之一, 该主办方已经成功主办了 5 年, CROHME2011 到 CROHME2016<sup>[6,21-22]</sup>, CROHME 竞赛分为 4 个任务, 其中孤立手写公式符号识别是其任务之一。数据库中的符号种类繁多, 结构复杂, 共有 101 种, 而且数据集的数量会随年份的增加而增加, 挑战手写公式符号识别的难度也因此而增大。本文采用目前最新的数据集 CROHME2016, 训练集中共有 85802 张公式符号, 公式符号中的所有标签均以 Latex 作为标记。为了体现模型框架的泛化性能, 使用 CROHME2014 和 CROHME2016 测试集来评估模型; 在验证算法上, 将 CROHME2013 的测试集作为验证集来评估模型框架的性能, 实验数据分布如表 1 所示。

表 1 CROHME 实验数据集的分布

Table 1 Distribution of CROHME experimental datasets

Dividing dataset	Dataset category	Image size / (cm×cm)	Scale	
			Previous quantity	Twisted quantity
Train	CROHME2016 train	48×48	85802	321301
Validation	CROHME2013 test	48×48	6082	—
Test	CROHME2016 test	48×48	10019	—
	CROHME2014 test	48×48	10061	—

实验所有的环节在 ubuntu16.04 系统上操作, 使用 Tensorflow 深度学习工具包运行框架, 硬件配置 12 线程 Corei7-8700CPU@3.2 GHz GTX1060, 显存为 6 G。初始的学习率为 0.001, 每经过 5000 步学习率衰减 5%, 批量的大小为 100, 采用的优化器是 Adam 算法<sup>[23]</sup>, 迭代次数为 50。

### 4.2 实验结果分析

为了验证 DenseNet-SE 的性能, 评估了

DenseNet 模型构架和 DenseNet-SE 模型构架, 并对这两种框架进行对比分析。前提条件是 DenseNet-SE 和 DenseNet 在相同的环境下运行, 参数设置: growth-rate 为 16, epoch 为 50, DenseNet 和 DenseNet-SE 均采用了三个密集块。使用相同的数据库来检验两种框架的有效性, 图 5 所示为 DenseNet-SE 和 DenseNet 在 CROHME2013 验证集的准确率, 从两条折线的走向和趋势来看, 在 10

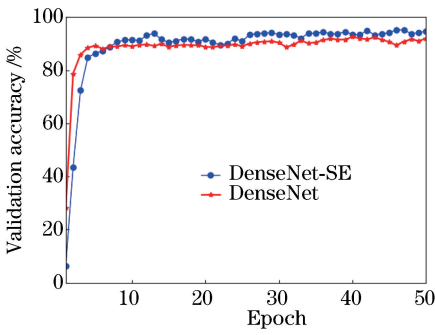


图5 DenseNet 和 DenseNet-SE 准确率的比较

Fig. 5 Validation accuracy comparison of DenseNet and DenseNet-SE

个 epoch 之前, DenseNet 准确率的增长速度大于 DenseNet, 因为稠密残差块模块叠加在 DenseNet 网络上, 从而在一定程度上增加了原始模型的复杂度, 因此模型训练的前期变得更加缓慢; 在 10 个 epoch 之后, 随着迭代次数的增加, 两条曲线开始趋于收敛的状态, 而且 DenseNet-SE 整体准确率高与 DenseNet。此外, 还测试了两个框架在 CROHME2016 测试集的准确率, DenseNet-SE 的准确率为 92.93%, 比 DenseNet 的准确率高 3%, 这表明引入稠密残差块可以提高模型的整体性能。图 6 所示为 DenseNet-SE 在 CROHME2016 的训练集和验证集的准确率, 两条折线基本趋于吻合的状态, 说明了 DenseNet-SE 在训练过程中高效地利用了滤波器参数设置, 提高了网络的稳健性, 避免了训练过程中出现过拟合的问题。为了验证稠密残差块是轻量型的模块和网络的复杂度, 测试了在一个 epoch 内, DenseNet-SE 和 DenseNet 模型在训练集中所消耗的时间和批量数据所需的时间以及验证集趋于稳定的准确率, 如表 2 所示。虽然本文算法的运行时间比 DenseNet 有所增加, 但最终的准确率有着明显的提高。由于受到计算资源的限制[图形

