

# 基于导师-学生网络的驾驶行为识别方法

褚晶辉, 张姗, 汤文豪, 吕卫\*

天津大学电气自动化与信息工程学院, 天津 300072

**摘要** 提出了一种基于导师-学生网络的驾驶行为识别模型。考虑到驾驶动作是在局部区域发生的, 将驾驶行为识别任务拆分成动作定位和动作分类两个子任务。针对动作定位任务, 设计了网络层数较浅和接收高分辨率图像输入的导师网络, 导师网络通过特征图的响应对动作区域进行弱定位; 在动作定位基础上, 针对动作分类任务, 设计了网络层数较深的接收低分辨率动作区域图像输入的学生网络, 学生网络根据深层网络提取的高层次语义特征实现高准确率分类。实验结果证明, 导师-学生网络模型能带来较高的识别准确率, 具有强稳健性。

**关键词** 图像处理; 深度学习; 驾驶行为识别; 卷积神经网络; 弱定位; 动作分类; 导师-学生网络

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.061019

## Driving Behavior Recognition Method Based on Tutor-Student Network

Chu Jinghui, Zhang Shan, Tang Wenhao, Lü Wei\*

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

**Abstract** This paper presents a driving behavior recognition model based on tutor-student network. Considering that driving behavior occurs in a local area, this paper divides the task of driving behavior recognition into two sub-tasks: action location and action classification. Aiming at the task of action location, a tutor network with shallow network layer receiving high-resolution image input is designed. The tutor network weakens the action area according to the response of feature map. On the basis of action location and action classification task, a student network with deeper network layer is designed to receive the input of low-resolution action area image. High-level semantic features which are extracted from student network are used to achieve high accuracy classification. Experimental results show that the tutor-student network model can bring high recognition accuracy and strong robustness.

**Key words** image processing; deep learning; driving behavior recognition; convolutional neural network; weak location; action classification; tutor-student network

**OCIS codes** 100.2960; 100.4996; 100.5010

## 1 引言

基于视觉的驾驶行为识别方法被认为是一种有发展前景的通向实际应用的方法。基于视觉的方法大致上可以被分为提取人工特征的方法和基于深度学习的方法。提取人工特征的方法是指使用手工特征[如 HOG(Histogram of Oriented Gradients)<sup>[1]</sup>特征, LBP(Local Binary Pattern)<sup>[2]</sup>特征等]并结合分类器(如支持向量机<sup>[3]</sup>等)实现有效分类。在驾驶行为识别的任务中, 李长勇等<sup>[4]</sup>基于贝叶斯置信网

络进行多信息融合实现了疲劳检测, 打破了单一信息特征判别的约束, 提高了动作识别模型的稳健性。李此君等<sup>[5]</sup>根据从图像中提取出的纹理、颜色以及梯度方向特征并基于协方差流形进行多特征融合驾驶行为判断, 打破了基于单一特征识别的不足, 提高了驾驶行为检测的准确率。

虽然传统的手工特征可以有效地解决一些基本的分类问题, 但是遇到复杂的分类模式, 手工特征往往不能很好地区分所识别物体, 导致任务失效。与此不同的是, 深度网络所提取的深度特征因其具有

收稿日期: 2019-09-11; 修回日期: 2019-10-16; 录用日期: 2019-11-02

基金项目: 国家自然科学基金(61572356)

\* E-mail: luwei@tju.edu.cn

高层次语义,可以克服手工特征在某些复杂情况下的局限性,因而在一些图像识别的任务中手工特征逐渐被深度特征所取代。卷积神经网络是一种常应用于提取图像深层特征的模型,近些年 AlexNet、VggNet、Inception、ResNet、DenseNet<sup>[6-10]</sup>等网络模型被提出以更好地解决图像分类和识别等问题。

