中图分类号：TP391

南昌航空大学

硕士学位论文

题　目

高效脱机手写汉字识别网络研究

作者姓名　许兴淼

指导老师　杨词慧、王磊

学科、专业　控制工程

2022 年 5 月

学校代码：10406
分类号：TP391　　学号：1904085210104

南昌航空大学
硕 士 学 位 论 文
（专业学位研究生）

高效脱机手写汉字识别网络研究

硕士研究生： 许兴淼

导　　师： 杨词慧，王磊

申请学位级别： 硕 士

学科、专业： 控制工程

所在单位： 信息工程学院

答辩日期： 2022 年 5 月

授予学位单位： 南昌航空大学

Research on Efficient Offline Handwritten Chinese Character Recognition Network

A Dissertation

Submitted for the Degree of Master

On the Control Engineering

By Xu Xingmiao

Under the Supervision of

A.Prof. Yang Cihui
A.Prof. Wang Lei

School of Information Engineering

Nanchang Aeronautical University, Nanchang, China

May, 2022

摘要

手写体汉字在人们的日常生活中随处可见，对脱机手写体汉字识别(Handwritten Chinese Character Recognition, HCCR)的研究与应用有利于银行、税务、邮政及教育等行业中信息处理的发展。随着近几年卷积神经网络的快速发展，许多高精度脱机 HCCR 方法被提出且获得了显著的效果。然而，此类高精度网络均存在参数量与计算量过大等问题，因此无法满足深度学习模型在移动端高性能部署的需求。针对上述问题，本文基于高性能卷积与注意力机制的思想，提出高效脱机 HCCR 网络，以实现高精度网络模型的高性能部署。主要研究工作包括：

(1) 基于高性能卷积块与跨层信息融合(Cross-layer Information Fusion, CIF)的脱机 HCCR 网络。高性能卷积块通过使用改进 ShuffleNetV2 卷积块并引入注意力机制来构建，该卷积块不仅解决了轻量化卷积过程中通道间特征稀疏的问题，也加强了网络对重要全局特征提取的效率。此外，通过跨层信息融合操作将网络低层与高层视觉信息进行融合，仅使用少量卷积便获得更深层次的特征信息。该网络有效降低了模型的参数量并提升了模型识别精度。最终，在竞赛数据集 ICDAR-2013 上与其他方法的对比中，该网络模型在识别精度与模型参数方面已经达到了优秀算法模型的标准。

(2) 基于多尺度卷积混洗(Multiple Scale Convolution Shuffle, MSCS)模块和注意力特征空间聚集(Attention Features Spatial Aggregation, ASA)模块的脱机 HCCR 网络。MSCS 模块可以获取局部和全局的多个感受野特征信息以增强关键特征提取，同时该模块包含了轻量级网络卷积块的结构思想来减小模型尺寸。此外，该网络还提出了一个 ASA 模块来生成重要的注意力特征，从而减少不同通道特征在空间降维中的信息损失。同时，引入联合损失函数以增强模型区分类间和类内注意特征差异的能力。最终实验结果表明，仅在手写数据集上训练的网络模型识别字符图像只需 3.97 毫秒，准确率高达 97.63%，且只需要 22.9MB 的存储空间。因此，该方法在考虑存储空间、计算量和推理时间方面达到了单一网络模型中较为先进的水平。

关键词：脱机手写汉字识别，跨层信息融合，注意力机制，空间聚合，联合损失函数

Abstract

Handwritten Chinese character can be seen everywhere in people's daily life. The research and application on offline handwritten Chinese character recognition can promote the development of information processing in banking, taxation, postal services and education industries. Since then, with the rapid development of convolutional neural networks in recent years, many high-precision offline HCCR methods have been proposed and achieved remarkable results. However, this kind of high-precision network has the problem of excessive parameter and calculation, so it can not meet the demand of high-performance deployment of deep learning model in mobile terminals. To solve the above problems, this paper proposes an efficient offline HCCR network based on the idea of high-performance convolution and attention mechanism to realize the high-performance deployment of high-precision network models. The main research work includes:

(1) Offline HCCR network based on high-performance convolution block and Cross-layer Information Fusion(CIF). The high-performance convolution block is constructed by improving ShuffleNetV2 convolution block and introducing attention mechanism, which not only solves the problem of sparse features between channels in the process of lightweight convolution, but also enhances the efficiency of network in extracting important global features. In addition, the network fuses the low-level and high-level visual information of the network through cross-level information fusion operation, and only uses a small amount of convolution to obtain deeper feature information. The network effectively reduces the parameters of the model and improves the accuracy of the model recognition. Finally, by comparing with other methods on the competition dataset ICDAR-2013, the network model can reach the standard of excellent algorithm model in recognition accuracy and model parameters.

(2) Offline HCCR network based on multiple scale convolution shuffle (MSCS) module and attention features spatial aggregation (ASA) module. MSCS module can obtain local and global multiple receptive fields feature information to enhance key feature extraction, and it contains the structure idea of the lightweight network conv-block for reducing the model size. Moreover, this network also proposes an attention features spatial aggregation (ASA) module to generate important attention features for reducing the information loss of different features of channel in spatial dimension reduction. Meanwhile, the joint loss function is introduced to strengthen the ability of the model to

distinguish the difference of attention features between inter-class and intra-class. The final experimental results show that the network model trained only on handwritten dataset only takes 3.97ms to recognize a character image and achieves 97.63 % accuracy just requires 22.9 MB for storage. Therefore, this method reaches the state-of-the-art in a single network model particularly in view of storage space, computation cost and reference time.

Keywords: Offline handwritten Chinese character recognition, Cross-layer information fusion, attention mechanism, spatial dimension reduction, joint loss function

目　录

第1章 绪 论

1.1 课题研究背景及意义

在银行、税务、邮政及教育等行业和领域，每天都需要人工处理海量手写体文档。若采用人力对文档内容进行记载，不仅降低了工作效率，而且容易出现错误。近几年，计算机视觉领域发展迅速，手写体汉字识别(Handwritten Chinese Character Recognition, HCCR)技术的应用场景变得十分广泛。通过该技术能够对手写体文字图像进行自动识别，很大程度上解决了人工处理时的效率问题。因此，若采用一个高性能手写体汉字识别技术，不仅能够有效减少人工成本，而且可以降低人工识别误差。

手写体汉字数据集根据其不同的获取方式，分为联机和脱机手写体汉字数据集。手写体汉字识别分类如图 1-1 所示。其中，联机手写体汉字数据是通过数字笔、数字手写板、触摸屏等设备在线书写采集的文字信号，脱机手写体汉字数据则是经过摄像头或扫描仪等设备采集的二维图像[1]。因此，基于循环神经网络(Recurrent Neural Network, RNN)[2]的联机手写体汉字识别方法凭借着对数据时序与位置信息的掌握取得了不错的成果。然而，脱机手写体汉字识别只能依靠基本的图像信息进行判断，丢失了类似于联机手写体数据的先验信息，是一个类别众多的分类任务，识别难度相对较高，且在研究过程中存在许多问题。

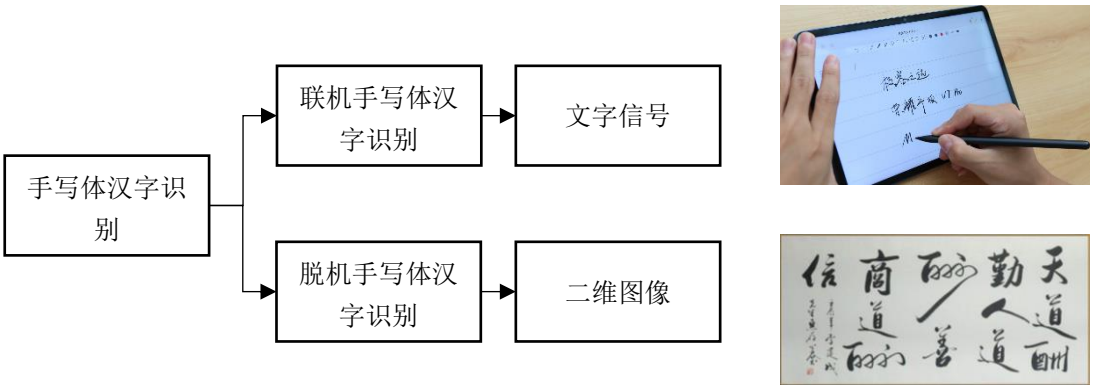
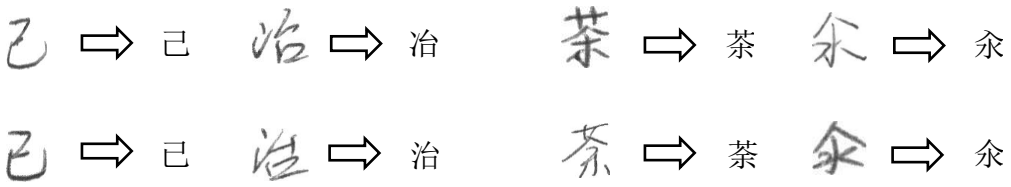


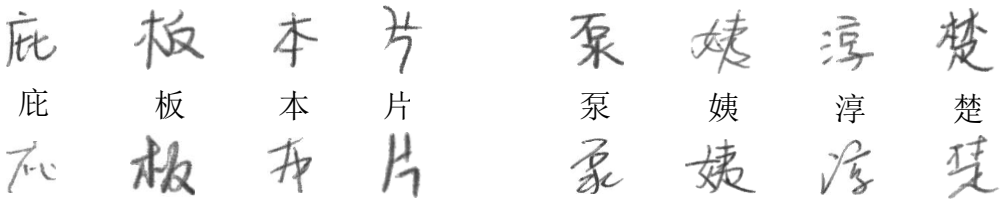
图 1-1 手写体汉字识别分类

如今，由于卷积神经网络(Convolutional Netural Networks, CNN)在图像分类中的突出表现，因此基于 CNN 改进的脱机 HCCR 方法也层出不穷，并且取得了显

著性的突破[3-5]。然而，在脱机 HCCR 任务中仍然面临着诸多亟待解决的问题，如众多类别分类、随机风格书写识别和形似字识别等，如图 1-2 所示。此时，若存在一种高效的方法能够解决上述问题，不仅可以提升模型的识别性能，一定程度上也利于脱机 HCCR 模型的部署。近几年，大多数脱机 HCCR 性能提升的研究都集中在输入数据预处理和骨干网络优化上，如方向特征图的提取[6, 7]，结合传统特征预处理方法有效提升输入图像表现力；图像形态归一化[3, 8]，增强模型对不规则大小、方向字符的识别鲁棒性；正则化的引入[9, 10]，有效避免训练过拟合并加快模型训练速度；网络复杂度的提升[3, 5, 11]，对网络横向与纵向的加深以增强非线性表达能力与拟合效果。尽管文献在竞赛数据集上的识别准确率已经超过了人类表现水平[12]，但如今大多数脱机 HCCR 网络模型在移动端高性能部署上还存在着精度过低且模型尺寸过大等问题。于是，为建立一个高效的脱机 HCCR 方法，研究人员通过对常规 CNN 的计算量占比和参数量分布的研究与分析，得出卷积层承担了整个计算工作的 90%以上，且 80%以上的参数来自全连接层。因此，脱机 HCCR 的研究主要在于降低卷积层计算成本和减少全连接层的冗余参数方面，方法如网络剪枝[13-15]、量化[16, 17]、低秩扩展[18, 19]和知识蒸馏[20-22]等。此后，学术界又涌现出一些高性能模型[3-5, 11]，实现了网络压缩和加速，然而在移动终端高性能部署上还存在较大的提升空间。



形似字示例



不同书写风格的汉字示例

图 1-2 形似字和不同风格汉字示例

为了解决上述问题，本课题将引入高性能卷积网络及网络优化算法，在确保模型精度水平提升的同时，能够有效减少移动端部署时空间与时间损耗，旨在提出一个精度高、速度快、参数量小的高效脱机 HCCR 解决方案。

1.2 国内外研究现状

上世纪 80 年代，国内脱机 HCCR 领域缺乏完备的研究体系。然而，汉字在国内广泛使用，很多场所都需要人工对手写体汉字进行识别与记载，因此研究者们开始着力于脱机 HCCR 相关技术的研究与应用。1985 年，国家总局发布了包含 2355 个通用手写体汉字的 GB2312-80 数据集，极大地推动了手写体汉字识别的研究。此后，刘迎建提出采用笔端为基元的联机手写体汉字识别技术，将手写正楷字体的准确率提升到了 95%以上。1996 年，中科院自动化研究所研发出了一套脱机手写体汉字识别系统，在特定环境中，其识别准确率可达 93.6%[23]。

进入二十一世纪，计算机视觉任务中的深度学习方法层出不穷，通过不断地更新与发展，在脱机 HCCR 的研究与应用中起到了不可忽视的作用。

1.2.1 脱机 HCCR 研究现状

由于智能终端产业的飞速发展，目前国内对于联机 HCCR 的研究已经趋于饱和，商业界也涌现出了许多优秀的 HCCR 软件产品，例如 HP 实验室的 TesseractOCR[24]、百度的 PaddleOcr[25]、旷视科技的 OCR[26]等。然而，脱机 HCCR 的研究历程相比于联机 HCCR 要更加艰辛。其中，二者训练数据量的差别是一个重要原因，同时脱机 HCCR 数据缺少了像联机 HCCR 数据的笔划先验信息，仅凭二维图像数据集进行处理与训练。此外，手写体汉字识别又普遍存在书写随意、形似字繁多等问题。因此，研究出一个识别精度高、速率快的脱机 HCCR 方法一直都是模式识别领域的一个重点与难题。

在过去的几十年里，研究人员提出了众多提高脱机 HCCR 性能的分类器，如隐马尔可夫模型(Hidden Markov Model, HMM)[27]、支持向量机(Support Vector Machines, SVM)[28]、修正二次判别函数(Modified Quadratic Discriminant Function, MQDF)[29]和判别学习的二次判别函数(Discriminative Learning Quadratic Discriminant Function, DLQDF)[30]等。由于近几年计算能力、非线性激活函数和数据集的快速发展，CNN 在脱机 HCCR 任务中得到了很多的改进和应用。首次应用于 HCCR 任务的网络是多列深度神经网络(Multi Column Deep Neural Network, MCDNN)[31]，它将不同数据集训练的 8 个网络输出平均值作为最终结果，其性能优于 DLQDF 等传统分类算法。为进一步提高脱机 HCCR 识别精度，富士通的研究团队通过增加卷积层层数提出了一种 CNN 方法[12]，且取得了不错的效果。此外，Zhong 等[3]又提出一种基于 GoogLeNet 的 19 层网络，并通过使用 Gabor 特征、Hog 特征、梯度特征作为网络的输入来增强 CNN 性能，且在 ICDAR-2013 脱机 HCCR 竞赛数据集上的最低错误率仅 3.26%，其存储空间也低于当时所有 CNN 模

型，同时该网络识别精度是当时第一个超越人类表现的方法。

此后，研究人员开始从多个方面进一步提高脱机 HCCR 性能。其中，在输入数据预处理中结合传统领域算法取得了较为显著的成果。Zhong 等[6]在深度残差网络(Deep Residual Networs，DRN)[32]中引入 STN[33]模块，该模块主要是对输入字符的形状进行归一化，使得训练后的模型可以适应不同方向、位置、大小的输入数据。实验结果表明，该模型对不规则手写汉字具有很强的鲁棒性。此外，Zhang 等[8]结合深度神经网络与传统的归一化方向分解特征图，并引入了一个新的适应层以减少特定层上训练数据和测试数据之间的不匹配。同时，损失函数的创新与发展对脱机 HCCR 任务也产生了极大的推动力。Xiao 等[34]使用字符模板来处理字符之间的内部相似性，并结合实例损失来增加类间方差，从而扩大类与类之间的距离以提升模型分类精度。Liu 等[35]引入了一种改进的 Inception-ResNet 网络，提出交叉熵判别加权方法来减少训练阶段的识别错误，并采用稀疏训练与权重剪枝来降低模型参数。

