DOI:10.16644/j.cnki.cn33-1094/tp.2020.10.017

基于通道域自注意力机制的图像识别算法*

何海洋[1]，孙　南[2]

(1. 河南大学软件学院，河南 开封 475004；2. 河南大学计算机与信息工程学院)

摘　要：为了提高自注意力网络的图像识别效果，对通道域自注意力机制存在的梯度死亡的问题，提出了一种优化算法。首先聚集特征在空间范围上的全局特征响应，然后使用新的激活函数对通道关系建模，构造通道权重响应。将改进后的通道域自注意力模块使用于神经网络分类器中，在CIFAR-10和CIFAR-100数据集上进行实验，结果显示，和传统模型相比，图像识别准确率提高了1.3%和1.4%，验证了所提算法的有效性。

关键词：通道域自注意力；神经网络；激活函数；图像识别

中图分类号：TP391　　**文献标识码**：A　　**文章编号**：1006-8228(2020)10-64-05

Image recognition algorithm based on channel domain self-attention mechanism

He Haiyang[1], Sun Nan[2]

(*1. Henan University College of Software, Kaifeng, Henan 475004, China; 2. Henan University School of Computer and Information Engineering*)

Abstract: In order to improve the image recognition of the self-attention network, an optimized algorithm is proposed for the problem of gradient death that exists in the channel domain self-attention mechanism. Firstly, the global feature response over the spatial extent of the feature is aggregated, and then the new activation function is used to model the channel relationship and to construct the channel weight response. The improved channel-domain self-attention module is used in neural network classifier, and experiments are carried out on CIFAR-10 and CIFAR-100 datasets. The experiment results show that the accuracy of image recognition is improved by 1.3% and 1.4% compared with the traditional model, which verifies the effectiveness of the proposed algorithm.

Key words: channel domain self-attention; neural network; activation function; image recognition

0 引言

图像识别是计算机视觉领域中的三大任务之一，其是通过分析图片的整体特征，找到目标图片的所属类别。目前，深度学习在图像识别领域取得了突破性的进展。

随着NLP中的注意力机制广受欢迎[1]，注意力机制引起了计算机视觉领域的重视，开始慢慢渗透到研究的主体结构中，以补充现有的CNN体系结构或完全替代它们。

注意力机制是网络架构的一个组成部分，负责管理和量化信息之间的相互依赖关系。对于在输入和输出元素之间的关系，一般称它为General Attention，而输入元素内部的关系叫做自注意力(Self-Attention)。自注意力是一种与单个序列的不同位置相关联的注意力机制，它接受n个输入，并返回n个输出，其目的是计算同一序列的表示形式。大量文献表明自注意力机制在机器阅读，抽象概念总结和图像描述及生成中非常有用[2-5]。

空间域自注意力[6]提出了一个叫空间转换器(spatial transformer)的模块，将图片中的的空间域信息做对应的空间变换，从而能将关键的信息提取出来，但是因为卷积层之后，每一个卷积核(filter)产生的通道信息，所含有的信息量以及重要程度其实是不一样的，都用同样的transformer其实可解释性并不强，它忽略了通道域中的信息。文献[7]将卷积核的卷积类似于信号做了傅里叶变换，从而能够将这个特征一个通道的信息给分解成64个卷积核上的信号分量，给每

收稿日期：2020-06-01

***基金项目**：国家自然科学基金资助项目(61602525)；河南省科技发展计划项目“大数据环境下基于语义情感分析的个性化推荐算法研究”(182102210229)

作者简介：何海洋(1996-)，男，河南信阳人，硕士研究生，主要研究方向：计算机视觉。

通讯作者：孙南(1992-)，男，河南省安阳市人，硕士研究生，主要研究方向：目标检测。

个通道上的信号都增加一个权重，来代表该通道与关键信息的相关度的话，这个权重越大，则表示相关度越高，也就是越需要去注意的通道了，但是加权函数出现了梯度死亡，不能还原通道间的复杂相关性。本文针对加权函数梯度死亡的缺点，对相关函数做出了相关的改进，避免了网络出现梯度死亡，更好的还原通道间的复杂相关性。实验结果表明，经过优化后的模型具有更加准确的识别率。

1 研究基础

通过使用软自注意力对卷积特征的通道之间的相互依赖性进行建模，对于CNN特定层中的通道响应重新加权。基于这种想法，研究人员提出了Squeeze-And-Excitation 模块，如图1所示。

