第 9 卷 第 6 期
2020 年 11 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 9 No. 6
Nov. 2020

引文格式：
张卫, 古林燕, 刘嘉. 基于区域分解的快速卷积神经网络学习策略研究 [J]. 集成技术, 2020, 9(6): 48-58.
Zhang W, Gu LY, Liu J. A study on domain decomposition inspired fast convolutional neural network learning strategy [J]. Journal of Integration Technology, 2020, 9(6): 48-58.

基于区域分解的快速卷积神经网络学习策略研究

张 卫[1,2,3] 古林燕[1,3] 刘 嘉[1,3]

[1](中国科学院深圳先进技术研究院 深圳 518055)
[2](中国科学院大学深圳先进技术学院 深圳 518055)
[3](深圳市 E 级工程与科学计算重点实验室 深圳 518055)

摘 要 为加快卷积神经网络的训练，该研究提出一种受区域分解方法启发的新型学习策略。将该方法应用于残差网络(ResNet)进行图像分类时，使用 ResNet32 可获得最佳结果。进一步地，将 ResNet32 分成 4 个子网络，其中每个子网具有 0.47 M 参数，此为原始 ResNet32 的 1/16，从而简化了学习过程。此外，由于可以并行训练子网络，因此在使用 CIFAR-10 数据集进行分类任务时，计算时间可以从 8.53 h (通过常规学习策略)减少到 5.65 h，分类准确性从 92.82% 提高到 94.09%。CIFAR-100 和 Food-101 数据集也实现了类似的改进。实验结果显示，所提出的学习策略可以大大减少计算时间，并提高分类的准确性。这表明所提出的策略可以潜在地应用于训练带有大量参数的卷积神经网络。

关键词 学习策略；区域分解；并行训练；加速
中图分类号 TP 399 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20200418001

A Study on Domain Decomposition Inspired Fast Convolutional Neural Network Learning Strategy

ZHANG Wei[1,2,3] GU Linyan[1,3] LIU Jia[1,3]

[1](*Shenzhen Institutes of Advanced Technology*, *Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[2](*Shenzhen College of Advanced Technology*, *University of Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[3](*Shenzhen Key Laboratory for Exascale Engineering and Scientific Computing*, *Shenzhen* 518055, *China*)

Abstract We propose a novel learning strategy inspired by domain decomposition methods to accelerate the training of convolutional neural network (CNN). The proposed method is applied to residual networks (ResNet) for image classification tasks. The best result is achieved with ResNet32. In this case, we split ResNet32 into 4 sub-networks. Each sub-network has 0.47 M parameters which is 1/16 of the original

收稿日期：2020-04-18 **修回日期：**2020-07-26
基金项目：国家自然科学基金项目(81661168015, 81871447)；深圳市科学技术创新委员会项目(ZDSYS201703031711426)
作者简介：张卫，硕士研究生，研究方向为深度学习、医学图像分类和分割；古林燕，助理研究员，研究方向为机器学习、计算数学；刘嘉(通讯作者)，研究员，研究方向为生物医学工程， E-mail：jia.liu@siat.ac.cn。

ResNet32, thereby facilitating the learning process. Moreover, because the sub-networks can be trained in parallel, the computational time can therefore be reduced to 5.65 h from 8.53 h (by the conventional learning strategy) in the classification task with the CIFAR-10 dataset. We also find that the accuracy of the classification is improved to 94.09% from 92.82%. Similar improvements are also achieved with the CIFAR-100 and Food-101 datasets. In conclusion, the proposed learning strategy can reduce the computational time substantially with improved accuracy in classification. The results suggest that the proposed strategy can potentially be applied to train CNN with a large amount of parameters.

Keywords　learning strategy; domain decomposition; parallel training; speed up

1　引　言

卷积神经网络(Convolutional Neural Network，CNN)[1]凭借其丰富的表示功能极大地提高视觉任务的性能，具有广泛的应用，如图像识别[1]、语义分割[2-3]和目标检测，并在这些领域取得了重大突破。为提高 CNN 的性能，现有研究主要集中在网络的深度和宽度[4-6]。虽然增加网络的深度和宽度可以使网络具有强大的表达能力，但对于大规模数据集[7-9]来说，该操作将导致更长的训练时间并占用更多的计算资源。例如，一个医学图像数据通常包含几百甚至几千张切片，每张切片的大小为 512×512，这种 3D 图像数据训练往往面临内存不足，训练时间较长的问题[10]。如何解决此类似问题在医学图像应用中具有重要的意义。

对于深度学习的应用，较大的数据集和更深层的网络结构可以提高准确性，但同时需付出更长的训练时间。此外，诸如财务风险、自动驾驶和医学图像[11]等应用需要大量的数据来训练网络。目前，增加图形处理器(Graphics Processing Unit，GPU)数量以实现数据并行性是网络训练中最常用的方法。例如，在单个 GPU 上用 ResNet 完成 ImageNet-1k 训练需要 14 天[11]，而将 GPU 增加到 1 024 个时，训练时间只需 20 min。GPU 是一种昂贵的计算资源，故在实际应用中很难获得上百甚至上千个 GPU 来训练 CNN。

