文章编号:1671-4598(2023)07-0199-08 DOI:10.16526/j.cnki.11—4762/tp.2023.07.029 中图分类号:TP753 文献标识码:A

一种改进残差深度网络的多目标分类技术

陈 超[1,2],吴 斌[1]
(1.西南科技大学 信息工程学院,四川 绵阳 621010;
2.四川省高等学校数值仿真重点实验室,四川 内江 641000)

摘要:由于受场景、视角、光照、尺度变化以及局部变形等因素的影响,对重叠目标、拥挤目标、小目标的识别精度较低,提出了一种改进多支路的残差深度卷积神经网络来提高多目标识别的准确度;在第一个卷积残差块 layer1 后保留恒等映射的同时,增加一个 1×1 的短接分支尽可能多的保留原始特征;再平行嵌入一个修改激活函数 ReLU6 的空间_通道注意力机制模块(CBAM);融合以上 3 个特征图;融合后的特征层着重关注空间和通道中比较显著的信息,从而增强特征图的特征表达能力,以至于卷积神经网络(CNN)获得更多的判别特征,从而大大提高物体识别精度;在 FashionMNIST 和 Cifar10 两个数据集的对比性实验显示改进的 resnet50 算法是准确性—速度较为折中的目标识别模型。

关键词:残差深度卷积神经网络;短接分支;CBAM;激活函数 ReLU6;多目标分类

A Multi-objective Classification Technique Based on Improved Residual Deep Network

CHEN Chao[1,2], WU Bin[1]
(1. School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China;
2. Numerical Simulation Key Laboratory of Sichuan University, Neijiang 641000, China)

Abstract: Due to the influences of scene, visual angle, illumination, scale change and local deformation, the recognition accuracies of overlapping target, crowded target and small target are low, an improved multi-branch Resnet50 convolutional neural network is proposed to improve the accuracy of multi-objective recognition. While retaining the constant maps after the first convolutional residual block Layer1; The short branch of one by one is added to preserve as many original features as possible; A space_channel attention mechanism module (CBAM) is embed to modify the activation function ReLU6 in parallel; The last three feature graphs are fused. The fused feature layer focuses on the significant information in space and channels, thus enhancing the feature expression ability in the feature graphs, so that the convolutional neural network (CNN) can obtain more discriminant features, thus greatly improving the accuracy of object recognition. The comparative experiment on the FashionMNIST and Cifar10 data sets shows that the improved Resnet50 algorithm is suitable for a target recognition model with a medium accuracy and speed.

Keywords: resnet; short circuit branch; CBAM; Activate function ReLU6; multi-objective classification

0 引言

图像识别的应用在现代社会的应用越发广泛,其重要程度越发明显。图像识别是计算机视觉中的一个重要环节,属于有监督学习类别,即在图片集或者视频中快速识别出其类别。近些年来一直是研究热点。传统的图像识别的方法是基于人工设置的一些特征进行匹配,然而识别速度较低,且对于多目标,遮挡,拥挤等场景中的目标识别效果很差。VGG[1]是牛津大学的 Visual Geometry Group 的组提出的。该网络是在 ILSVRC 2014 上的相关工作,主要工作是证明了增加网络的深度能够在一定程度上影响网络最终的性能。VGG 有两种结构,分别是 VGG16 和 VGG19,两者并没有本质上的区别,只是网络深度不一样。使用了 3 个 3×3 卷积核来代替 7×7 卷积核,使用了 2 个 3×3 卷积核来代替 5×5 卷积核,这样做的主要目的是在保证具有相同感知野的条件下,提升了网络的深度,在一定程度上提升了神经网络的效果。在 VGG[1]将 ImageNet 的 top-1 分类准确率提高到 70%以上后,在使卷积高性能变得复杂方面

收稿日期:2022-10-28; **修回日期**:2022-12-08。
基金项目:国家自然科学青年基金项目(11502121);四川省项目应用基础研究计划(2021JY0108);四川省科技研究计划项目(njfh20-003)。
作者简介:陈 超(1985-),男,四川内江人,博士研究生,讲师,主要从事目标识别和控制科学与工程方向的研究。
通讯作者:吴 斌(1964-),男,四川绵阳人,博士,教授,博士生导师,主要从事计算机视觉和控制科学与工程方向的研究。
引用格式:陈 超,吴 斌.一种改进残差深度网络的多目标分类技术[J].计算机测量与控制,2023,31(7):199-206.

