# 基于深度神经网络的非正常驾驶检测

骆明远, *406130917323, M.S., CST,* 刘茜, *406130917318, M.S., CST*

**摘要—近些年来，大量交通事故由非正常驾驶造成。针对非正常驾驶检测日益增长的需求，我们在神经网络领域提出改进的融合深度模型（WGD、WGRD1 和 WGRD2）并将其运用到非正常驾驶检测，以此提供一个检测方案。融合深度模型基于目前优秀的有卓越表现的深度学习模型，包括基本的卷积神经网络、宽卷积神经网络、分组卷积神经网络、残差网络和稠密网络。将每个模型最主要的技巧融合到一个模型作为最基本的改进，在此基础上将残差网络按元素相加的连接方式运用到稠密网络作为提高性的改进。实验表明，WGD 和 WGRD2 的表现处于未改进模型的中间水平。**

**关键词—非正常驾驶检测；神经网络；深度学习**

## I. 引言

随着机动车保有量的快速增加以及道路通车里程逐年增长，发生的大量交通事故给世界各国造成了巨大的财产损失和人员伤亡。如何减少交通事故已经成为世界性的难题，在社会经济快速发展的今天，人们生活节奏不断加快，随之出现的非正常驾驶成为影响驾驶安全的重要因素之一。

根据国际标准化组织（ISO）的定义，非正常驾驶是指注意力指向与正常驾驶不相关的活动，从而导致驾驶操作能力下降的一种现象。从驾驶人行为发生源头分析，非正常驾驶行为可以分为三类：一是满足驾驶人身体舒适需求的分心驾驶行为，如抽烟、饮水、进食、调节空调等行为；二是满足驾驶人心情愉悦需求的分心驾驶行为，如化妆、刮胡子、聊天等，也包括使用手机拨打电话、收发信息等；三是周围环境引发的分心驾驶行为，如照顾小孩、长时间关注某个车外突发情况等。处于移动互联网快速发展的今天，各类手机应用、通讯软件的盛行，使用手机已成为非正常驾驶最主要的表现形式。美国科学家通过模拟实验表明，开车打电话导致驾驶人注意力下降 20%，如果通话内容很重要，则驾驶人的注意力会下降 37%。驾驶人边开车边发短信，发生车祸的概率是正常驾驶时的 23 倍。[1]

针对于此问题，非正常驾驶检测技术得到了关注并推广，其目的是判断驾驶员在驾驶机动车的过程中是否存在潜在的危险。通常来说，非正常驾驶检测通过分析安装在驾驶室的摄像头获取的图像来检测可能存在的危险。

## II. 相关工作

### A. 非正常驾驶检测

目前常用的检测驾驶员非正常驾驶状态的方案包括基于人体生理信号的检测（具体如脑电图、眼电图、血流及呼吸变化、呼吸气流等基于多种传感器的检测方法）、基于人脸细节的检测（具体如眼部动作变化、嘴部运动变化、头部运动变化和手部特征等检测方法）和方向盘运动特征。基于人体生理信号的检测实时性好，准确率高，但缺点是会影响到驾驶员的正常驾驶，且人体生理信号又因个体、环境的不同而差异非常大，很难给出一个统一的量化标准。基于人脸细节的检测比较著名的有 PERCLOS 监测法，这种方法监测单位时间内眼睛闭合时间所占的百分比。目前，PERCLOS 检测法已被公认为最有效的、车载的、实时的非正常驾驶测评方法。虽然利用 PERCLOS 原理能够很好地监测驾驶员状态，但在实际运用中有许多问题。比如，不同体质和生活习惯的驾驶员的眼睛状态有很大不同，比如有些人睡觉时眼睛不闭合，因此误判率较高；PERCLOS 是一种多重的非严格的对象跟踪，车内照明条件的变化和头部的移动可导致预测不准甚至失败；当驾驶员头部没有正对摄像头时或者头发遮住了部分眼睛，无法识别眼部特征。这些问题使得采用 PERCLOS 方法产生了一定困难。[2]

### B. 深度学习

深度学习（Deep Learning，DL）由机器学习中的人工神经网络（Artificial Neural Network，ANN）发展而来，包含多隐层的神经网络即深度学习结构。对神经网络而言，深度指的是网络学习得到的函数中非线性运算组合水平的数量。传统神经网络的学习算法多是针对较低水平的网络结构，将这种网络称为浅层神经网络，如一个输入层、一个隐层和一个输出层的神经网

