文章编号:1671-4598(2022)07-0194-07 DOI:10.16526/j.cnki.11-4762/tp.2022.07.029 中图分类号:TP3 文献标识码:A

基于三维注意力机制的车辆重识别算法

方彦策[1]，张宏江[2]，谢雨成[1]，刘培顺[1]
(1. 中国海洋大学 信息科学与工程学部，山东 青岛 266100；
2. 中国运载火箭技术研究院 研究发展部，北京 100091)

摘要：为解决套牌车识别难度大的问题，通过深度学习的技术，基于ResNet-50，结合通道注意力机制和位置注意力机制，设计了一种三维注意力机制对近似车辆进行精确识别；解决了当前大部分注意力算法都关注于一维的通道注意力和二维的位置注意力，而处理的图像数据是三维的，不能将注意力集中在所有需要关注的区域，造成部分关键信息遗失的问题；该三维注意力机制在多种视觉任务下均有很好的效果，在Cifar100数据集上，相比SENet有1.12%的提升，在PKU VehicleID数据集上，相比SENet平均有2%的提升。

关键词：注意力机制；深度学习；套牌车识别；交通管理；

Vehicle Recognition Algorithm Based on 3D Attention Mechanism

FANG Yance[1], ZHANG Hongjiang[2], XIE Yucheng[1], LIU Peishun[1]
(1. College of Information Science and Engineering, Ocean University of China, Qingdao 226019, China;
2. R&D Department, China Academy of Launch Vehicle Technology, Beijing 100076, China)

Abstract: In order to solve the problem in recognizing the fake-licensed vehicles, by using deep learning technology and combining the channel attention mechanism and position attention mechanism, based on ResNet-50, a accurate identification method of similar vehicles based on three-dimensional attention mechanism is proposed. At present, most attention algorithms focus on one-dimensional channel attention and two-dimensional positional attention, while the images are processed in three dimensions. Therefore, these attention mechanisms cannot focus on all attention areas, which result in the loss of some key information. The three-dimensional attentional mechanism works well in a variety of visual tasks, with the improvement of 1.12 % over the SENet on the Cifar100 dataset, and the improvement of 2% over the SENet on average on the PKU VehicleID dataset.

Keywords: attention mechanism; deep learning; fake-licensed vehicles; traffic management computer

0 引言

车辆重识别技术能给我们的生活带来便利，提供安全保障，例如智慧安防、智慧交通方面：公安涉车侦查、套牌车比对、无牌车管理、违法抓拍、道路拥堵路况检测、交通异常行为分析、停车收费管理等。

近年来，随着机器学习技术的提升，用机器学习算法实现车辆重识别逐渐成为研究热点。2015年，文献[2]通过提取车辆前脸的HOG特征，采用线性判别的分析技术提取车辆梯度方向直方图特征进行车辆识别，但无法处理遮挡问题。文献[3]采用主成分分析PCA算法，通过特征提取，实现对车辆的识别，PCA是非监督的机器学习算法，车辆重识别率较低。文献[4]提出对车辆图像提取的VAR特征图像，统计图像的LBP直方图特征，利用SVM进行分类。文献[5]针对SVM模型识别性能不足且训练时间过长的问题，提出了一种基于Haar类特征和改进的级联分类器的图像分割方法，但该方法未考虑数据不平衡导致的分类精度下降的问题。文献[6]基于Alexnet网络构建了9层的深度卷积神经网络用于识别车辆。该方法与机器学习的车辆识别方法相比，具有更高的准确率。文献[8]等采用自适应算法优化BP神经网络。对车辆识别方面具有较高识别准确率和鲁棒性。文献[9-10]利用卷积神经网络提取的特征实现对车型的识别，显著提高了车辆识别的准确率。文献[11]提出使用PCANet网络对车型进行识别，该网络模型具有较强的抗畸变能力。目前这些车辆重识别的算法大部分是基于车辆的整体特征进行识别，对于某些外观几乎一样的套牌车识别率效果比较差，例如图1所示的套牌车。从图中可以可以看出套牌车最明显的差别位置在车窗，同一车型的车辆在使用一段时间之后，因为使用人的习惯不同，前车窗最容易产生明显的差别，因此本文主要根据前车窗的特征进行车辆重识别，与人脸识别类似，本文称之为车脸识别。

“车脸识别”与“人脸识别”类似，不是靠车牌识别车

收稿日期:2022-03-26； **修回日期**:2022-04-11。
基金项目:国家重点研发计划(2017YFC0806200)。
作者简介:方彦策(2001-),男,浙江杭州人,大学本科,主要从事计算机视觉方向的研究。
通讯作者:刘培顺(1975-),男,山东巨野人,博士,副教授,主要从事人工智能方向的研究。
引用格式:方彦策,张宏江,谢雨成,等.基于三维注意力机制的车辆重识别算法[J].计算机测量与控制,2022,30(7):194-200.