有了很多创新。此时提升网络性能最直接的办法就是增加网络深度会引发诸多问题[2]：比如网络层数太多，一般计算机或者手持设备无法胜任，因为网络越深、参数越多，计算复杂度越大，难以应用；增加网络深度同时引入全连接变成稀疏连接来减少参数量。但在实现时计算所消耗的时间却很难减少。Google 研究人员提出了 Inception 的方法[2]，如当代 GoogLeNet[2]和后来的 Inception 模型[3-5]采用了精心设计的多分支架构，Inception 模块中并列提供多种卷积核的操作，网络在训练的过程中通过调节参数自己去选择使用，同时，由于网络中都需要池化操作，所以此处也把池化层并列加入网络中，使得网络的宽度得到了前所未有的扩展，因而得到了更为全面的特征提取网络结构，效果明显好于 VGG。何凯明的 ResNet[6]提出了简化的残差架构，在宽度和深度上都同时扩展了网络结构，尤其是增加了恒等连接，使得网络提取的特征始终不低于浅层特征，保证了网络可以加深加宽，在当年的图像识别，图像检测上取得了压倒式的成就。DenseNet[7]通过将低层与大量的高层连接起来，使得拓扑结构更加复杂，各层之间的特征进行了融合，保证了特征的丰富性。除了实现上的不便外，复杂的模型可能会降低并行度[8]，从而降低推理速度。提出了一种初始化方法[9]来训练极深普通卷积网络，保证了网络的鲁棒性。最近的工作[10-11]结合了几种技术，包括 Leaky _ ReLU 和 max-norm 等操作，保证了网络特征图提取的自然过渡，使得提取的特征更为真实。然而，Resnet 50 网络不能提取重叠目标、拥挤目标、小目标的某些特征。相反，改进算法的目标是建立一个具有合理深度和良好的精度－速度权衡的优秀模型，它只涉及最常见的组件（例如：卷积，批归一化和激活函数）。Resnet50 结构仅由 3×3 卷积、1×1 卷积和 sigmod 激活函数的堆栈组成。可以取得更好的效果受近年残差模块、1×1 卷积[2-6],[11]等思想，在第一个残差模块后再增加了 1×1 短接分支来提取更多的原始特征和调整与之深层匹配的通道数。以上网络中各层次之间的特征图没有信息的交互[13-15]，于是引入 convolutional block attention module（CBAM）表示卷积模块的注意力机制模块，是一种结合了空间（spatial）和通道（channel）的注意力机制模块[16-19]。但是现实图像中的特征彼此之间有位置，领域等信息，CBAM 提取空间注意力是用局部卷积，只能捕获局部的信息，无法获得长程依赖，基于此作者提出了 CA，充分利用了位置信息，而且控制了计算开销[20]。充分利用 channel 和 spatial 之间的关系，有人提出提出了 shuffle attention 注意力机制，首先将输入的特征分为 g 组，然后每一组的特征进行 split，分成两个分支，分别计算 channel attention 和 spatial attention，两种 attention 都使用全连接结合 sigmoid 的方法计算[20]。受多种注意力机制的启发，在 CNN 的第一个残差块层特征之后引入了空间 _ 通道注意力机制，并修改了激活函数，从而增强了特征的表示，贡献总结如下。

首先，为融合更多的原始特征，在 7×7 卷积特征 FeatureMap1 后添加一个 1×1 卷积残差，提取特征 FeatureMap2。

其次，在 Resnet50 的第一个残差块层特征之后，嵌入 CBAM 模块，使得 CNN 获得交互特征 FeatureMap3。继而综合融合 3 个特征，即：

（FeatureMap1＋FeatureMap2＋FeatureMap3）。

第三，在 CBAM 模块中，为了融合显著特征和缓解梯度消失，将 CBAM 模块中的激活函数 ReLU 修改为激活函数 ReLU6。

最后，在 FashionMNIST 数据集和 Cifar10 数据集上进行了实验，改进后的 Resnet50 达到了较好的识别精度和检测速度，用实验结果展示改进的 Resnet50 在分类方面的有效性和效率。

1 相关工作

Resnet50 将靠前若干层的某一层数据输出直接跳过多层引入到后面数据层的输入部分。保证后面的特征层的特征图含有的信息至少不会比前一层差，因此是最常用的残差结构的深度神经网络结构，这种深度残差网络的设计就很好地克服由于网络深度加深而产生的学习效率变低与准确率无法有效提升的问题[6]。Resnet50 比起 Resnet34 来说，深度更深，提取的特征更优质。比起 Resnet101，Resnet152 深度更浅，但是效果相差不是很大。根据 FashionMNIST 数据集和 Cifar10 数据集的特性以及训练模型的实际 GPU 性能，本文选择对 Resnet50 作为基础算法进行改进。为了更好地识别这两个数据集中的目标，首先要保证浅层特征的完整，然后进行多特征的融合，最后嵌入空间 _ 通道注意力机制模块（CBAM）进行多层特征之间的信息交互。于是在此进行相关模块的介绍。

1.1 短接模块

因为浅层卷积提取最浅层的特征，也是最真实，最能反映图像的底层特征，所以就要更好地保留原始特征，为后期的关键特征提取，提供的良好的特征层。此时由何凯明提出的残差网络很好的解决了梯度消失的问题，结构如图 1 所示[6,11]。

受近年残差模块、1×1 卷积[2-6,11]等思想激发，在残差网络的第一个大模块之后增加一个 1×1 短接分支来保留原始特征，也避免了梯度消失。添加 1×1 短接分支后的 Resnet50 结构如图 2 所示。

1.2 注意力机制

为了加强各空间特征层之间的相互作用。对图像中的空间域信息做相应的空间变换，从而提取出关键信息。空间注意模块强调空间像素的重要程度；同样每个通道上的信号添加一个权重，以表示通道与关键信息的相关性。权重越大，相关性越高。生成 channel 的掩码并得分。代表作有《SENET》和《Channel Attention Module》。注意机制表示特征数据中各部分的临界程度，并对其进行学习和训