目前网络加速的方法主要是通过数据并行，即增加 GPU 的个数使得每次训练选取的样本数更大。随着网络模型的加深，参数之间的相互依赖增大，需要传输的参数量也增大，使得 GPU 之间的传入时间和通信时间增加。当模型参数以及计算产生的中间变量无法放入 GPU 设备时，数据并行会遇到训练困难的问题。本文从网络结构方面对加速网络的训练进行研究，提出了一种基于区域分解[12]的快速卷积神经网络学习策略。本文的主要创新点为：(1) 受到区域分解的启发，将卷积神经网络拆分成多个子网络，其中子网络的参数比大网络参数少很多，从而达到加速训练的目的；(2) 每个子网络能够独立并行地训练相应的子样本数据，CNN 通过子网络初始化更加注重对局部特征的学习，提高了 CNN 的分类结果。

2　卷积神经网络的分解和组合

基于区域分解的快速卷积神经网络学习策略研究受区域分解方法启发。区域分解是通过迭代求解较小子域上的子问题来求解全局问题的计算方法。分解和组合网络的训练过程为：(1) 将卷积神经网络拆分成 k 个子网络 ($k=2^n(n\geq 2)$)；(2) 将每个图片分成 K 份子区域，其中子区域的个数和

子网络相同，分别用子网络并行独立地训练子样本；(3)将训练好的子网络权重参数重新组合作为大网络的初始权重，然后使用原样本数据对大网络进行微调。

2.1 损失函数

假设在图像分类任务中，将 CNN 表示为 $F(x;\Theta)$，其中 x 为网络的输入，Θ 为网络参数的集合。定义网络参数集合 $\Theta=\{\Theta_1,\Theta_2,\cdots,\Theta_L\}$，其中 $\Theta_i(i=1,2,\cdots,L)$为第 i 层的参数集，并定义第 i 层的参数，用集合 $\Theta_i=\{\theta_1,\theta_2,\cdots,\theta_l\}$表示。同时，将 x_i、f_i 和 F_i 分别表示为第 i 层输出、第 i 层的卷积操作和第 i 层中网络的第一个卷积操作。$x_i=f_i(x_{i-1};\theta_i)=F_i(x_0;\Theta_i)$，其中 $x_0=x$。对于特定的 x 和 θ，定义 $\hat{y}=\left[\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_c\right]^{\mathrm{T}}=F\left(x;\theta\right)\in\left[0,1\right]^C$，$C$ 为类别个数，$\sum_{c=1}^{C}\hat{y}_c=1$。即网络的输出 $\hat{y}_c$ 是输入为第 c 类的后验概率。更具体地，假设 CNN 具有 L 层，并且最后一个全连接层的输出为 $x_L=\left[x_{L_1},x_{L_2},\cdots,x_{L_C}\right]^{\mathrm{T}}\in\mathbb{R}^C$，后验概率通过 Softmax 函数获得：

$$\hat{y}_c=\frac{e^{x^{L_c}}}{\sum_{i=1}^{C}e^{x^{L_i}}},c=1,\cdots,C \tag{1}$$

相应地，样本 x 对应的标签为 $y\in\{0,1\}^c$，则其损失函数定义为：

$$J\left(x,y,\theta\right)=-\sum_{c=1}^{C}y_c\log\hat{y}_c \tag{2}$$

在图像分类任务中，对于样本 $X=\{x_1,x_2,\cdots,x_n\}$ 且其对应的标签 $Y=\{y_1,y_2,\cdots,y_n\}$来说，训练网络解决优化问题为：

$$\theta^*=\arg\min_{\theta}\sum_{i=1}^{n}J\left(x_i,y_i,\theta\right) \tag{3}$$

2.2 分解网络

本文通过具有 2 个卷积层、2 个池化层和 2 个全连接层的简单 CNN 来说明区域分解的快速卷积神经网络学习策略，具体如图 1 所示。其中，图 1(a)为分解的 CNN 结构，其输入图片大小为 32×32。第一个卷积层的卷积核大小为 5×5，深度为 16，步长为 1。输入图片经过卷积运算后，输入图片输出的特征图定义为 conv1，