近年来，伴随着脱机 HCCR 的不断发展，出现了众多识别精度较高的方法。然而在终端设备的深度学习模型部署中，还有许多尚待处理的问题与难点。因此研究者开始关注脱机 HCCR 在实际场景中的应用，其中模型计算成本高和存储空间大成为最需要解决的关键问题。为此，Xiao 等[4]提出全局监督低秩扩展(Global Supervised Low Rank Extension, GSLRE)和自适应降权方法，在一个 9 层的 CNN 基础上，使得最新模型压缩到基础模型原始大小的 1/18，且该网络性能方面几乎不受剪枝影响。此外，Li 等[11]通过对 SqueezeNet[36]中 fire 模块的改进，不仅大幅降低了卷积层的计算参数，而且确保了模型识别精度几乎不受影响。同时，他们还提出了全局加权平均池(Global Weighted Average Pooling, GWAP)的概念。具体来说，GWAP 是对全局平均池化(Global Average Pooling, GAP)中的加权参数进行优化，既实现了全连接层的特征降维，一定程度上也降低了因维度减少所带来的精度损失。基于上述思路，Melnyk 等[5]进一步提出了全局加权输出平均池(Global Weighted Output Average Pooling, GWOAP)，同时通过类激活图(Class Activation Map, CAM)将网络注意力可视化，从而提高了网络的可解释性。与此同时，在特征提取部分中引入 VGG16[37]的 backbone 改进版本，对 VGG16 原先的卷积块中加入了 bottleneck 结构思想，且无残差连接，其性能优于当时其他基于残差网络的脱机 HCCR 模型，且进一步兼顾了模型精度与参数损耗。不同于传统的深度学习脱机 HCCR 方法，Xu 等[38]提出了一种基于概念学习的方法，它通过提取汉字笔画并结合贝叶斯程序来学习并建立中文概念模型，然后采用蒙特卡罗马尔可夫链抽样建立每个概念模型的字符生成器。此外，Min 等[39]采用了一种改进的浅层 GoogLeNet，在不损失精度的情况下减少了训练参数量。同时，该网络还提出了一种误差消除算法，可以准确计算样本在测试结果中的置信度，然后通过多次识

别从而减少相似单词对模型精度的影响。

1.2.2 高效 CNN 研究现状

在早期的一些 CNN 结构中[37, 40, 41]，网络计算成本过高和参数冗余一直是 CNN 迈向智能化部署的棘手问题。因此，研究学者近年来在脱机 HCCR 模型压缩和加速领域进行了大量的研究，其中一个研究重点便是对高效卷积块的设计，该类卷积块在大幅降低卷积计算参数的同时，也确保了模型精度的稳定。

模型精度和参数计算量一直是高效卷积块必须考虑的重要因素。SqueezeNet[36]引入大量 1×1 卷积核取代部分 3×3 卷积核，从而降低了卷积层的参数量，同时该网络通过压缩和扩展卷积层的方式保留模型表现力。此外，通过延迟网络的下采样时间能够获得更丰富的特征图，从而避免模型精度损失。此外，谷歌近几年对轻量级网络的研究也作出了重要贡献，他们提出了 MobileNet V1-V3[42-44]，目前已经在许多移动设备中成功部署并投入到商业化使用。MobileNetV1 卷积块的核心便是深度可分卷积的应用，通过将卷积计算分为通道和空间两个部分，并采用相加的方式得出计算成本，从而显著降低了模型参数量。之后，考虑到 MobileNetV1 存在的一些缺陷与不足，MobileNetV2 在第一次深度卷积之前引入点卷积以进一步扩张通道，因此保留了更多有效的特征信息。与此同时，MobileNetV2 在最后一层点卷积后采用线性激活函数，并使用快捷连接来有效避免了梯度发散问题。随着注意力机制的兴起，MobileNeV3 提出的卷积块引入了压缩和激励(Squeeze and Excitation, SE)注意力模块[45]，并提出了一个 Hard-switch 激活函数来增强模型的非线性。值得注意的是，作者在网络的开头和结尾添加了一些 5×5 卷积以提升网络模型精度。

之后，旷视科技提出了一个全新的高效 CNN 架构：ShuffleNetV1-V2[46, 47]，该类网络架构在准确率、参数规模、内存消耗、模型大小和推理时间等方面均取得了优异的成绩。ShuffleNetV1 的卷积块基于瓶颈结构，且该卷积块采用组卷积代替标准卷积操作以减少参数计算量，同时利用通道混洗增强了不同组通道之间的交互能力。然而，组卷积操作将增加模型宽度从而导致模型容量增加，为避免此类情况，ShuffleNetV2 提出的卷积块在输入部分使用通道分离代替组卷积操作，并在通道融合中引入拼接操作替换加法运算，最终执行通道混洗。实验结果表明，ShuffleNetV2 的卷积块优化了模型的计算和内存访问开销能力。华为诺亚实验室最近提出了另一种高效的神经网络 GhostNet[48]，该网络是在 MobileNetV3 上改进的轻量级网络，其主要作用便是利用廉价的线性运算来生成更多的输出特征图，从而显著降低模型参数。

1.2.3 注意力机制研究现状

注意力机制可表示为人们在视觉上对事物的关注度，当观察一个图像或者一个事物时，不同区域所产生的关注度是不同的，即重要部分所获取到的关注会更多。简而言之，图像中引入注意力机制就是对图像各通道的像素值作加权，从而有利于模型对图像深更层次的理解。注意力机制作为目前计算机视觉领域中的研究热点，在图像分类、目标检测、图像分割中已成功应用，并取得了突破性的进展。

注意力机制最早是由 Google 的 Mnih 等[49]在 2014 年提出，当时该方法主要是为了提升 RNN 图像分类的精度。经过训练后的模型能够很好地定位到图像中重要的特征的相关区域，成为首个应用于图像分类的注意力模型。此后，Bahdanau 等[50]在机器翻译任务上使用了类似于注意力的机制，得到的模型能够自动地搜索句子中与预测目标词相关的部分，因而增强了模型对于文本段的理解能力，该模型也是注意力机制在自然语言处理中的首次应用。由于传统 RNN 模型计算受方向顺序限制的问题，且该问题容易导致 RNN 模型的并行能力不足，同时引发信息的丢失，导致模型的表现力不足。因此，Google 的 Vaswani 等[51]提出 Transformer 模型，通过引入大量 Self-Attention 模块和残差网络中的 short-cut 结构，有效解决了上述问题。

鉴于注意力机制在自然语言处理应用中的成功案例，研究者们开始尝试将其处理计算机视觉领域任务。2015 年，Jaderberg 等[33]受普通 CNN 学习平移不变性与旋转不变性原理的启发，设计了一个显式处理各种图形变换的空间注意力模块，该模块使用向量或矩阵对图像像素点进行乘积运算，使得网络能够自动调整特征图大小。同时，该模块是一个可以加载到 CNN 任意位置的轻量化注意力模块。然而，对于一个二维图像，除了其长宽维度，还有一个通道维度。SENet[45]取得了 2017 年 ImageNet 分类竞赛的第一名，其核心便是基于通道注意力的 SE 模块。该模块首先通过 squeeze 操作对图像空间维度特征进行聚合与降维，使得每个通道都由一个象征性实数表示。其次，使用 Excitation 操作对特征图进行升维并获得每个通道所对应的权重参数。最终，将这些权重参数映射到原特征图，即网络所学习到的相关注意力特征。SE 模块的可塑性很强，且适用于目前任何现有网络，因此，对于 SE 模块的创新与改善具有较强的研究意义。

由于前述的空间注意力模块与通道注意力模块在计算机视觉任务中分别都取得了显著效果，因此研究者对两模块的共同作用也产生了浓厚的兴趣。2018 年，Woo 等[52]提出了一种基于空间与通道的卷积块注意力模块(Convolutional Block Attention Module，CBAM)。该模块首先对输入特征图分别使用全局最大池化和全局平均池化得到两个一维特征，接着经过两个多层感知器(Multilayer Perceptron,

MLP)层得到变换后的结果，最后将两个一维特征相拼接并对其使用 sigmoid 函数得出通道注意力参数。之后，对经过通道注意力的特征图各通道分别取最大值和平均值得到两个空间特征，再将二者拼接后使用卷积层进行空间融合，最终形成了拥有通道和空间注意力的重要特征。注意力机制方法种类较多，针对不同的任务使用的方法也是有利有弊。因此，对于本课题的多分类任务，应当选择一个最佳的注意力方法来提高脱机 HCCR 网络对字符图像的理解力。

1.3 主要研究内容

设计一个高性能且拥有较强部署能力的脱机 HCCR 方法，是本文研究的主要内容。该方法使得脱机手写体汉字识别不仅要具备优异的鲁棒性，有着较高的识别精度，而且还需要有着移动端部署的轻量化形态。本课题首先对脱机手写体汉字数据进行预处理，对输入层字符图像进行归一化操作，使得字符在每个图像中分布均匀。之后，提出了一个基于高性能卷积与跨层信息融合(Cross-layer Information Fusion, CIF)的脱机 HCCR 模型，在模型参数量处于一个较低水平的同时，一定程度上还能够提高模型识别精度。最后，为满足模型在轻量化与高精度两方面的要求，进一步提出了基于多尺度卷积混洗(Multiple Scale Convolution Shuffle，MSCS)与注意力特征空间聚合(Attention Feature Spatial Aggregation，ASA)的脱机 HCCR 方法。本文的研究内容可以分为以下几个部分。

(1) 数据预处理。观察脱机 HCCR 数据集 CASIA-HWDB1.0-1.2 原图像，可以发现图像中字符的形态各异，位置分布差异较大，且部分图像中的文字模糊不清，导致网络模型的学习空间与范围大幅降低。因此，本文通过扩充字符图像边界，进行图像阈值反转等预处理操作实现了图像增强。

(2) 基于高性能卷积与跨层信息融合的方法。本文对轻量化网络 ShuffleNetV2 中的卷积块进行了改进，并首次应用到脱机 HCCR 任务中。同时，还使用了残差连接与密集连接作为跨层信息融合的桥梁，从而进一步增强了网络的表现力。

(3) 基于多尺度卷积混洗与空间聚合的方法。为兼顾移动端应用时的各项指标，本文设计的多尺度卷积混洗模块作为卷积块内的瓶颈层不仅降低了网络模型中的参数、计算量与推理时间，同时还有效提升了网络模型的识别精度。此外，本文基于注意力机制对全局平均池化作了改进，提出了一个空间聚合模块来压缩特征。最后，联合 softmax 损失函数与中心损失函数对特征进行监督。

1.4 论文章节安排

第一章绪论，该章节介绍脱机 HCCR 的研究背景及意义，同时对于脱机 HCCR 相关研究方法作了介绍。

第二章相关背景知识，该章节主要介绍了脱机 HCCR 相关技术、脱机 HCCR 数据集以及脱机 HCCR 评价指标三个部分。

第三章提出的脱机 HCCR 方法基于高性能卷积与跨层信息融合，首先介绍了网络的总体结构及相关模块，其次分析了网络参数与损失函数的实现细节，然后介绍了网络训练的环境与参数设置，并在 ICDAR-2013 数据集上与近几年提出的方法进行了对比，最终通过控制变量法对提出的模块进行了消融实验。

第四章提出的脱机 HCCR 方法基于多尺度卷积混洗与空间聚合，在第三章方法的基础上，对其存在的缺陷进行了优化，主要表现在模型精度、计算量、参数量与推理时间等方面。

第五章总结和展望，总结本文研究内容与实验结果，提炼出创新点与优势点。与此同时，针对本课题方法中存在的一些问题进行分析，对脱机手写体汉字识别的下一步研究工作进行展望。

第2章 相关背景知识

当前众多脱机手写体汉字识别方法虽取得了较为理想的识别精度，但还有一定的提升空间。此外，在深度学习模型移动端部署中，模型的容量大小与计算速度也是一个重点问题。因此，针对上述问题，本文首先对输入端数据进行图像增强预处理操作，随后基于轻量化卷积神经网络思想、注意力机制与特征聚合等方法构建了脱机 HCCR 网络模型。

2.1 脱机 HCCR 相关技术

卷积神经网络作为目前图像视觉领域内的主流发展趋势，已经成功应用于众多行业，如制造业、教育行业、娱乐行业等。同时，随着对脱机 HCCR 的不断深入研究，近几年涌现出了诸多实用性较高的算法并实现了部署。本文将在轻量级卷积神经网络架构、注意力机制、特征聚合等方面来构建网络整体架构，最终实现一个性能与部署能力兼备的脱机手写体汉字识别算法。

2.1.1 卷积神经网络

卷积神经网络模型最早由 Lecun 等[53]在上世纪 80 年代末提出，该模型采用了局部连接、权值共享方案，并采用反向传播(Back Propagation，BP)进行训练，最终成功应用于当时手写邮政编码的识别。此后，Lecun 等在上世纪 90 年代末提出了 LeNet-5 模型[54]，该网络模型共计 5 层，前两层由卷积层和池化层交替构成，接着是两个全连接(Full Connected，FC)层，最后通过 Softmax 层输出结果。该模型在当时美国众多银行支票识别系统被广泛应用，同时其独特的设计思想也推动了之后众多优秀卷积神经网络的发展。

之后的几年内，由于深度学习模型训练成本十分高昂且难以收敛，因此卷积神经网络的发展较为缓慢。直到 Hinton 等[55, 56]提出深度置信网络(Deep Belief Network，DBN)，才有效解决了深层神经网络陷入局部最值导致模型收敛较慢的问题。在 2012 年，Hinton 等设计了 AlexNet[40]网络模型，获得了当年 ImageNet 比赛的冠军，其识别精度相比于第二名高出了 10 个百分点。该模型首次采用了修正线性单元(Rectified Linear Unit，ReLU)激活函数代替 Tanh 激活函数，有效缓解了梯度消失的问题。此外，该网络还引入了修剪网络隐层连接的 Dropout 技术，使用了双 GPU 网络结构进行加速运算，促进了深度学习在手写体汉字识别领域的发展。2014 年，Simonyan 等提出了 VGGNet 网络模型[37]，此模型包含多个网络结

构，如 VGG-11、VGG-13、VGG-16 以及 VGG-19。与当时其他卷积神经网络不同的是，VGG 采用了大量的 3×3 的卷积核代替了先前 5×5 卷积核与 7×7 卷积核，该做法大幅降低了卷积层的参数量，并增强了网络的鲁棒性。此外，VGGNet 使用了多个卷积层与激活函数组合的方法，使得其非线性表达能力得到增强。VGG-16 网络整体框架如图 2-1 所示。