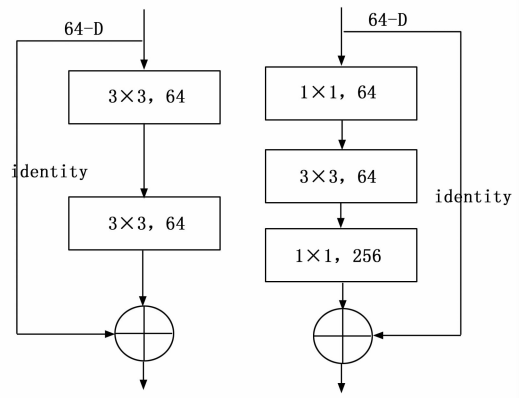

图 1　残差结构

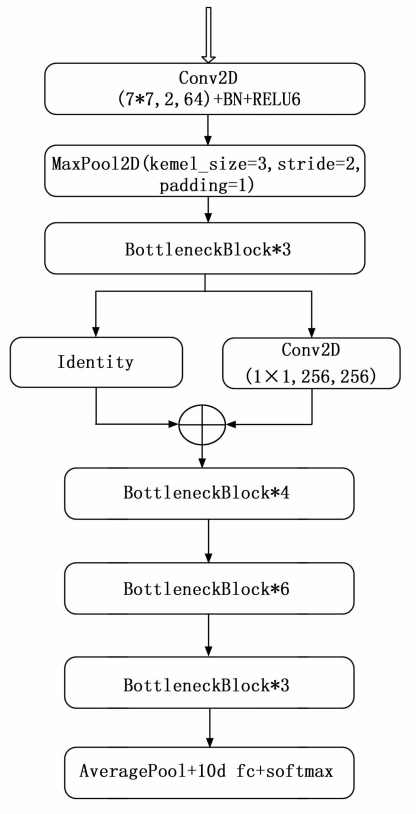

图 2　添加短接分支的 Resnet50 结构

练。注意机制的本质是利用相关的 FeatureMap 进行权值学习，然后将学习权值应用于原始 FeatureMap 进行加权求和，得到增强的 FeatureMap。根据注意域的不同，CV 中的注意机制可分为空间域和通道域两类，本文采用二者结合后的 CBAM 注意机制进行提取特征，具体如图 3 所示。

CBAM 包含 CAM（Channel Attention Module）和 SAM（Spartial Attention Module）两个子模块，分别进行通道和空间上的注意。能够节约参数和计算力，同时保证了其能够做为即插即用的模块集成到现有的网络架构中去。因为小目标本来像素就很少，假如多次卷积与池化会丢失大量的信息，此时通道注意力模块就可以提取更为显著的

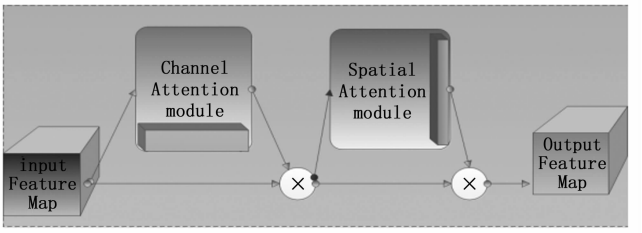

图 3　两个注意力模块图

特征，主要表现在：通道维度不变，压缩空间维度。该模块关注输入图片中有意义的信息，保证后期的识别效果更佳。空间注意力机制更加关注空间层面更需要注意的地方。二者对应的注意力机制通过神经网络的操作生成一个掩码，然后在掩码上给出评价的得分，然后指导后期卷积提取特征的侧重点。

CBAM 是一个轻量级的通用模块，可以集成到任何经典的 CNN 骨干网中，可以利用骨干网进行端到端的训练。CBAM 模块的主要结构如图 3 所示，CBAM 中通道注意力机制主要涉及的公式如式（1）所示。

$$M_c(F)=\sigma(MLP(AvgPool(F))+MLP(MaxPool(F)))=\sigma(W_1(W_0(F^c_{avg}))+W_1(W_0(F^c_{\max})) \quad (1)$$

其中：$W_0\in\mathbb{R}^{C/r\times C}$，$W_1\in\mathbb{R}^{C\times C/r}$。

将输入的特征图，分别经过全局最大池化和全局平均池化，然后分别经过 MLP。将 MLP 输出的特征进行基于点加操作，再经过 sigmoid 激活操作，生成最终的 channel attention featuremap 与输入的特征图做点乘操作，生成 spatial attention 模块需要的输入特征。

CBAM 中空间注意力机制主要涉及的公式，如式（2）所示，其中，σ 为 sigmoid 操作，7×7 表示卷积核的大小，7×7 的卷积核比 3×3 的卷积核效果更好，可以提取更多的空间信息。

$$M_s(F)=\sigma(f^{7\times7}([AvgPool(F);MaxPool(F)]))=\sigma(f^{7\times7}(F^s_{avg};F^s_{\max})) \quad (2)$$

在第一卷积模块层 layer1（即：BottleNeck Block ＊3）后添加了通道间注意机制和空间注意机制来融合通道和空间的重要显著信息，如图 6 所示。

2　改进的 Resnet50 算法

2.1　改进激活函数

为了更好地保留原始特征，替换了原来的 ReLU 激活函数。因为 sigmoid 函数比 ReLU6 计算昂贵得多；ReLU6 比 ReLU 激活函数也有一个很好的缓冲，可防止梯度消失。将 ReLU6 函数作为激活函数，可以很好的保留图像的原始特征。常用激活函数及其导数曲线图如图 4～5 所示（因为激活函数曲线有重叠的部分，为显示所有函数，在此把 ReLU，ReLU6，Swish 函数横坐标平移了 1 个像素）。

