基于改进卷积神经网络的手写体汉字识别

徐奇
（安徽工商职业学院　安徽省合肥市　230001）

摘　要：本文针对传统脱机手写体汉字识别特征提取非常困难的问题，文章在 GoogLeNet 网络的基础上搭建了一个适合脱机手写体汉字识别的卷积神经网络。文章首先介绍了卷积神经网络的基本原理和 GoogLeNet 网络中 Inception 模块的特点，然后通过激活函数，批量归一化，加入注意力机制等方法对网络进行优化。实验结果表明，改进后的神经网络准确率达到 98.1%，相比于 AlexNet，Xinception 等卷积神经网络模型的识别准确率有明显的提高。

关键词：卷积神经网络；注意力机制；汉字识别

随着信息技术和网络技术的进步，手写汉字识别在各类使用汉字作为信息传递的应用场景中有非常巨大的潜在需求，因此成为众多学者研究的热点之一。从 70 年代开始至今，汉字识别研究经过了多年研究，从最简单的印刷体，过渡到手写体识别、从单机版的脱机识别过渡到连接云端的联机识别、从单个汉字的识别过渡到长篇的识别，汉字识别技术已经应用在了车牌识别，支票签字识别等多个领域，明显提高了工作效率。但由于手写体汉字的识别率因为难度较大，一直没有达到实际应用的水平。手写汉字识别之所以被公认为是模式识别的难点，可以归纳为以下几点：

（1）以目前流传最广的英文为例，英文字母数量不到 30 个，而日常生活中大多数情况下使用的一级字库中的汉字就有 3700 多个，数量非常庞大.

（2）不像印刷体汉字都是四四方方的，每个人手写汉字时都有自己的书写习惯，风格各异，同一个人不同时期所写的汉字都会有所区别。

（3）汉字中存在很多字体结构相近的相近字，这些相近字字体相似，只是在细微处有差别，增加了计算机识别和特征提取的难度。

（4）统一，全面的数据是设计和训练模型的基础，但是目前还缺乏权威统一的手写体汉字数据库。

近年来，由于神经网络非常适合应用于手写体汉字识别，在该方向成果不断，值得深入研究。 神经网络是一个含有多个隐层的非线性网络模型，通过不断的大规模训练，神经网络能从原始数据中自动提取特征，据此做出预测或者分类。在图像识别、行为预测等领域，深度卷积神经网络 (CNN) 被认为是最适合的算法之一。 卷积神经网络利用不同的卷积提取多维特征。相对于 BP 神经网络等传统深度神经网络，CNN 可以直接针对二维图像进行处理，避免了将二维图像转换成一维信号时丢失特征，因而提取的特征更加全面，是当前准确率最高的图像识别方法之一 [1]。近几年 CNN 在汉字识别领域的研究方向包括结合基于统计的特征提取 [2]、利用迁移学习提高识别率 [3] 等。 以上方法虽然取得了较好的识别结果，但还是存在模型结构复杂，参数调整难度较大，训练过程极为耗时等问题。同时，注意力机制 [4] 也成为了提高分类准确率的方法之一。当大量的二维图像信息被输入到卷积神经网络中时，尽管卷积操作减少了计算量，但网络提取特征时对所有的信息赋予同样的权重，而没有考虑到某些局部信息的重要性。而注意力机制可以识别特征的重要性，将较大的权值赋给较重要的特征，从而提高准确率。为了提高手写体汉子识别率，本文提出了一种基于双重注意力机制（CBAM）的卷积神经网络模型（Convolutional Neural Network Based on Attention, ATT-CNN），用于手写体汉字识别，关注图像关键特征，进而提高分类的准确率。

1 相关基础

与传统的根据字体结构提取特征的手写体汉字识别方法相比，卷积神经网络提取的汉字特征并没有明显的含义，但其提取速度，系统自适应等方面却有非常明显的优势。

1.1 卷积神经网络

CNN 是一种深度学习方法，它能够通过应用可学习的权重和与对象不同的偏差来提取对象的不同特征。如所述，该方法在图像分类领域得到了成功的证明。

CNN 模型将图像作为输入，处理图像并将其分类到预定义的类中。CNN 模型首先使用大量不同类别的图像进行训练。在此阶段，将建立每个类别的通用模型。然后，在测试阶段，根据不同类别的通用模型对图像进行测试，并确定图像属于哪个类别。