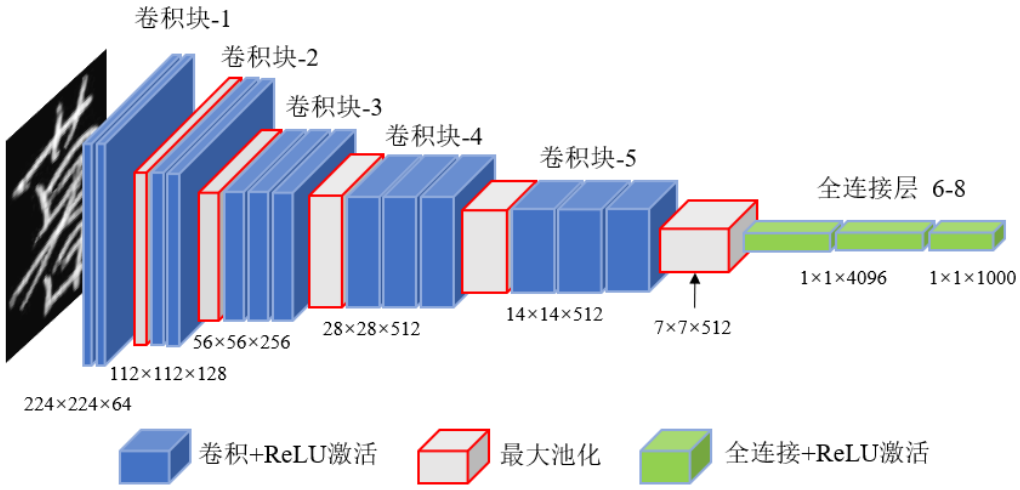


图 2-1 VGG16 网络整体框架

目前，随着深度学习在实际场景任务中的不断应用与发展，以及计算设备的不断更新迭代，卷积神经网络成为了计算机视觉领域任务中重要的解决方案。一般情况下，分类任务中的卷积神经网络一般分为输入层、卷积层、下采样层、全连接层以及 Softmax 层。首先，整个网络的入口是输入层，输入的图像数据一般为四个维度，分别为一次训练的样本数(batchsize)、通道(channel)、宽度(width)、高度(height)。此外，为增强网络模型对数据的拟合能力，在输入层数据使用一些数据增强算法也是常见的操作。卷积层通过使用卷积核在输入图像上下滑动并进行加权计算，从而提取图像中的特征信息。随着卷积层数量的增多，更深层次的图像信息就能够很好地被发掘。下采样层即池化层，该层主要通过像素点之间的相关性对卷积后的图像特征进行提取与压缩，在扩大感知野的同时减少了网络学习的参数量。同时，下采样层具有平移不变性，该层能够使得训练出来的网络模型具有更强的鲁棒性。全连接层类似于一个分类器，且一般出现在网络的末端，其主要功能便是将卷积后的分布式特征映射到真实的样本空间。此后，Softmax 层对全连接层输出的特征向量进行归一化操作，使得所有的特征值都转变成了 0-1 之间的概率问题。最终，概率最大的结点所对应的类别便是预测的目标。

2.1.2 高效卷积网络

随着用户对性能要求的不断提高，普通浅层卷积神经网络已经无法满足市场需求。于是，研究者们相继提出了许多高性能卷积神经网络，如 VGG、GoogLeNet、ResNet 等。然而，随着网络性能的提升，卷积层的深度也在不断累加，从 19 层 VggNet 到 22 层 GoogLeNet，再到之后 152 层的 ResNet，这对网络模型部署中模型存储与预测速度方面造成了很大的阻碍。此后，为提高深度学习模型在移动终端设备部署的能力，研究者提出了一些高效的卷积神经网络，例如 SqueezeNet、MobileNet、ShuffleNet 等。该类网络不仅解决了网络模型参数冗余问题与模型预测速度问题，同时也有效缓解了轻量级网络模型表现力不足的问题。

例如 Google 首次提出的 MobileNet 轻量级高效网络，其核心深度可分离卷积(Depthwise Separable Convolution，DSC)在网络参数压缩方面取得了非常显著的成果。传统的卷积计算是对所有输入特征采用一个完整卷积核进行计算，而 DSC 则是对卷积核进行了因式分解，实现分步计算并累加，有效降低了计算成本。该卷积计算分为逐深度卷积(Depthwise Convolution，DC)和逐点卷积(Pointwise Convolution，PC)。首先采用逐深度卷积对输入特征图提取特征，其次使用逐点卷积将各通道特征信息进行线性融合，最终在精度损失可控范围内，大幅降低了模型参数。

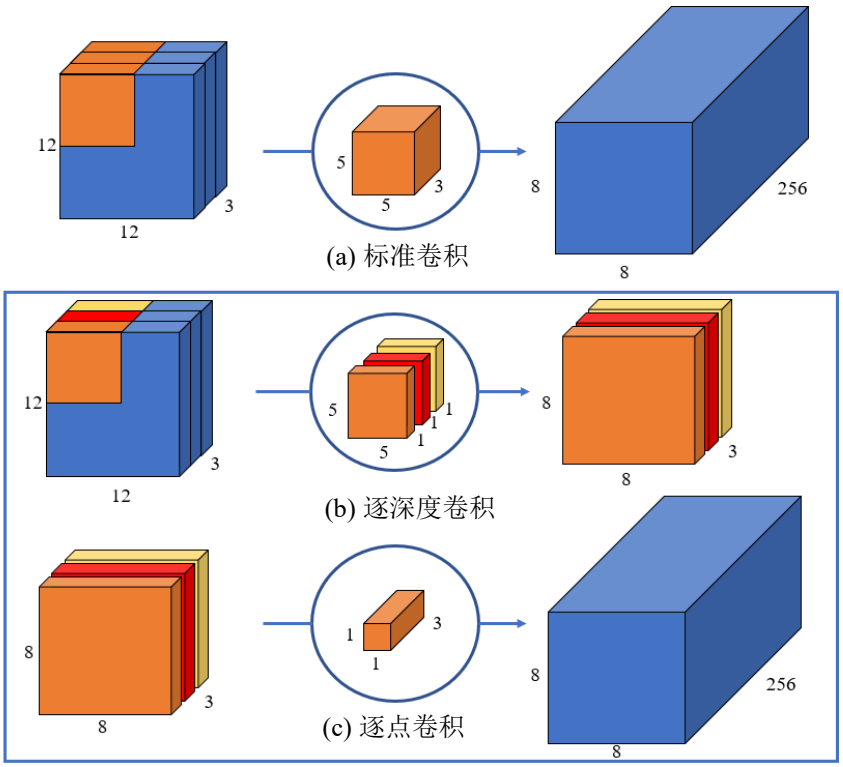


图 2-2 标准卷积和深度可分离卷积

标准卷积与深度可分离卷积计算过程如图 2-2 所示。假设 DSC 的输入图像尺寸为 K*K*C，卷积核为 N*N*C，且输出图像为 M*M*Z。首先，将卷积核分解为 N*N*1 与 1*1*C 的卷积核。接着，使用 N*N*1 卷积核进行逐深度卷积，对输入图像的每一个通道进行卷积操作，最后输出一个 M*M*C 尺寸的特征，其中 padding 为 0，且步长为 1，则 M 为$(K-N)/1+1$，该卷积计算量为CM^2N^2。此后，对逐深度卷积的输出采用 Z 个 1*1*C 的卷积核进行逐点卷积，且输出一个 M*M*Z 尺寸的特征图，该卷积计算量为CZM^2。因此，DSC 计算总量为$CM^2(N^2+Z)$。此时，若直接采用 N*N*C 尺寸的标准卷积核进行计算，则会产生CZM^2N^2的计算量。可以发现，采用 DSC 的参数计算量仅为标准卷积的$1/Z+1/N^2$。简而言之，若采用的卷积核尺寸为 5*5，输出特征图个数为 256，通过计算可以发现，DSC 的计算量是标准卷积计算量的 1/25，因此表明了 DSC 极大地降低了模型卷积时的参数量。

2.1.3 注意力机制

目前，注意力机制已经被视为深度学习领域中十分有利的工具。类似于人眼观察事物的过程，该机制首先对图像进行全局扫描，之后获取图像中关键特征所在的区域，随后对此类区域信息重点分析以提高网络模型的性能。如今，注意力机制已经广泛应用于自然语言处理与计算机视觉任务。同时，由于其强大的学习能力，注意力机制对脱机 HCCR 的发展也有着重要的推动作用。

2016 年，Zhong 等[6]在深度残差网络输入层引入 STN 空间注意力模块，该模块通过在图像像素点中进行归一化矩阵运算，增强了网络模型对不规则字符图像的学习能力，从而有效提升了训练模型在不规则数据中的表现力。2017 年，Yang 等[57]提出了一种迭代细化模块，该模块通过结合低层视觉信息与高层视觉信息，形成了一个多尺度的残差块级联，其功能类似于循环神经网络中的注意力机制。2019 年，Xu 等[58]提出一种多重比较注意力网络(Multiple Comparative Attention Network, MCANet)，该网络中的多重注意力模块首先对卷积后的特征图划分通道，然后通过 SE 注意力模块分别学习各通道注意特征并聚合，最后在聚合特征上通过对比损失和中心损失函数来预估模型，从而得到了当时脱机 HCCR 最先进的准确性。

计算机视觉任务中的注意力机制分为两种类型，强注意力机制和软注意力机制。强注意力机制侧重于随机的预测过程，更加侧重动态的变化，且不可微，训练需要通过增强学习来完成。然而，软注意力机制则是一个连续的分布问题，更加关注空间或者通道，且该注意力特征在网络学习后可以生成，并且是可微的。由于可微的注意力能够通过计算梯度来得到注意力权重，因此，软注意力机制是目前主流的注意力研究方向。

注意力机制可以分为通道、空间以及通道空间混合的三种形式。通道注意力机制是指每个通道内所有元素使用相应方法聚合后，对得到的一维具有通道表现力的特征向量(1×1×C)归一化后的结果。其中，当某个通道聚合后的归一化参数值越大，则表明该通道的特征关注度越高。空间注意力机制则是忽略了通道之间的信息交互，根据通道维度上的元素进行压缩聚合，对压缩后具有空间表现力的特征向量(H×W×1)归一化后的结果。Woo 等[52]提出的 CBAM 模块是具有通道与空间双重注意力机制的经典模块。研究者通过多次实验，发现只有将两机制按顺序并行组合，且前置通道注意力机制时，所取得的实验效果最佳。该模块的实现方式如图 2-3 所示。

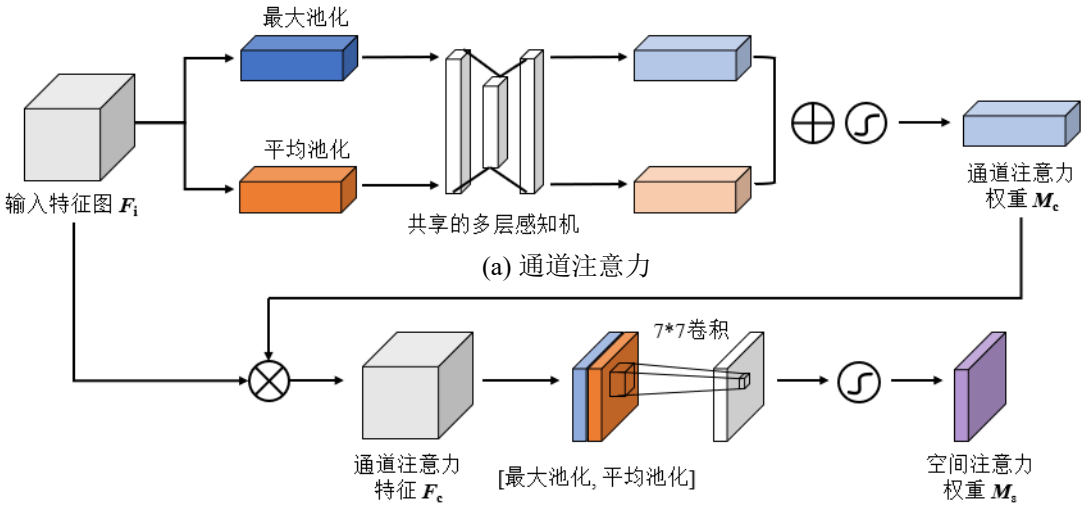


图 2-3 通道空间注意力模块 CBAM

CBAM 模块首先对输入特征图 F_i 在空间维度上分别采用全局最大池化(*MaxPool*)和全局平均池化(*AvgPool*)以获得两个一维的向量特征。之后通过一个共享的多层感知机(Multilayer Perceptron, MLP)对特征向量进行降维与升维，该操作能够增强网络的非线性，从而提高通道之间的相关性。最终对输出的特征依次进行加操作和 sigmoid 激活操作 σ，即可生成基于通道的注意力权重 M_c，如图 2-3 (a)所示。

$$M_c(F_i)=\sigma(MLP(AvgPool(F_i))+MLP(MaxPool(F_i))) \tag{2-1}$$

在得到通道注意力权重参数 M_c 后，将其与输入的特征图相乘，得到一个具有通道注意力特性的特征图 F_c。此后，在 F_c 通道维度上分别进行全局最大池化和全局平均池化以获得双通道特征，将其按通道拼接后经过一个 7×7 卷积 f^{7*7}，并通过 sigmoid 操作 σ 得到空间注意力权重 M_s，如图 2-3 (b)所示。最终将该参数与

输入特征图 F_c 相乘以生成双重注意力机制的新特征图。

$$M_s(F_i) = \sigma(f^{7*7}([AvgPool(F_i), MaxPool(F_i)])) \tag{2-2}$$

通过分析以上不同注意力机制，针对不同的任务和目标，选择一个合理的注意力机制应用于脱机 HCCR 网络是一个需要重点考虑的问题。

2.1.4 特征聚合方法

在分类任务中，特征聚合指的是卷积层部分输出的特征图采用一些方法聚合成一个类别分值向量，即降维操作。在得到聚合特征后，将其输入到 softmax 层以计算每个类别的得分。目前常见的特征聚合方法有全连接、全局平均池化、全局最大池化(Global Max Pooling, GMP)、Gem 池化[59]等。

全连接是近几年最常使用的特征聚合方法，即输出的特征图每个神经元与输入的所有神经元相连接，就是将三维输入特征图拉伸为一维向量再进行矩阵乘法运算。该聚合方法虽然能够复现较为完整的特征信息，但其权重参数十分冗余，因此很容易导致过拟合问题。为解决上述问题，Lin 等[60]提出使用 GAP 来代替全连接层，将卷积层输出的特征图中每个通道元素取均值，从而形成一个具有空间意义的一维特征向量。该操作不仅实现了特征图维度的降低，也极大减少了网络模型参数的损耗。全连接与全局平均池化实现细节如图 2-4 所示。

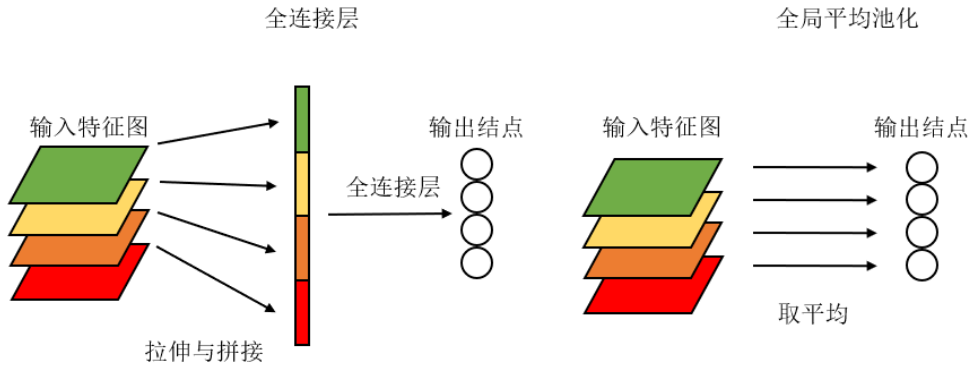


图 2-4 全连接与全局平均池化实现

此后，类似于 GAP 的各种降维方法开始尝试在特征聚合中应用。GMP 通过提出特征图中各通道最大元素来获取重要特征，该操作同样能够实现特征维度与网络模型参数的降低。此外，GMP 在单词识别[61]领域中也有较好的效果。然而，由于 GMP 在噪声过滤上的不足，从而导致其对噪点图像信息提取较差。随后，考虑到 GAP 和 GMP 池化特征的优劣，进一步提出 Gem 池化。该池化操作介于 GAP 与 GMP 之间，通过参数调节双方比重，从而可以关注到不同细粒度的区域。Gem