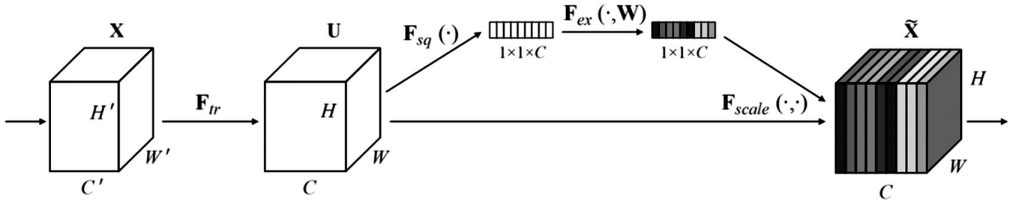

图1 Squeeze-And-Excitation模块

Squeeze-And-Excitation模块的工作原理如下：对于特征F_{tr}从X到U的任何一个转换（例如卷积操作），都会有一个转换操作F_{sq}，它聚集了特征在空间范围(H,W)上的全局特征响应，这是Squeeze操作。在Squeeze操作之后是Excitation操作F_{ex}，这和循环神经网络中门机制[8-9]很像，通过给通道信息一个系数，抑制差的通道信息，激励重要的通道信息，来构造通道权重响应。随后F_{tr}的输出在通道方向上乘以Excitation操作的结果(即图中的F_{scale}操作)。Squeeze操作的数学公式可以表示为：

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{1}$$

其中，u_c是F_{tr}操作的输出，(i,j)是信息在特征图上的位置表示。Squeeze操作是通过使用全局平均池化来创建信道中全局信息的聚合。

另一方面，Excitation操作的数学公式可以表示为：

$$S = F_{ex}(Z,W) = \sigma(\varphi(Z,W)) = \sigma(W_2\delta(W_1 Z)) \tag{2}$$

Excitation操作将Squeeze操作的输出乘以网络学习到的权重W_1，将结果通过ReLU函数δ，接着输出与另一组权重W_2相乘，最后使用sigmoid函数以确保所有的通道权重都是正数。

在Excitation操作过程中，W_1会除以因子r来减少通道数，而W_2将乘以r再次将通道维数增大到原始通道数。这样是为了减少网络的计算量。最后，将F_{tr}的通道特征乘以从Excitation操作中获得的权重，就可以看成在通道上使用了全局信息的自注意力机制。

挤压和激励块的主要思想是在神经网络的决策过程中可以读取全局信息。卷积操作只可以查看特定范围内的局部空间信息，而Squeeze-And-Excitation模块可以汇总来自整个特征图的信息。

2 针对通道域自注意力的改进方案

2.1 通道域自注意力机制存在的问题

在通道域自注意力的模型中，可注意到Excitation操作中使用ReLU函数和全连接层来模拟通道之间的相互依赖性。网络经过全连接层和ReLU函数可以模拟出卷积层的非线性，但是使用ReLU激活函数时，不会得到非常小的值，只能得到0。当计算梯度时，如果太多数值小于0，梯度更新会等于零，因此得到了一堆未更新的权重和偏差，这就是ReLU死亡问题。

假设一个简单的网络结构，存在一个输入神经元a_0，在经过权重w_1和偏置b_1后，神经元a_0生成了神经元a_1，以此类推，在经过权重w_2, w_3, w_4和偏置b_2, b_3, b_4后依次生成了神经元a_2, a_3, a_4，最后a_4输出了C。

根据神经网络的梯度公式，可以很简单算出b_1的梯度，见公式(3)。

$$\frac{\partial C}{\partial b_1} = \sigma'(w_1 a_0 + b_1) w_2 \sigma'(w_2 a_1 + b_2) w_3 \sigma'(w_3 a_2 + b_3) w_4 \sigma'(w_4 a_3 + b_4) \frac{\partial C}{\partial a_4} \tag{3}$$

可以发现如果等式中某个神经元经过ReLU函数变成了0，那么偏置b_1的梯度也会归零，无法进行。同理其他与该神经元相关的偏置都不会得到梯度更新。这种问题直接会影响到网络的梯度更新，降低网络识别效果，因此要对该问题进行针对性改进。