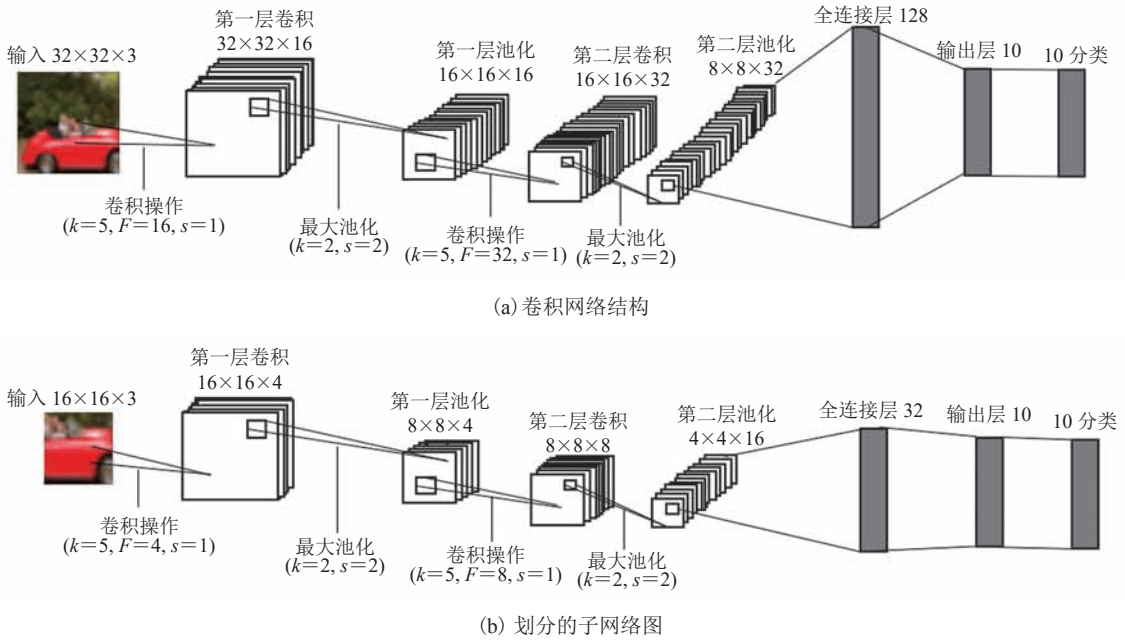

图 1 卷积网络结构和划分的子网络图

Fig. 1 The schema of convolutional neural networks and the CNN into 4 sub-CNNs

大小为 32×32×16(特征图个数为 16)。第一个池化层的卷积核大小为 2×2，步长为 2。通过池化操作后，第一层卷积输出的特征图定义为 pool1，大小为 16×16×16。同样地，第二层 conv2 的卷积特征图为 16×16×32，pool2 的大小为 8×8×32。最后两个全连接的节点数为 128 和 10(其中 10 代表 10 个分类)。

将 CNN 分解为具有相同结构的 $2^n(n\geqslant 2)$ 个子网络。此例中共 4 个子网络，每个子网络具有与图 1(b)相同的结构。在子网络中输入图片的大小为 16×16×3。第一个卷积层的卷积核大小为 5×5，深度为 4，步长为 1。通过卷积运算后，输入图片输出的特征图定义为 conv1_sub，大小为 16×16×4(特征图个数为 4)。第一个池化层的卷积核大小为 2×2，步长为 2。通过池化操作后，第一个池化层输出的特征图定义为 pool1_sub，大小为 8×8×4。同理，第二层卷积特征图 conv2 为 8×8×8，pool2 的大小为 4×4×8。最后两个全连接的节点数为 32 和 10。

2.3　组合网络

首先定义子网络的张量，然后将子网络的张量合成 CNN 的张量。在图 1(b)的示例中，有 4 个相同的子网络，可以定义为相应的 4 个张量 $(\boldsymbol{W}_1^n,\boldsymbol{W}_2^n,\boldsymbol{W}_3^n,\boldsymbol{W}_4^n)\in\mathbf{R}^{k\times k\times c_{\text{in}}\times c_{\text{out}}}(n=1,2,3,4)$。其中，$\boldsymbol{W}_i(i=1,2,3,4)$中的 i 表示子网络的数量；n 为层数；$k\times k$(卷积层中 $k=3$，全连接层中 $k=1$)为一个卷积核的大小；c_{in} 和 c_{out} 分别为该层的输入和输出特征图的数量。池化层卷积核大小通常为 2×2，步长为 2。接下来，可以在图 1(a)定义 CNN 的张量，采用 $\boldsymbol{W}^n\in\mathbf{R}^{k\times k\times c_{\text{in}}\times c_{\text{out}}}$ 表示该张量。本节中使用训练好的 $\boldsymbol{W}_1$、$\boldsymbol{W}_2$、$\boldsymbol{W}_3$、$\boldsymbol{W}_4$ 初始化 CNN 的初始权重 $\tilde{W}$。对 CNN 的初始化权重参数定义如下：

(1)采用张量 $\boldsymbol{W}^1$ 表示第一个隐含层，其中 $k=3$、$c_{\text{in}}=c_{\text{out}}=3$(3 是 RGB 图像的通道数)和 $\tilde{c}_{\text{out}}=4c_{\text{out}}$，第一层 CNN 初始参数可由 $\boldsymbol{W}_1$、$\boldsymbol{W}_2$、$\boldsymbol{W}_3$、$\boldsymbol{W}_4$ 组合得到；

(2)采用张量 $\boldsymbol{W}^4$ 表示最后一层，其中 $k=3$、$\tilde{c}_{\text{in}}=4c_{\text{in}}$ 和 $\tilde{c}_{\text{out}}=4c_{\text{out}}=n$($n$ 为图像分类的类别数)；

(3)采用张量 $\boldsymbol{W}^2$ 和 $\boldsymbol{W}^3$ 表示其他层，并且 $k=3$、$\tilde{c}_{\text{in}}=4c_{\text{in}}$ 和 $\tilde{c}_{\text{out}}=4c_{\text{out}}$，CNN 初始化权重由 $\boldsymbol{W}_1$、$\boldsymbol{W}_2$、$\boldsymbol{W}_3$、$\boldsymbol{W}_4$ 组合得到。

基于区域分解的快速卷积神经网络学习策略先将 CNN 分解为多个子网络。每个子网络由相应的子样本独立训练，并且子网络和 CNN 使用相同的优化器进行训练。此外，每个子网络可以通过多个 GPU 同时并行训练。CNN 初始化权重过程如图 2 所示。

卷积神经网络的分解和组合学习策略步骤如下：

(1)假设输入图片定义为 $X(H,W,C)$，采用 $x_k=(h_k,w_k,c)$表示第 k 个子样本数据，且 $(h_k,w_k,c)\in(H,W,C)$；

(2)将 CNN 分解成 K 个子网络，每个子网络优化参数定义为 $F_k\left[D(x_k);\theta_k\right]$；

(3)使用随机梯度下降算法来优化子网络，可得子网络所需优化的问题为$\theta^*=\arg\min\limits_{\theta}\sum\limits_{i=1}^{n}J(x_k;y_i,\theta)$；