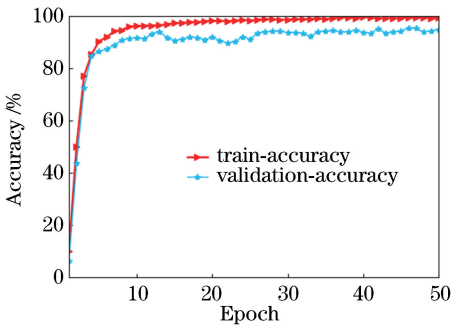


图6 DenseNet-SE 测试集和验证集的准确率比较

Fig. 6 Accuracy comparison of DenseNet-SE test set and validation set

处理器(GPU)的个数和性能], 运行 50 个 epoch, 共耗时 4 h。

表2 测试一个 epoch 所需的时间和准确率

Table 2 Time consumption and accuracy for each epoch test

Model	Train		Validation
	time /s	batch /s	accuracy /%
DenseNet	307	0.112	91.08
DenseNet-SE	406	0.123	95.31

### 4.3 本文算法与目前主流算法比较

将目前在手写数学公式符号识别上已经提出的算法与本文算法进行分析和比较, 结果如表 3 所示。可以看出, 本文算法在数据集 CROHME2014 和 CROHME2016 中, 均实现目前最高的识别率, 分别为 93.83% 和 92.93%。CROHME2014 数据集的准确率略高于 CROHME2016, 这是因为 CROHME2016 的数据集中存在一些空白的符号图像, 在收集数据时, 没有将轨迹信息保存到 inkml 文件中。本文算法和文献[7]、[8]均采用离线的特征对公式符号进行分类, 并分别选择 CNN 和 VGG 这两种卷积神经网络。在模型设计的方面, DenseNet-SE 能够达到更深的网络结构, 同时有效地避免过拟合的问题, 获取了更加稳健和可区分的符号特征, 识别率高于原系统, 表明了所提框架优于其他两种框架。并且, 文献[3, 4, 6]同时采用了在线特征和离线特征对手写公式符号进行分类, 而离线特征获取系统采用手工提取特征且需要调节参数, 无法保证所提特征适合网络的训练。DenseNet-SE 能够自动获取图像中的特征, 通过反向传播算法, 自动调节网络的参数, 达到模型的最优化, 捕获较强判别符号的离线特征。综上, 从主观识别率的角度以及客观的因素(时效性、框架的选取、特征提取)评价了所提系统, 验证了所提框架的实效性。

表3 所提方法与不同种类系统的比较

Table 3 Comparison between proposed method and different types of systems

System	CROHME	CROHME	Feature used
	2014 test accuracy /%	2016 test accuracy /%	
Ref. [6]	91.04	92.81	Online+offline
Ref. [5]	91.28	92.27	Online+offline
Ref. [3]	91.24	—	Online+offline
Ref. [4]	88.66	88.85	Online+offline
Ref. [8]	87.72	—	Offline
Ref. [7]	91.82	92.42	Offline
Proposed	93.38	92.93	Offline

#### 4.4 错误样本分析

DenseNet-SE 能够有效地提取符号形状特征,在很大程度上提升了识别率。但依旧存在一些错误判别的样本,即在识别符号的过程中,错误判别了其他符号的类型。为此,对所有符号的类别进行了测试,列出了 CROHME2016 测试集中排行前 10 的符号错误类型,如表 4 所示。在测试集中测试了一些常见容易混淆的符号,主要在于手写风格上具有易混淆的地方,如“X-x- $\times$ ”、“| - ) - }”、“S-s-5”、“0-O”。并在测试过程中对误判类型的符号进行了检验,如图 7 所示,取得了很好的识别率。

表 4 CROHME2016 中排行前 10 的符号错误判别类型

Table 4 Symbols of TOP-10 error discrimination types in CROHME2016

No.	Symbol label	Total symbols	Percentage of number of misclassified symbols / %
1	o	11	100
2	prime	11	100
3	C	31	96.77
4	times	72	88.89
5	Y	13	76.92
6	COMMA	82	76.83
7	s	21	71.43
8	.	21	71.43
9	tn	3	66.67
10	r	40	65.00

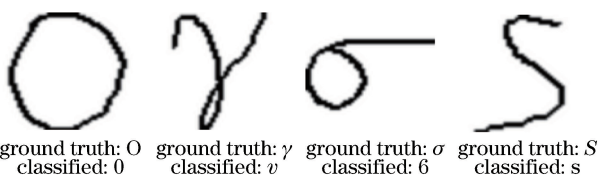


图 7 CROHME2016 测试集中误判类型的符号

Fig. 7 Symbols of misjudgment types in CROHME2016 test set

## 5 结 论

针对于手写公式符号识别率提升的问题,提出了基于脱机下的多特征稠密卷积神经网络手写公式符号识别的算法,采用主流的网络框架 DenseNet 对其进行扩展,引入了 SeNet 这种轻量级的模块叠加在 Dense 模块上,构架出稠密残差块新的网络模块,该模块具备了两种模块的特点,能够避免了网络随深度的增加而出现拟合的问题,同时能够捕获深层次特征,抑制与符号形状无关的特征,通过