2.2 通道域自注意力的激活函数优化

针对ReLU函数可能会出现的死亡问题，本文基于ReLU激活函数做了改进，提出FeLU函数，解决了ReLU函数出现的问题。FeLU函数的数学公式如下：

$$FeLU(x)=\begin{cases}\dfrac{e^{x}-1}{1+e^{-x}}, & x<0\\ \mathrm{x}, & x>0\end{cases} \tag{4}$$

FeLU激活函数图像如图2所示，可以发现在$F(x)$函数中如果输入的x值大于零，则输出值与ReLU相同，是等于x的值。但是如果输入值x小于0，输出结果将得到一个略低于零的值。这样可以保证梯度不会出现死亡，停止更新。同时，和ReLU函数相比，FeLU函数存在负值，激活神经元的输出平均值可以更加靠近0，神经元的输出平均值与0越近，激发单元的偏移效应越低，网络的梯度更靠近自然梯度，同时也能起到批量正则化的效果。

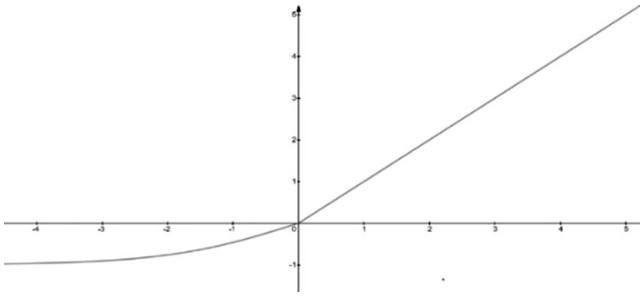

图2 FeLU激活函数

对ReLU激活函数改进之后，Squeeze–And–Excitation模块可以表示为图3中的结构。

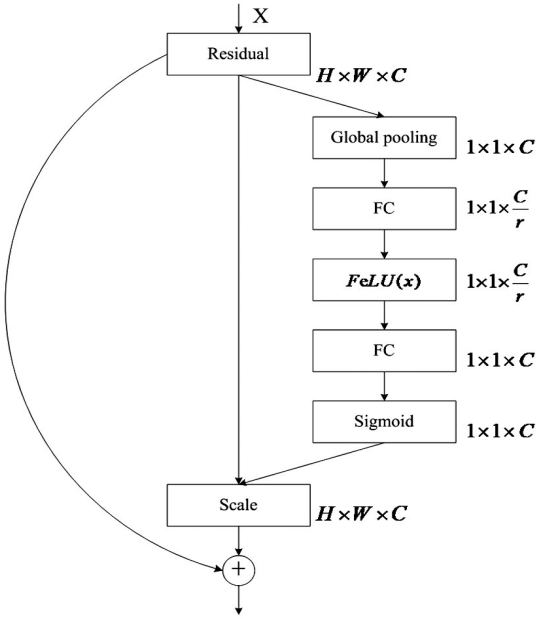

图3 改进Squeeze–And–Excitation模块

当形状为$H\times W\times C$的特征图传入到Squeeze–And–Excitation 模块，首先会进行一次全局池化变成$1\times 1\times C$的特征图，也就是Squeeze操作。然后进行了2次全连接操作，用来模拟通道间的复杂关系，在两次全连接操作中加入了FeLU激活函数用来避免ReLU死亡问题，第一次全连接层会将特征图的通道数放缩16倍，这可以减少神经网络的计算量，降低运行时间，第二次全连接层会将特征图还原到原始通道数。最后将全连接输出的特征值传到Sigmoid函数中，放缩到0和1之间，等待加权。这是Excitation操作。将输入特征信息乘以从Excitation操作中获得的权重，就完成了在通道上的自注意力机制。

3 实验模型建立及结果分析

3.1 实验模型的建立

实验建立了一个小型神经网络，包含一个输入层，一个添加了通道域自注意力模块的卷积层和一个全连接输出层。实验的数据集选用了MNIST数据集，并对数据集分别进行了归一化和标准化。为了减少不可避免的随机性的影响，每组实验被测试5次，每次网络会迭代100次。这些实验的训练精度将取平均值，这将有助于确保在实验中看到更具代表性的结果。

实验中的评价标准分为四个，为了方便标识，本文使用A、B、C、D四个选项分别表示这四个实验评价标准。其中A表示5次实验中最后一次迭代的验证集精度平均值；B表示5次实验中训练集精度平均值最高的验证集精度；C表示5次实验中最后一次迭代的验证集精度最大值；D表示所有实验中的验证集精度最大值。