(4)使用子网络的参数$\left\{\theta_k^*\right\}_{k=1}^{K}$ 作为 CNN 的初始训练权重；

(5)将 CNN 的初始参数设置为 $\theta=\tilde{\theta}$，并对 CNN 进行微调。由此可得，CNN 的优化问题为 $\theta^*=\arg\min\limits_{\theta}\sum\limits_{i=1}^{n}J(x_i;y_i,\theta)$。

3　实验结果

图像分类是计算机视觉领域研究的重要部分，其目的是使用计算机代替人类去识别处理各种图像信息。图像分类的过程分为图像的预处理、特征提取及网络模型的选择。其中，特征提取及网络模型选择对图像分类的结果具有很大

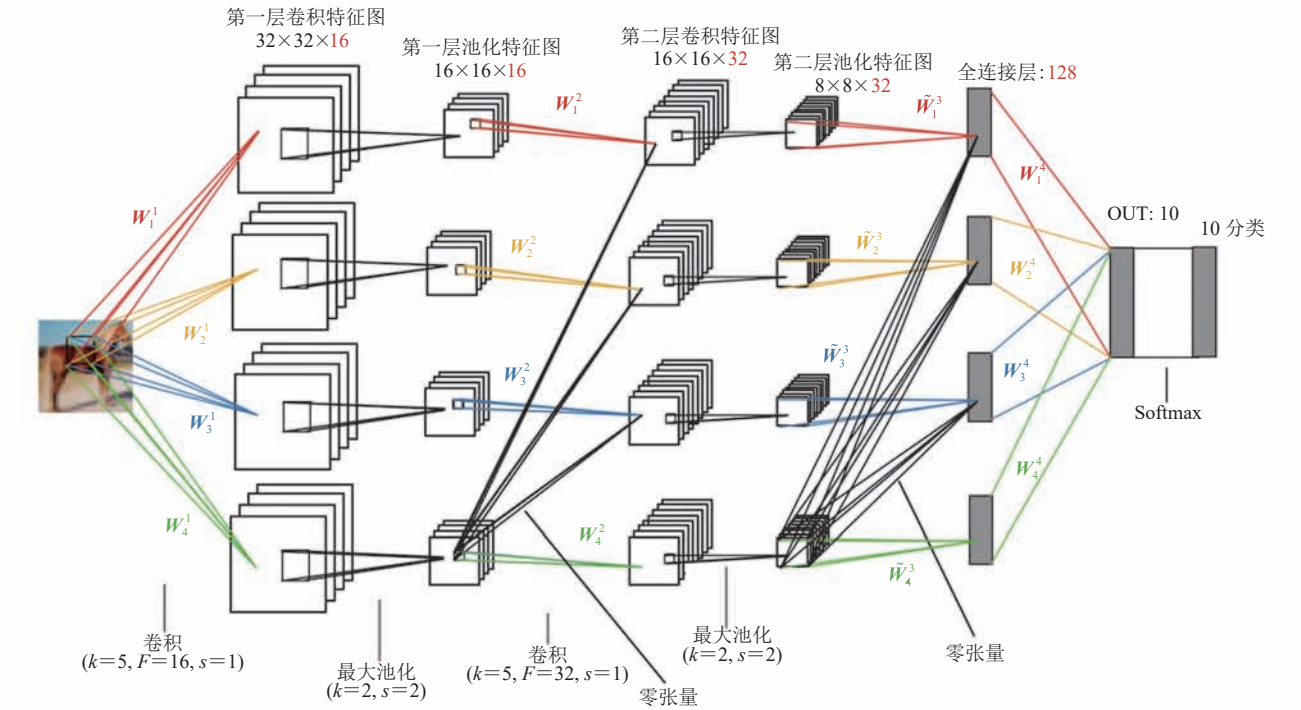

图 2 卷积神经网络迁移学习的初始化权重

Fig. 2 Initial weights for CNN transfer learning

的影响。对于基于区域分解的快速卷积神经网络学习策略的实验，采用 CIFAR-10/CIFAR-100 和 Food-101 等公开数据集进行分类实验验证。

3.1 不同深度 ResNet 分解

为了评估分解和组合学习策略的有效性，本文使用公开数据集 CIFAR-10 进行实验。其中，CIFAR-10 数据集中所有图片大小均为 32×32，共 10 类，每类中包含 50 000 张训练集和 10 000 张测试集[4]。具体数据集如图 3 所示。

本文网络结构选择的是 ResNet[4]，其中 ResNet 结构是图像分类任务中最常用的网络。训练图像的大小为 32×32，第一个层是 3×3 的卷积。然后，网络分别在大小为{32,16,8}的特征图上使用具有 3×3 卷积的 $6n$ 层堆栈，每个特征图尺寸为 $2n$ 层。卷积核的数量分别为{64,128,256}。二次下采样通过步长为 2 的卷积核执行操作。网络采用全局平均池化操作，共 $6n+2$ 个堆叠的加权层[1]。本文对比了 n=5, 9, 8 的实验结果，即 ResNet32、ResNet56 和