从图 4、图 5 可以看出，ReLU6 比 ReLU 激活函数有一个很好的缓冲，更符合特征的变换过程。

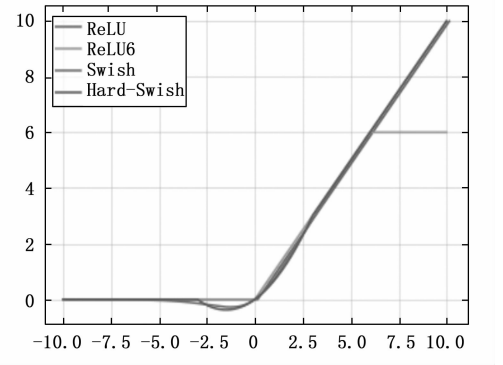

图 4 4 种常用激活函数

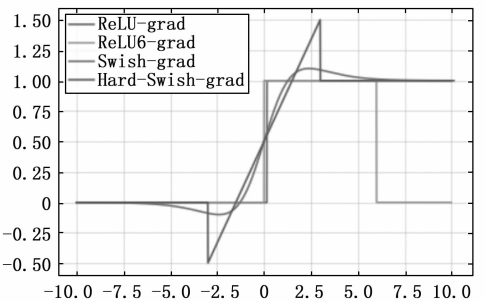

图 5 4 个激活函数对应的导数

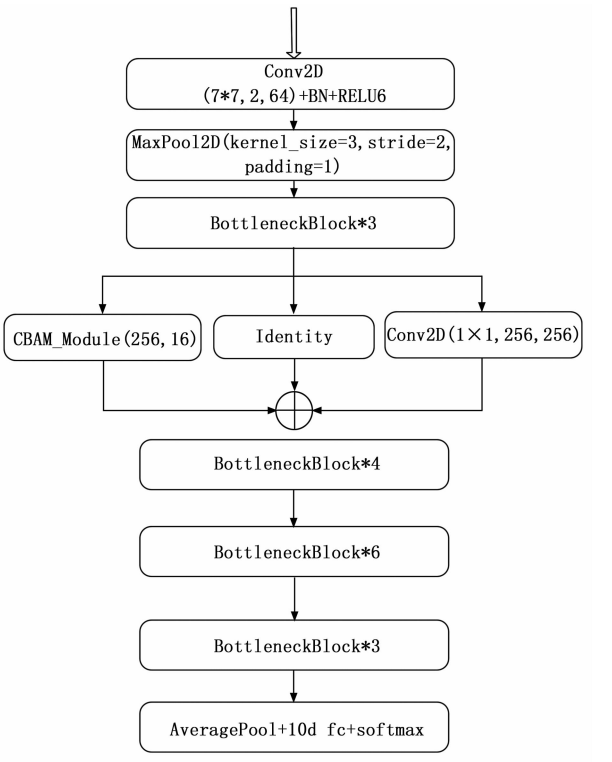

图 6 改进算法的流程图

2.2 改进的 Resnet50 算法

结合以上 3 个有效的模块，在此设计了一个改进的 Resnet50。将第一个 7×7 卷积中的激活函数修改为 ReLU6 函数；再在第一个大型残差块后，添加了一个 1×1 卷积来融合显著特征。在 Resnet50 的第一层特征之后，再添加 CBAM，最后再与原始的第一个大模块的特征图进行大融合，可以获得更多的判别特征。CBAM 是一个轻量级的通用模块，可直接放在 CNN 骨干网任意层，但需保证输出通道数和下一层的特征图通道数匹配，为此添加通道间注意机制和空间注意机制，改进后 Resnet50 的框架，如图 6 所示。

改进后的算法融合了三路特征（CBAM 特征、前一层特征 Identity，新增的 1×1 卷积特征），这样加强了浅层和深层特征的交互信息，很好的保留了显著特征，也避免了梯度消失。后面依次保留 Resnet50 网络原来的三个大型卷积模块。最后通过平均池化和十分类的 Softmax 进行物体的识别，输出对应的类别和概率。

3 实验

3.1 数据集介绍

3.1.1 FashionMNIST 数据集

为了验证算法的识别速度和识别精度，在 FashionMNIST 数据集上做了对比性实验，Fashion-MNIST 样品图片如图 7 所示。

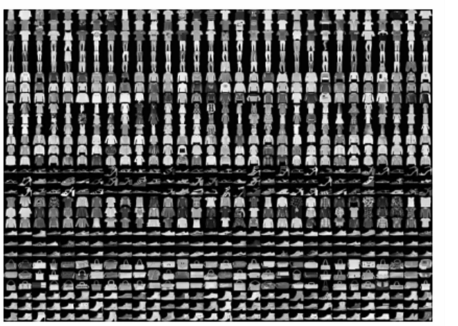

图 7 FashionMNIST 的样本图片

与 MNIST 数据集相比，FashionMNIST 数据集有以下差异：

1）FashionMNIST 的图像尺寸也为 28×28，但特征明显多于 MNIST。

2）与 MNIST 数据集相比，FashionMNIST 包含现实世界中的真实物体，不仅有大量的噪声，而且物体的比例和特征也不同，给识别带来很大的困难。

3）60 000 张图片用于训练，10 000 张用于测试。模型的规模扩大了。

4）同时 FashionMNIST 不再是抽象符号，而是更加具象化的人类必需品（服装），共 10 大类，具体如表 1 所示。

表 1 FashionMNIST 对应类别

序号	名称	序号	名称
0	T 恤(T-shirt)	5	凉鞋(Sandal)
1	裤子(Trouser)	6	衬衫(Shirt)
2	套头衫(Pullover	7	运动鞋(Sneaker)
3	连衣裙(Dress)	8	包(Bag)
4	外套(Coat)	9	靴子(Ankle boot)

综上所述，对 Fashion-MNIST 数据集进行识别要难得多，所以正好用来检验改进算法的有效性。

3.1.2　Cifar10 数据集

Cifar10 是由亚历克斯·克里日夫斯基和伊利亚·萨斯克弗整理的一个小数据集，用于识别普遍存在的物体。共有 10 种 RGB 颜色图像：飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船、卡车。图像大小为 32×32，数据集包括 50 000训练图像和 10 000 个测试图像。Cifar10 的示例图片如图 8 所示。

图 8　Cifar10 数据集中的样本图像

与 MNIST 数据集相比，Cifar10 有以下区别：Cifar10 是一个三通道的彩色 RGB 图像，而 MNIST 是一个灰度图像。Cifar10 的图像尺寸为 32×32，而 MNIST 的图像尺寸为 28×28，略大于 MNIST。与手写字符相比，Cifar10 包含现实世界中的真实物体，不仅噪声大，而且物体的比例和特征不同，识别存在很大困难，正好可用来检验改进算法的优劣。