在这样的背景下,基于深度学习的方法在驾驶行为识别的任务上也得到了广泛的研究应用。例如 Hssayeni 等<sup>[11]</sup>通过将 HOG 特征和 SIFT (Scale-Invariant Feature Transform)描述子与支持向量机相结合实现驾驶员行为动作分类,并将此基于人工特征的方法和基于 AlexNet、VggNet 和 ResNet 的卷积神经网络的分类方法进行对比,实验结果显示,基于深度学习的模型的泛化性能要远超过基于人工特征的模型。Yan 等<sup>[12]</sup>开发了一个卷积神经网络 (CNN)模型,通过无监督学习和迁移学习的方法实现了正常驾驶、驾驶员接打手机、驾驶员饮食和驾驶员吸烟的分类。Majdi 等<sup>[13]</sup>通过利用 CNN 和随机决策森林来提高分类性能,提出了一种新的检测模型,称为 DriveNet。Colbran 等<sup>[14]</sup>结合 Vgg16 和 GoogleNet 的特征得到集成模型以实现分类,实验结果证明集成的深度网络模型可以提高分类的准确率。Hu 等<sup>[15]</sup>探索了融合策略,将从具有不同卷积核大小的多流 CNN 中提取的多尺度深度特征结合起来以实现多尺度的驾驶行为识别。2016 年 Kaggle 举办了一个对驾驶员行为图像进行分类的比赛,数据集包含了驾驶员在驾驶过程中的 10 种不同的分心动作,在该竞赛中集成模型被大量应用以提升分类性能。2018 年,Abouelnaga 等<sup>[16]</sup>创建了一个 10 分类的驾驶行为数据集,考虑到这些驾驶分心动作主要是源于人脸和手部的状态,在原图像上进行了人脸检测和手部检测,针对人脸的图像数据训练出面部状态分类器,并使用基于手部的数据训练出一个手部行为分类器,然后结合这两部分的数据训练出一个综合分类器,使用原图训练出一个基于原始图像动作分类器,最后集成这 4 个分类器,得到效果更佳的驾驶行为识别模型。

集成深度学习模型往往因为网络规模过大以及运算量过大,在性能不高的硬件上实时性不足,因此基于卷积神经网络的驾驶行为分析算法在实际应用方面依然存在很大问题。近些年,人们将研究重点放在网络结构的优化方面。与集成模型相比,使用一些优化技巧不仅可以避免较大的模型规模还可以提高网络的性能。黄友文等<sup>[17]</sup>基于卷积神经网络和

长短期记忆 (LSTM)神经网络,通过融合浅层特征与深层特征获得丰富的信息表示,从而提高模型的识别行为的精度。Baheti 等<sup>[18]</sup>提出了一种利用正则化技术改进的 VGG (Visual Geometry Group)模型,该模型使用较少的参数获得较高的准确率。与原模型相比,该模型大大减少了参数,节省了计算时间。

立足于这个角度,本文基于导师-学生网络框架提出了一种驾驶行为识别模型。该模型通过构建指导与被指导的两个网络来拆分驾驶行为识别的动作定位和动作分类这两个任务。具体地,导师网络负责响应关键动作区域,并通过特征响应映射来指导裁剪关键区域,最终得到关键动作图像块。学生网络负责根据输入的关键动作图像块来学习关键动作区域的特征,从而实现高精度驾驶行为识别。第一个网络为第二个网络提供位置信息以指导第二个网络学习的输入,发挥了导师的作用,因此第一个网络称作导师网络 (T-Net)。第二个网络在导师网络的指导下,在输入信息中排除了冗余信息以专心学习不同驾驶动作的本质区别,被称作学生网络 (S-Net)。

## 2 算法原理

传统意义上的教师-学生网络是通过一个深层的教师网络学习知识,然后将知识教授给一个较浅的学生网络,这样便可得到一个轻量级的网络,从而实现分类。如图 1(a)所示,教师网络传授给学生网络的过程是通过教师网络输出的结果指导学生网络的输出,从而实现知识蒸馏<sup>[19]</sup>。传统意义上的教师-学生网络主要用于简化和压缩模型。和传统意义上的教师-学生网络不同的是,如图 1(b)所示,本文提出的导师网络是一个输入高分辨率图像且网络层数较浅的结构,导师网络不需要学到完全的知识去教导学生,只需要弱定位出关键动作区域,从而在输入图像上给予学生网络指导,让学生去充分地学习关键的局部特征。最后通过联合导师网络的全局判别和学生网络的局部判别共同实现了高准确率的识别任务。实验结果证明,所提出的模型通过导师网络和学生网络的共同学习和相互配合提高了驾驶行为识别的准确率。