络；与此相反，将非线性运算组合水平较高的网络称为深度神经网络，如一个输入层、三个隐层和一个输出层的神经网络。深度学习与浅层学习相比具有许多优点，说明了引入深度学习的必要性：

a) 在网络表达复杂目标函数的能力方面，浅层神经网络有时无法很好地实现高变函数等复杂高维函数的表示，而用深度神经网络能够较好地表征。

b) 在网络结构的计算复杂度方面，当用深度为 $k$ 的网络结构能够紧凑地表达某一函数时，在采用深度小于 $k$ 的网络结构表达该函数时，可能需要增加指数级规模数量的计算因子，大大增加了计算的复杂度。另外，需要利用训练样本对计算因子中的参数值进行调整，当一个网络结构的训练样本数量有限而计算因子数量增加时，其泛化能力会变得很差。

c) 在仿生学角度方面，深度学习网络结构是对人类大脑皮层的最好模拟。与大脑皮层一样，深度学习对输入数据的处理是分层进行的，用每一层神经网络提取原始数据不同水平的特征。

d) 在信息共享方面，深度学习获得的多重水平的提取特征可以在类似的不同任务中重复使用，相当于对任务求解提供了一些无监督的数据，可以获得更多的有用信息。

深度学习比浅层学习具有更强的表示能力，能够从数据中学习到潜在的抽象高层次特征。2006 年，*Hinton* 等人提出的用于深度置信网络（Deep Belief Network，DBN）[3]的无监督学习算法，解决了深度学习模型优化困难的问题。2012 年，*Hinton* 课题组使用设计的深度卷积神经网络模型 AlexNet[4] 在当年的 ImageNet 大规模视觉识别挑战大赛[5]上夺得冠军，使得深度卷积神经网络受到关注。之后深度学习进入爆发期，各种深度学习模型（VGG[6]、GoogleNet[7]、ResNet[8] 和 DenseNet[9] 等）在 ImageNet 各个比赛上取得越来越好的成绩，使得深度学习受到广泛关注。深度学习模型在图像分类、目标检测等多领域都能够取得卓越的效果。

## III. 方法

### A. 深度学习模型

在本文中，我们分析了下列现有的卓越的深度学习模型并在此基础上设计融合深度模型用于非正常驾驶检测。

*1) 卷积神经网络（CNN）：* 最原始的深度卷积神经网络是 2012 年提出的 AlexNet 模型。AlexNet 在当年的 ImageNet 大规模视觉识别挑战大赛（ILSVRC-2010）上夺得冠军，证明了卷积层在复杂模型下的有效性，深层卷积网络能够学习到数据的潜在高层次特征，使得卷积神经网络受到关注。

*2) 宽卷积神经网络（Wide CNN）：* 宽卷积神经网络的想法来自于宽残差网络（WRN）[10]。宽残差网络是在残差网络（ResNet）的基础上增加卷积层的卷积核数量，因为随着模型深度的加深，梯度反向传播时，并不能保证能够流经每一个残差模块（residual block）的权值，以至于它很难学到东西，因此在整个训练过程中，只有很少的几个残差模块能够学到有用的表达，而绝大多数的残差模块起到的作用并不大。因此宽残差网络希望使用一种较浅的，但是宽度更宽的模型，来更加有效的提升模型的性能。在此基础上，我们直接关注宽的卷积层这个特点，作为宽卷积神经网络。

*3) 分组卷积神经网络（Group CNN）：* 分组卷积神经网络的想法来自于残差网络的改进 ResNeXt[11] 模型。ResNeXt 采用 VGG 堆叠的思想和 Inception[12] 的“拆分-变换-合并”策略，可扩展性比较强，在不增加参数复杂度的前提下提高准确率，同时还减少了超参数的数量。ResNeXt 提出了不同与深度和宽度的“基数”这个维度，并说明了增加基数比增加深度和宽度更有效。同时，ResNeXt 模型这种聚合残差结构完全等价于分组卷积。