3.2　实验环境

实验环境为百度 AI Studio 云计算平台，具体参数如下，CPU：4 核，内存：32 GB；GPU：V100 32 GB；硬盘：100 GB；编程语言：python 3.7；框架：Paddle Paddle2.0.2。

3.3　训练参数

在此选择 Resnet50 作为模型的基础网络架构。通过对 FashionMNIST 做简单的预处理，主要包括：随机调整裁剪，Color Jitter，随机水平翻转，Normalize 等操作，不使用过多的数据增强，主要更能体现改进算法的鲁棒性和通用性。

本次在 FashionMNIST 和 Cifar10 数据集上分别训练了 90 轮和 100 轮，即：epoch = 90（100），bitchsize =128，base _ lr=1e−2，boundaries= [60，70]，wamup _ steps = 2，momentum= 0.9，weight _ decay = 5e−3，延迟学习率 learning _ rate = paddle.optimiz er. lr. piece wise Decay (boundaries = boundaries，values = values)，学习率 learning _ rate = paddle.optimizer.lr.Linear Warmup.paddle.Optimizer。优化方法为 Adam，损失函数设置为交叉熵 CrossEntropyLoss，每轮评价检验一次：eval _ freq=1。

3.4　相关指标

最常用的分类问题使用的是交叉熵损失，计算如式（3）所示：

$$loss = \frac{1}{N}\sum[y_i\log(p_i)+(1-y_i\log(1-p_i)) \tag{3}$$

正确率（acc）是指使用测试集正确分类的记录占分类记录总数的比例，计算如式（4）所示：

$$acc = \frac{TP}{TP+FP} \tag{4}$$

其中：TP 表示正确分类的记录数，FP 表示错误分类的测试数据数。ImageNet 大约有 1 000 个类别，当模型预测某一张图片时，它会给出概率从高到低的 1 000 个类别排名。所谓 top-1 Accuracy 是指排名中第一类与实际结果一致的正确率。top-5 Accuracy 指在前 5 个类别中包含实际结果的准确性。在此加载在 ImageNet 数据集上的预训练模型。

3.5　时间复杂度和空间复杂度

时间复杂度就是模型的运算次数，可用 FLOPs 衡量，也就是浮点运算次数（Floating-point Operations），卷积网络的时间复杂度如式（5）所示。

$$Time \approx O(\sum_{d=1}^{D} M_d^2 \cdot K_d^2 \cdot C_{d-1} \cdot C_d) \tag{5}$$

D 是神经网络所具有的卷积层数，也即网络的深度。d 表示神经网络第 d 个卷积层；C_d 神经网络第 d 个卷积层的输出通道数 C_{out}，也即该层的卷积核个数。对于第 d 个卷积层而言，其输入通道数 C_{in}，就是第 $d-1$ 个卷积层的输出通道数。

空间复杂度严格来讲包括两部分：总参数量＋各层输出特征图。参数量：模型所有带参数的层的权重参数总量；特征图：模型在实时运行过程中每层所计算出的输出特征图大小，具体如式（6）所示。

$$Space \approx O(\sum_{d=1}^{D} K_d^2 \cdot C_{d-1} \cdot C_d + \sum_{l=d}^{D} M^2 \cdot C_d) \tag{6}$$

总参数量只与卷积核的尺寸 K、通道数 C、层数 D 相关，而与输入数据的大小无关。

输出特征图的空间占用比较容易，就是其空间尺寸 M^2 和通道数 C 的连乘。增加了 1×1 卷积知识常数级别的复杂度，但保证了梯度始终不消失，为保证后期的关键特征保留做出了准备。时间复杂度决定了模型的训练/预测时间。如果复杂度过高，则会导致模型训练和预测耗费大量时间，既无法快速的验证想法和改善模型，也无法做到快速的预测。本轮着重在时间复杂度和空间复杂度、识别准确度、检测速度等指标之间找到一个折中的临界点。即：时间复杂度和空间复杂度增加的成本，可以使用识别速度和精度

的提升。

3.5.1 FashionMNIST 数据集的实验结果

为表示简便，以此使用代号如下。a 表示 Resnet50 网络；b 表示 Resnet50 在第一个大模块后增加一个 CBAM 模块的网络；c 表示在 Resnet50 网络只增加 1×1 短接分支；d 表示 Resnet50 网络只增加了 1 个 CBAM 模块，同时也增加了 1×1 的短接分支，修改了 ReLU6 激活函数的 CBAM 模块。FashionMNIST 数据集上检验的损失如图 9 所示。

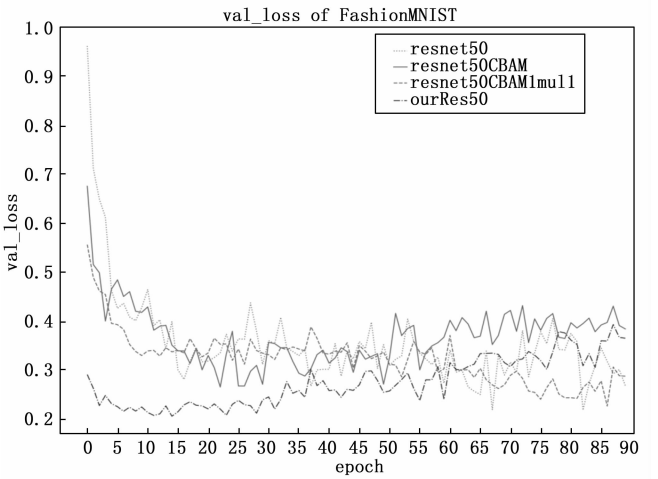

图 9 FashionMNIST 数据集上检验损失

在迭代的 eval 过程中，可以看到改进后的 Resnet50 对应的损失在不断下降，没有反弹现象，且收敛速度比其他 4 种算法都快。优势就很好的体现出来了，可以在一些手持设备或者野外监控设备上得到更好的应用。可在 resnet 50 网络中增加了 1×1 的短接分支后，增加的恒等映射可以保证以后提取的特征总是在当前最好的特征图基础上提取的，实验结果也证实增加了一个恒等分支后确实提取了更多的原始特征，保证了算法的鲁棒性和收敛性。改进后的算法对应的实验评价指标之损失是最先趋于收敛的，在后期也是较稳定的。