对于这种训练和测试，每个输入图像都经过一系列具有不同内核的卷积层。卷积层还包括 BN 层、ReLU 层和最大池层。在卷积层之后，有用于概率分布的全连接层和 softmax 层。在卷积层，通过将输入图像与不同的核进行卷积，从图像中提取特征。在这个操作中，内核从左到右、从上到下扫描整个图像，并在输入图像的像素值和内核之间执行点积。每一种核对应着每一种特征。该卷积层的最终提取结果是一个特征向量图。

在卷积层，卷积核执行卷积运算如下式所示：

$$y_j^l = \sum_i W_{ij}^l x_j^{l-1} + b_i^{l(j)} \tag{1}$$

●基金项目：安徽高校自然科学研究重点项目（项目号：KJ2020A1094）。

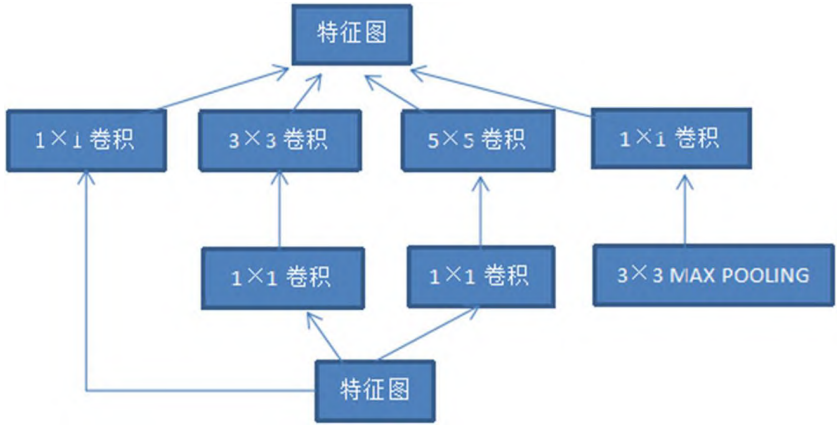

图 1：Inception 模块

式中，b_j^i 为偏置项，W_{ij}^l 为权值项，y_j^l 是第 L 层卷积层特征，x_j^{l-1} 是 L-1 层的第 i 个特征值。

ReLU 层用于将非线性引入特征映射，ReLu 激活函数，如下式所示：

$$\mathrm{ReLu}(x)=\begin{cases}x & x>0\\ 0 & x<0\end{cases} \quad (2)$$

然后池化层用于对特征映射进行下采样。这用于二维性缩减，在保持重要信息和像素之间的空间关系不变的同时，减少了特征图的空间大小。最后，将特征映射展平并转化为向量的全连通层馈入使用反向传播算法的传统前馈网络。接下来，softmax 层执行基于其对图像进行分类。

1.2 Inception模块

传统卷积神经网络采用的大多是较大尺寸的卷积核，提取的特征较为单一，而 Inception 结构并行使用较小的卷积核，从多维度来提取特征，使得提取后的特征更加的全面和完整，能够包含更多的输入信息。本文采用了基于 Inception 模块构建的 GoogLeNet 卷积神经网络，Inception 模块的结构如图 1 所示。

2 算法优化及模型结构

2.1 PReLU激活函数

激活函数对于深度卷积神经网络必不可少。应用 ReLU 激活函数可能会出现震荡，不收敛，过拟合等现象，为了提高效率，本文采用 PReLU 激活函数，解决了 Relu 函数在 x 负半轴为 0 的问题，避免了过拟合。

改进后的神经网络虽然训练速度较慢，但是收敛性很好，准确率也有所提高。PReLU 激活函数定义如下：

$$f(y_i)=\begin{cases}y_i, & \text{if } y_i>0\\ a_i y_i, & \text{if } y_i, 0\end{cases} \quad (3)$$