*4) 深度残差网络（ResNet）：* 深度残差网络是何恺明于 2015-2016 年提出的深度学习领域最新技术概念，可以用于退化问题求解并加快深层网络的收敛速度。在 ImageNet 和 COCO 2015 竞赛中，共有 152 层的深度残差网络在图像分类、目标检测和语义分割各个分项都取得最好成绩，相关论文更是连续两次获得 CVPR 最佳论文。残差结构图如图 1(a) 所示。该方法的核心思想是：在原始网络层的基础上增加并行的恒等映射层，形成一个残差学习结构，具体来说：假设需要学习的潜在映射为 $H(x)$，那么构造非线性堆叠网络层来对应另一个残差映射 $F(x) := H(x) - x$ 这样的话潜在映射就可写为 $F(x) + x$ 而原始网络层优化残差映射比优化潜在映射更容易，映射 $F(x) + x$ 由加入直连连接的前馈神经网络实现。

*5) 稠密网络（DenseNet）：* 稠密网络主要参考了残差网络、Highway Network[13] 以及 GoogleNet。稠密网络的构建建立在卷积层如果离输入层/输出层更近，那么它更容易收敛，所以它的做法比残差网络更暴力，残

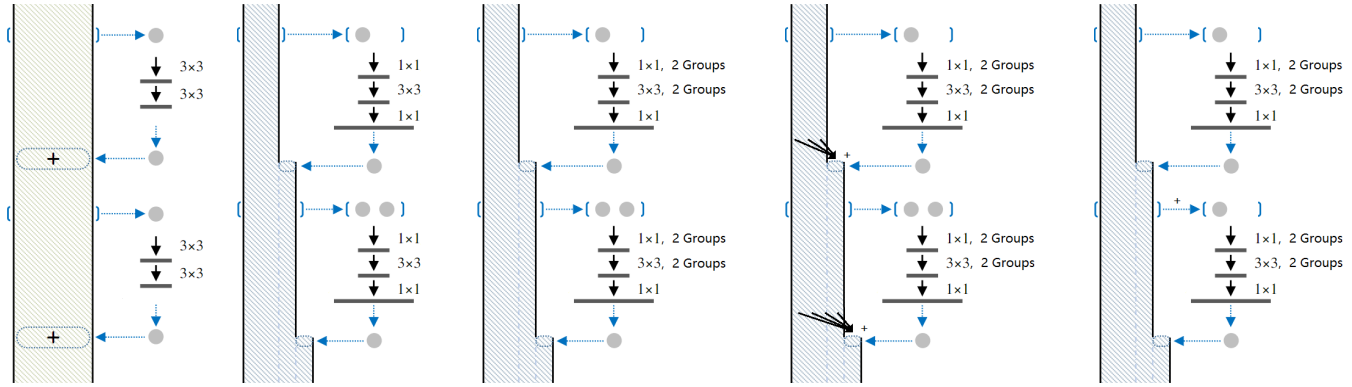


图 1: 模型基本结构图，每个图中长条是模型直连连接的数据流，每个基本结构从中取出数据，经过卷积的处理后，将输出加到直连的数据流中。(a) 深度残差网络（ResNet）(b) 稠密网络（DenseNet）(c) 加宽+分组+稠密卷积神经网络（WGD）(d) 加宽+分组+残差+稠密卷积神经网络1（WGRD1）(e) 加宽+分组+残差+稠密卷积神经网络2（WGRD2）

差网络只是前后两层之间的输出按元素相加，而稠密网络则是要求各个层之间都有关联，如图 1(b) 所示。即残差网络如果有 $L$ 层，则有 $L-1$ 个连接，而稠密网络每两个层之间都有一个连接，即 $L*(L+1)/2$ 个连接。稠密网络改变了传统网络反向传递时，梯度（信息）传播方式，由线性变成树状反向，这样的好处就在于减少了梯度消失的可能，并且加速训练，有利于更深层网络的训练。

### *B. 融合深度模型*

我们设计了三个融合深度模型（WGD、WGRD1 和 WGRD2），首次将其用于非正常驾驶检测。通常来说，非正常驾驶检测的主要目标是检测出驾驶员的驾驶状态，并判断其状态是否正常。即拍摄一张图像作为模型的输入，根据输出的标签判断其中驾驶员的状态是否正常。