Concat 层将细粒度特征和深度特征进行融合,获得一种更加稳健的特征来对公式符号进行分类。与目前主流的手写公式符号识别的算法相比,本文算法在 CROHME2014 和 CROHME2016 数据集中均取得了较好的识别率。

## 参 考 文 献

- [1] Blostein D, Grbavec A. Recognition of mathematical notation [M]. Singapore: World Scientific, 1997: 557-582.
  - [2] Chan K F, Yeung D Y. Mathematical expression recognition: a survey [J]. International Journal on Document Analysis and Reconition, 2000, 3(1): 3-15.
  - [3] Álvaro F, Sánchez J A, Benedí J M. Offline features for classifying handwritten math symbols with recurrent neural networks [C] // International Conference on Pattern Recognition, 2014: 2944-2949.
  - [4] Davila K, Ludi S, Zanibbi R. Using off-line features and synthetic data for on-line handwritten math symbol recognition [C] // International Conference on Frontiers in Handwriting Recognition, 2014: 323-328.
  - [5] Nguyen H D, Duc le A, Nakagawa M. Recognition of online handwritten math symbols using deep neural networks [J]. IEICE Transactions on Information and Systems, 2016, E99.D(12): 3110-3118.
  - [6] Mouchère H, Viard-Gaudin C, Zanibbi R, et al. ICFHR2016 CROHME: competition on recognition of online handwritten mathematical expressions [C] // International Conference on Frontiers in Handwriting Recognition, 2016: 607-612.
  - [7] Dong L F, Liu H C. Recognition of offline handwritten mathematical symbols using convolutional neural networks [M]. Cham: Springer International Publishing, 2017: 149-161.
  - [8] Ramadhan I, Purnama B, Faraby S A. Convolutional neural networks applied to handwritten mathematical symbols classification [C] // International Conference on Information and Communication Technology, 2016: 1-4.
  - [9] Feng X Y, Mei W, Hu D S. Aerial target detection based on improved faster R-CNN [J]. Acta Optica Sinica, 2018, 38(6): 0615004.
- 冯小雨,梅卫,胡大帅.基于改进 Faster R-CNN 的空中目标检测 [J]. 光学学报, 2018, 38(6): 0615004.

- [10] Guo C C, Yu F Q, Chen Y. Image semantic segmentation based on convolutional neural network feature and improved superpixel matching [J]. *Laser & Optoelectronics Progress*, 2018, 55(8): 081005.  
郭呈呈, 于凤芹, 陈莹. 基于卷积神经网络特征和改进超像素匹配的图像语义分割[J]. *激光与光电子学进展*, 2018, 55(8): 081005.
- [11] Wang M, Liu K X, Liu L, *et al.* Super-resolution reconstruction of image based on optimized convolution neural network [J]. *Laser & Optoelectronics Progress*, 2017, 54(11): 111005.  
王民, 刘可心, 刘利, 等. 基于优化卷积神经网络的图像超分辨率重建[J]. *激光与光电子学进展*, 2017, 54(11): 111005.
- [12] 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.
- [13] Hinton G E, Srivastava N, Krizhevsky A, *et al.* Improving neural networks by preventing co-adaptation of feature detectors [J]. *arXiv preprint arXiv: 1207.0580*, 2012.
- [14] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [15] Huang G, Liu Z, Maaten L V D, *et al.* Densely connected convolutional networks [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2261-2269.
- [16] W3C Recommendation. Ink makeup language [EB/OL]. (2017-04-06) [2018-08-06]. <http://www.w3.org/TR/InkML/>.
- [17] Simard P Y, Steinkraus D, Platt J C. Best practices for convolutional neural networks applied to visual document analysis [C] // *International Conference on Document Analysis and Recognition*, 2003: 958-963.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132-7141.
- [19] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C] // *32<sup>nd</sup> International Conference on Machine Learning*, 2015, 37: 448-456.
- [20] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [21] Mouchère H, Zanibbi R, Garain U, *et al.* Advancing the state of the art for handwritten math recognition: the CROHME competitions, 2011-2014 [J]. *International Journal on Document Analysis and Recognition*, 2016, 19(2): 173-189.
- [22] Mouchère H, Viard-Gaudin C, Zanibbi R, *et al.* ICFHR 2014 competition on recognition of on-line handwritten mathematical expressions (CROHME 2014) [C] // *International Conference on Frontiers in Handwriting Recognition*, 2014: 791-796.
- [23] Kingma D P, Ba J. Adam: a method for stochastic optimization [J]. *arXiv preprint arXiv: 1412.6980*, 2014.