计算公式如下：

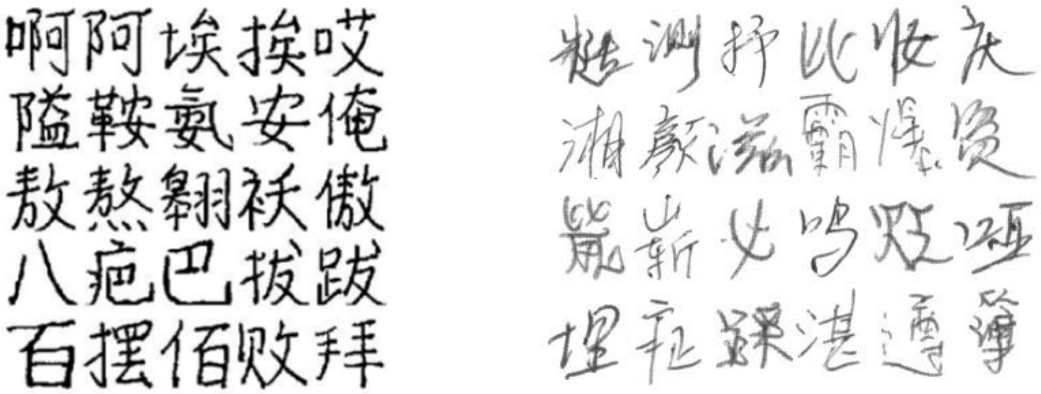
$$f_k^{(g)} = \left(\frac{1}{|X_k|}\sum_{x\in X_k} x^{P_k}\right)^{\frac{1}{P_k}} \tag{2-3}$$

其中，X_k 表示第 k 个通道内的像素点，f_k 为第 k 个通道聚合后的特征图。P_k 是一个可学习参数，当 P_k 为 1 时表示全局平均池化，当 P_k 趋于无穷大时为全局最大池化操作。

2.2 脱机 HCCR 数据集

二十一世纪后，深度学习能够如此迅速地崛起，离不开大数据的推进。具体而言，数据集大小决定了深度学习模型所能达到的下限，就算拥有再好的算法模型与强劲的计算资源，没有一定数据量的积累也会导致模型效果不佳，甚至不如一些简单的线性模型。

目前，常见的脱机手写体汉字数据集有北京邮电大学模式识别实验室的 HCL2000 数据集[62]与中科院自动化研究所模式识别国家实验室的 CASIA-HWDB 数据集[63]。HCL2000 数据集包含了常用的 3755 种汉字，由 1000 位参与者撰写。为便于研究文字书写时的影响因素，该数据集不仅包含字符图像，而且还包含了书写参与者的信息，例如书写者的年龄、职业、文化程度等。然而相比于 CASIA-HWDB 数据集，HCL2000 数据集的样本丢失了现实中手写汉字的随机性与差异性，导致训练出来的模型缺乏一定的鲁棒能力。因此，目前主流的脱机手写体汉字识别数据集采用的都是 CASIA-HWDB，本文也采用该数据集进行实验。HCL2000 与 CASIA-HWDB 数据集样本如图 2-5 所示。



(a) HCL2000 数据样本图　　(b) CASIA-HWDB 数据样

图 2-5 HCL2000 与 CASIA-HWDB 数据样本图对比

CASIA-HWDB1.0-1.2 子集为脱机手写体单字数据集，该数据集是对文本行图像数据采用注释工具进行分割与标记而得，一般用以训练或验证搭建好的网络模型。CASIA-HWDB1.0 包含 3866 个汉字数据及 171 个字母、数字和符号，由 420 名参与者撰写，其中包含了 GB2312-80 规定的 3740 个汉字，共计 1,609,136 个字符样本。此后，CASIA-HWDB1.1 引入了 GB2312-80 规定的所有常用汉字(共计 3755 种)与 171 个字母、数字和符号，由 300 名参与者书写，且每个汉字的样本数在 300 左右，共计 1,121,749 个字符样本，同时该子集也是目前脱机 HCCR 领域中使用最为广泛的数据集。CASIA-HWDB1.2 由 3319 个汉字与 171 个字母、数字、符号组成，被 300 名参与者书写，其汉字集与 CASIA-HWDB1.0 无交集。CASIA-HWDB1.0-1.2 共计 7185 个汉字类别，可用于大类别脱机手写汉字识别模型的研究。同时，CASIA-HWDB 还包含了一个竞赛测试集 ICDAR-2013，该数据集一般用于评估训练好的网络模型，其种类与 CASIA-HWDB1.1 的种类相同，由 60 名参与者书写，共计 224,419 个字符样本数。本文主要使用 CASIA-HWDB1.1 子集作为训练数据集，并采用 ICDAR-2013 数据集进行测试。CASIA-HWDB 数据集分析如表 2-1 所示。

表 2-1 CASIA-HWDB 数据集分析表

数据集	参与者人数(个)	汉字类别(个)	样本总数(个)
CASIA-HWDB1.0	420	3,866	1,609,136
CASIA-HWDB1.1	300	3,755	1,121,749
CASIA-HWDB1.2	300	3,319	990,989
CASIA-HWDB1.0-1.2	1,020	7,185	3,721,874
ICDAR-2013	60	3,755	224,419

2.3 脱机 HCCR 评价指标

对于相同的输入数据，要考察不同算法模型或是同一算法模型在使用不同参数的效果时，需引入一个能够判别模型优劣的定量指标，即评价指标。本文研究的脱机 HCCR 是一个多分类任务，因此本文选用分类任务的评价指标。目前，在分类任务中使用最为广泛的评价指标即准确率(Accuracy)，其可简单定义为预测正确的结果在总样本中的占比。

在二分类任务中，混淆矩阵是准确率指标定义的基础，该矩阵对数据集中的记录按照真实的类别与分类模型预测的类别判断两个标准进行汇总。混淆矩阵如表 2-2 所示。矩阵的行表示实际结果，矩阵的列表示预测的结果。其中，真正例(True Positive, TP)，表示模型预测为正类的正类样本数，即预测正确；真负例(True Negative, TN)，表示模型预测为负类的负类样本数，即预测正确；假正例(False Positive, FP)，表示模型预测为正类的负类样本数，即预测错误；假负例

(False Negative, FN)，表示模型预测为负类的正类样本数，即预测错误。

表 2-2 混淆矩阵

		实际结果	
		1	0
预测结果	1	TP	FP
	0	FN	TN

因此，所有预测正类样本数总和除以所有样本预测数总和即为二分类任务中准确率的定义。其定义公式如下：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2-4}$$

然而，针对脱机 HCCR 多分类任务，二分类评价标准明显不能够适用。虽然可以将多分类转换为多个二分类进行解决，但其计算过程较为复杂与冗余，不利于模型评估。因此，分类准确度(categorical accuracy)和稀疏分类准确度(sparse categorical accuracy)被研究者提出用以解决多分类任务。分类准确度用于检查真实结果中最大值所对应的标签与预测值中最大值所对应的标签是否相等，其应用的模型与标签均为向量形式，仅适用于多分类单标签的任务。例如一个三分类任务，真实标签为[[0,1,0], [1,0,0], [1,0,0]]，模型的预测结果为[[0.1,0.6,0.3], [0.3,0.4,0.3], [0.5,0.1,0.4]]。随后根据真实标签和预测结果里元素的具体位置转换为标量格式，真实标签为[1,0,0]，预测结果为[1,1,0]，因此该模型准确率为 0.67。同理，稀疏分类准确度与分类准确度效果类似，但其真实标签已经转换为上述的标量格式。综上，多分类任务的评价指标 $Accuracy_{ca}$ 定义公式如下：

$$Accuracy_{ca} = \frac{\sum_{i=1}^{m} num(equ(\max(y_i\_ture), \max(y_i\_pred)))}{m} \tag{2-5}$$

其中，m 表示数据总量。num 代表满足括号内条件的数据个数。equ 为相等条件，当满足条件时，返回 1，否则为 0。max 表示取向量中元素的最大值。因此，多分类指标指的是满足真实结果向量中元素的最大值与预测向量中元素最大值相等的数据个数在所有预测数据总数的占比。

2.4 本章小结

本章主要介绍了脱机 HCCR 的相关背景知识，包含脱机 HCCR 的相关技术、

数据集及其评价指标三个方面。首先，介绍了卷积神经网络的发展与原理，并详细介绍了目前常用的分类网络 VGG-16。然后，根据目前深度学习模型移动端的广泛应用，介绍了一些兼顾模型参数与模型预测速度的高效卷积网络，其中详细介绍了 MobileNet 卷积块的网络结构。接着，为提升脱机 HCCR 网络模型的性能与可解释性，介绍了注意力机制在计算机视觉领域的发展，并详细介绍了通道与空间注意力机制。此外，为解决全连接层参数冗余问题，介绍了一些高性能的特征聚合方法，包括 GAP、GMP、Gem Pooling。最后，详细介绍了脱机 HCCR 常用的数据集与评价指标。

第3章 基于高性能卷积与跨层信息融合的脱机 HCCR

3.1 引言

目前，基于卷积神经网络的脱机 HCCR 技术已经趋于成熟，网络模型的规模与性能得到了巨大提升。然而，当前业务场景与市场需求也变得更加丰富，很多高精度网络模型均有计算量与空间容量过大等弊端，该类模型很难满足现实场景下的高性能部署要求。因此，本章提出了一个基于高性能卷积与跨层信息融合的脱机 HCCR 网络。首先，对于轻量化卷积性能不足的问题，将引入 ShuffleNetV2 中的卷积块并进行优化，以确保网络模型获取到更加丰富的全局性特征。其次，对优化的卷积块及其单元组采用跨层信息融合(Cross-layer Information Fusion, CIF)操作，该操作增强了通道之间的特征传播，复用了不同层次的特征，从而仅使用少量的卷积便能获取到更加丰富的特征。最后，采用全局平均池化进行特征降维操作，该操作不仅获得了各通道的全局性特征，而且一定程度上减少了模型的冗余参数。提出的网络在公共数据集上与目前各类方法的实验结果对比表明，该网络的参数容量和模型精度已经达到优秀的水平。本章节的主要内容如下：

(1) 首先，对 ShuffleNetV2 轻量化卷积块进行结构改进，在降低模型参数的同时，进一步增强卷积块对图像全局特性的学习能力。

(2) 此外，引入跨层信息融合操作，该操作主要由残差连接与密集连接组成，进一步加强通道间特征的流通，且实现了特征的重复利用。

3.2 网络结构设计

3.2.1 网络总体架构

本章总体网络主要由三部分组成，输入图像预处理操作、基于高性能卷积与跨层信息融合模块和特征降维操作。本章提出的网络总体架构如图 3-1 与表 3-1 所示。由于卷积神经网络有时无法学习到某些关键的局部特征，导致特征提取时产生的偏差会影响到最终的模型精度。因此，在输入层中添加图像预处理操作对网络整体效果的提升有着至关重要的作用。近几年，图像预处理算法[3, 6, 8]在脱机 HCCR 领域的应用取得了不错的效果。该类方法对输入层图像进行增强，将文字中的一些详细的局部特征作为网络的输入，然而该操作增加了图像输入层的维度，从而导致模型计算成本变大。此外，考虑到手写体汉字数据集中字符分布不均、

字符大小与像素不一等问题，本章对输入图像采用边界填充与像素值反向操作，如图 3-1 所示。该操作将文字处于整个图像的中心区域，规范所有输入图像位置。同时该操作起到了图像增强的作用，降低了卷积层特征提取的难度。实验结果表明，使用了本章预处理操作的网络模型较原图作为输入的网络性能有所提升，因此表明了该预处理操作的有效性。

表3-1 基线网络与优化后网络的整体框架

“Baseline”代表基线网络模型；“Proposed”代表优化后的网络模型。“LiteNet”代表轻量级网络 ShuffleNetV2；“HENet”代表改进的高性能网络。

Layers	Baseline	Proposed	Output Shape
Input	original image	preprocessed image	96 × 96 × 1
Conv_Layers	3×3 conv. 64, BN+ReLu 3×3 conv. 64, BN+ReLu	3×3 conv. 64, BN, ReLu 3×3 conv. 64, BN, ReLu	96 × 96 × 64 96 × 96 × 64
Downsampling	LiteNet_Downsampling, 96	HE_Downsampling, 96	48 × 48 × 96
Conv_Block1	LiteNet_Conv. 96, BN+ReLu LiteNet_Conv. 96, BN+ReLu LiteNet_Conv. 96, BN+ReLu	HENet_Conv. 96, BN+ReLu HENet_Conv. 96, BN+ReLu HENet_Conv. 96, BN+ReLu	48 × 48 × 96 48 × 48 × 96 48 × 48 × 96
Downsampling	LiteNet_Downsampling, 128	HE_Downsampling, 128	24 × 24 × 128
Conv_Block2	LiteNet_Conv. 128, BN+ReLu LiteNet_Conv. 128, BN+ReLu LiteNet_Conv. 128, BN+ReLu	HENet_Conv. 128, BN+ReLu HENet_Conv. 128, BN+ReLu HENet_Conv. 128, BN+ReLu	24 × 24 × 128 24 × 24 × 128 24 × 24 × 128
Downsampling	LiteNet_Downsampling, 256	HE_Downsampling, 256	12 × 12 × 256
Conv_Block3	LiteNet_Conv. 256, BN+ReLu LiteNet_Conv. 256, BN+ReLu LiteNet_Conv. 256, BN+ReLu	HENet_Conv. 256, BN+ReLu HENet_Conv. 256, BN+ReLu HENet_Conv. 256, BN+ReLu	12 × 12 × 256 12 × 12 × 256 12 × 12 × 256
Downsampling	LiteNet_Downsampling, 448	HE_Downsampling, 448	6 × 6 × 448
Conv_Block4	LiteNet_Conv. 448, BN+ReLu LiteNet_Conv. 448, BN+ReLu LiteNet_Conv. 448, BN+ReLu	HENet_Conv. 448, BN+ReLu HENet_Conv. 448, BN+ReLu HENet_Conv. 448, BN+ReLu	6 ×6 × 448 6 ×6 × 448 6 ×6 × 448
Feature Aggregation	Global Average Pooling	Global Average Pooling	448
Output	3755-dim Softmax		3755