图 1　套牌车

辆，而是识别车辆的前部特征来区分不同车辆。在车脸识别算法研究中，本文用到了注意力机制。注意力机制通过模仿人类对于信息的注意力分配的不均衡，聚焦有用信息，来获得性能提升。近年来，将通道注意力，位置注意力等引入深度神经网络，在提升卷积网络的图像分类、目标检测、语义分割等领域均取得了很大的进展。

大部分注意力机制例如 SENet[12]，CBAM[13]，GCNet[14]等都通过叠加卷积，池化和激活函数等获得通道或位置特征。SENet 获得的是 1 维的通道注意力，首先通过压缩操作使用一个平均池化来嵌入全局信息，再通过激励操作使用一个具有两个全连接层的结构来获得通道之间的关系。CBAM 通过两个模块串行的方式获得了同时具有通道和位置特征的矩阵，以串行的方式来获取同时具有通道和位置注意力的图像。是一个 2 维的注意力机制。GCNet 使用了和 SE 块相似的结构，将 Non-local 块和 SE 块结合的方式使 GCNet 得到了比两者网络都更好的效果。SimAM[15]提出了 3 维注意力的概念，但实现的方法与其他注意力模块完全不同，引入神经科学的概念，用一种无参的方式获得 3 维注意力，但效果与 SENet 类似。可以看出大部分注意力算法都关注于一维的通道注意力和二维的位置注意力，而处理的图像是三维的，因此这些注意力机制往往不能将注意力集中在所有需要关注的区域，造成部分关键信息遗失。

本文介绍一种新的三维注意力机制，通过结合一维的通道注意力和二维的位置注意力机制，得到三维的图像注意力权重矩阵，计算后得到经注意力分配的新图像。相比 SENet，在 ResNet18[16]和 ResNet34 上参数量几乎没有增长，在 ResNet50 和 ResNet101 上增长了 2%左右。在 Cifar100[17]数据集上，以 ResNet50 作为主网络加入注意力后，相比 SENet 有 1.12%的提升。在 ImageNet[18]数据集上，比 SENet 有接近 4.5%的提升。

1　车脸识别网络模型

1.1　车脸识别网络模型概述

为了实现车脸识别，本文在 ResNet 网络基础上添加三维注意力，并对 ResNet 网络的大量结构进行重新设计，形成 TDANet 作为主干网络。

在主干网络结构中，通过实验对比分析发现网络中的 block 对于结果的影响是最重要的，因此将 ResNet 的（3，4，6，3）的 block 数量比例改为了 Swin-Transformer[19]所使用的（3，3，9，3）的比例，对于更大的模型，使用了（1，1，9，1）的 block 比例。

为了提高网络的性能，借鉴了 ResNeXt[20]中分组卷积的方式，使用了深度卷积的方式，即分组数等同于输入通道数的卷积方式，即将原本的大卷积拆成多个小卷积的并行计算，再将结果结合。

在 Transformer[21]所使用的网络中，总共使用了一个激活层，而在 ResNet 中则可以看到大量的激活层，参考 Transformer 结构我们将 TDANet 中的部分激活层去除。

我们在主干网络中使用了反转模块的方式，与原本的从多通道——少通道——多通道的方式不同，我们采用了少通道——多通道——少通道的方式来减少计算量，同时尽可能地避免了从多通道到少通道的转换带来的信息损失。

网络结构如图 2 所示，车脸图像输入到 TDANet 网络中，得到尺寸为 $C\times H\times W$ 的特征图 X，其中 C 表示通道数量，H 表示高度，W 表示宽度。特征图 X 输入到主干网络中的三维注意力模块，然后将得到的注意力特征和原特征图 X 融合在一起，得到新的注意力分配后的特征图。网络的最后一层是分类层，由 1 000 个神经元的全连接层组成，和 ResNet 相同，将前面经过多次卷积后高度抽象化的特征进行整合，归一化，对每一种类别都输出一个概率，代表图片属于该类别的可能性。

1.2　三维注意力机制

1.2.1　通道注意力

图像的每一个通道都具有不同样的特征。因此，通道注意力是关注什么样的特征是有意义的。通道注意力是一个一维的注意力，它对不同通道区别对待，对所有位置同等对待。如图 2 所示。

通道注意力的计算包括 3 个步骤：

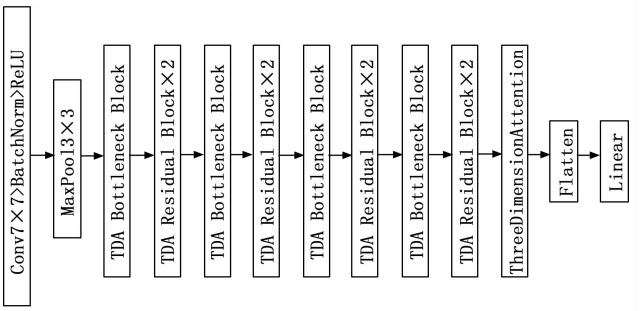
1）对于输入特征图 X，本文首先使用均值池化从每个通道的 $H\times W$ 图像中提取特征，相比于最大值池化，均值池化在计算量接近的情况下，能表示更多信息，同时实验证明均值池化注意力比使用最大值池化效果好。均值池化的公式为：

$$x_i^j=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{ij},x_i^j\in R^C,X\in R^{C\times H\times W}\qquad(1)$$