本文所提的模型如图 2 所示,整个模型大致分为导师网络和学生网络两个部分,其中导师网络负责为学生网络的学习提供位置指导,学生网络则是在导师网络指导基础上深刻挖掘本质的特征。导师网络指导学生网络的过程可以分为特征响应映射过程和指导裁剪图块过程。

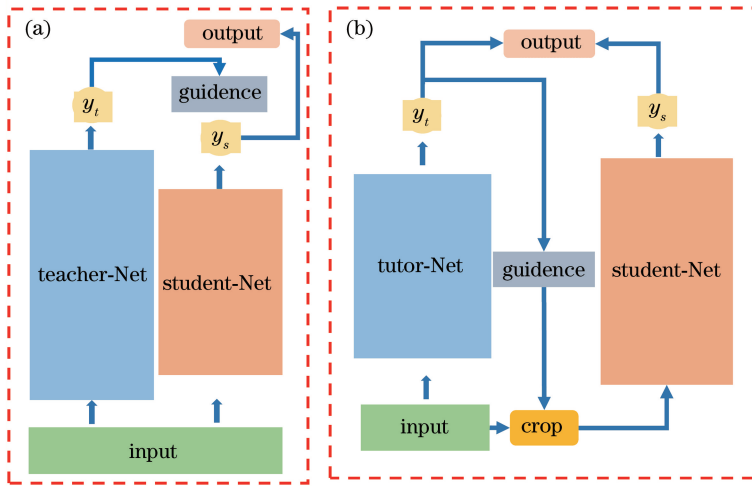


图1 传统教师-学生网络和所提网络。(a)传统的教师-学生网络;(b)所提网络

Fig. 1 Traditional teacher-student network and proposed method. (a) Traditional teacher-student network; (b) proposed method

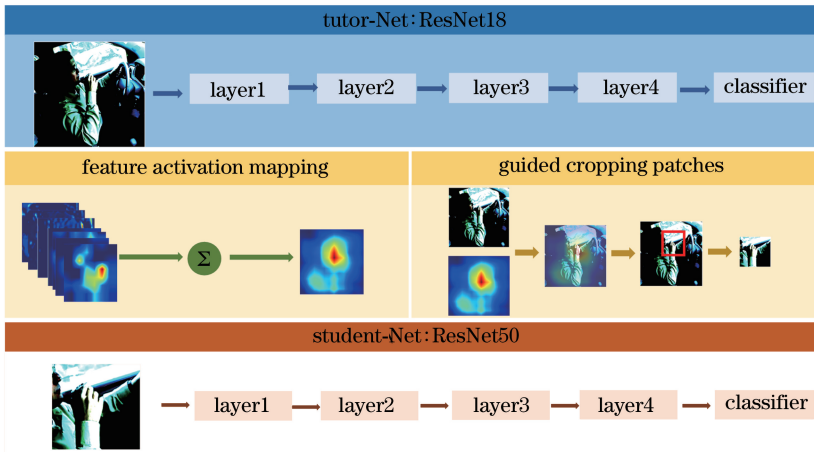


图2 所提模型结构图

Fig. 2 Structural diagram of proposed model

### 2.1 导师网络

受到卷积神经网络具有定位能力的启发,本文设计了导师网络以对动作区域进行弱定位。

卷积神经网络通过多个卷积层级联不断增加感受野,直至获得全局信息。在这个过程中,卷积层没有破坏输入的空间位置关系。与一般意义上的用卷积神经网络实现目标定位不同的是,本文中训练导师网络的过程中不需要使用目标位置框的信息或像素级标注信息,而只使用图像级标注信息就可以实现关键区域的弱定位,这是因为卷积层保留了位置信息并可以敏感地响应目标对象,通过观察特征图的响应位置可以定位关键区域。

具体地,导师网络的主干网络是 ResNet18。这里使用迁移学习,将 ImageNet 预训练好的 ResNet18 网络用驾驶行为数据集进行训练微调。

在结构上,去掉了 ResNet18 后面的全连接层,使用全局平均池化层(Global Average Pooling)连接分类层,全局平均池化层可以降低网络参数量,从而缓解过拟合现象。另外,导师网络的输入选择高分辨率的全局图像,使得网络能够从全局图像中挖掘出局部的关键动作区域。