其中，y_i 是第 i 个通道上的非线性激活输入，a_i为控制负数部分的斜率，a_i的下标 i 表明，允许非线性激活在不同通道有不同取值。

2.2 批量归一化（Batch Normalization）

Inception 模块中引入批量归一化（Batch Normalization）层，解决在训练过程中，中间层数据分布发生改变的问题，防止梯度消失或爆炸，加快训练速度，主要计算方法是：首先求出该批次数据的均值，再求出该批次数据的方差，对样本数据进行标准化处理，最后引入两个可学习的参数 β，γ，使网络能够学习原始网络的特征分布。具体操作过程如下式所示：

$$\mu_B=\frac{1}{m}\sum_{i=1}^{m}x_i \quad (4)$$

$$\sigma_B^2=\frac{1}{m}\sum_{i=1}^{m}(x_i-\beta_B)^2 \quad (5)$$

$$\hat{x}_i=\frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\epsilon}} \quad (6)$$

$$y_i=\gamma\hat{x}+\beta\equiv BN_{\gamma,\beta}(x_i) \quad (7)$$

式中σ_B^2，β_B 分别表示每批次数据的方差和均值；$\hat{x}$为每批次数据经过标准化处理后的数据；∈ 为常数项；γ，β 为可学习参数。

2.3 注意力机制

随着信息技术的发展，我们处在一个需要快速处理大量数据的时代。但我们处理信息的能力是有限的，研究发现，人类视觉系统的接收能力是有限的，但它同时具有强大的视觉信息处理能力。在视觉处理的过程中，眼睛接收数据后，视觉系统将注意力迅速聚焦于局部重要的区域，这种筛除机制可以很大程度上降低视觉系统需要处理的数据量，使我们在处理复杂视觉信息时可以过滤不重要的信息，把资源更合理 的分配到局部的重要区域，为后续的分析和推理提供了更精确的信息。因此，研究人员提出了重点关注局部关键信息的注意力机制。

对于图像信息的特征来说，它们的重要性不是平均分配的，这也意味着卷积网络中每个特征图的重要性是各不相同的。注意力机制的核心思想是根据特征图重要性的不同分配不同的权重，让神经网络的计算资源更多投入在更关键的特征图上，使用结果导向反向指导特征图的权重更新，以推高关键信息对于最终分类结果的影响。