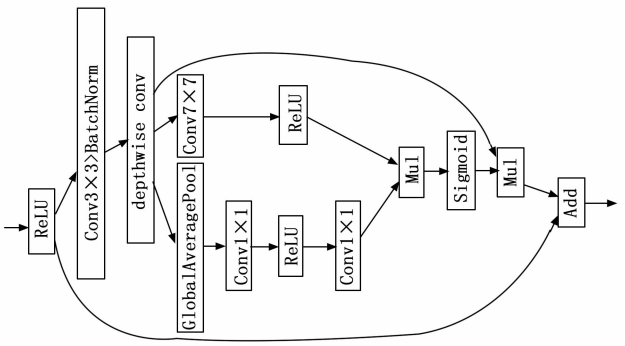
其中：$x_i^j\in R^C$ 代表进行均值池化操作后所获得的向量，$X\in R^{C\times H\times W}$代表输入的图像，$H$ 和 W 分别代表了图像的高度和宽度。

2）上述操作之后可以得到尺寸为 $C\times 1\times 1$ 的特征向量，它们具有所有通道的全局感受野。因为在注意力模块上，全连接层相比卷积层进行了很多额外且效果不明显的计算，全连接层的降维也给通道注意力的预测带来了一定的副作用。借鉴了 ECANet 的结构，本文用卷积层代替了全连接层。这里本文设计了两个连续的输入和输出通道相差 16 倍的卷积层，增强网络的非线性，实践证明这种方式可以更合理地描述图像的特征。其计算公式为：

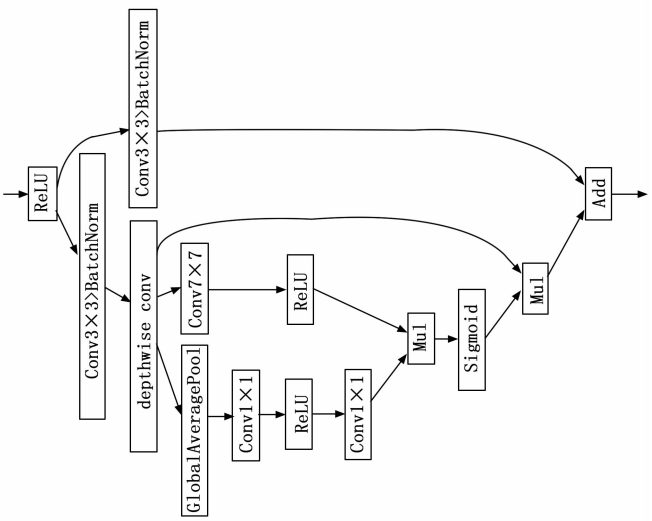
$$y_i^j=\sigma\Big(\sum_{j=1}^{k}\omega_i^jx_i^j\Big),y_i^j\in R^{\frac{C}{16}},x_i^j\in R^C\qquad(2)$$



(a)TDANet



(b)TDA Bottleneck Block



(c)TDA Residual Block

图 2 网络总体结构

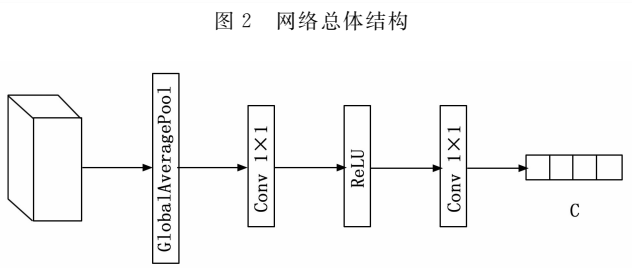


图 3 通道注意力

σ 代表了激活函数 ReLU，ω_i^j 代表了第一个卷积层中的参数，$y_i^j \in R^{\frac{C}{16}}$ 代表了特征向量 x_1 通过卷积操作和 ReLU 函数后所得到的一个只含原特征图 1/16 通道数的中间特征向量。