### 2.2 特征激活映射模块

导师网络的响应位置体现在网络最后的卷积层输出的特征图。处于不同通道的特征图分别针对不同空间位置的信息进行响应,比如有的通道对驾驶员的手臂进行响应,而有的通道对驾驶员手中握着的电话进行响应等。为了得到一个关键动作发生的区域,对不同通道的特征图沿着通道方向求均值,得到平均特征图。设最后一层卷积层输出的特征图  $X = \{x_1, x_2, x_3, \dots, x_n\}$ ,其中特征图的通道数为

$n$ , 特征激活映射图  $x_{fam}$  可表示为

$$x_{fam} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

导师网络中由不同深度的卷积层得到的特征图所展示的特征具有不同尺度,因而从不同尺度特征图得到的特征激活映射图可以表示不同层次特征的响

应情况。图 3 表明伴随着网络从浅层到深层,网络的感受野不断变大,特征激活映射图从轮廓响应逐步转化为对所识别动作的响应,实现了从低层次语义特征到高层次语义特征提取的过程。本文发现最后一层的特征激活图能反映动作区域,因而使用最后一层的特征激活图去定位驾驶动作区域,实现区域指导。

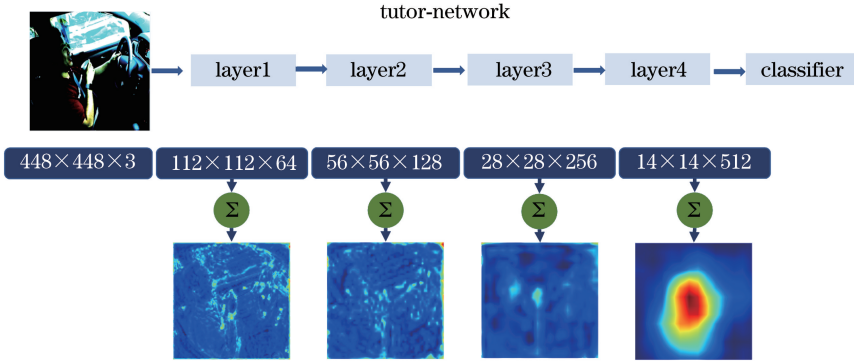


图 3 由导师网络中不同卷积层的特征图得到的特征激活映射图

Fig. 3 Feature activation maps obtained from feature maps of different convolution layers in tutor network

### 2.3 指导裁剪模块

在获得特征激活映射图的基础上,可将特征激活映射图对应到全局图像,得到全局图的关键动作区域。具体地,首先得到特征激活映射图  $x_{fam}$  的最大值  $\text{Max}(x_{fam})$ ,并根据这个最大值确定二值化的阈值  $r$ :

$$r = 0.5 \times \text{Max}(x_{fam}) \quad (2)$$

根据阈值将特征激活映射图进行二值化:

$$x_{fam_{ij}} = \begin{cases} 0 & (x_{fam_{ij}} < r) \\ 1 & (x_{fam_{ij}} > r) \end{cases} \quad (3)$$

式中: $i$  代表图像的第  $i$  行; $j$  代表图像的第  $j$  列; $x_{fam_{ij}}$  代表第  $i$  行第  $j$  列的图像像素点。根据二值化的图进行最大连通区域提取操作,然后得到连通区域的矩形包围框。最后通过截取矩形包围框内的图像块得到关键动作区域的图像块。

具体的图像处理流程图如图 4 所示,包括输入特征激活映射图部分、提取最大响应值部分、二值化

操作部分、最大连通区域提取部分以及输出部分,通过指导裁剪模块最终得到关键动作区域图像,用来指导学生网络更好地完成任务。

### 2.4 学生网络

在获得关键动作区域的图像块的基础上,用关键动作区域的图像块训练学生网络,可以使得学生网络排除冗余信息以专注学习不同动作的本质,获得准确的知识,从而更好地完成驾驶行为识别的任务。

具体地,学生网络的主干网络为 ResNet50 网络。这里使用迁移学习,将 ImageNet 预训练好的 ResNet50 网络用驾驶行为数据集进行训练微调。另外,学生网络的输入选择较低分辨率的局部区域图像,这样做一方面是因为局部区域不需要输入高分辨率图像进行训练,另一方面是因为输入较小分辨率的图像可以降低计算量,最后使得学生网络能够专注于局部动作信息,从而实现更好的分类性能。