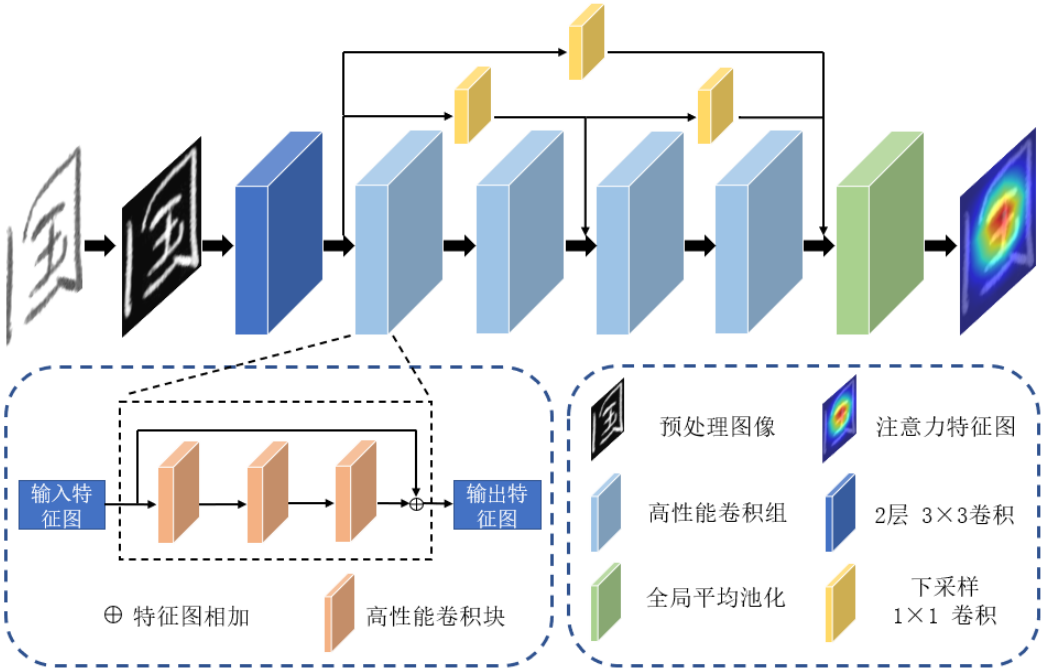
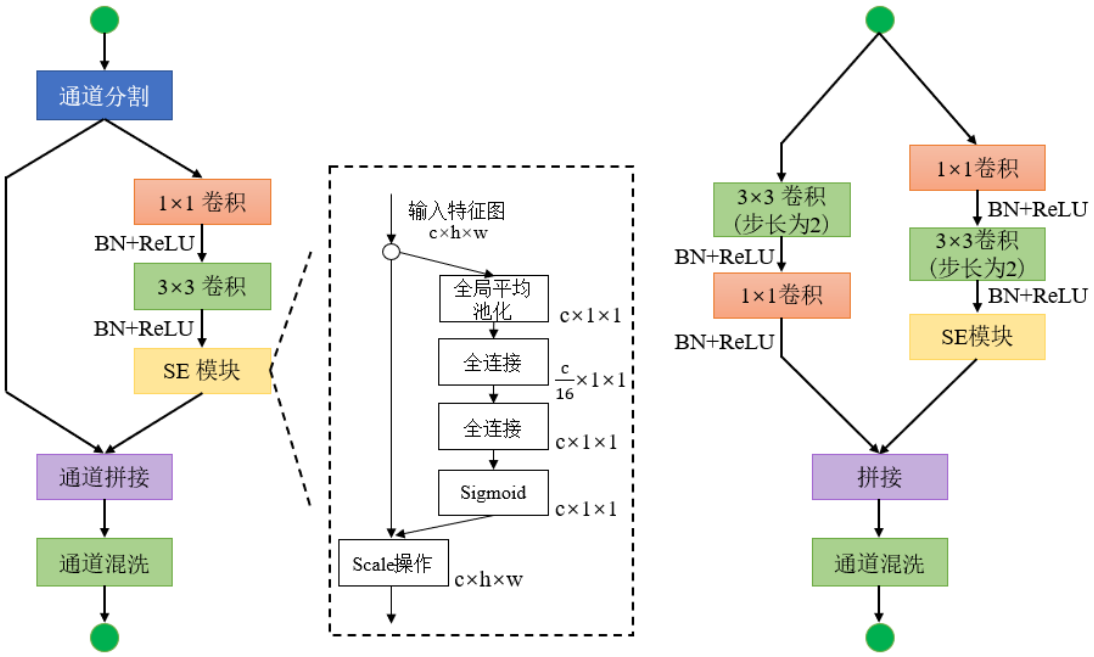


图 3-1 跨层信息融合的高性能卷积网络架构

此外，对于轻量化卷积网络 ShuffleNetV2 在文字识别中存在精度不足的问题，本章对该网络卷积块结构进行了优化。引入注意力机制以增强模型对图像特征的学习能力，同时结合跨层信息融合操作来加强轻量化卷积间的通道信息交流来弥补其性能的不足。其次，针对特征降维方法中常见的冗余参数问题，采用全局平均池化(GAP)来提取各通道全局特征，该操作一方面降低了模型参数损耗，另一方面也保留了较多关键性特征。

3.2.2 高性能卷积模块

目前，一些常见的高性能卷积神经网络[36, 44, 47]已经成功应用于计算机视觉领域。此类网络在不影响精度的同时，其较低的参数损耗与计算量满足了实际应用端部署的需求。因此，本章对轻量级卷积网络 ShuffleNetV2 的卷积模块进行改进，在实现轻量化参数的同时填补了模型性能上的空缺。



(a) 卷积块　　(b) SE 注意力模块　　(c) 下采样模块

图 3-2 高性能卷积模块

虽然ShuffleNetV2中的深度可分离卷积(DSC)大幅降低了模型容量，但是由于DSC 将标准卷积分成了空间与通道的两部分卷积，导致通道与空间之间的特征关联度变得十分稀疏，因此其在脱机 HCCR 任务的表现力还有提升空间。改进的高性能卷积模块主要由卷积和下采样两部分组成，如图 3-2 (a)所示。首先采用通道分割算法将输入特征平均分为两部分特征，一部分特征进行恒等映射，另一部分依次通过两种卷积核尺寸的标准卷积提取特征。同时对提取后的特征使用SE注意力模块以获取重要通道特征，SE 模块如图 3-2 (b)所示。然后，采用拼接(Concat)

操作对恒等映射的特征与卷积后的特征进行融合，最终使用通道混洗模块(Channel Shuffle)以加强两分支特征通道间的信息交流。

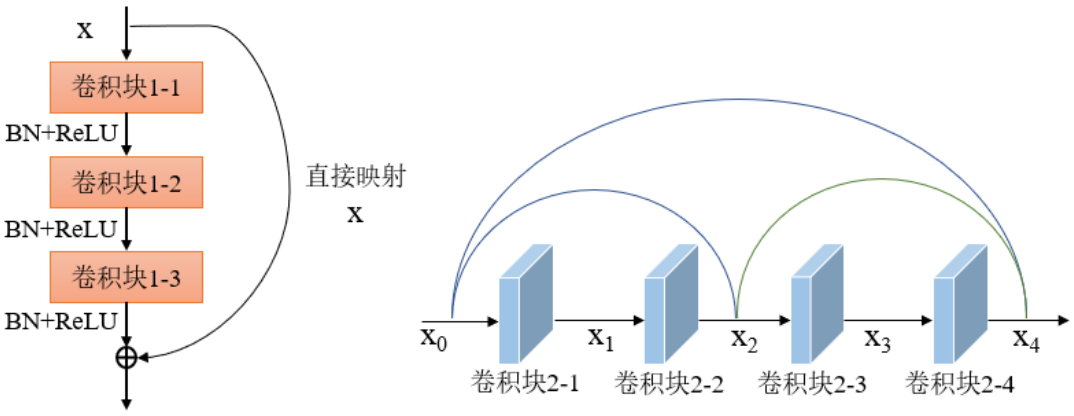
此外，特征压缩部分采用改进的 ShuffleNetV2 下采样模块，如图 3-2 (c)所示。该模块直接将两份相同的特征分别进行下采样，输入图像大小变为原来的一半。为了避免图像特征信息的丢失，通道数则提升为原来的两倍，因此该模块未进行通道分割。然后，使用拼接操作与通道混洗操作进行特征融合。类似于前述卷积块改进的思想，将下采样中存在的深度卷积采用标准卷积进行替换。实验结果表明，采用改进后的下采样网络性能优于其他下采样或是池化操作。

3.2.3 跨层信息融合操作

本章在提出的跨层信息融合操作主要由残差连接和密集连接[64]两部分组成。其中，高性能卷积组间使用了密集连接操作，高性能卷积块间采用了残差连接。在图 3-1 中，改进的高性能卷积模块中的残差连接通过对低层特征信息的复用，显著缓解了因网络层数加深而发生网络退化等问题。此外，残差连接一定程度上能够避免过深网络中出现的梯度消失、梯度爆炸以及模型过拟合等问题。常见的残差连接如图 3-3 (a)所示，采用快捷连接将输入特征 x 直接映射到最后一层作为输出结果，输出结果计算公式如下：

$$F(x) = H(x) - x \tag{3-1}$$

其中，输入特征为 x，中间的卷积块计算为 $F(x)$，期望输出为 $H(x)$。因此，神经网络的学习目标便成了学习 $H(x)$ 和 x 的差值，即所谓的残差。当残差部分训练趋于 0 时，模型精度不会因为网络深度的增加而降低。



(a) 残差连接　　(b) 密集连接

图 3-3 跨层信息融合操作

此外，在图 3-1 中，将两个高性能卷积组为单元进行密集连接操作。由于前

后的卷积组特征图的大小与通道数不一致，因此在密集连接中引入下采样卷积模块使得特征能够正常融合，同时该卷积还能使得网络模型获得更大的感受野。密集连接通过对低层特征的不断重用，使得网络每层的输入都会融合前面所有层的输出，从而仅采用少量的卷积便能达到深层网络的性能。常见的密集连接如图 3-3 (b)所示，该操作不仅通过较少的卷积获取到了更加高效的网络性能，而且提升了梯度的反向传播，从而加速了网络模型的训练。第 n 层输入特征 x_n 计算公式如下：

$$x_n = W([x_0, x_1, ..., x_{n-1}]) \tag{3-2}$$

其中，W 为卷积块中的所有非线性操作，如卷积、批归一化、激活等。

3.3 网络参数与实验细节

3.3.1 网络参数设置

网络的整体结构参数如表 3-1 所示。首先，本章提出一个基于 ShuffleNetV2 轻量级卷积的基线网络，该网络前两层采用 3×3 卷积提取特征，之后使用 ShuffleNetV2 的下采样与卷积块进行特征压缩与提取，由 4 个下采样操作与 4 个轻量级卷积块组成。卷积层所有的输出通道为 64-96-128-256-448，且最终输出 6×6×448 大小的特征图。此外，每次卷积后都采用批归一化操作和 ReLU 激活函数来加快网络训练速度。最终，网络采用全局平均池化操作对图像特征进行降维。

同时，对于基线网络性能不足的问题，本章对轻量化卷积块进行了改进，结合跨层信息融合操作进一步弥补了模型性能的缺陷。改进后的卷积块采用 1×1 与 3×3 卷积核获取多尺度特征，且 SE 模块中的减速比参数设置为 16。此外，卷积层输出的通道数、输出的特征图大小以及最终的降维操作都与基线网络一致，同时每层卷积后仍采用批归一化与 ReLU 激活函数优化网络，最终分类类别总数为 3755。

3.3.2 损失函数设置

脱机 HCCR 任务其实是一个多分类任务，该任务通过神经网络对各样本进行计算得出一个多维数组，该数组中每个维度对应一个类别，维度总数便是所有类别数，且数组中对应概率最大的维度便是预测的类别。因此，本章使用多分类任务中最常见的交叉熵损失函数，该损失函数避免了 sigmoid 型函数导数易发生饱和的情况。同时，在使用交叉熵损失函数进行梯度下降计算时可以防止网络出现

导致学习率下降的梯度弥散问题。此外，该损失函数一定程度上增大了类间的距离。交叉熵损失函数 $Loss$ 如下：

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} \ln p_{i,k} \tag{3-3}$$

其中，N 表示一批数据的样本总数；K 表示样本真实标签值所在的位置；$y_{i,k}$ 表示第 i 个样本的真实标签 k；$p_{i,k}$ 表示第 i 个样本预测为第 k 个标签值的概率。

3.4 实验结果与分析

3.4.1 训练环境与参数设置

在一定程度上，合理调整输入图像的大小能够有效提高特征提取的效率，从而能够改善识别精度。通过脱机 HCCR 研究[5, 11]的相关经验，同时考虑网络模型的识别性能和计算成本等因素，将输入的字符图像尺寸设为 96×96。此外，初始学习率设置为 0.01，且在精度停止提升时，学习率将变为原始学习率的 1/10。同时，通过向每个卷积层中添加批量归一化操作来加快网络模型训练，并在模型优化中使用随机梯度下降法进行训练，将动量、最小批量样本数和所有样本训练次数分别设置为 0.9、128 和 20。此外，通过使用 L2 正则化策略来避免实验中出现的过拟合问题，并将权重衰减设置为 10^{-3}。同时，在 softmax 层之前添加了 dropout[10]，并将输出比例设置为 0.5。使用基于 tensorflow 框架的 Keras 深度学习库来构建所提出的模型，在 11GB 内存的 NVIDIA GeForce GTX 2080 Ti 上进行实验。

3.4.2 ICDAR-2013 数据集上的性能对比

表 3-2 显示了近几年提出的方法在 ICDAR-2013 公共数据集上的性能比较。其中，本章提出的方法相比于其他先进方法在模型识别精度与尺寸方面处于中上水平，因此该方法仍有一定的提升空间。

通过下表的对比数据可以发现，提出的网络模型识别精度与容量的利用都要优于 Zhong 等[3]提出的基于多方向特征图的单网络和集成网络。此外，基于 STN 模块的残差网络[6]对输入层图像进行了预处理操作，其最佳的识别精度较提出的网络增加了 0.15%。然而，由于该网络输入特征的复杂性，其模型容量是提出模型容量的 3.94 倍。之后，Cheng 等[65]提出了一个联合损失函数用以增大分类时的类间距离，然而提出的网络相比于其单网络提升了 0.15%识别精度，且减少了 12.8MB 空间损耗。与前述的 STN 模块的残差网络类似，Zhang 等[8]针对多方向特

征图输入维度的复杂性，压缩了网络的参数量，同时引入了一个适应层来提升网络训练与测试时的性能。该网络的参数量与提出的网络参数量相近，同时其识别精度也有着 0.15%的优势。

表 3-2 ICDAR-2013 竞赛数据集上不同方法性能的比较

"Ensemble"代表模型集成策略。所有方法都在 HWDB1.0-HWDB1.1 数据集下进行训练。

Method	Size (MB)	Accuracy (%)	Ensemble
Human Level Performance[12]	n/a	96.13	n/a
HCCR-Gabor-GoogLeNet[3]	27.7	96.35	no
HCCR-GoogLeNet-Ensemble-10[3]	270.0	96.74	yes (10)
Residual-34[6]	92.2	97.36	no
STN-Residual-34[6]	92.3	97.37	no
DCNN-Similarity ranking[65]	36.2	97.07	no
Ensemble DCNN-Similarity ranking[65]	144.8	97.64	yes (4)
DirectMap + ConvNet[8]	23.5	96.95	no
DirectMap + ConvNet + Ensemble-3[8]	70.5	97.12	yes (3)
DirectMap + ConvNet + Adaptation[8]	23.5	97.37	no
M-RBC + IR[57]	n/a	97.37	no
HCCR-CNN9Layer[4]	41.5	97.30	no
HCCR-CNN12Layer[4]	48.7	97.59	no
Cascaded Model (Quantization)[11]	**3.3**	97.11	no
Cascaded Model[11]	20.4	97.14	no
Melnyk-Net[5]	24.9	97.61	no
MCANet[58]	>500	**97.66**	no
Our Network	23.4	97.22	no

为满足移动端深度学习模型高性能部署的需求，使得模型更加高效。Xiao 等[4]提出了全局有监督低秩扩展(Global Supervised Low-rank Expansion, GSLRE)和自适应降权(Adaptive Drop-weight, ADW)方法对模型参数进行剪枝，其最佳的网络 HCCR-CNN12Layer 相比于提出的网络虽然提升了 0.37%的精度，但其空间容量却比提出的模型多出了 25.3MB。Li 等[11]提出了一种基于多尺度卷积模块与全局加权平均池化的级联网络，该网络不仅丰富了特征提取的信息，而且优化了对特征降维时有效参数的利用，同时该网络的量化模型相比于提出的模型节省了 20.1MB 的存储空间。然而，其模型精度较提出的网络仍有 0.11%的误差。