3）然后再经过第二个卷积层，如公式（3）所示，就得到代表了通道注意力的 $\omega_i^{ca} \in R^C$ 特征向量，它是一个含有各通道权重的 $C\times1\times1$ 的向量。

$$w_i^{ca} = \sum_{j=1}^{k} \varphi_i^j y_i^j, W_i^{ca} \in R^C, y_i^j \in R^{\frac{C}{16}} \tag{3}$$

w_i^{ca} 代表了最终得到的各通道的权重，φ_i^j 代表了第二个卷积层中的参数。

1.2.2 位置注意力

在通道注意力模块之后，我们再引入空间注意力模块来关注哪里的特征是有意义的。和通道注意力不同，位置注意力机制通过一层卷积将所有通道的信息提取成一个通道，得到一个 $1\times H\times W$ 的特征图，代表图像中每个像素所拥有的注意力权重。在 ReLU 层的激活函数操作之后，可以得到代表每个通道内的图像中每个像素的权重的矩阵。位置注意力的公式：

$$w_i^{ca} = \sigma(f^{7\times7}(X)), w_i^{ca} \in R^{H\times W}, X \in R^{C\times H\times W} \tag{4}$$

$X \in R^C$ 代表输入的图像，σ 代表了激活函数 ReLU，代表了卷积核大小为 7 的卷积操作。

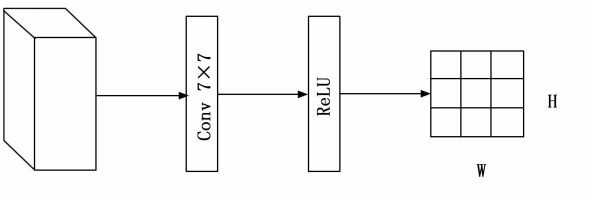


图 4 位置注意力

1.2.3 三维注意力模型

CBAM 发现在获得的特征图相加的情况下，通道注意力和位置注意力并行的结合方式不如串行的结合方式，并且由实验发现先通道后位置的方式拥有更好的效果[13]。在 DANet 中采用位置注意力和通道注意力两种注意力得到两种特征图，然后分别和原图相乘后利用相加的方式进行结合[22]这样的结合方式均只是考虑了简单的加法，位置和通道的联系并不紧密。在多注意力结合的领域，几乎所有的多注意力结合方式都使用了最原始的加法来对特征进行融合。对于 $C\times H\times W$ 的特征图，每一个点的像素，在通道注意力和位置注意力中，并没有独立的权重，而是由有通道权重而无位置权重的特征图 X_1，和有位置权重而无位置权重的特征图 X_2 相加而得，这样相当于假设通道权重和位置权重对每个点具有相同的影响。如果点 a，b 在位置的中的权重分别接近于 0，a，而 a，b 所在通道的权重接近于 $2a$，a，在相加的情况下，这两个点所得到的权重是相似的，而事实上点 a 所受到的关注应该远小于 b。实验中我们发现通道权重和位置权重对于每个点的影响是独立的并且很难做到相同。

因此我们提出了一种全新的，更加符合图像本质的融合方式，这种结合方式的公式如下：

$$\omega_i = X * sigmoid(w_i^{ca} * w_i^{sa}),$$

$$\omega_i \in R^{C\times H\times W}, X \in R^{C\times H\times W}, w_i^{ca} \in R^{C\times H\times W}$$

由通道注意力得到的尺寸为 $C\times1\times1$ 的张量代表了每个

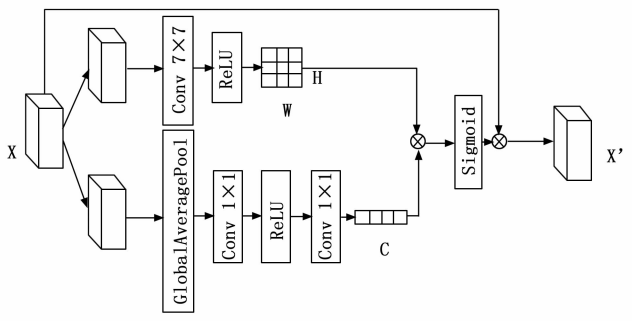


图 5　三维注意力

通道的权重，由位置注意力得到的尺寸为 $1\times H\times W$ 的张量代表了位置中每个像素的权重，将两个张量相乘可以得到 $C\times H\times W$ 的张量，每个点所具有的权重等于该点所在通道的权重和该点在位置中的权重之积。在此种结合方式中，注意力能够更加集中，权重高的像素点意味着它在通道和位置上都受到重视。这种结合方式更符合点的实际权重。

Sigmoid 函数的添加是为了防止出现特别大的数，而不会影响像素之间的相对关系。同时也能防止乘法带来的极小的数，对此，我们研究了 $sigmoid\ (w_i^{ca} * w sa_i)$ 的值分布，所有的值均分布在 0～1 中，随机挑选 100 张照片统计后，可以得到如下的 W 值分布，极小的值的出现次数比较少，大多数值的分布聚集在 0.5～0.6 和 0.9～1.0 的区间之中。三维注意力模型去除了约三分之一不重要的区域，而把更关注图中其他三分之二的区域，和对所有区域视为同样重要的无注意力的方法，提高了很多的性能。