## 3 实验结果与分析

### 3.1 驾驶行为识别数据集

#### 3.1.1 Kaggle 驾驶行为数据集

2016 年,Kaggle 举办了驾驶动作检测的比赛,这是首个公开的考虑到多种干扰驾驶员行为的数据集。Kaggle 的驾驶动作数据集是通过采集 26 名驾驶员的多种动作图像构成的。经整理,数据集提供 22424 张图像,其中包含 10 类驾驶行为,分别是:正常驾驶,喝水,右手接打手机,左手接打手机,右手发

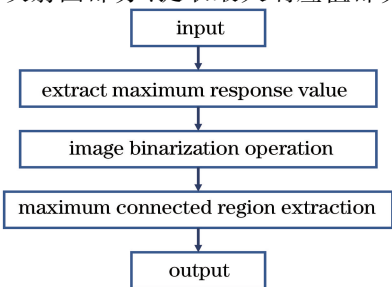


图 4 指导裁剪模块流程图

Fig. 4 Flow chart of guiding tailoring module

送短信,左手发送短信,操作收音机,整理仪表,向后座接东西以及和乘客对话。数据集的 10 个类别的图像样本如图 5 所示。在 Kaggle 的数据集中,驾驶员编号分别为 2,12,14,15,16,21,22,24,26,35,39,41,42,47,49,50,51,52,56,61,64,66,72,75,



图 5 Kaggle 数据集的 10 类动作图像样本

Fig. 5 Ten kinds of action image samples from Kaggle dataset

### 3.1.2 AUC 驾驶行为数据集

2017 年,Abouelnaga 创建了新的驾驶行为数据集[AUC(American University in Cairo)驾驶行为数据集],这是一个公开的具有 10 类驾驶动作的数据集。

AUC 数据集包括 17308 张图像,其中 12977 张用于训练,4331 张用于测试。该数据集是通过使用 ASUS Zenphone(Z00UD 型)后摄像头收集得到的。考虑到驾驶行为的多样性,该数据集收集了来自 7

81。在此基础上,根据不同的驾驶员对训练图像和测试图像进行了划分,取第 14,35,51 号驾驶员的动作图像作为测试集,其余编号对应的驾驶员的动作图像作为训练集。最后得到了 19780 张训练图像和 2644 张测试图像。

个国家的 31 名参与者[埃及(24)、德国(2)、美国(1)、加拿大(1)、乌干达(1)、巴勒斯坦(1)和摩洛哥(1)]在 4 种不同车辆环境中的驾驶图像。同 Kaggle 的驾驶行为数据集一样,AUC 数据集包含 10 种驾驶动作,包括正常驾驶,喝水,右手接打手机,左手接打手机,右手发送短信,左手发送短信,操作收音机,整理仪表,从后座接东西以及和乘客对话这些动作。来自 AUC 数据集的每个行为的示例图像如图 6 所示。



图 6 AUC 数据集的 10 类动作图像样本

Fig. 6 Ten kinds of action image samples from AUC data set

## 3.2 实验设置

实验中使用英伟达的 GTX 1080Ti 显卡进行训练。导师网络的输入尺寸固定为 448,学生网络的输入尺寸固定为 224。所选择的优化器为随机梯度下降(SGD)算法。为了使训练变得稳定,设置的初

始学习率为  $10^{-3}$ ,在迭代次数中,每进行 2 次迭代学习率下降为原来学习率的 90%。实验中的迭代次数为 80 次,每次训练的批量数据个数为 16。依次交替训练导师网络和学生网络,即训练一次导师网络,再训练一次学生网络,然后重复上述过程。

### 3.3 实验结果分析

#### 3.3.1 Kaggle 数据集实验结果分析

首先分析数据增强对准确率的影响,本文选择的数据增强策略是随机裁剪(random crop),通过表1中的结果,发现使用随机裁剪操作进行数据增强能显著提高模型准确率。表中 PPI 代表 pixel per inch (1 inch=2.54 cm)