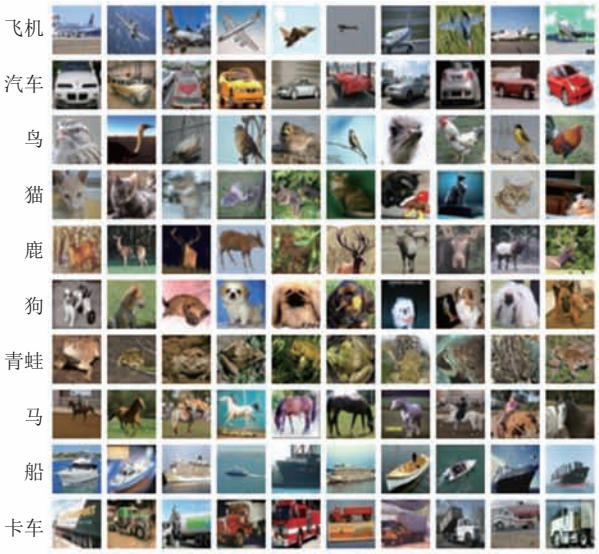

图 3 10 分类数据集

Fig. 3 CIFAR-10 dataset

ResNet110。在实验过程中，将图像分解为 4 个具有重叠的子图像[13-14]，其中子图像的大小为 20×20。因此，对于子网络，特征图的大小分别为{20,10,5}，卷积核的数量分别为{16,32,64}。需要说明的是，所有子网络都是从头开始进行

训练的。网络使用训练集训练后，在测试集上测试 top-1 和 top-5 准确率。此外，本文还使用 CIFAR-100 数据集对分解和组合的迁移学习策略进行实验验证。该数据集包含 100 个类别的 50 000 张训练图像和 10 000 张测试图像。数据集图像是大小为 32×32 的 RGB 图像。实验结果显示，与 CNN 的随机初始化相比，分解和组合迁移学习策略的准确性有一定程度的提升。

本文中，每个子网络和大网络都使用相同的优化方案进行训练。本文采用图像增强的标准做法，通过执行随机水平翻转来执行数据增强。所有网络模型都在深度学习服务器上进行训练，该服务器可以同时训练多个网络模型。网络优化器使用 Adam[15]进行优化，批量训练图像的大小为 128。对于随机初始化，初始学习率设置为 0.001，网络训练到 80、120 和 160 个 Epoch 时，学习率减少 10 倍。对于分解和组合迁移学习方法，子网络训练 120 个 Epoch， 初始学习率为 0.001；网络训练 80、90 个 Epoch 时，学习率减少 10 倍。大网络训练 80 个 Epoch，网络初始学习率为 0.000 1，网络训练到 20、40 和 60 个 Epoch，学习率减少 10 倍。

首先将分解和组合的学习策略使用 ResNet32、ResNet56、ResNet110 等网络，并采用 CIFAR-10 和 CIFAR-100 数据进行训练和测试。对于训练时间方面，从表 1 可以看出分解和组合的学习策略在 3 个 ResNet 网络的训练时间分别是 5.65 h、9.61 h 和 18.78 h，相对于随机初始化学习时间加速比是 1.51、1.45 和 1.42。这表明分解和组合的学习策略能够加快网络的训练时间，同时还在一定程度上提高网络的分类结果。由于 CIFAR-100 数据集的大小和 CIFAR-10 是一样的，故未在表中记录 CIFAR-100 的时间。在 2017 年 ILSVRC 比赛中，SENet 以 top-5 错误率 2.25% 夺得冠军[16]，相比于 2015 年的冠军 ResNet(3.57%)提高了 1.32%[4]。这表明在一些公开数据集的分类中，分类的错误率(或准确率)达到一定程度时，继续提高具有一定的挑战。从表 2 可知，迁移学习在 ResNet56 的性能提升效果最佳，其中在 CIFAR-10 数据集上相较于随机初始化提升了 1.25%，在 CIFAR-100 数据集上，ResNet32 提升了 4.23%。另外，迁移学习的方法在 ResNet32 相较于 2015 年 ResNet32 文章中的结果[4]提升了 1.6%。

表 1　10 分类数据集训练时间

Table 1　Training time of 10 classification data sets

方法	CIFAR-10 数据集训练时间（h）		
	ResNet32	ResNet56	ResNet110
分解和组合的学习策略	5.65	9.61	18.78
随机初始化	8.53	13.98	26.74

表 2　10 分类和 100 分类数据集实验结果

Table 2　CIFAR-10 and CIFAR-100 datasets classification experimental results

CNN 训练策略	CIFAR-10 数据集准确率(%)			CIFAR-100 数据集准确率(%)		
	ResNet32	ResNet56	ResNet110	ResNet32	ResNet56	ResNet110
sub_net_1	81.25	81.18	81.15	52.30	53.57	51.60
sub_net_2	81.40	80.78	81.53	52.28	53.34	49.33
sub_net_3	79.99	80.26	80.10	52.88	53.04	51.80
sub_net_4	79.92	79.97	79.81	52.94	52.44	51.44
transfer	88.63	89.03	88.86	60.85	61.54	59.14
transfer_T	**94.09**	**94.34**	**94.14**	**73.14**	**73.20**	**73.75**
random_T	92.82	93.09	92.80	68.91	72.84	73.48
ResNet_paper	92.49	93.03	93.57	--	--	--

注：“ResNet_paper”表示 2016 年 ResNet 文章中的结果；“--”表示 ResNet 文章中出现没有 CIFAR-100 的结果

为了进一步分析分解和组合的学习策略对卷积神经网络模型优化的影响，将 CIFAR-10 数据集在 ResNet32 网络的测试结果曲线进行整理展示(图 4)。从图 4 可以看出，迁移学习方法一开始的准确率比随机初始化的结果要大很多，迁移学习方法训练 10 多个 Epoch 时，网络已收敛，而随机初始化网络需要训练 80 多个 Epoch 才达到收敛。