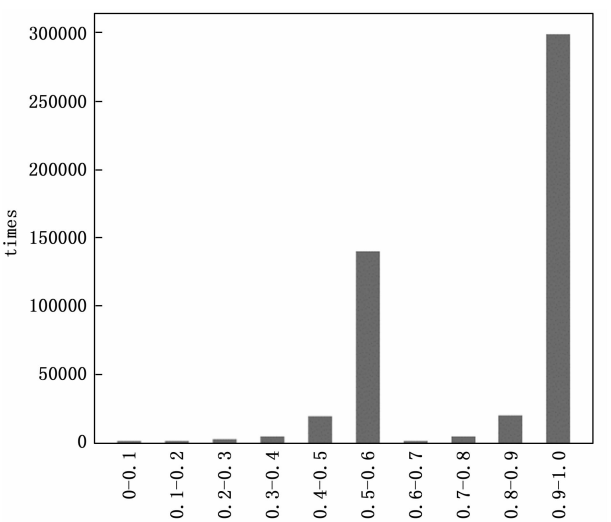


图 6　W 值分布统计

1.2.4　损失函数

本文设计的模型的损失函数采用了 Arcface loss[23] 损失函数，它由 Softmax loss 改进而来。

$$L_1=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i+b_{y_i}}}{\sum_{j=1}^{n}e^{w_w^T x_i+b_j}} \tag{6}$$

Softmax loss 函数中，$x_i\in R^d$ 代表了第 i 个样本的深度特征，属于第 y_i 类。$W_j\in R^d$ 代表了权重 W 的第 j 列，$b_j\in R^N$ 代表偏置项。

$$L_2=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}}+\sum_{j=1,i\neq y_i}^{n}e^{s\cos\theta_j}} \tag{7}$$

通过 $W_j^T x_i=\|W_j\|\|x_i\|\cos\theta_j$ 的转换，θ_j 代表了 W_j 和 x_i 之间的角度，通过 L2 标准化将 $\|W_j\|$ 的初始权重设置为 1，$\|x_i\|$ 设置为 s，得到了标准化后的 Softmax loss 函数。

$$L_3=-\frac{1}{N}\sum_{i=1}^{N}log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s}(\cos(\theta_{y_i+m}))+\sum_{j=1,i\neq y_i}^{n}e^{s\cos\theta_j}} \tag{8}$$

ArcFace loss 在基于中心和特征归一化的基础上，将所有样本看作一个个分布在超球面上的点，在角度空间上进行分割，它的角度间隔，也就是在超球面曲面上的最小距离。

在 Arcface loss 损失函数中，在 x_i 和 W_j 之间的 θ 上加上角度间隔 m，以加法的方式惩罚深度特征与其相应权重之间的角度。

$$L_5=L_2+\frac{1}{\pi N}\sum_{i=1}^{N}\theta_{y_i} \tag{9}$$

通过减少样本和中心之间的角度，Arcface loss 可以提高类内的紧密性，内部的损失函数可以有效压缩类内的变化，但也会带来类间角度较小的缺点。

$$L_6=L_2-\frac{1}{\pi N(n-1)}\sum_{i=1}^{N}\sum_{j=1,j\neq 1}^{n}arccos(W_{y_i}^T W_j) \tag{10}$$

通过增加不同中心之间的角度，可以增加类间差异性。

2　实验验证

为了验证三维注意力模块的有效性，选用了汽车之家数据集，PKU VehicleID 数据集，CIFAR-100 数据集和 ImageNet 数据集，分别进行了对比实验。我们的算法使用 PyTorch 实现，实验结果表明，三维注意力模块在以上数据集都达到了最先进的性能。在接下来的章节中，将具体介绍数据集和实验细节，并对结果进行分析。

2.1　汽车之家数据集

车脸识别数据集对于每辆车均只采用前车窗的部分进行识别。测试的方法借鉴于人脸识别，目的是对于给定任意两张车窗图片，都能够判断是否属于同一辆车。

图 7　汽车之家数据集

数据来源于汽车之家的公开数据，如上图所示。训练集共有 1 757 辆车，测试集共有 522 辆车。每辆车的车窗都有三张照片，分为左前，正前，右前 3 个视角。测试的共有 6 000 对，其中 5 000 对为不同车辆的车窗图片，1 000 对为相同车辆的车窗图片。

使用了 ResNet-50 作为主干网络，初始学习率为 0.1，在第 80，160，240 个 epoch 的时候学习率分别变成 0.01，

0.001，0.000 1。batchsize 为 64，momentum 为 0.9，weight decay 为 0.000 5，总共训练 200 个 epoch。进行对比实验得到的准确率结果如下：

表 1 前车窗数据集测试对比分析

Attention Model	ResNet-18	ResNet-34	ResNet-50	ResNet-101
Baseline	12.50	11.83	10.60	9.97
＋SE	9.79	9.09	8.14	8.01
＋CBAM	9.58	9.34	8.96	8.12
＋ECA	8.53	8.37	7.54	7.23
＋GC	8.77	8.30	7.22	7.34
＋TDA	8.10	7.87	5.56	5.30