在数据增强的基础上,本文探究分辨率对结果的影响。根据表2的结果,发现通过输入较高分辨率的图像来训练模型,能获得较高的准确率。这是因为低分辨率图像损失了一些信息,而高分辨率图像有更多信息可供模型学习,因此,使用不同分辨率的图像进行训练会导致模型性能上有差别。通过统计模型参数量和每秒浮点运算次数,可以发现网络模型加深的过程伴随着参数量和运算量的增加,而分辨率的增加会大幅度提升运算量,因而依靠增大输入图像的分辨率去提高网络性能的方法是以增加计算量为代价的。

表1 Kaggle 数据集中有无数据增强的对比实验结果

Table 1 Comparative experimental results with and without data enhancement in Kaggle dataset

Model	Random cropping	Image resolution/PPI	Accuracy / %
ResNet18	No	224	89.30
ResNet18	Yes	224	92.56
ResNet50	No	224	90.17
ResNet50	Yes	224	94.93

表2 Kaggle 数据集中设置不同分辨率的实验结果

Table 2 Experimental results of setting different resolutions in Kaggle data set

Model	Parameter quantity	Image Flops	Image resolution/PPI	Accuracy / %
ResNet18	11,181,642	1.82G	224	92.56
ResNet18	11,181,642	7.28G	448	94.21
ResNet50	23,528,522	4.12G	224	94.93
ResNet50	23,528,522	16.47G	448	96.48

表4 Kaggle 数据集中模型联合判别的对比实验结果

Table 4 Comparative experimental results of model joint discrimination in Kaggle data set

Model	Parameter quantity	Flops	Image resolution/PPI	Accuracy / %
S-Net (ResNet18)	22,363,284	9.10G	224	92.78
S-Net (ResNet50)	34,710,164	11.40G	224	96.29
ResNet18+ResNet50(ensemble)	34,710,164	5.94G	224(T-Net)+224(S-Net)	95.95
ResNet18+ ResNet50(ensemble)	34,710,164	11.40G	448(T-Net)+224(S-Net)	96.10
T-Net(ResNet18)+S-Net(ResNet18)	22,363,284	9.10G	448(T-Net)+224(S-Net)	95.99
T-Net(ResNet18)+S-Net(ResNet50)	34,710,164	5.94G	224(T-Net)+224(S-Net)	96.56
<b>T-Net(ResNet18)+S-Net(ResNet50)</b>	<b>34,710,164</b>	<b>11.40G</b>	<b>448(T-Net)+224(S-Net)</b>	<b>97.92</b>

本文设计的导师网络输入为 448PPI 的分辨率,选择使用较浅的网络去实现动作定位。将关键动作区域的图像大小固定为 224,然后去训练学生网络,学生网络使用较深的网络去实现准确率高的分类。本文设计层数较少且接收高分辨率图像输入的网络去捕获图像中动作发生的位置,设计层数较多且接收低分辨率图像输入的网络模型去分析局部动作区域的信息。两个网络的分工和配合能稳定有效地提高识别准确率。实验结果如表3所示,通过分析可得学生网络(ResNet50)的准确率能在导师网络的指导下、在低分辨率的图像训练下提高近2个百分点。

更进一步地,虽然学生网络能专心挖掘局部关键区域的特征以获得更好的判别能力,然而学生网络因为不能学到图像的全局信息而导致提升动作识别准确率受到限制。为了摆脱这种限制,本文将联合导师网络的预测结果和学生网络的预测结果以完成最终的驾驶行为识别。实验结果如表4所示,发现导师网络和学生网络的联合判别可以进一步提高准确率。

另外,导师网络和学生网络联合判别的时候,导师网络负责把握全局信息,而学生网络负责把握局

表3 Kaggle 数据集中不同模型的对比实验结果

Table 3 Comparative experimental results of different models in Kaggle data set

Model	Parameter quantity	Image Flops	Image resolution/PPI	Accuracy / %
ResNet18	11,181,642	1.82G	224	92.56
ResNet34	21,289,802	3.67G	224	94.67
ResNet50	23,528,522	4.12G	224	94.93
ResNet101	42,520,650	7.84G	224	95.69
S-Net (ResNet18)	22,363,284	9.10G	224	92.78
<b>S-Net (ResNet50)</b>	<b>34,710,164</b>	<b>11.40G</b>	<b>224</b>	<b>96.29</b>