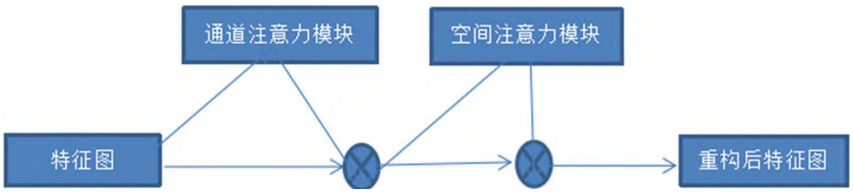

图 2：CBAM 注意力机制示意图

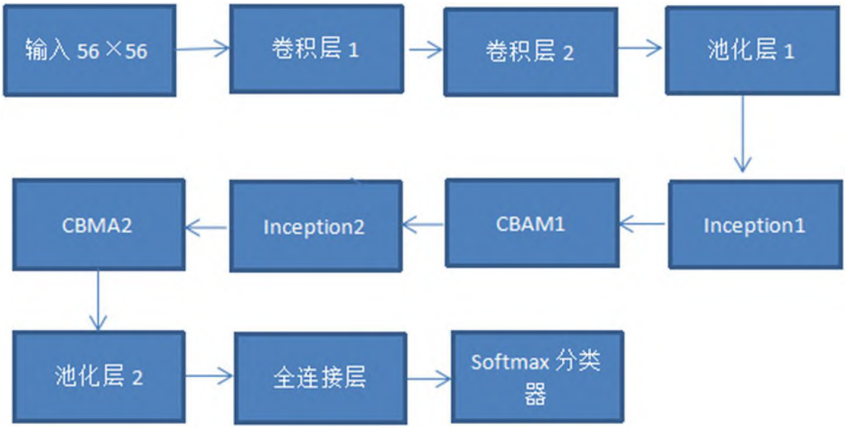

图 3：ATT-CNN 网络模型的结构图

在实际应用中，因为计算机的性能有限，导致无法构建符合理论参数的神经网络去处理和存储全部的信息，因此需要对信息进行筛选，以提高效率，因此，有研究人员将注意力机制引入到 CNN 中。注意力机制是当前深度学习领域中一个比较流行的方向，其模仿人的视觉注意力模式，每次只关注与当前任务最相关的信息，使得信息的使用更为高效。注意力机制已在语言识别、图像标注等诸多领域取得了突破性的进展，并且研究人员发现将注意力机制与 CNN 结合的可以让网络关注最重要的特征，提高重要特征的权重，并抑制不必要的特征，减小不必要特征的权重，并且缓解了传统 CNN 卷积操作时局部感受野缺乏全局信息的缺点，进一步提升了 CNN 的特征提取能力。

CBAM（Convolutional Block Attention Module）双重注意力机制与只关注通道特征的 SENet 相比，CBAM 是一种结合了空间和通道的注意力模块，增强特征图中的有用特征，抑制无用特征，在实际应用当中可以取得更好的效果。图 2 为 CBAM 注意力模块的示意图，首先通过通道注意机制模块对输入特征图进行处理，在通道注意力模块当中分别进行全局平均池化，全局最大池化提取更为丰富的高层次特征，接着分别通过 MLP（Multilayer Perceptron），将通道数压缩为 C/r，再扩张回 C，其中 C 为通道数，r 为衰减比率，取 r=16，然 后将 MLP 输出的特征经过 sigmoid 激活操作，生成 最终的通道注意力模块输出，将该输出对应元素相乘。计算流程如下式：

$$M_c(F)=\sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)))+\sigma(\mathrm{MLP}(\mathrm{MaxPool}(F)))$$
$$=\sigma(\mathrm{w}_1(\mathrm{w}_0(\mathrm{F}^c_{avg})))+\sigma(\mathrm{w}_1(\mathrm{w}_0(\mathrm{F}^c_{\max}))) \quad (8)$$

式中 F 表示输入特征；σ 表示 sigmoid 激活函数。空间注意力模块主要探讨在空间层面特征图的内在关系，即突出区域的重要性，与通道注意力模块 相辅相成。空间注意力模块在算法上相对简单些， 把通道注意力模块输出作为空间注意力模块所需要 的输入，经过卷积核大小为 7×7 的标准卷积层后获得空间注意力模块的特征图。计算流程如下式：

$$\sigma\left(f^{7\times7}([\mathrm{Avg}\,Pool(F);\mathrm{Max}\,Pool(F)])\right)=\sigma\left(f^{7\times7}\left(\left[F^s_{\mathrm{avg}};\mathrm{F}^s_{\max}\right]\right)\right) \quad (9)$$

式中 $f^{7\times7}$ 表示卷积核大小为 7×7。

2.4 基于CNN结合CBAM注意力机制的脱机手写体汉字识别

本文在选用模型时，以 GoogleNet 网络作为此次实验的基本网络，在 Inception 模块后加入了 CBAM 注意力模块。图 3 为 ATT-CNN 网络模型的结构图。表 1 为 ATT-CNN 网络模型的参数。

3 实验

文中使用的测试数据由 3 755 类汉字组成，每一类都有 45 个样本，每个样本数据大小为 56×56。在实验中，使用 90% 的汉字进行训练，剩下的用于测试。实验将批量大小 (batch size) 设置为 56，训练周期设置为 30 次，学习率为 0.01。在模型训练的过程中，神经网络的参数值都会更新，直到网络收敛。将使用本文提出的改进后的卷积神经网络的测试结果与其他卷积神经网络的测试结果进行比较，结果如表 2 所示。

将本论文提出方法与典型网络模型如 AlexNet、GoogleNet 等进行测试结果比较。AlexNet 模型在测试集上的准确率为 84.2% ，GoogleNet 模型在测试集上的准确率为 87. 4%， Xception 在测试集上的准确率为 94. 0% 。而本论