对于图像识别模型的比较实验，在ResNet神经网络[10]中添加了改进后的通道域自注意力模块，和其他神经网络在CIFAR–10和CIFAR–100数据集[11]上进行了实验精度和时间上的比较。实验中参数设定了mini–batch为128，Momentum为0.9，learningrate为0.1。

3.2 实验结果对比与分析

对于自注意力模块的激活函数的比较实验，为了减少梯度下降时间，实验中对数据集分别进行了归一化和标准化，实验结果见表1和表2。

从实验结果可以发现，无论是归一化数据集还是标准化数据集，FeLU激活函数的实验精度基本都高于其他激活函数。而同种激活函数之间数据比较可以发现，经过标准化数据训练的网络的性能稍微差一点。但是总的来说，本文FeLU激活函和其他的激活函数在通道域自注意力机制神经网络中的表现相比，实验的验证集精度都要高一些，证明了经过改进后的FeLU激活函数的可行性和有效性。

表1 归一化数据集实验结果

激活函数	A	B	C	D
SoftPlus	97.63	97.76	97.68	97.89
Sigmoid	97.53	97.61	97.60	97.76
SoftSign	97.37	97.46	97.48	97.60
ReLU	97.42	97.54	97.52	97.77
Leaky_ReLU	97.36	97.55	**97.97**	97.73
FeLU	**97.89**	**98.01**	97.92	**98.16**

表2 标准化数据集实验结果

激活函数	A	B	C	D
SoftPlus	97.16	97.61	97.34	97.78
Sigmoid	97.33	97.44	97.46	97.63
SoftSign	97.43	97.50	97.55	97.66
ReLU	97.12	97.45	97.37	97.73
Leaky_ReLU	97.31	97.35	**97.97**	97.52
FeLU	**97.77**	**97.91**	97.88	**98.10**

对于图像识别模型，本文和DSN[12]，文献[13]，文献[14]，Highway Network[15]，All-CNN[16]，SENet进行测试精度的比较，实验结果如表3所示。

表3 数据集实验结果

神经网络模型	CIFAR-10testerror	CIFAR-100testerror
DSN	7.97	34.57
Highway_Network	7.6	32.24
ALL-CNN	7.25	33.71
文献[13]	**2.92**	19.52
文献[14]	3.21	16.65
SENet	4.46	17.67
本文模型	3.24	**16.17**

从表3中可以发现，在两个数据集上，包括DSN、Highway Network和All-CNN神经网络在内的无注意力机制神经网络的实验准确度明显低于自注意力神经网络。而在SENet、文献[13]和文献[14]等自注意力神经网络的实验数据中，本文模型在CIFAR-10数据集上排名第二，测试集实验误差为3.14%，但是在CIFAR-100数据集上的实验表现最佳，测试集实验误差仅为16.27%。其中和包含Squeeze-And-Excitation模块的SENet网络相比，本文模型实验中准确率更高，表现的更好，这证明了本文模型在针对Squeeze-And-Excitation 模块的激活函数改进是有效果的，体现了本文算法优化策略的价值和有效性。同时本文和SENet，文献[13]，文献[14]等自注意力神经网络进行比较了神经网络运行时间的比较，如表4所示。

从表4中，可以发现本文模型和SENet，文献[14]等传统神经网络相比，本文模型网络训练和测试的时间高了15%~23%，这是因为本文模型为ResNet网络架构，神经网络的整体架构较为复杂，网络的层数深，和其他神经网络模型相比，消耗了大量的网络计算时间。但是就单张图像识别而言，本文模型的图像识别时间是可以满足识别实时性要求的，在提高图像识别准确率的条件下，这些消耗的时间仍然是可以容忍的。

表4 CIFAR数据集实验时间

算法	CIFAR-10数据集		CIFAR-100数据集	
	训练时间/s	测试时间/s	训练时间/s	测试时间/s
SENet	**10924**	**97**	11021	98
文献[13]	14458	118	**10845**	**97**
文献[14]	10956	98	11047	107
本文模型	13730	113	14500	114

4 结论

本文根据通道域自注意力模块的缺点提出了FeLU激活函数，并进行了激活函数相关实验对比，通过实验数据的分析，证明了FeLU激活函数是可行的。并且在ResNet神经网络的基础上添加了基于FeLU的通道域自注意力机制，在实验对比中，本文算法取得了不错的效果。但是模型在通道压缩阶段使用的全局池化会导致一些空间特征信息的丢失，在接下来的工作中我们将重点研究通道信息压缩，实现空间特征信息的有效保留。

参考文献(References):

[1] Firat O, Cho K, Bengio Y. Multi-way, multilingual neural machine translation with a shared attention mechanism[J]. arXiv preprint arXiv:1601.01073,2016.