部信息,两个网络联合判别的时候侧重的是全局与局部特征相互补充以丰富信息表达。与一般的独立训练的集成模型不同的是,导师网络和学生网络立足的角度不同,因而两个模型所提取的特征具有多样性,因此这两个网络组成的集成模型效果更好。表4中集成模型实验是构建由两个独立关系网络所集成的模型,将它和导师-学生网络模型的结果进行对比,实验表明,本文所提出的模型与一般的集成模型相比具有一定的优势。为了降低运算量以更适用于实际应用,本文降低导师网络输入图像的分辨率,得到了较高的准确率,这表明本文提供的指导-学习的模式具有一定的泛化性能,一定程度上提高了驾驶行为识别准确率。

本文的实验可视化结果如图7所示,图7中的第一列为原始图像,第二列是得到的特征激活映射图,从第三列可以看出特征激活映射图可以描述出关键局部区域位置,第四列是通过指导裁剪模块得到的动作区域图块。从可视化的结果可以看出,所提模型具有能够有效提取关键动作位置的能力。

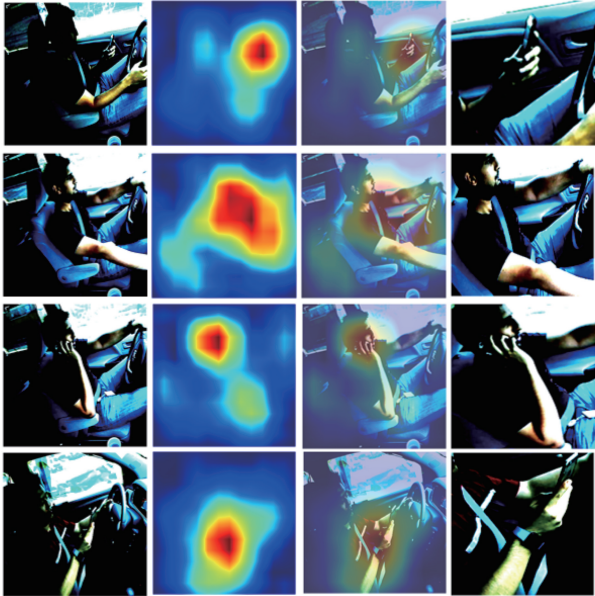


图7 Kaggle数据集上的可视化结果

Fig. 7 Visualization results on Kaggle data set

### 3.3.2 AUC数据集实验结果分析

本文在AUC驾驶行为识别数据集上进行实验,实验结果如表5所示。考虑到驾驶动作大都体现在脸部和手部,文献[16]通过截取手部和脸部区域的图像训练相应的分类器。比如使用数据集的脸部区域图像训练Inception V3网络,最终模型在测试集的准确率为84.28%,而使用手部区域图像进行模型训练得到的准确率为89.52%。从表中可以发

现利用不同来源的数据训练出的模型性能是不相同的。而本文所提出的模型不需要定位脸部或者手部,仅需要导师网络提供的对位置区域的指导,学生网络(ResNet50)就能够达到95.20%的准确率,这说明本文所提供的定位方法更适合驾驶动作分类任务。当联合导师网络和学生网络共同判别动作类别时,在测试集得到了95.71%的准确率。

表5 AUC数据集上的模型对比结果

Table 5 Model comparison results on AUC dataset

Model	Source	Accuracy / %
AlexNet <sup>[16]</sup>	Original	93.65
AlexNet <sup>[16]</sup>	Skin segmented	93.60
AlexNet <sup>[16]</sup>	Face	84.28
AlexNet <sup>[16]</sup>	Hands	89.52
AlexNet <sup>[16]</sup>	Face+hands	86.68
Inception V3 <sup>[16]</sup>	Original	95.17
Inception V3 <sup>[16]</sup>	Skin segmented	94.57
Inception V3 <sup>[16]</sup>	Face	88.82
Inception V3 <sup>[16]</sup>	Hands	91.62
Inception V3 <sup>[16]</sup>	Face+hands	90.88
ResNet50	Original	94.87
S-Net(ResNet50)	Original	95.20
T-Net(ResNet18)+ S-Net(ResNet50)	Original	95.71