为进一步增强模型性能，Yang 等[57]提出了一种能够获取低层与高层视觉信息的残差注意力模块，且该模块相比于提出的模型有着 0.15%的精度优势。此后，Melnyk 等[5]提出了一个基于高效卷积块与改进的加权平均池化的可视化网络，相比于提出的模型不仅有着 0.39%的精度优势，而且仅损耗了 1.5MB 的存储空间。此外，Xu 等[58]基于多重注意力与联合损失提出了 MCANet，该网络模型的识别精度达到了目前最为先进的水平，即 97.66%。然而其特征聚合部分使用了全连接操作，因此该网络模型容量十分巨大。

3.4.3 消融实验

通过控制变量的方式对网络模型进行了消融实验，实验结果如表 3-3 所示。其中，采用 ShuffleNetV2 卷积与下采样的基线网络虽然有着较低的参数量，然而其识别准确度仅仅只有 96.53%。因此，本章对基线网络存在的缺陷进行了改进，从表 3-3 中可以发现改进前后的性能对比。

表 3-3 提出的网络在 ICDAR-2013 测试集上不同配置的性能对比

“Preprocess”表示图 3-1 中的预处理操作。“Conv3”表示标准 3×3 卷积。“SE”表示 SE 注意力模块。“CIF”表示跨层信息融合操作。

Method	Parameters	Accuracy
Baseline	**2,591,873**	96.53%
Baseline (Preprocess)	2,591,873	96.58%
Baseline (Preprocess + Conv3)	5,933,025	96.88%
Baseline (Preprocess + Conv3 + SE)	5,961,864	97.12%
Baseline (Preprocess + Conv3 + SE + CIF)	6,131,848	**97.22%**

通过上表可以发现，当采用预处理后的图像替换原图时，网络输入层数据均是大小、像素、位置分布均匀的图像，因此该网络模型在基线模型的基础上提升了 0.05%，然而其识别精度仍然处于一个较低的水平。其次，将 ShuffleNetV2 中的深度可分离卷积采用标准 3×3 卷积替换，尽管网络模型的参数量较原模型有所提升，但其识别精度却上升到了 96.88%。为进一步提升网络性能，在替换的标准卷积后采用SE注意力模块以获取文字重要局部特征，此时的网络模型识别精度又提升了 0.24%，且已经能够与近几年优秀的网络进行比较。最终，通过在卷积组间与卷积块间引入跨层信息融合操作，使得网络获得了 97.22%的识别精度。

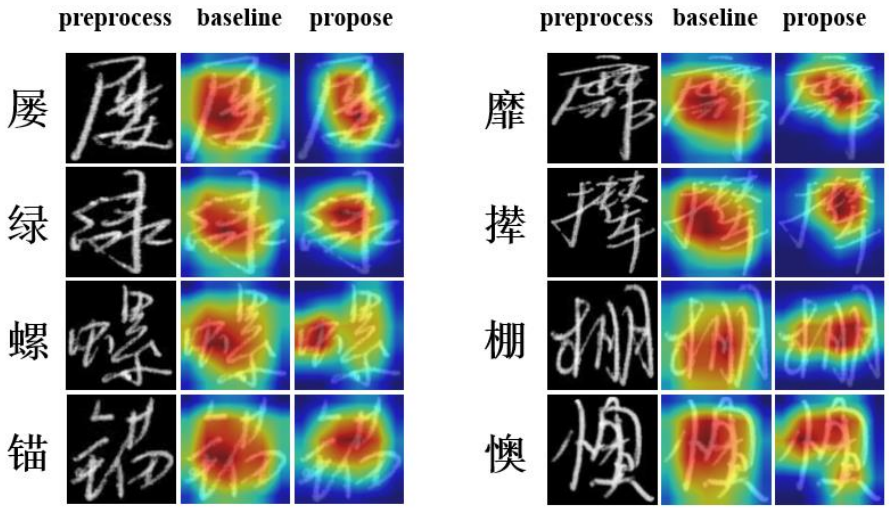


图 3-4 视觉注意力特征图

其中，“preprocess”表示预处理后的字符图像。“baseline”代表由基线模型预测时的视觉注意力图。“propose”代表提出的模型预测时的视觉注意力图。

为进一步观察网络对手写文字特征学习的注意力分布，将训练好的模型识别字符图像，并展示可视化效果，如图 3-4 所示。通过观察网络学习的特征分布图可以发现，基线网络识别文字的注意力热图相比于提出的网络要更加分散，同时提出的网络识别文字的热图则要更加聚焦于文字的关键特征。因此，提出的网络对于重要字符特征的关注相比于基线网络要更加准确，因此证实了高性能卷积与跨层信息融合操作的有效性。

3.5 本章小结

本章主要提出了一个基于高性能卷积与跨层信息融合的脱机 HCCR 方法。针对目前深层网络参数量普遍冗余的情况，通过引入轻量级卷积神经网络 ShuffleNetV2 中的卷积块来构建网络模型。同时对于轻量级卷积块性能不足的问题，加入了注意力机制来优化特征提取。为进一步加强脱机 HCCR 识别性能，该模型在高性能卷积组间与高性能卷积块间引入了跨层信息融合的操作。最终，在使用 CASIA-HWDB1.0-1.1 数据集训练的单网络模型在 ICDAR-2013 数据集上获得了 97.22%的准确度，且仅占有 23.4MB 的存储空间。

第4章 基于多尺度卷积混洗与空间聚合的脱机 HCCR

4.1 引言

针对第三章中模型在识别精度与存储容量方面的缺陷，本章进一步设计了一个基于多尺度卷积混洗与空间聚合的高效脱机 HCCR 网络。首先，为获取输入图像的多个感受野的特征信息，提出一个新的模块：多尺度卷积混洗(Multiple Scale Convolution Shuffle，MSCS)模块。其次，结合注意力机制与空间聚合特性，提出了一个注意力特征空间聚合(Attention Feature Spatial Aggregation，ASA)方法来代替全连接层的降维操作。该模块不仅体现了特征图中重要的空间映射，而且大幅降低了冗余参数。此外，为提高识别相似字符的准确率，引入了中心损失函数(center loss function)[66]来学习每个字符类中的注意力特征。实验结果表明，在不增加卷积层参数计算量的情况下，该网络能够有效提升模型精度。提出的网络在 ICDAR-2013 竞赛数据集上具有 97.63%的高准确率，该准确率下的存储空间仅为 22.9MB，同时单个字符图像识别时间仅为 3.97ms。简而言之，本章主要内容如下：

(1) 首先，提出了一个多尺度卷积混洗模块，该模块能够学习多个感受野特征信息，从而获取到更加丰富的字符特征，最终取得更好的鲁棒性和可解释性。

(2) 此外，提出了一种注意特征空间聚合降维方法，它不仅可以在整个特征图区域中获取重要特征分布来提高模型的学习能力，同时还能大幅降低特征降维时带来的参数冗余。

(3) 最终，将 softmax 损失函数[67]与中心损失函数相结合，增强了类间和类内特征的区分度，从而有效提高了模型的分类性能。

4.2 网络结构设计

4.2.1 网络总体架构

目前，基于脱机 HCCR 的高效卷积神经网络已经被广泛应用。Melnyk 等[5]采用了一种先进的单网络特征提取框架，该框架使用两个卷积层和四个高效卷积块进行特征提取。其中，每个卷积块的中间层包含了压缩卷积模型参数的瓶颈层(Bottleneck layer)。为了进一步提高模型的识别性能，本章基于上述方法进行了一些改进，网络整体框架如表 4-1 和图 4-1 所示。

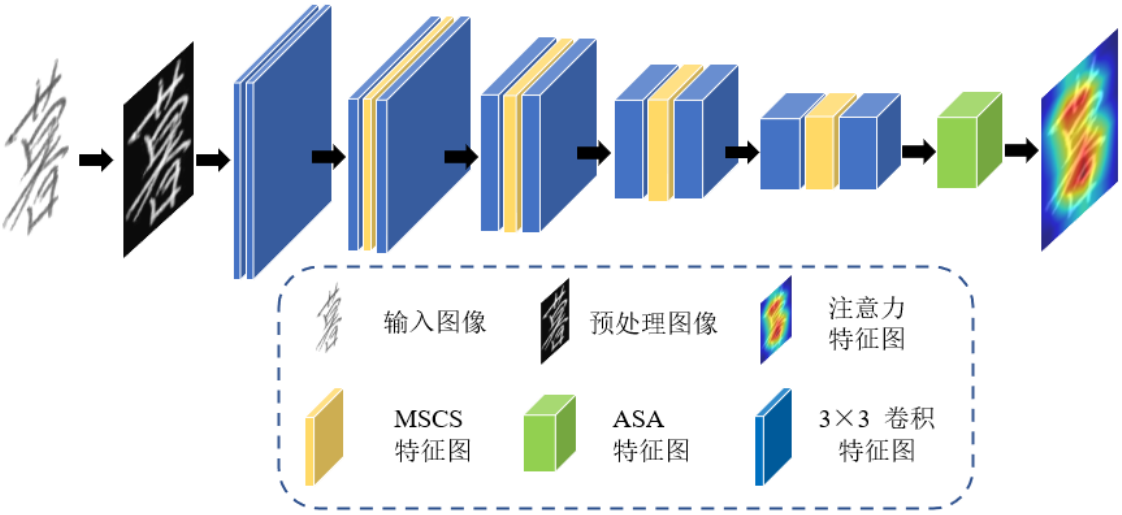


图 4-1 多尺度卷积混洗与空间聚合网络总体架构

表4-1 基线网络和提出网络的整体框架

“Baseline”代表基线模型；“Proposed”代表提出的模型。“feature Aggregation”表示特征聚合层。

Layers	Baseline	Proposed	Output Shape
Input	preprocessed image	preprocessed image	96 × 96 × 1
Conv_Layers	3×3 conv. 64, BN, ReLu 3×3 conv. 64, BN, ReLu	3×3 conv. 64, BN, ReLu 3×3 conv. 64, BN, ReLu	96 × 96 × 64 96 × 96 × 64
AvgPool1	3×3 avg-pool stride 2	3×3 avg-pool stride 2	48 × 48 × 64
Conv_Block1	3×3 conv. 96, BN, ReLu 3×3 conv. 64, BN, ReLu 3×3 conv. 96, BN, ReLu	3×3 conv. 96, BN, ReLu MSCS module. 64 3×3 conv. 96, BN, ReLu	48 × 48 × 96 48 × 48 × 64 48 × 48 × 96
AvgPool2	3×3 avg-pool stride 2	3×3 avg-pool stride 2	24 × 24 × 96
Conv_Block2	3×3 conv. 128, BN, ReLu 3×3 conv. 96, BN, ReLu 3×3 conv. 128, BN, ReLu	3×3 conv. 128, BN, ReLu MSCS module. 96 3×3 conv. 128, BN, ReLu	24 × 24 × 128 24 × 24 × 96 24 × 24 × 128
AvgPool3	3×3 avg-pool stride 2	3×3 avg-pool stride 2	12 × 12 × 128
Conv_Block3	3×3 conv. 256, BN, ReLu 3×3 conv. 192, BN, ReLu 3×3 conv. 256, BN, ReLu	3×3 conv. 256, BN, ReLu MSCS module. 192 3×3 conv. 256, BN, ReLu	12 × 12 × 256 12 × 12 × 192 12 × 12 × 256
AvgPool4	3×3 avg-pool stride 2	3×3 avg-pool stride 2	6 × 6 × 256
Conv_Block4	3×3 conv. 448, BN, ReLu 3×3 conv. 256, BN, ReLu 3×3 conv. 448, BN, ReLu	3×3 conv. 448, BN, ReLu MSCS module. 384 3×3 conv. 448, BN, ReLu	6 ×6 × 448 6 ×6 × 256 / 384 6 ×6 × 448
Feature Aggregation	FC 1024 Dropout	ASA module 448 Dropout	1024 / 448
Output	3755-dim Softmax		3755

受到 shuffleNetV2[47]中轻量级卷积块设计思想的启发，本章设计了一个 MSCS 高效卷积模块，它能够代替基线模型中卷积块的瓶颈层，从而进一步提高特征提取器的效率。此外，该模块通过加深最后一个 MSCS 卷积块的输出通道数，来获取图像中的深层特征。受文献[5, 58]对 GAP 的加权改进与注意力机制应用的影响，进一步提出 ASA 特征聚合方法来处理空间降维中的参数冗余和性能不足等问题。本章的研究流程如下:

(1) 首先使用 MSCS 模块提取器，获取字符图像中较为深层的重要特征，其次

将特征提取器输出的特征图引入 ASA 模块以获取全局特征并进行特征压缩。

(2) 接着使用 softmax 损失和 center 损失作为联合损失函数，以学习类间和类内注意特征的差异，从而提高模型在特征分类时的判别能力。

4.2.2 多尺度卷积混洗模块

卷积层作为CNN中最关键的特征提取器，其参数量与计算量的大小往往决定着网络模型的性能与部署能力。为搭建一个高性能特征提取模块，本文提出了一个多尺度卷积混洗(MSCS)模块，类似于文献[5]中瓶颈层的作用，如图 4-2 所示。该模块首先进行通道分割操作，然后对两通道特征采用 1×1 与 3×3 卷积分别提取特征，接着将二者进行拼接并混洗各通道。该模块不仅优化了卷积冗余的参数量，还增强了通道特征之间的信息交流，从而提升了网络的表现力。

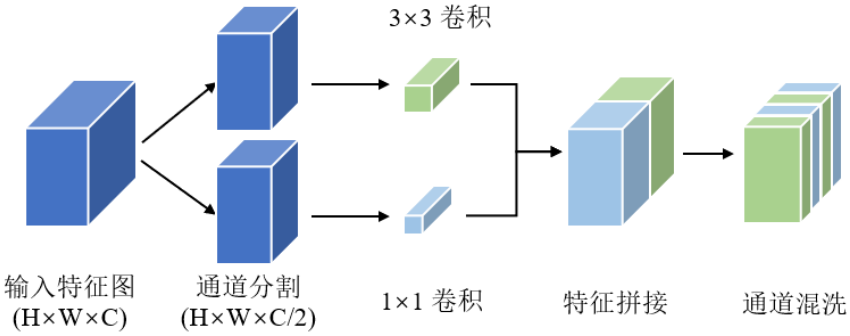


图 4-2 MSCS 模块结构

模型存储空间的计算量可以通过卷积块的乘法累加数(Multiply Accumulate, MAC)进行判断，MAC 是通过将输入特征图的高度和宽度、输入通道的数量、输出通道的数量以及卷积核的高度、宽度相乘而获得的。为了使输出特征图的大小在卷积后保持不变，卷积的填充值和步长被设置为 1。因此，单个卷积层的乘法累加数 N_{MAC} 计算如下:

$$N_{MAC} = 9 \cdot H \cdot W \cdot I \cdot O \tag{4-1}$$

其中 H 、W 和 I 表示输入特征图的高度、宽度和通道数。由于网络使用 3×3 卷积核，所以 9 代表卷积核的高度和宽度的乘积。另外，O 代表卷积核的个数。

接着，假设卷积块中卷积层的输出通道数一致，同时卷积层的图像大小不变，则未使用瓶颈层的标准卷积块的乘法累加数 N_{MAC_CB} 计算公式如下:

$$N_{MAC\_CB} = 9 \cdot H \cdot W \cdot I \cdot O + 18 \cdot H \cdot W \cdot O^2 \tag{4-2}$$