表 1：ATT-CNN 网络模型的参数

编号	层种类	输出尺寸
1	输入层	56×56×1
2	卷积层 1	52×52×5
3	BN 层 1	52×52×5
4	激活函数层 1（PreLu）	52×52×5
5	卷积层 2	48×48×15
6	BN 层 2	48×48×15
7	激活函数层 2（PreLu）	48×48×15
8	池化层 1	24×24×15
9	Inception	24×24×15
10	CBAM	24×24×15
11	池化层 2	10×10×15
12	Dropout 层	10×10×15
13	全连接层	1×3755
14	BN 层 3	1×3755
15	激活函数层（softmax）	1×3755

表 2：不同算法的比较

模型	准确率 /%
AlexNet	84.2
GoogleNet	87.4
Xception	94
本文方法	98.1

文中的模型在测试集上的准确率为 98.1%，优于其它卷积神经网络模型。

4 结语与展望

本文提出了一种基于双向注意力机制（CBAM）的卷积神经网络模型的脱机手写体汉字识别方法，模型中的 Inception 模块可以提取多维度特征，增强了模型的特征提取能力；BN 层提升了模型的收敛速度；CBAM 注意力机制模块提高了关键信息的权重。实验结果表明，本文算法的检测精度达到了 98.1%，相比于 AlexNet，xinception 等卷积神经网络模型的识别准确率有明显的提高。

卷积网络中的注意力机制的核心在于以赋予不同权重的方法来深度挖掘特征图中所含有的信息，但目前所发现的注意力获取渠道相对单一，急需改进。注意力机制已经被广泛证明其针对数据量大，计算量大的深度学习任务来说，不仅具有参数量小，即插即用的便捷性，还可以较为明显地提升特征提取的效果，非常值得继续深入研究。

参考文献

[1] 张驰，郭媛，黎明．人工神经网络模型发展及应用综述 [J]．计算机工程与应用，2021,57(11)．
[2] 郑延斌，韩梦云，樊文鑫．基于二维主成分分析与卷积神经网络的手写体汉字识别 [J]．计算机应用，2020,40(8):2465-2471.
[3] HELCL J，LIBOVICKY J. Neural monkey:an open-source tool for sequence learning[J].Prague Bulletin of Mathemati-cal Linguistics,2017,107(1):5-17.
[4] 张宸嘉，朱磊．俞璐．卷积神经网络中的注意力机制综述 [J]．计算机工程与应用，2021,57(20)．
[5] 武子毅，刘亮亮，张再跃．基于集成注意力层卷积神经网络的汉字识别 [J]．计算机技术与发展，2018,8.
[6] 江培营，陶青川，艾梦琴．基于注意力机制和深度学习的钢板表面缺陷图像分类 [J]．计算机应用与软件，2021,9.
[7] 姚齐水．一种结合改进 Inception V2 模块和 CBAM 的轴承故障诊断方法 [J]．振动工程学报，2022.1.17.
[8] 柴伟佳，王连明．卷积神经网络的多字体汉字识别 [J]．中国图象图形学报，2018,23(3):0410-0417.
[9] 丁蒙，戴曙光，于恒．卷积神经网络在手写字符识别中的应用 [J]．软件导刊，2020,1.
[10] 谢东阳，李丽宏，苗长胜．基于改进 AlexNet 卷积神经网络的手写体数字识别 [J]．河北工程大学学报（自然科学版），2021.12.
[11] JIANG L, ZHANG L, LI C, et al. A correlation-based feature weighting filter for naive Bayes[J]. IEEE transactions on knowledge and data engineering, 2019.
[12] FANG P, ZHOU J, ROY S K, et al.Attention in attention networks for person retrieval[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
[13] ZHU M, JIAO L, LIU F, et al.Residual spectral-spatial attention network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021: 449-462.
[14] XING X, YUAN Y, MENG M Q H.Zoom in lesions for better diagnosis: attention guided deformation network for WCE image classification[J].IEEE Transactions on Medical Imaging, 2020: 4047-4059.
[15] GAO Y, GONG H, DING X, et al.Image recognition based on mixed attention mechanism in smart home appliances[C] IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference(IAEAC),2021:1501-1505.

作者简介

徐奇（1982-），男，安徽省合肥市人。实验师。研究方向为人工智能，模式识别，物联网技术。