## 4 结 论

提出了一种基于导师-学生网络的模型来实现驾驶行为的识别。此模型由导师网络和学生网络构成,导师网络通过特征激活映射模块定位出关键动作区域,之后经过指导裁剪模块得到关键动作区域的图像块,并用这些动作区域的图像块去训练学生网络。由这两个网络构建出了一种指导与学习的模式:导师网络挖掘动作位置信息来指导学生网络,学生网络受到指导后学习高层次语义的判别特征,最终实现了较高准确率的驾驶行为识别。

下一步工作应该为导师-学生网络设计更为合适的损失函数,引起学生网络判别困难的样本应该在导师网络的训练中被重视,引起导师网络判别困难的样本应该在学生网络的训练中被重视,旨在通过设计这样的损失函数更好地完成驾驶行为识别任务。

## 参 考 文 献

- [1] Dalal N, Triggs B, Schmid C. Human detection using oriented histograms of flow and appearance[M] //Leonardis A, Bischof H, Pinz A. Computer vision-ECCV 2006. Lecture notes in computer science.

- Berlin, Heidelberg: Springer, 2006, 3952: 428-441.
- [2] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971-987.
- [3] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers [J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [4] Li C Y, Wu J Q, Fang A Q. A multi-information-based fatigue state recognition method[J]. *Laser & Optoelectronics Progress*, 2018, 55(10): 101503.  
李长勇, 吴金强, 房爱青. 基于多信息的疲劳状态识别方法[J]. *激光与光电子学进展*, 2018, 55(10): 101503.
- [5] Li C J, Liu Y P. Abnormal driving behavior detection based on covariance manifold and Logit Boost [J]. *Laser & Optoelectronics Progress*, 2018, 55(11): 111503.  
李此君, 刘云鹏. 基于协方差流形和 Logit Boost 的异常驾驶行为识别方法[J]. *激光与光电子学进展*, 2018, 55(11): 111503.
- [6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C] // *Advances in Neural Information Processing Systems*, December 3-6, 2012, Lake Tahoe, Nevada, United States. Canada: NIPS, 2012: 1097-1105.
- [7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J/OL]. (2015-04-10) [2019-09-10]. <https://arxiv.xile sou. top/abs/1409.1556>.
- [8] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C] // *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 1552-3970.
- [9] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [10] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks [C] // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2261-2269.
- [11] Hssayeni M, Saxena S, Ptucha R, et al. Distracted driver detection: deep learning vs handcrafted features[J]. *Electronic Imaging*, 2017, 2017(10): 20-26.
- [12] Yan C, Coenen F, Zhang B L. Driving posture recognition by convolutional neural networks[J]. *IET Computer Vision*, 2016, 10(2): 103-114.
- [13] Majdi M S, Ram S, Gill J T, et al. Drive-net: convolutional network for driver distraction detection [C] // *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, April 8-10, 2018, Las Vegas, NV, USA. New York: IEEE, 2018: 69-72.
- [14] Colbran S, Cen K, Luo D. Classification of driver distraction [EB/OL]. [2019-09-10]. <https://pdfs.semanticscholar.org/cb49/ac9618bb2f8271409f91d53254a095d843d5.pdf>.
- [15] Hu Y C, Lu M Q, Lu X B. Driving behaviour recognition from still images by using multi-stream fusion CNN [J]. *Machine Vision and Applications*, 2019, 30(5): 851-865.
- [16] Abouelnaga Y, Eraqi H M, Moustafa M N. Real-time distracted driver posture classification [J/OL]. (2018-11-29) [2019-09-10]. <https://arxiv.xile sou. top/abs/1706.09498>.
- [17] Huang Y W, Wan C L, Feng H. Multi-feature fusion human behavior recognition algorithm based on convolutional neural network and long short term memory neural network[J]. *Laser & Optoelectronics Progress*, 2019, 56(7): 071505.  
黄友文, 万超伦, 冯恒. 基于卷积神经网络与长短期记忆神经网络的多特征融合人体行为识别算法[J]. *激光与光电子学进展*, 2019, 56(7): 071505.
- [18] Baheti B, Gajre S, Talbar S. Detection of distracted driver using convolutional neural network [C] // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 1145-1151.
- [19] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network [J/OL]. (2015-03-09) [2019-09-10]. <https://arxiv.xile sou. top/abs/1503.02531>.