从实验数据来看，相比 SENet 等注意力机制，三维注意力机制拥有相对来说更好的效果。相比 SENet，CBAM 有约 3%的提升，相比 ECANet 和 GCNet 有约 2%的提升。

2.2 PKU VehicleID 数据集

北大 VehicleID 数据集由北京大学视频技术国家工程实验室（NELVT）在国家基础研究计划和国家自然科学基金委员会的资助下构建。

“VehicleID”数据集包含分布在中国一个小城市的多个现实世界监控摄像头在白天捕获的数据。整个数据集中有 26 267 辆汽车（共 221 763 张图像）。每个图像都附有与其在现实世界中的身份相对应的 id 标签。

图 8 PKU VehicleID 数据集

对于 PKU VehicleID 数据集，采用了训练 100 个 epoch，batchsize 为 64，学习率初始值为 1e-5，使用 adam 优化器，weight decay 为 0.000 4 的设置。

在公开的 VehicleID 车辆重识别数据集上，表 2 给出了验证的详细结果。本文设计的三维注意力模块取得了很好的效果，在 Small，Medium，Large 三种上均获得了最佳效果。

2.3 数据集 CIFAR-100 分类任务

CIFAR-100 数据集由 100 个类别的 60 000 张 32×32 彩色图像组成，每类有 600 张图像，其中有 500 张训练图像和 100 张测试图像。

对于 CIFAR-100 测试，用 ResNet 作为主干网络，验证了 ResNet-18，ResNet-34，ResNet-50，ResNet-101 四种网络作为基础网络时，三维注意力相比于不添加注意力和添加其他注意力机制都有效果提升。

表 2 PKU VehicleID 数据集测试对比分析

Attention Model	Small		Medium		Large	
Err	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
Baseline	13.58	5.91	16.11	7.66	20.13	9.37
＋SE	12.21	4.67	13.97	6.32	17.80	7.80
＋CBAM	12.69	5.22	14.86	6.77	19.01	8.32
＋ECA	11.03	4.70	13.27	6.24	17.07	7.63
＋GC	16.34	7.05	19.86	9.22	24.47	11.68
＋TDA	10.32	4.44	12.19	5.91	15.59	7.06

训练 200 个 epoch，开始的学习率为 0.1，在第 60，120，160 个 epoch 的时候分别变成 0.02，0.004，0.000 8。其中 Baseline 和添加了 SE 模块的效果使用了其提供的数据，添加了 CBAM，ECA 和 GC 注意力的代码均来自其官方论文所提供的代码。

表 3 CIFAR-100 数据集测试对比分析

Attention Model	ResNet-18	ResNet-34	ResNet-50	ResNet-101
Baseline	24.39	23.24	22.61	22.22
＋SE	23.56	22.07	21.42	20.98
＋CBAM	23.47	22.85	21.19	20.85
＋ECA	23.80	22.31	20.58	20.21
＋GC	23.24	22.41	21.68	20.82
＋TDA	23.18	22.19	20.30	20.20

表 3 给出了以上这些注意力模块大多数采用了与我们相同的主干网络 ResNet。可以看出，本文提出的三维注意力模块在 ResNet-50 相对于其他注意力模块有比较明显的效果，相比不添加注意力的 ResNet-50 有 2.31%的提升，相比其他的注意力有约 1%的提升。值得注意的是，即使是在采用了更加深层的主干网络的时候，三维注意力与基准相比，提升仍然是显著的，这表明了我们的三维注意力模块的有效性。结论表明，三维注意力方法也有较好的鲁棒性。

2.4 大规模分类数据集 ImageNet

ImageNet 是一个用于大规模图像分类的基准数据集，包含来自 1 000 个类的 128 万张训练图像和 5 万张验证图像。我们基于带有注意力机制的 ResNet50，对三维注意力模块进行了测试。在 ILSVRC2012 _ train 数据集上进行训练，在 ILSVRC2012 _ val 上测试准确率。ImageNet 上的所有训练都使用了 8×Tesla V100。

为了加快测试的速度，使用了渐进式图像大小调整来进行分类——在训练开始时使用小图像，然后随着训练的进行逐渐增加大小。图像的宽度从 160 像素开始训练 15 个 epoch，到 320 像素训练 12 个 epoch，到图像本身的宽度训练 1 个 epoch。同时使用了 LARS（Layer-wise Adaptive Rate Scaling）[24] 的优化方法，通过学习率的动态调整，加快模型收敛速度。Batchsize 采用 512，在第 13，25 个 epoch 时，分别变成 224，