[2] Shen T, Zhou T, Long G, et al. Disan: Directional self-attention network for rnn/cnn-free language understanding[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[3] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 3146-3154

[4] Zhu Y, Ko T, Snyder D, et al. Self-Attentive Speaker Embeddings for Text-Independent Speaker Verification [C]//Interspeech,2018:3573-3577

[5] Ambartsoumian A, Popowich F. Self-attention: A better building block for sentiment analysis neural network classifiers[J]. arXiv preprint arXiv:1812.07860,2018.

[6] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial

(下转第72页)

应用[J].计算机工程与设计,2006.21:4153–4156
[2] 胡树煜.医学图像中粘连细胞分割方法研究[J].计算机仿真,2012.29(2):260–262,27
[3] 苏士美,吕雪扬.骨髓细胞图像的小波变换与K–means聚类分割算法[J].郑州大学学报(工学版),2015.36(4):15–18
[4] 刘应乾,曹茂永.基于Gabor滤波与区域生长的细胞分割[J].山东科技大学学报(自然科学版),2012.31(2):99–103
[5] Ronneberger O, Fischer P, Brox T. U–net: Convolutional networks for biomedical image segmentation[C]//Proceedings of the International Conference on Medical image computing and computer–assisted intervention. Berlin, Germany:Springer,2015:234–241
[6] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified Real–Time Object Detection[C]. IEEE Conference on Computer Vision and Pattern Recogniton. IEEE,201:779–788
[7] Liu W, Anguelovd E, et al. SSD: Single Shot Multi Box Detector[J]. Computer Vision–ECCV 2016. Springer International Publishing,2016:21–37
[8] R. Girshick, Fast R–CNN, in IEEE International Conference on Computer Vision (ICCV),2015.
[9] S. Ren, K. He, R. Girshick, and J. Sun. Faster R–CNN: To–wards real–time object detection with region proposal net–works[C]. In NIPS,2015.
[10] He K, Gkioxari G, Dollar P, et al. Mask r–cnn[C]. International Conference on Computer Vision. New York:IEEE,2017:2980–2988
[11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014.39(4):640 – 651
[12] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. Conference on Computer Vision and Pattern Recognition, 2016: 936–944
[13] Kaiming He, Xiangyu Zhang, et al. Deep Residual Learning for Image Recognition[C]. Conference on Computer Vision and Pattern Recognition,2015.
[14] Fisher Yu, Vladlen Koltun. Multi–Scale Context Aggregation by Dilated Convolutions[C]. Conference on Computer Vision and Pattern Recognition,2016.
[15] Chen L C, Papandreou G, Kokkinos I, et al. Deep Lab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Trans on Pattern Analysis & Machine Intelligence,2016.40(4):834–848
[16] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[C]. European Conference on Computer Vision,2014:740–755

(上接第67页)

Transformer Networks[J]. 2015. Simonyan K, Zisserman A. Very deep convolutional networks for large–scale image recognition[J]. arXiv preprint arXiv:1409.1556,2014.
[7] 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
[8] 赵鹏,刘扬,刘慧婷等.基于深度卷积–递归神经网络的手绘草图识别方法[J].计算机辅助设计与图形学学报,2018.2:217–224
[9] 刘礼文,俞强.循环神经网络(RNN)及应用研究[J].科技视界,2019.32.
[10] 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
[11] Liang M, Hu X. [IEEE 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) – Boston, MA, USA (2015.6.7–2015.6.12)] 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) –Recurrent convolutional neural network for object recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society,2015:3367–3375
[12] Lee C Y, Xie S, Gallagher P, et al. Deeply–supervised nets[C]//Artificial intelligence and statistics,2015:562–570
[13] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:3156–3164
[14] 袁嘉杰,张灵,陈云华.基于注意力卷积模块的深度神经网络图像识别[J].计算机工程与应用,2019.55(8):9–16
[15] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks[C]//Advances in neural information processing systems,2015:2377–2385
[16] Springenberg J T, Dosovitskiy A, Brox T, et al.Striving for simplicity: The all convolutional net[J]. arXiv preprint arXiv:1412.6806,2014.