Melnyk 等[5]在卷积块的中间层添加了一个输出为 $H \times W \times O_b$ 的瓶颈层。基于上述条件，则改进的卷积块的乘法累加数 N_{MAC_CBb} 与标准卷积块计算参数 N_{MAC_CB} 的比值如下:

$$N_{MAC\_CBb} = 9 \cdot H \cdot W \cdot I \cdot O + 18 \cdot H \cdot W \cdot O_b \cdot O \tag{4-3}$$

$$N_{MAC\_CB} / N_{MAC\_CBb} = (I + 2O) / (I + 2O_b) \tag{4-4}$$

假设 MSCS 模块的输出通道数也为 O_b，则提出的 MSCS 模块的乘法累加数 N_{MAC_CBm} 与标准卷积块计算参数 N_{MAC_CB} 的比值如下:

$$N_{MAC\_CBm} = 9 \cdot H \cdot W \cdot I \cdot O + \frac{5}{2} \cdot H \cdot W \cdot O_b \cdot O + 9 \cdot H \cdot W \cdot O_b \cdot O \tag{4-5}$$

$$N_{MAC\_CB} / N_{MAC\_CBm} = (I + 2O) / (I + 23O_b / 18) \tag{4-6}$$

比较公式 4-4 和公式 4-6，可以得出 MSCS 模块进一步减少了卷积层中的参数量。此外，其混洗操作还能够为多个组卷积层提供跨组信息流[46]，从而增强了特征图通道之间的信息交流。因此，该操作使得网络获得了更加丰富的特征信息，并在一定程度上能够提高网络精度。

4.2.3 注意力特征空间聚合方法

分类任务中特征降维方法的好坏是影响特征提取优劣的重要因素。目前，基于全连接层完全映射的降维方法已经很少使用，因为该方法积累了大量的冗余参数。为了解决上述问题，研究者们开始在全局平均池化等降维方法中找寻最佳的特征聚合方案。GWAP，GWOAP[5, 11]通过引入可训练权重来代替全局平均池化中的空间平均参数，该类方法增强了网络在特征降维时对关键特征的提取，一定程度上提升了网络性能，同时解决了传统特征降维时参数冗余的问题。

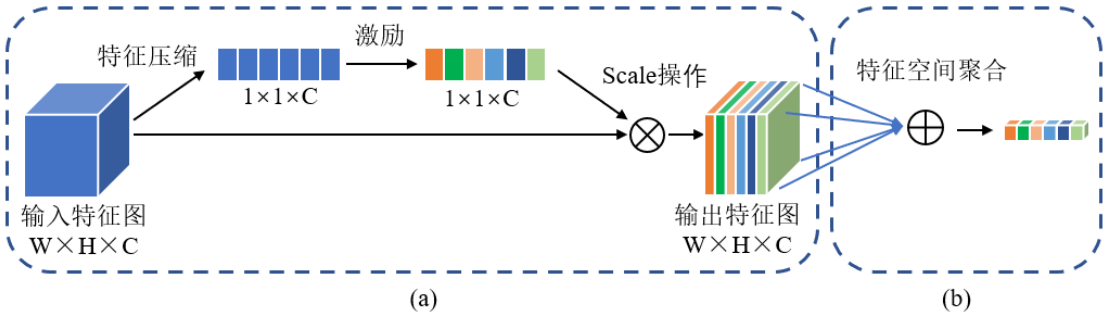


图 4-3 ASA 模块结构 (a) SE 模块 (b) 空间聚合操作

基于上述思想，本章提出注意力特征空间聚集模块，该模块在特征聚合方法中引入SE注意力模块，以加强模型在特征降维时对全局重要特征的学习能力。如图 4-3 (a)所示，首先，采用全局平均池化操作来压缩网络，对输入特征图各通道上的所有元素求取平均值以实现空间特征聚合。该操作不仅实现了特征图降维效果，而且大幅减少了特征映射过程中的参数冗余。通过以上描述，输入特征图第 c 通道的聚合特征值 g_c 的计算公式如下:

$$g_c = 1/(H \times W)\sum_{i=1}^{H}\sum_{j=1}^{W} f_{ijc} \tag{4-7}$$

其中，H 和 W 表示输入特征图的高度和宽度；f_{ijc} 代表特征图第 c 通道中第 i 行第 j 列的像素值。g_c 则表示特征图经过全局平均池化后第 c 通道的特征节点。

随后，使用激励操作来恢复特征图原始通道数以增强特征通道之间的信息交互，这类似于循环神经网络中的门控机制。通过以上描述，输入特征图第 c 通道特征的关注度 s_c 的计算公式如下:

$$s_c = \sigma(W_2\delta(W_1 g_c)) \tag{4-8}$$

其中，s_c 、σ 和 δ 表示第 c 通道特征图注意力、sigmoid 激活函数和 relu 激活函数。$W_1 \in R^{c/r\times c}$, $W_2 \in R^{c\times c/r}$ 表示神经网络中的权重参数。r 表示减速比，通常设置为 16。

与 GWAP 和 GWOAP 方法中引入的可训练权重不同，在图 4-3 (b)中可以看出，通过将公式(4-8)的注意力权重代替普通全局平均池化中的平均参数来进行最终的空间聚合，使得特征降维中的信息损失达到最低。综上，输入特征图在经过 ASA 操作后第 c 个通道的聚合特征值 x_c 的定义如下：

$$x_c = \sum_{i=1}^{H}\sum_{j=1}^{W} s_c \cdot f_{ijc} \tag{4-9}$$

通过以上 ASA 降维方法的详解，可以发现该方法有效结合了注意力机制与降维特性，不仅在降维操作中较少了分类精度的损失，同时也避免了特征聚合时参数冗余的问题。

4.3 网络参数与损失函数设计

4.3.1 网络参数设置

网络整体结构参数如表 4-1 所示。首先，本章同样引入了一个基线网络用以进行对比试验，该网络卷积层部分主要由 3×3 卷积核组成，且每次卷积后都采用

批归一化操作和 ReLU 激活函数来加快网络训练速度，有效避免了梯度弥散与梯度消失问题。此外，使用平均池化进行全局特征的采集，且池化后的特征图长宽变为原来的一半。特征提取部分主要由 2 层卷积层和 4 个卷积块构成，且随着层数的不断加深，网络最终输出 6×6×448 大小的特征图。在特征聚合部分则采用了全连接层进行降维。

与此同时，本章提出的网络采用 MSCS 模块替换原先卷积块中的瓶颈层，该模块使用 1×1 与 3×3 卷积以获取多尺度特征。此外，卷积层与池化层的参数设置跟基线模型中的相同，且各层输出通道维数也基本一致。最终，为实现模型参数的压缩，提出了 ASA 降维方法替代传统的全连接操作，该方法输出的特征通道数为 448。

4.3.2 损失函数设置

为了进一步优化脱机 HCCR 网络模型的分类精度，本章结合 Softmax 损失函数和中心损失函数对输入特征进行联合监督并学习区分性特征。首先，使用 Softmax 损失函数来计算预测值与真实样本之间的误差，该损失函数本质上是一个结合了 Softmax 函数的交叉熵损失函数。因此，Softmax 损失函数 L_S 表达式为：

$$L_S = -\frac{1}{m}\sum_{i=1}^{m}\log(e^{W_{y_i}x_i+b_{y_i}} / \sum_{j=1}^{n}e^{W_j x_i+b_j}) \tag{4-10}$$

其中，m 表示一批量样本的数量，y_i 表示第 i 个样本的真实值。x_i，W_{y_i} 和 b_{y_i} 分别表示第 i 个样本的特征图、神经网络的权值项和偏置项。此外，字符类总数采用 n 表示，即输入特征通道维数。因此，分子的指数部分 $W_{y_i}x_i+b_{y_i}$ 代表通过神经网络预测为第 y_i 类样本的第 i 个样本的输出分数。同时，分母的指数部分 $W_j x_i+b_j$ 表示通过神经网络预测为其他类别的分数。简而言之，上述公式给出了一批量样本中类间损失的总和。

虽然上述的 softmax 损失函数具有很强的类间特征区分能力，但对于汉字识别任务中较多的相似字而言就不够全面。因此，提出的网络引入了中心损失函数，通过缩小输入特征与同一类的特征中心之间的距离来区分类内中的非同类特征。中心损失函数 L_C 的计算公式如下:

$$L_C = \frac{1}{2}\sum_{i=1}^{m}\left\|x_i - c_{y_i}\right\|_2^2 \tag{4-11}$$

其中 c_{y_i} 是第 y_i 类样本中所有特征的平均值(特征中心)，其维度与注意力特征 x_i 一致。此外，c_{y_i} 的更新随着不同样本的中心特征而改变。

然而，当单独使用 softmax 损失函数进行监督时，类内特征的可变性使得模

型很难区分存在于类内特征中的非同类特征。此外，仅使用中心损失函数来监督网络模型时，深层特征和中心将减少到零[66]，这将导致分类性能过度。因此，网络将分类优化 softmax 损失函数与度量学习的中心损失函数相结合，通过权重 λ 来控制中心损失函数的占比，从而有效提升脱机 HCCR 中相似字识别的准确率。综上所述，联合损失 L 的表达式为:

$$L = L_S + \lambda L_C \tag{4-12}$$

4.4 实验结果与分析

4.4.1 训练环境与参数设置

与第三章节类似，将输入的字符图像尺寸设为 96×96。此外，将初始学习率设置为 0.1，同样在精度停止提升时，学习率将变为原始学习率的 0.1 倍。在加速模型训练方面采用批归一化方法，并使用 L2 正则化操作避免模型过拟合，并将权重衰减参数设置为 10^{-3}。随后，使用随机梯度下降法进行模型优化，将动量、最小批量样本数和所有样本训练次数分别设置为 0.9、128 和 20。同时，在 softmax 层之前添加了 dropout[10]，并将输出比例设置为 0.5。此外，使用基于 tensorflow 框架的 Keras 深度学习库来构建所提出的模型，并在 11GB 内存的 NVIDIA GeForce GTX 2080 Ti 上进行了实验。

4.4.2 ICDAR-2013 数据集上的性能对比

表 4-2 显示了近年来研究人员对 ICDAR-2013 数据集提出的不同方法的比较。可以发现，本章提出的网络模型在准确性、存储空间、计算量和推理时间方面几乎是所有模型中最佳的。

由于 2017 年之前基于深度学习的脱机 HCCR 方法只考虑了模型精度和参数容量，因此早期某些方法在模型部署时计算成本和推理时间方面没有相关数据记录。由表 4-2 可以发现，基于方向特征图预处理操作提出的 GoogLeNet-Ensemble-10[3] 和 DirectMap-ConvNet-Adaption[8]虽然取得了显著的效果，但它们增加了输入数据的维数，从而增大了特征提取时的复杂度。然而，提出方法的输入特征在仅使用了简单的图像填充与阈值反转操作后，其准确率在上述两种方法的基础上分别提高了 0.89%和 0.26%。另外，STN-Residual-34[6]在特征预处理中使用了方向归一化，取得了 97.37%的高精度。尽管如此，提出的模型在初始角度下，其精度较该方法仍然提升了 0.26%，并减少了 69.4MB 的空间损耗。

为优化模型参数，使得模型能够更加轻量化。Xiao 等[4]提出了 GSLRE 和

ADW 的参数剪枝方法获得了较为显著的成果。然而，提出的模型在没有剪枝策略的情况下，其容量仅有22.9MB。此外，该模型还具有更高的精度和更快的推理时间。表中 Cascaded Model[11]的结构与 MSCS 模块类似，它们都使用了多尺度卷积来获取丰富性的特征信息。然而相比于提出的模型，它仍然有 0.49%的精度劣势。

表 4-2 在 ICDAR-2013 竞赛数据集上不同方法性能的比较

"FLOPs"表示每秒的浮点运算，即计算量。"Inference time"代表单个样本的预测时间。所有方法都在 HWDB1.0-HWDB1.1 数据集下训练。

Method	Size (MB)	Accuracy (%)	FLOPs ($\times 10^8$)	Inference time (ms)
Human Level Performance[12]	n/a	96.13	n/a	n/a
HCCR-Gabor-GoogLeNet[3]	27.7	96.35	3.58	n/a
HCCR-GoogLeNet-Ensemble-10[3]	270.0	96.74	35.8	n/a
Residual-34[6]	92.2	97.36	n/a	n/a
STN-Residual-34[6]	92.3	97.37	n/a	n/a
DCNN-Similarity ranking[65]	36.2	97.07	n/a	n/a
Ensemble DCNN-Similarity ranking[65]	144.8	97.64	n/a	n/a
DirectMap + ConvNet[8]	23.5	96.95	2.63	2.46 (GPU)
DirectMap + ConvNet + Ensemble-3[8]	70.5	97.12	7.89	n/a
DirectMap + ConvNet + Adaptation[8]	23.5	97.37	2.67	n/a
M-RBC + IR[57]	n/a	97.37	n/a	**0.623 (GPU)**
HCCR-CNN9Layer[4]	41.5	97.30	5.94	492 (CPU)
HCCR-CNN12Layer[4]	48.7	97.59	12	n/a
Cascaded Model (Quantization)[11]	**3.3**	97.11	**1.35**	6.90 (CPU)
Cascaded Model[11]	20.4	97.14	1.41	6.93 (CPU)
Melnyk-Net[5]	24.9	97.61	21.44	4.15 (GPU)
MCANet[58]	>500	**97.66**	21.51	n/a
Our Network	22.9	97.63	17.62	3.97 (GPU)

近年来，基于注意力的脱机 HCCR 方法开始逐渐普及。Yang 等[57]提出了一种基于多尺度残差块和注意力的迭代细化模块，该模块能够融合低层和高层视觉信息以获取丰富的特征信息。尽管如此，其字符分类精度较提出模型还是低了0.26%。此外，MCANet[58]是一个将多重注意力机制与联合损失函数相结合的方法，其精度已经达到了目前最为先进的水平。然而，该网络模型的空间降维部分仍采用完全映射的拉伸操作，从而导致模型中出现了较多的冗余参数，因而使得模型参数空间巨大。同时，MCANet 与提出的模型相比，其损耗了较多的卷积层计算量。MelnykNet[5]作为一个具有高效卷积模块和降维方法的卷积网络，其性能相比于提出的模型依旧有着 0.02%的精度和 2MB 的空间的劣势。此外，提出的网络模型还具有较低的计算成本和推理时间。

4.4.3 消融实验

为证明提出模型的可行性，本章进行了广泛的消融实验。其中，基线模型如表 4-1 所示。在特征提取后，使用两个全连接层进行特征压缩，最终在测试集上获得 97.47%的准确率。通过表 4-3 可以观察不同网络设置在测试集上的性能对比。

表 4-3 提出的网络在 ICDAR-2013 测试集上不同配置的性能对比
“MSCS”表示图 4-2 中的多尺度卷积混洗模块。“ASA”表示公式(4-9)中的注意力特征空间聚集模块。“SCL”表示由 softmax 和中心损失函数组成的联合损失函数。

Method	Parameters	Accuracy
Baseline	25,491,092	97.47%
Baseline (MSCS)	24,866,292	97.52%
Baseline + ASA	6,523,819	97.57%
Baseline (MSCS) + ASA	6,011,591	97.59%
Baseline (MSCS) + ASA + (SCL)	**6,011,591**	**97.63%**