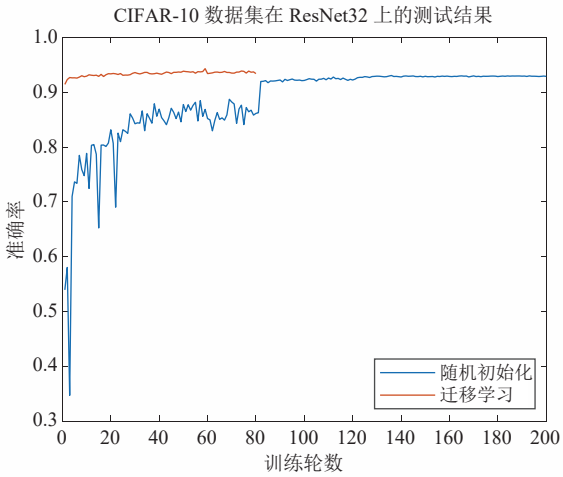

图 4　10 分类数据集结果

Fig. 4　Results of the dataset CIFAR-10

分解和组合的学习策略相较于随机初始化训练方法在准确率方面有一定的提升。为了研究准确率提升的原因，本文将第一层卷积的特征图进行可视化处理(图 5)。从图 5 可以看出，子网络提取的特征能很好地被 CNN 学习到。通过实验验证分析，子网络训练的子样本的感受野较小，卷积操作过程中能够注意一些细节的特征，而这些细节特征是 CNN 所忽略的。通过子网络训练提取的特征，该细节特征通过迁移学习方法被 CNN 所学习，从而提高了网络的分类准确率。

3.2　相同深度 ResNet 分解

基于区域分解的快速卷积神经网络学习策略的一个关键步骤是如何划分子网络个数。若网络划分的个数较多，则导致子网络结构比较简单，无法处理复杂数据分类任务。若网络划分个数太少，则网络结构太大，不仅需要占用内存资源，而且需要耗费大量的时间去训练。因此，子网络的划分对于区域分解的快速卷积神经网络学习策略十分重要。

本小节实验数据采用 Food-101 数据集进行。该数据集包含 101 个类别中的 90 900 张训练图像和 10 100 张测试图像[17-19]。与其他分类任务相比，该分类任务具有一定挑战性。这是由于每个图片都包含多个目标，并且不同类别之间存在相似之处，这会对目标的分类造成一定的干扰。对于数据预处理，采用标准做法——将图像尺寸大小重置为 224×224，并对数据集进行增强，如对图像进行随机水平翻转。

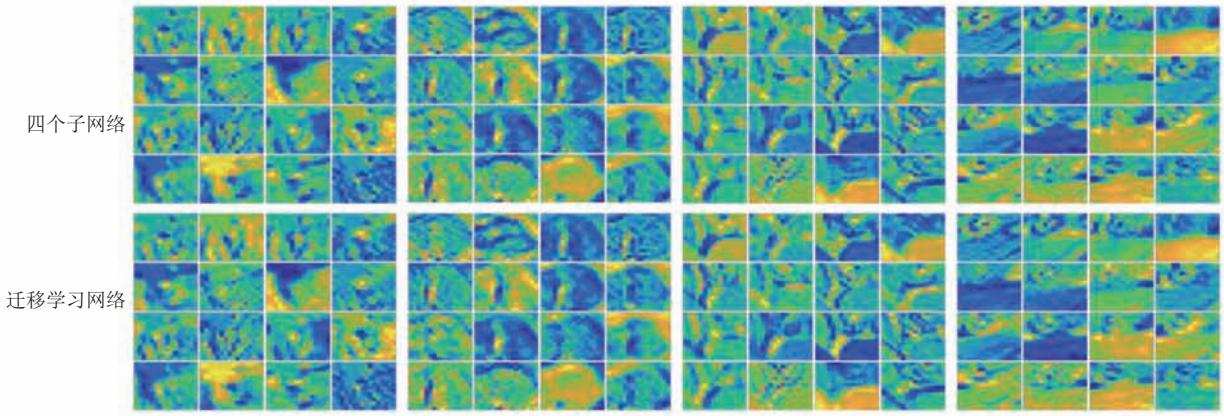

图 5　四个子网络第一层的特征图和迁移学习策略网络第一层的特征图

Fig. 5　The feature map of the first layer of the 4 sub-networks and the feature map of the first layer of the transfer learning strategy network

基于区域分解的快速卷积神经网络学习策略，在子网络的表达能力与子网络的并行训练效率之间需要衡量。具体地，随着网络分解次数的增加，子网络训练的并行效率提高，但由于网络规模减少，子网络表达能力可能不足，这就导致子网络训练学习得不好，无法很好地提取子样本的样本特征，从而影响 CNN 结果。在本节中，使用分解和组合的学习策略将 CNN 分解为不同数量的子网络，并进行实验对比。其中，CNN 使用的是 ResNet50[4]，且每个子网络都是从头开始训练的，而不是由预先训练好的 ResNet 进行训练。实验中，将 CNN 划分为 4 个子网络和 8 个子网络，并分别训练相应的子样本，其网络架构如图 6 所示。

在不同分块的实验中，网络划分为 4 个子网络和 8 个子网络，分别使用 ResNet50 和 ResNet101[4]对 Food-101 数据集进行实验。从表 3 可以看出，8 个子网络的结果优于 4 个子网络的结果。从表 4 可知，在训练时间方面，4 个子网络和 8 个子网络花费的训练时间都比