同时，为了切实比较改进后算法对于 FashionMNIST 数据集较为适用，选择国际上公认的排名第一的和排名前五的类别来评价算法的优劣，在此展现了准确率 top1 和 top5 曲线的实际效果，如图 10～11 所示。

在迭代的 eval 过程中，改进后 Resnet50 的 acc _ top1 和 acc _ top5 继续以相对稳定的趋势上升。它的精度超过了 Resnet 50，Resnet50 _ CBAM 等算法。模型的参数决定了数量空间复杂度，具体来说还分析了权重等参数、浮点计算量等相关参数详见表 2。

表 2 4 种算法的相关参数比较

	训练和测试时长	PARAM	FLOPs
a	2:15:27	23 581 642	81 669 632
b	2:51:34	24 282 438	83 113 408
c	2:26:52	23 347 974	89 527 232
d	2:13:23	23 655 714	88 123 360

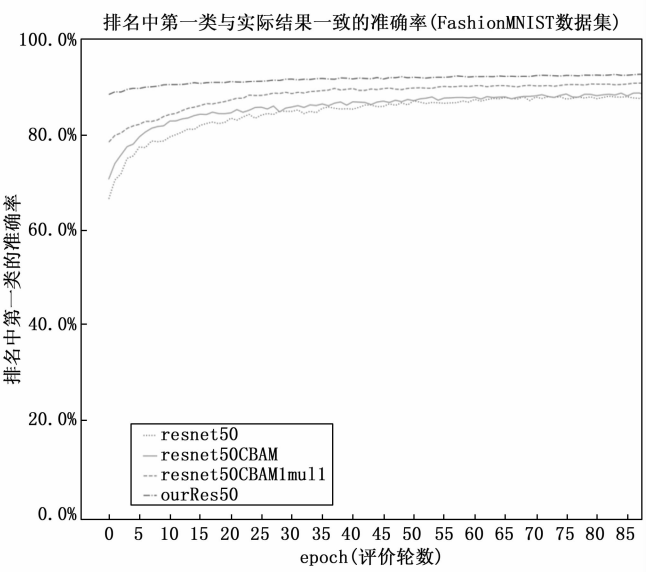

图 10 检验的 top1 准确度

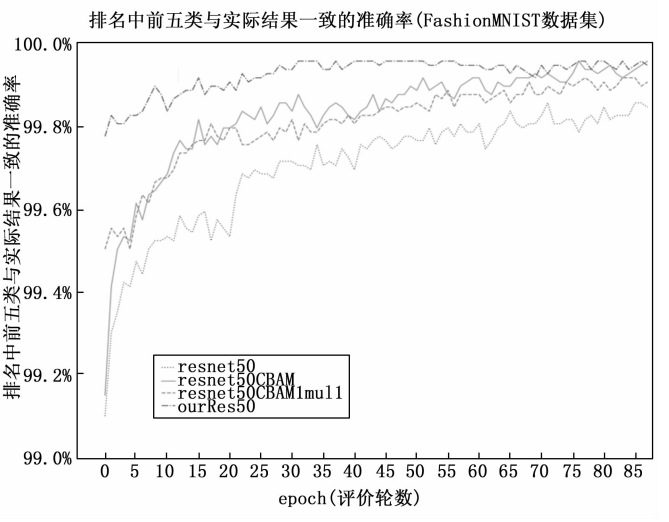

图 11 检验的 top5 准确度

由表 2 可以看出，改进后的模型的训练和评估时间减少了 2 分钟。相关参数仅增加 4.8%，FLOPs 仅增加 2%。实际准确率和检测速度如表 3 所示。

表 3 4 种算法的比较结果

	loss	Acc_top1/%	Acc_top5/%	train/eval/ms
a	0.189 7	88.13	99.83	340/207
b	0.396 9	88.76	99.94	346/214
c	0.215 2	90.60	99.95	360/240
c	0.156 2	92.47	99.99	342/210

其中，比起 Resnet50，acc _ top1 的准确率提高了约 4.3%，但批量（batch=128）训练和测试时间分别只增加 2 ms 和 3 ms。即使与先进的模型如 Res-net50 _ CBAM 相比，准确性和速度都有优势。在处理图像识别时，可提高

识别速度。

3.5.2　Cifar10 数据集的实验结果

为了提高 Cifar10 的实际效果，将图像归一化为 32×32，修改了部分实验参数 bitchSize＝256，训练轮数 epoch＝100，shuffle＝True，Drop _ last＝True，Cifar10 的损失值如图 12 所示。