如表 4-3 所示，当使用 MSCS 模块替换瓶颈层时，该模型可以获得丰富的多尺度卷积特征和通道交流信息。因此，基线模型精度提升了 0.05%，然而模型的参数仍然处于过度冗余状态。为了避免这种情况，在基线模型中使用 ASA 操作代替全连接层映射进行特征空间降维。与此同时，该操作中的聚合特征维度的减少仅伴有少量的特征信息损失。最终，达到了 97.57 %的准确率，但模型参数却减少到原来的 1/40。之后，结合 MSCS 模块和 ASA 方法构建了一个新的网络，其准确率高达 97.59%，且参数下降到了 600 万。考虑到脱机 HCCR 中出现的相似字问题，通过引入 softmax 损失和中心损失联合监督来增强类间和类内注意特征的学习能力，并通过一个权重参数来合理控制中心损失的比例。最终，提出的模型实现了 97.63%的准确率，同时其参数处在一个较为合理的水平，且与仅使用 softmax 损失函数的模型相比提高了 0.04%的准确度。

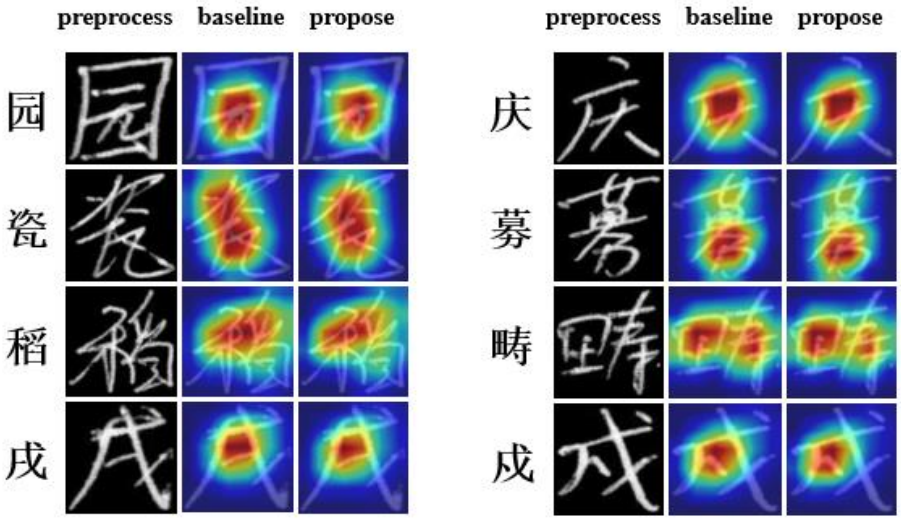


图 4-4 视觉注意力特征图

其中，“preprocess”表示预处理后的字符图像。“baseline”代表由本章基线模型预测时的视觉注意力图。“propose”代表本章提出的模型预测时的视觉注意力图。

通过图 4-4 所展示的可视化注意力特征图可以发现，提出的模型相比于基线模型对文字特征的定位热图要更加精确与直观，表明了 MSCS 模型、ASA 方法和联合损失函数对脱机 HCCR 有效性。

4.5 本章小结

针对第三章高效脱机 HCCR 模型在终端部署时识别精度与存储容量的不足，本章进一步提出了一个更加高效的脱机 HCCR 网络框架，同时在实验部分添加了两个评价指标：计算量与推理时间。首先，提出一个多尺度卷积混洗(MSCS)模块来获取丰富的感受野特征，以提高脱机 HCCR 模型的性能。接着，进一步提出了注意力特征空间聚合(ASA)方法，有效改善了特征空间降维中的信息丢失和参数冗余。然后，结合 softmax 损失和中心损失作为联合损失函数，以增强类间和类内注意特征的学习能力。最终，在考虑存储空间的情况下，使用 CASIA-HWDB1.0-1.1 数据集训练的单网络模型在 ICDAR-2013 数据集上取得了较为先进的结果。

第5章 总结与展望

5.1 总结

本文分析了脱机手写体汉字识别应用中的模型部署存在问题，主要包括模型识别精度较低、存储容量大、计算量大与推理时间较长等情况。因此，本文根据高效卷积模块、注意力机制及特征融合的思想提出了高性能脱机 HCCR 解决方案。首先，针对脱机 HCCR 训练数据集字符图像大小、位置及像素分布不均的问题采用预处理操作，一定程度上提升了模型性能。其次，基于高性能卷积与跨层信息融合操作初步优化网络模型的空间占用与识别精度。同时，为满足移动端脱机 HCCR 模型高性能部署的需求，提出了基于多尺度卷积混洗模块与空间聚合方法，进一步完善并提升了网络模型在终端部署时的各项指标。本文主要工作如下：

(1) 通过观察输入层字符图像数据，可以发现训练集图像中的字符存在大小不一、位置与像素分布不均的情况。因此，本文通过采用边界填充与像素值反向进行预处理操作，使得输入层图像中的字符均处于相对居中的位置，类似于字符像素的归一化处理。最终的实验结果表明，使用预处理图像的网络模型性能要强于采用原始图像的网络模型。

(2) 通过对目前现有的高效卷积网络进行研究，本文提出了基于高性能卷积与跨层信息融合的网络，有效降低了网络模型空间占用并提升了模型识别精度。该网络针对 ShuffleNetV2 网络中轻量级卷积块的结构进行了改进，并引入注意力机制改善了网络提取特征时的关注度。同时，通过在改进的卷积块间与卷积组间使用残差连接与密集连接等跨层信息融合的操作，对低层特征信息进行复用，从而采用少量卷积便能获取深层特征信息，一定程度上提升了脱机 HCCR 模型的识别精度。

(3) 针对第三章提出的脱机 HCCR 网络在移动端部署中可能存在的不足，本章提出了基于多尺度卷积混洗与空间聚合的网络，以进一步完善并优化模型高性能部署时的各项指标。该网络中的多尺度卷积混洗模块不仅能够获取到多层感受野特征信息，而且其轻量化结构大幅降低了网络模型的参数量。其次，本章提出注意力特征聚合方法来进行特征降维，该方法结合了全局平均池化与注意力机制的思想，进一步优化了网络在空间聚合时对重要特征的提取。

此外，本文的研究取得了如下成果及优势：

(1) 本文所提出的高性能卷积与跨层信息融合网络首次将轻量化卷积应用于

脱机 HCCR 任务中，推动了 HCCR 模型在移动端的高性能部署。该网络模型在公共数据集ICDAR-2013上获得了97.22%的准确率，且仅占有23.4MB的内存空间。

(2) 本文所提出的多尺度卷积混洗与空间聚合网络进一步增强了模型在移动端部署时的性能，该网络在公共数据集上获得了 97.63%的精度且仅占有 22.9MB 的存储空间。相比于近几年其他顶尖脱机 HCCR 算法模型，该网络在模型识别精度、存储空间、计算量与推理时间等指标上均有着不错的成绩。

5.2 展望

本文提出的高性能脱机 HCCR 网络模型虽然有效改善了网络模型的参数量、计算量与识别精度，并获得了目前较为先进的水平，然而在移动端高性能部署上提出的算法模型还存在一些不足，需进一步研究：

在第三章改进的轻量级卷积中，采用了普通的标准卷积替换深度可分离卷积以提升网络模型性能，该方法仍然存在一些冗余计算。为进一步提升网络性能并优化卷积参数量，可考虑采用扩张卷积来代替标准卷积以扩大提取的特征范围。同时，可以对本文高性能卷积块中的SE注意力模块进一步优化，采用1×1卷积替换对特征压缩与扩张的全连接层，从而能够确保图像空间结构的完整性。

在第四章提出的多尺度卷积混洗模块中，仅采用 1×1 与 3×3 两尺寸的卷积进行特征提取，因此可考虑引入 5×5 尺寸卷积进一步获取多层感受野特征。此外，特征聚合方法可引入多重注意力机制联合进行决策，从而增强网络对空间特征的学习能力。

参考文献

[1] Casey R, Nagy G. Recognition of printed Chinese characters[J]. IEEE Transactions on Electronic Computers, 1966, (1): 91-101.

[2] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.2329, 2014.

[3] Zhong Z, Jin L, Xie Z. High performance offline handwritten chinese character recognition using googlenet and directional feature maps[C]. 2015 13th International Conference on Document Analysis and Recognition (ICDAR), 2015. 846-850.

[4] Xiao X, Jin L, Yang Y, et al. Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition[J]. Pattern Recognition, 2017, 72: 72-81.

[5] Melnyk P, You Z, Li K. A high-performance CNN method for offline handwritten Chinese character recognition and visualization[J]. soft computing, 2020, 24 (11): 7977-7987.

[6] Zhong Z, Zhang X-Y, Yin F, et al. Handwritten Chinese character recognition with spatial transformer and deep residual networks[C]. 2016 23rd international conference on pattern recognition (ICPR), 2016. 3440-3445.

[7] 肖正欣. 基于深度学习的离线手写汉字识别算法研究与实现[D]. 电子科技大学, 2021.

[8] Zhang X-Y, Bengio Y, Liu C-L. Online and offline handwritten chinese character recognition: A comprehensive study and new benchmark[J]. Pattern Recognition, 2017, 61: 348-360.

[9] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International conference on machine learning, 2015. 448-456.

[10] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15 (1): 1929-1958.

[11] Li Z, Teng N, Jin M, et al. Building efficient CNN architecture for offline handwritten Chinese character recognition[J]. International Journal on Document Analysis and Recognition (IJDAR), 2018, 21 (4): 233-240.

[12] Yin F, Wang Q-F, Zhang X-Y, et al. ICDAR 2013 Chinese handwriting recognition competition[C]. 2013 12th international conference on document analysis and recognition, 2013. 1464-1470.

[13] Guo Y, Yao A, Chen Y. Dynamic network surgery for efficient dnns[J]. arXiv preprint arXiv:1608.04493, 2016.

[14] 林泽建, 李珍妮, 谢影, 等. 基于尺度不变的结构稀疏化神经网络剪枝方法[J]. 中国自动化大会论文集, 2021.

[15] 张明明, 卢庆宁, 李文中, 等. 基于联合动态剪枝的深度神经网络压缩算法[J]. 计算机应用, 2021, 41 (6): 1589.

[16] Jacob B, Kligys S, Chen B, et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. 2704-2713.

[17] 龚成, 卢冶, 代素蓉, 等. 一种超低损失的深度神经网络量化压缩方法[J]. Journal of Software, 2021, 32 (8): 2391-2407.

[18] Jaderberg M, Vedaldi A, Zisserman A. Speeding up convolutional neural networks with low rank expansions[J]. arXiv preprint arXiv:1405.3866, 2014.

[19] 余沁茹, 卢桂馥, 李华. 自适应图正则化的低秩非负矩阵分解算法[J]. 智能系统学报, 2021, 17 (2): 325-332.

[20] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J]. arXiv preprint arXiv:1503.02531, 2015.

[21] 刘昊, 张晓滨. 基于关系型蒸馏的分步神经网络压缩方法[J]. 计算机系统应用, 2021, 30 (12): 248-254.

[22] 孙显, 杨竹君, 李俊希, 等. 基于知识自蒸馏的轻量化复杂遥感图像精细分类方法[J]. 指挥与控制学报, 2021, 7 (4): 365-373.

[23] 林恒青. 基于深度卷积神经网络的脱机手写汉字识别系统的设计与实现[J]. 湖北理工学院学报, 2021, 2019 (2): 31-34.

[24] Smith R W. Hybrid page layout analysis via tab-stop detection[C]. 2009 10th International Conference on Document Analysis and Recognition, 2009. 241-245.

[25] Du Y, Li C, Guo R, et al. Pp-ocr: A practical ultra lightweight ocr system[J]. arXiv preprint arXiv:2009.09941, 2020.

[26] Peng C, Xiao T, Li Z, et al. Megdet: A large mini-batch object detector[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018. 6181-6189.

[27] Kim H J, Kim K H, Kim S K, et al. On-line recognition of handwritten Chinese characters based on hidden Markov models[J]. Pattern Recognition, 1997, 30 (9): 1489-1500.

[28] Mangasarian O L, Musicant D R, "Data discrimination via nonlinear generalized support vector machines[M]," in Complementarity: Applications, Algorithms and Extensions. Springer, 2001, pp. 233-251.

[29] Kimura F, Takashina K, Tsuruoka S, et al. Modified quadratic discriminant functions and the application to Chinese character recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987, (1): 149-153.

[30] Liu C-L, Sako H, Fujisawa H. Discriminative learning quadratic discriminant function for handwriting recognition[J]. IEEE Transactions on Neural Networks, 2004, 15 (2): 430-444.

[31] Cireşan D, Meier U. Multi-column deep neural networks for offline handwritten Chinese character classification[C]. 2015 international joint conference on neural networks (IJCNN), 2015. 1-6.

[32] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 770-778.

[33] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[J]. Advances in neural information processing systems, 2015, 28: 2017-2025.

[34] Xiao Y, Meng D, Lu C, et al. Template-instance loss for offline handwritten Chinese character recognition[C]. 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019. 315-322.

[35] Chen L, Peng L, Yao G, et al. A modified inception-ResNet network with discriminant weighting loss for handwritten chinese character recognition[C]. 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019. 1220-1225.

[36] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[37] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[38] Xu L, Wang Y, Li X, et al. Recognition of handwritten Chinese characters based on concept learning[J]. IEEE Access, 2019, 7: 102039-102053.

[39] Min F, Zhu S, Wang Y. Offline Handwritten Chinese Character Recognition Based on Improved Googlenet[C]. Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition, 2020. 42-46.

[40] 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.

[41] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 1-9.

[42] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[43] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. 4510-4520.

[44] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019. 1314-1324.

[45] 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.

[46] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. 6848-6856.

[47] Ma N, Zhang X, Zheng H-T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. Proceedings of the European conference on computer vision (ECCV), 2018. 116-131.

[48] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020. 1580-1589.

[49] Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]. Advances in neural information processing systems, 2014. 2204-2212.

[50] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.

[51] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Advances in neural information processing systems, 2017. 5998-6008.

[52] Woo S, Park J, Lee J-Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV), 2018. 3-19.

[53] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1 (4): 541-551.

[54] 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.

[55] Hinton G E, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18 (7): 1527-1554.

[56] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60 (6): 84-90.

[57] Yang X, He D, Zhou Z, et al. Improving offline handwritten Chinese character recognition by iterative refinement[C]. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017,1. 5-10.

[58] Xu Q, Bai X, Liu W. Multiple Comparative Attention Network for Offline Handwritten Chinese Character Recognition[C]. 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019. 595-600.

[59] Berman M, Jégou H, Vedaldi A, et al. Multigrain: a unified image embedding for classes and instances[J]. arXiv preprint arXiv:1902.05509, 2019.

[60] Lin M, Chen Q, Yan S. Network in network[J]. arXiv preprint arXiv:1312.4400, 2013.

[61] Sudholt S, Fink G A. Phocnet: A deep convolutional neural network for word spotting in handwritten documents[C]. 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016. 277-282.

[62] Zhang H, Guo J, Chen G, et al. HCL2000-A large-scale handwritten Chinese character database for handwritten character recognition[C]. 2009 10th International Conference on Document Analysis and Recognition, 2009. 286-290.

[63] Liu C-L, Yin F, Wang D-H, et al. CASIA online and offline Chinese handwriting databases[C]. 2011 International Conference on Document Analysis and Recognition, 2011. 37-41.

[64] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. 4700-4708.

[65] Cheng C, Zhang X-Y, Shao X-H, et al. Handwritten chinese character recognition by joint classification and similarity ranking[C]. 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016. 507-511.

[66] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C]. European conference on computer vision, 2016. 499-515.

[67] Liu W, Wen Y, Yu Z, et al. Large-margin softmax loss for convolutional neural networks[C]. ICML, 2016,2 (3). 7.