Output size	Output size	ResNet50	Sub-ResNet50 for 4 decompositions	Sub-ResNet50 for 8 decompositions
Conv1	112×112 70×70{4 decomposition } 48×48{4 decomposition }	7×7, 64, stride 2	7×7, 16, stride 2	7×7, 8, stride 2
Conv2_*x*	56×56 35×35{4 decomposition } 24×24{4 decomposition }	$\begin{Bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{Bmatrix}\times3$	$\begin{Bmatrix}1\times1,16\\3\times3,16\\1\times1,64\end{Bmatrix}\times3$	$\begin{Bmatrix}1\times1,8\\3\times3,8\\1\times1,32\end{Bmatrix}\times3$
Conv3_*x*	28×28 17×17{4 decomposition } 12×12{4 decomposition }	$\begin{Bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{Bmatrix}\times4$	$\begin{Bmatrix}1\times1,32\\3\times3,32\\1\times1,128\end{Bmatrix}\times4$	$\begin{Bmatrix}1\times1,16\\3\times3,16\\1\times1,64\end{Bmatrix}\times4$
Conv4_*x*	56×56 8×8{4 decomposition } 6×6{4 decomposition }	$\begin{Bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{Bmatrix}\times23$	$\begin{Bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{Bmatrix}\times23$	$\begin{Bmatrix}1\times1,32\\3\times3,32\\1\times1,128\end{Bmatrix}\times23$
Conv5_*x*	56×56 4×4{4 decomposition } 3×3{4 decomposition }	$\begin{Bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{Bmatrix}\times3$	$\begin{Bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{Bmatrix}\times3$	$\begin{Bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{Bmatrix}\times3$
	1×1	average pool, 101-d FC softmax		

图 6　卷积神经网络和子网络的结构

Fig. 6　Structure of CNN and sub-network

表 3　101 类食物实验结果

Table 3　Experimental results of Food-101

CNN 训练策略	ResNet50 准确率(%)		ResNet101 准确率(%)	
	Top-1	Top-5	Top-1	Top-5
CNN_transfer_4	45.25	71.84	59.95	82.48
CNN_transfer_8	12.00	32.09	29.09	55.14
CNN_random_T	70.49	89.53	72.37	53.04
CNN_ transfer_4_T	**71.69**	90.26	**73.70**	90.49
CNN_ transfer_8_T	**72.92**	90.73	**74.53**	91.30
CNN_Food101	**70.41**	--	**--**	--

注：“--”表示 Food-101 文章中没有出现 ResNet101 的结果

随机初始化的少，相较于随机初始化时间加速比在 ResNet50 中是 1.61、1.64。同样地，在 ResNet101 中，时间加速比为 1.57、1.58。在准确率方面，分解和组合学习策略相较于随机初始化方法在 ResNet50 准确率提升了 2.51%，而在 ResNet101 提升了 2.16%。从图 7 可以看出，4 个子网络迁移学习初始的准确率比 8 个子网络的好，但随着网络的训练，8 个子网络的最终结果比 4 个子网络的结果要好。两个迁移学习的初始化结果都比随机初始化要好，这说明迁移学习给网络提供了一个较好的初始训练点，对网络进行微调后的结果相较于随机初始化有一定的提升。这表明基于区域分解的快速卷积神经网络学习策略能够加快网络的训练，同时还能提高网络的性能。

表 4　101 分类数据集训练时间

Table 4　Training time of Food-101 data sets

数据集	网络	训练时间（h）
ResNet50	大网络	12.59
	4 个子网络	7.79
	8 子网络	7.67
ResNet101	大网络	14.62
	4 个子网络	9.30
	8 个子网络	9.23