128。最终我们得到的结果如表 4 所示。

表 4　ImageNet 数据集测试对比分析

Attention Model	Top-1 err	Top-5 err	计算量 GFLOPS	参数值 Param/(M)
Baseline	24.80	7.48	3.86	25.56
＋SE	23.29	6.62	3.87	26.26
＋CBAM	23.32	6.83	3.88	26.22
＋ECA	23.25	6.55	3.86	25.55
＋SA	23.01	6.59	3.86	25.87
＋TDA	18.79	4.55	3.90	27.39

如表 4 所示，在以 ResNet-50 为主干网络时，添加了三维注意力模型有 6.01%的提升，相比于其他注意力模型有接近 5%的提升。

在参数量和计算量没有太大差距的情况下，三维注意力同时达到了轻量化和高效的目标。

2.5　注意力可视化

我们用 Grad-CAM[25] 对 SENet 和三维注意力机制进行可视化，并进行对比。

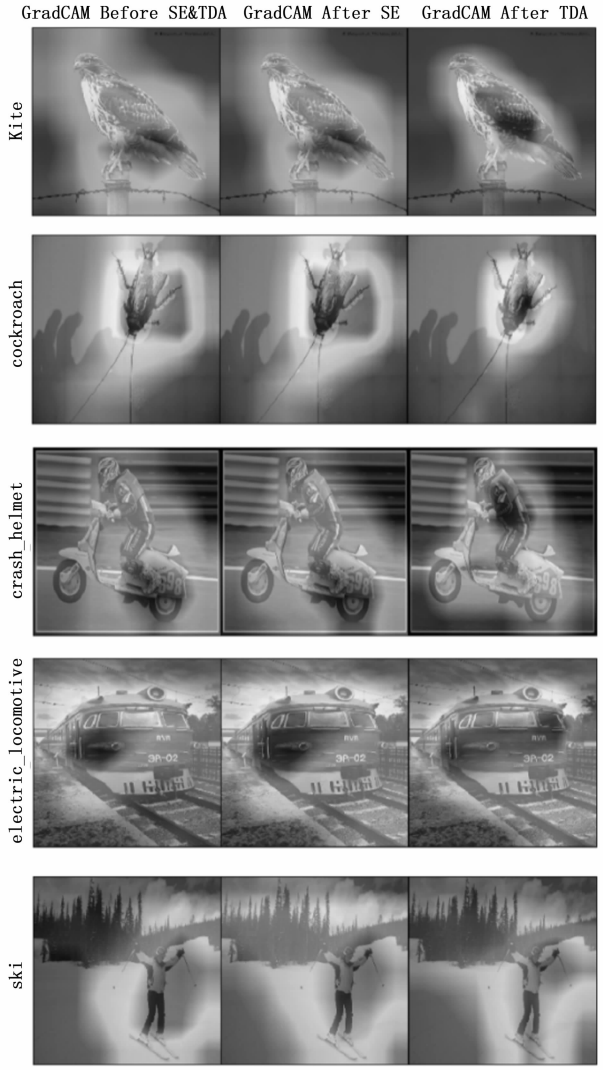


图 9　三维注意力的可视化

图中第一列表示未经注意力模块的图像，第二列表示经 SE 模块得到的图像，第三列表示经本文提出的三维注意力模块得到的图像。

在 SENet 中，得到的通道注意力处理的图像，并不能非常精确地聚焦图像中的重点信息，而将许多注意力放到了不需要关注的地方。

本文提出的三维注意力在聚焦重点信息上更加具有优势，由于三维注意力所关注的像素同时需要满足在通道和位置上都值得受到注意，所以相比其他注意力更加集中。

同时，在注意力机制上，卷积层比全连接层有更好的表现，所以和 SENet 的全连接层相比，使用卷积层来获得通道特征的三维注意力能够获得更加贴合物体本身的注意力分布。如图中所示，本文提出的三维注意力所得到的焦点区域相比 SENet 更加符合人类的注意力特征。

3　结束语

参考已有位置注意力和通道注意力机制，我们提出了一种新的三维注意力模块，能够更全面的反应三维图像中的关键信息，在做到轻量化的同时能有效地提高车辆识别的精确度。把该注意力机制应用到 ResNet-50 上，大量的实验表明在各类数据集上本文提出三维注意力模块在图像分类的视觉任务上都有很好的效果。

参考文献：

[1] Yang Shiqin. Design and implementation of vehicle identification based on full face feature [D]. Xi'an: Xidian University, 2019.

[2] Zhang Hongbing, Li Hailin, Huang Xiaoting, et al. Research and implementation of vehicle recognition method based on HOG feature of front face [J]. Computer Simulation, 2015, 32 (12): 119 - 123.

[3] Lu Xiangying. Research on vehicle face recognition system based on PCA algorithm [J]. Telecom World, 2018, 25 (3): 316 - 317.

[4] Zhu Shanwei, Li Yuhui. Vehicle face recognition based on SVM and VAR/LBP [J]. Electronic Science and Technology, 2018, 31 (7): 7 - 10.