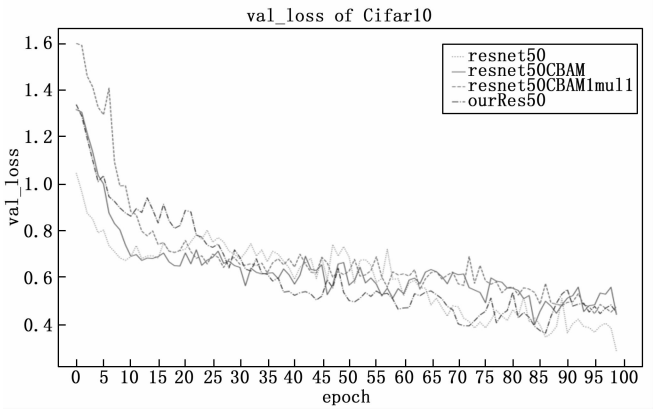

图 12　Cifar10 检验损失

在迭代过程中，可以看出，与 Resnet50 相对应的损失持续下降，收敛速度比其他 4 种算法都快。通过识别度观察实际效果，如图 13 和图 14。

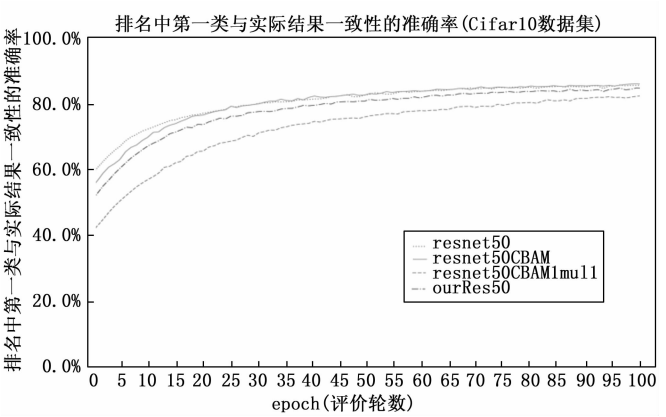

图 13　Cifar10 检验 top1 的准确度

在迭代的过程中，Resnet50 的 acc _ top1 和 acc _ top5 继续以相对稳定的趋势上升。它接近了 Resnet 50 的精度，并且明显优于其他 3 种算法，因为训练时长减短了。在精度方面，它明显优于另一个 Resnet 50。同时分析了对比效果的参数数量、计算等方面。具体详见表 4。为表示的简便，在此假设：（a：Resnet50，b：Resnet 50 _ CBAM，c：Resnet50 _ 1mul1，d：Our Resnet50）。

从表 4 中可以看到，改进后的算法与经典的 Resnet50 模型相比，效果不理想。而仅仅增加 1×1 卷积的模型在 Cifar10 数据集上训练的时间有所增加，PARAM 参数是最小的，FLOPs 稍微增加了。但是仍然体现出 1×1 卷积的明显作用，即从深度神经网络的浅层特征中提取了原始的

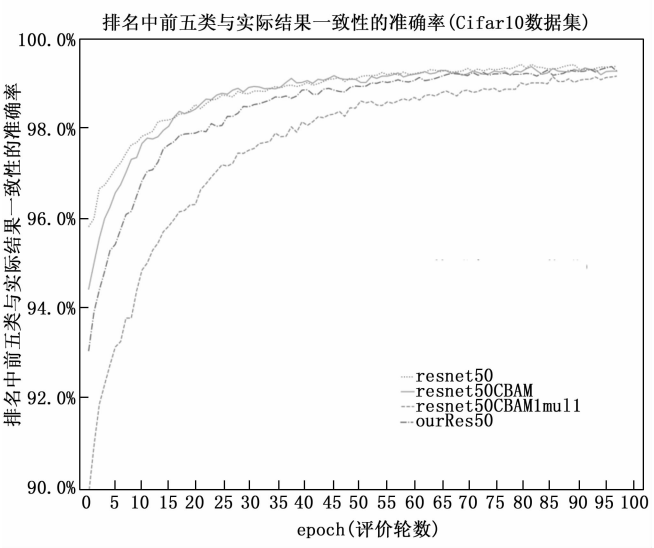

图 14　Cifar10 检验 top5 的准确度

特征，可以促进深层网络特征之间的信息交互，最终完成多层特征的融合，保证最终多目标的实际识别效果，本次实验的最终的 acc _ top1 和 acc _ top5 精度和检测速度详见表 5。

表 4　Cifar10 数据集的训练模型相关参数大小

	Time(hour)	PARAM	FLOPs
a	2:23:29	23 581 386	4 107 881 472
b	2:24:49	24 282 438	4 112 809 472
c	2:42:58	23 584 190	4 118 703 680
d	2:24:39	24 284 986	4 122 292 736

表 5　4 种算法训练 Cifar10'数据集的对比结果

	loss	Acc_top1(%)	Acc_top5(%)	train/eval(ms)
a	0.291 91	85.48	99.42	407/219
b	0.404 28	82.25	99.37	409/220
c	0.450 79	85.77	99.16	411/218
d	0.368 05	84.47	99.37	413/220

从表 5 中可以看到，与经典的 Resnet50 模型相比，而仅仅增加 1×1 卷积的模型在 Cifar10 数据集上，达到了超过 85.77%的 acc _ top1 精度。运算量减少的同时，还提高了 acc _ top1 的识别准确度 0.3%。

4　结束语

为了更好地识别多目标，首先，在第一个卷积残差块 layer1 后增加一个 1×1 的短接分支尽可能多的保留原始特征，保证了浅层特征的完整；然后与 layer1 特征进行融合，进行了多特征的融合；最后再嵌入一个修改激活函数 ReLU6 的空间 _ 通道注意力机制模块（CBAM），从而增强特征图的特征表达能力。对比实验表明在 FashionMNIST 数据集和 Cifar10 数据集上准确率有一定的提高，同时也保持

良好的检测速度。在未来，可能会对新的骨干网络进行修改以提取更好的特征，或者对损失函数进行修改以找到最优值。在未来可能会在引入新的注意力机制，新的激活函数、优化网络结构等来改进残差深度网络，保证当前空间复杂度和时间复杂度的基础上，类别识别准确度方面还能有一定的提高。

参考文献：

[1] SIMONYAN K, ZISSERMAN A . Very Deep Convolutional Networks for Large-Scale Image Recognition [J/OL]. Computer Science, 2015: 1 - 14. 2015. 04. 10. https://ar xiv. org/abs/1409. 1556.