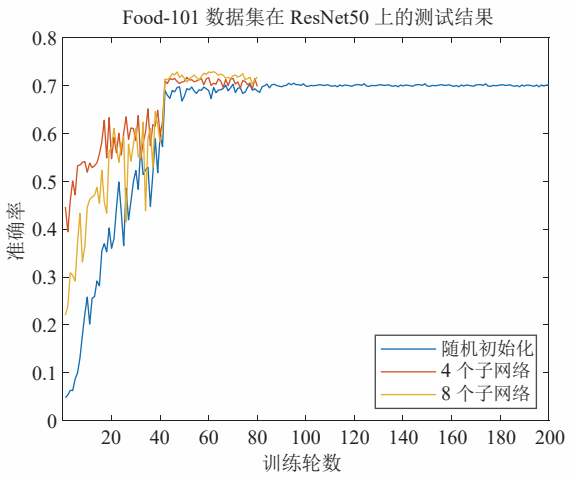

图 7　101 类食物分类测试结果

Fig. 7　Results of the dataset Food-101

4　讨论与分析

图像分类是计算机视觉领域中重要的研究方向，而卷积神经网络是处理图像分类任务最常用的方法。随着图像数据规模的增大，训练 CNN 需要耗费的时间过长和占用大量的计算资源。网络加速训练最常见的方法是数据并行，即增加 GPU 个数。Goyal 等[20]等使用 8 块 Tesla P100 GPUs 训练 ResNet50 需要 29 h，即使 GPU 增加到 256 块时，也需要 1 h。与前述方法相比，本文从网络模型方面来加速网络的训练，提出一种基于区域分解的快速卷积神经网络学习策略，也是一种迁移学习的方法。该方法先将一个大网络拆分成多个子网络并行训练，然后将子网络训练好的参数组合作为大网络的初始权重，并对大网络进行微调。在网络训练时间方面，ResNet32、ResNet56、ResNet110 在 CIFAR-10 数据集中使用分解和组合迁移学习方法的时间分别 5.65 h、9.61 h、18.78 h，相较于随机初始化的 8.53 h、13.98 h、26.74 h 训练时间有了一定的减少。其中在 ResNet110 训练时间减少得最多，达到了 7.96 h。在分类准确率方面，He 等[4]等使用 ResNet32、ResNet56、ResNet110 对 CIFAR-10 数据分类的准确率分别为 92.49%、93.03%、93.57%，而本文使用分解和组合卷积神经网络的迁移学习方法的结果是 92.86%、94.34%、94.14%，相较于 He 等[4]方法的准确率有一定的提高。其中，训练时间减少的原因是，分解和组合的学习策略中子网络的参数较少，子网络的参数组合作为大网络的初始权重，这样使用大网络训练很少的轮数就可以达到理想的结果。分类结果提升的原因是，通过将 CNN 划分成子网络，其中子网络更加关注图片的局部特征，随后通过子网络训练参数的组合作为 CNN 的初始权重，这使得 CNN 进行微调时更加注重局部特征的学习，从而提高 CNN 的分类结果。本文分解和组合学习策略也还存在一

些不足之处：(1) 子网络需划分多少合适。若子网络划分过多，则子网络的学习较差，这会影响最终 CNN 的训练。(2) 图片数据的划分策略，每个子网络要训练相应的子图片，若子图片包含的信息较少，则会影响子网络的训练结果。(3) 对 2 个网络时间还需要进一步的研究。对于以上问题，未来将对子网络的划分和图片的划分做进一步的研究，并完善基于区域分解的快速卷积神经网络学习策略不足之处。

5 结　论

本文提出一种基于区域分解的快速卷积神经网络学习策略，旨在有限计算资源条件下使用 CNN 完成大规模数据的训练并加快 CNN 的训练。在 CIFAR-10、CIFAR-100、Food-101 上的实验结果显示，基于区域分解的快速卷积神经网络学习策略比随机初始化的方法收敛得更快、分类结果更好。由此可知，卷积神经网络分解和组合的学习策略在图像分类任务中是有效的。在未来工作中，将对基于区域分解的快速卷积神经网络学习策略进行完善和改进，并尝试应用于图像分割和目标检测等任务。

参 考 文 献

[1] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks [C] // Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012: 1097-1105.

[2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.

[3] Chen LC, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.

[4] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.

[5] He KM, Zhang XY, Ren SQ, et al. Identity mappings in deep residual networks [C] // European Conference on Computer Vision, 2016: 630-645.

[6] Szegedy C, Liu W, Jia YQ, et al. Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.

[7] Keskar NS, Mudigere D, Nocedal J, et al. On large-batch training for deep learning: generalization gap and sharp minima [Z/OL]. arXiv Preprint, arXiv:1609.04836, 2016.

[8] Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [Z/OL]. arXiv Preprint, arXiv:1409.1556, 2014.

[10] Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016: 424-432.

[11] You Y, Zhang Z, Demmel J, et al. Imagenet training in 24 minutes [Z/OL]. arXiv Preprint, arXiv:1709.05011, 2017.

[12] Chan TF, Mathew TP. Domain decomposition algorithms [J]. Acta Numerica, 1994, 3: 61-143.

[13] Cai XC, Saad Y. Overlapping domain decomposition algorithms for general sparse matrices [J]. Numerical Linear Algebra with Applications, 1996, 3(3): 221-237.

[14] Cai XC. Overlapping domain decomposition methods [M] // Advanced Topics in Computational Partial Differential Equations, 2003: 57-95.

[15] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [Z/OL]. arXiv Preprint,

arXiv:1502.03167, 2015.

[16] Hu J, Shen L, Sun G, et al. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[17] Bossard L, Guillaumin M, Van Gool L. Food-101–mining discriminative components with random forests [C] // European Conference on Computer Vision, 2014: 446-461.

[18] Ciocca G, Napoletano P, Schettini R. Learning CNN-based features for retrieval of food images [C] // International Conference on Image Analysis and Processing, 2017: 426-434.

[19] Ciocca G, Napoletano P, Schettini R. CNN-based features for retrieval and classification of food images [J]. Computer Vision and Image Understanding, 2018, 176/177: 70-77.

[20] Goyal P, Dollár P, Girshick R, et al. Accurate, large minibatch SGD: training imagenet in 1 hour [Z/OL]. arXiv Preprint, arXiv:1706.02677, 2017.

[21] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors [J]. Nature, 1986, 323(6088): 533-536.

[22] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.

[23] Krizhevsky A. One weird trick for parallelizing convolutional neural networks [Z/OL]. arXiv Preprint, arXiv:1404.5997, 2014.

[24] Pan SJ, Yang Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(10): 1345-1359.

[25] Mishkin D, Matas J. All you need is a good init [Z/OL]. arXiv Preprint, arXiv:1511.06422, 2015.

[26] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[27] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.

[28] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.

[29] Smith B, Bjorstad P, Gropp W. Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations [M]. Cambridge: Cambridge University Press, 2004.

[30] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks [C] // Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011: 315-323.

[31] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [Z/OL]. arXiv Preprint, arXiv:1502.03167, 2015.

[32] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.

[33] Sudre CH, Li WQ, Vercauteren T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations [M] // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision, 2017: 240-248.