[5] Zhang T. The method of cutting image of vehicle face based on haar feature and improved cascade classisifier [C] // Guangzhou: Asia Pacific Institute of Science and Engineering Proceedings of 2019, The Third International Conference on Computer Graphics and Digital Image Processing, 2019.

[6] Luo X C, Shen R H, Huc J, et al. A deep convolution neural network model for vehicle recognition and face recognition [C] //Dalian: Internati on National Congress of Information and Communication Technology, 2017.

[7] Li Xiying, Yuan Minxian, Lü Shuo, et al. Vehicle brand identification based on LLC and weighted SPM [J]. Computer Engineering, 2017, 43 (5): 210 - 216.

[8] Wang Panpan, Li Yuhui. Vehicle re identification method based on feature fusion and LM algorithm [J]. Electronic Science and

Technology, 2018, 31 (4): 12 - 15.
[9] Deng Liu, Wang Zijie. Research on vehicle recognition based on deep convolution neural network [J]. Computer Application Research, 2016, 33 (3): 930 - 932.
[10] Peng Qing, Ji Guishu, Xie Linjiang, et al. Application of con - volutional neural network in vehicle recognition [J]. Comput- er Science and Exploration, 2018, 12 (2): 282 - 291.
[11] Soon F C, Khaw H Y, Chuah J H, et al. PCANet - based convolutional neural network architecture for a vehicle model recognition system [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20 (2): 749 - 759.
[12] 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.
[13] 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.
[14] Cao Y, Xu J, Lin S, et al. Gcnet: Non - local networks meet squeeze - excitation networks and beyond [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0 - 0.
[15] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter - free attention module for convolutional neural networks [C] //International Conference on Machine Learning. PMLR, 2021: 11863 - 11874.
[16] 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.
[17] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images [J]. Handbook of Systemic Autoimmune Diseases, 2009, 1 (4): 1 - 60.
[18] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge [J]. International journal of computer vision, 2015, 115 (3): 211 - 252.
[19] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012 - 10022.
[20] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks [C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492 - 1500.
[21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] //Advances in Neural Information Processing Systems. 2017: 5998 - 6008.
[22] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation [C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 3146 - 3154.
[23] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition [C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4690 - 4699.
[24] You Y, Gitman I, Ginsburg B. Large batch training of convolutional networks [J]. arXiv preprint, 2017: 1 - 8. [2017 - 8 - 13]. https://arxiv.org/pdf/1708.03888
[25] Selvaraju R R, Cogswell M, Das A, et al. Grad - cam: Visual explanations from deep networks via gradient - based localization [C] //Proceedings of the IEEE international conference on computer vision. 2017: 618 - 626.
[26] 杨述斌, 蒋宗霖, 刘 寒. 基于 Kalman 滤波的车位侧方距离修正方法 [J]. 计算机测量与控制, 2020, 28 (2): 220 - 223.
[27] Liu Z, Mao H, Wu C Y, et al. A ConvNet for the 2020s [J]. arXiv preprint, 2022: 1 - 15. [2022 - 1 - 10]. https://arxiv.org/pdf/2201.03545
[28] 彭煜民, 岳鹏超, 张 丹, 等. 结合注意力机制的火灾检测算法 [J]. 计算机测量与控制, 2021, 29 (8): 42 - 46.
[29] 李红俊, 韩冀皖. 数字图像处理技术及其应用 [J]. 计算机测量与控制, 2002, 10 (9): 620 - 622.
[30] 姚明海, 陈志浩. 基于深度主动学习的磁片表面缺陷检测 [J]. 计算机测量与控制, 2018, 26 (9): 29 - 33.
[31] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks [C] //Advances in Neural Information Processing Systems. 2015: 2017 - 2025.
[32] Zhao B, Wu X, Feng J, et al. Diversified visual attention networks for fine - grained object classification [J]. IEEE Transactions on Multimedia, 2017, 19 (6): 1245 - 1256.
[33] Mnih V, Heess N, Graves A. Recurrent models of visual attention [J]. Advances in neural information processing systems, 2014, 27: 2204 - 2212.
[34] 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.
[35] Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation [C] //European conference on computer vision. Springer, Cham, 2016: 483 - 499.
[36] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation [C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 3146 - 3154.
[37] Peng Y, He X, Zhao J. Object - part attention model for fine - grained image classification [J]. IEEE Transactions on Image Processing, 2017, 27 (3): 1487 - 1500.
[38] Wang X, Girshick R, Gupta A, et al. Non - local neural networks [C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794 - 7803.
[39] Du Y, Yuan C, Li B, et al. Interaction - aware spatio - temporal pyramid attention networks for action classification [C] //Proceedings of the European conference on computer vision (ECCV). 2018: 373 - 389.
[40] Fang P, Zhou J, Roy S K, et al. Bilinear attention networks for person retrieval [C] //Proceedings of the IEEE/CVF international conference on computer vision. 2019: 8030 - 8039.