[2] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [J/OL]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1 - 9. 2015. 06. 12. https://DOI:10. 1109/CVPR. 2015. 7298594.

[3] CHRISTIAN SZEGEDY, VINCENT VANHOUCKE, SERGEY IOFFE, et al. Rethinking the inception architecture for computer vision [J/OL]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 2818 - 2826. 2015. 12. 11. https://doi. org/10. 48550/arXiv. 1512. 00567.

[4] CHRISTIAN SZEGEDY, SERGEY IOFFE, VINCENT VANHOUCK. Inceptionv4, inception-resnet and the impact of residual connections on learning [J/OL]. In Thirty-first AAAI conference on artificial intelligence, 2017: 4278 - 4284. 2017. 02. 04. https://arxiv. org/abs/1602. 07261.

[5] SERGEY IOFFE, CHRISTIAN SZEGEDY. Batch normalization: Accelerating deep network training by reducing internal covariate shift [J/OL]. In International Confer-ence on Machine Learning, 2015: 448 - 456. 2015. 02. 11. https://arxiv. org/abs/1502. 03167.

[6] KAIMING HE, ZHANG X Y, REN S Q. Deep residual learning for image recognition [J/OL]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770 - 778. 2016. 12. 12. https://ieeexplore. ieee. org/document/7780459/citations/tabFilter=papers.

[7] HUANG G, LIU Z, Laurens vander Maaten Densely connected convolutional networks [J/OL]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261 - 2269. 2017. 08. 25. https://arxiv. org/abs/1608. 06993.

[8] MA N N, ZHANG X Y, ZHEN H T. Shufflenet v2: Practical guidelines for efficient cnn architecture design [J/OL]. In Proceedings of the European confe-rence on computer vision (ECCV), 2018: 116 - 131. 2018. 01. 01. https://link. springer. com/chapter/10. 1007/978 - 3 - 030 - 01264 - 9_8/coverhttps://link. springer. com/chapter/10. 1007/978 - 3 - 030 - 01264 - 9_8/cover.

[9] XIAO L C, YASAMAN BAHRI, JASCHA SOHL DICKSTE. Dynamical isometry and a mean field theory of cnns: How to train 10 000-layer vanilla convolutional neural networks [J/OL]. In International Conference on Machine Learning, 2018: 5393 - 5402. 2018. 06. 14. https://arxiv. org/abs/1806. 05393.

[10] OYEBADE K OYEDOTUN, DJAMILAAOUADA, BJÖRN OTTERSTEN, et al. Going deeper with neural networks without skip connections [J/OL]. In 2020 IEEE International Conference on Image Processing (ICIP), 2020: 1756 - 1760. 2020. 05. 30. https://ieeexplore. ieee. org/document/9191356.

[11] TAN M, LE Q V. EfficientNet: Rethinking Model Scal— ing for Convolutional Neural Networks [J/OL]. Interna tional Conference on Machine Learning, 2019: 1 - 11. 2019. 09. 11. https://arxiv. org/ abs/ 1905. 11946.

[12] ZHANG H, WU C, ZHANG Z, et al. ResNeSt: Split-Atten tion Networks [J/OL], 2020: 1 - 12. 2020. 11. 30. https://ar xiv. or g/abs/2004. 08955.

[13] WANG S, LIU Y, QING Y, et al. Detection of Insulator Defects With Improved ResNeSt and Region Proposal Network [J/OL]. IEEE Access, 2020. 08. 01. 2020:1848 41—184850. https://ieeexplore. ieee. org/document/921 8924.

[14] BROCK A, DE S, SMITH S L, et al. High-Performance Large-Scale Image Recognition Without Norm alization [J/OL]. 2021: 1 - 22. 2021. 02. 11. https://arxiv. org/abs/2102. 06171v1.

[15] WOO S, PARK J, LEE J, et al. CBAM: convolutional block attention module [J/OL]. in IEEE European Conference on Computer Vision (ECCV), 2018: 3 - 19. 2018. 10. 06 3—19. https://link. spr inger. com/chapter/10. 1007/978—3—030—01234—2_1.

[16] CAO Y, XU J, LIN S, et al. GCNet: Non-Local Net-works Meet Squeeze-Excitation Networks and Beyond [J/OL]. 2019 IEEE/CVF International Confe rence on Computer Vision Workshop (ICCVW). 2020: 1 - 21. 2020. 03. 05. https://ieee xplore. ieee. org/doc ument/902234.

[17] PETIT O, THOME N, RAMBOUR C, et al. U-Net Transformer: Self and Cross Attention for Medical Image Segmentation [J/OL]. 2021: 267 - 276. 2021. 03. 10. https://arxiv. org/ abs/2103. 061— 04.

[18] HOU Q, ZHOU D, FENG J. Coordinate Attention for Effi— cient Mobile Network Design [J/OL]. 2021: 1 - 10. 2021. 03. 04. https://arxiv. org /abs/2103. 02907.

[19] CHENG Q, LI H, WU Q, et al. BA^{2}M: A Batch Aware Attention Module for Image Classification [J/OL]. 2021. 03. 28. 2021:01—10. https://arxiv. org/abs/2103. 15099.

[20] YANG Y B. SA-Net: Shuffle Attention for Deep Con volutional Neural Networks [J/OL], International Con-ference on Acoustics, Speech and Signal Processing (ICASSP), 2021: 1 - 13. 2021. 01. 30. https://arxiv. org/ abs /2102. 00240.