融合深度模型基于目前取得卓越效果的稠密网络（DenseNet）模型，同时加入其他卓越模型的特点。具体细节如下：

*1) 加宽+分组+稠密卷积神经网络（Wide + Group + Densely CNN，WGD）：* 在稠密网络的基础上把卷积替换为分组并加宽的卷积，使模型具有较少的参数并同时具有较好的性能，如图 1(c) 所示。

*2) 加宽+分组+残差+稠密卷积神经网络1（Wide + Group + Residual + Densely CNN 1，WGRD1）：* 在 WGD 的基础上，改变输出并联的操作，先按元素相加再并联，可以看做把残差相加的方法引入模型，如图 1(d) 所示。具体来说：对于输入图像 $x_0$ 定义一个 $L$ 层的网络，他的第 $l$ 层是一个非线性变换（如 BN，ReLU，Conv 等），设为 $H_l(\cdot)$，输出为 $x_l$，那么我们一般有 $x_l = H_l(x_{l-1})$，在稠密网络和 WGD 中有

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \tag{1}$$

在 WGRD1 中则有

$$x_l = H_l([x_0, x_0 + x_1, \ldots, \sum_{i=0}^{l-1} x_i]) \tag{2}$$

其中 $\sum_{i=0}^{l-1} x_i$ 表示 $x_i (i = 0, 1, \ldots, l-1)$ 依次按元素相加。

*3) 加宽+分组+残差+稠密卷积神经网络2（Wide + Group + Residual + Densely CNN 2，WGRD2）：* 与 WGRD1 相似，不同的是对稠密块输出与输入的操作，先并联输出再按元素相加作为输入，如图 1(e) 所示。即在 WGRD2 中有

$$x_l = H_l(\sum_{i=0}^{l-1} x_i) \tag{3}$$

这与残差网络相似，若把残差网络中每个残差块之后的 ReLU 去掉，则相同。

## IV. 实验

### *A. 数据集和实验设置*

为了验证深度学习模型和融合深度模型在非正常驾驶检测上的效果，我们从 Kaggle 竞赛平台上获取了

相关竞赛数据集。数据集共包括 22424 张驾驶员图像，分别属于 10 个不同的类别。这10个类别包括安全驾驶、发短信、打电话、与乘客说话等等。我们对数据集中的每张图像进行标准化，并按照类别分别随机地划分到训练集和测试集中，即训练集和测试集分别地包含相当数量每个类别的图像。实验将深度学习模型和融合深度模型作比较，深度学习模型包括卷积神经网络（CNN）、宽卷积神经网络（Wide CNN）、分组卷积神经网络（Group CNN）、深度残差网络（ResNet）和稠密网络（DenseNet）；融合深度模型包括加宽+分组+稠密卷积神经网络（Wide + Group + Densely CNN，WGD）、加宽+分组+残差+稠密卷积神经网络1（Wide + Group + Residual + Densely CNN 1，WGRD1）和加宽+分组+残差+稠密卷积神经网络2（Wide + Group + Residual + Densely CNN 2，WGRD2）。

本论文的实验环境为 NVIDIA 提供的 CUDA 并行计算架构， cuDNN 深度学习加速库和 Facebook 提供的 Torch 深度学习框架，使用 Tesla K40m GPU 进行运算。实验控制每个模型都有一样多的参数（5.7M），我们采用批量梯度下降算法（Batch Gradient Descent，BGD）进行优化学习，进行了两个不同学习率的实验，学习率分别为 0.01 和 0.0001。我们的实验代码在 https://github.com/Lmy0217/AbnormalDrivingDetection 上开源。

### *B. 实验结果和分析*

*1) 学习率设置为 0.01:* 8 个模型都在训练集上学习了足够长的时间，最终在测试集上的性能表现如表 I 所示。我们发现 WGRD1 最终表现与瞎猜没有区别，这个模型是个失败的模型，失败的原因可能是先按元素相加再并联这种数据流操作方式使得数据过于复杂，网络难以学习出较好的特征。对于其他模型，Top 1 正确率都能达到 99% 以上，Top 5 正确率都能达到 99.9% 以上。我们看到最高的正确率由残差网络获得，最低的 Top 1 正确率由分组卷积神经网络获得，Top 5 正确率由卷积神经网络获得，我们设计的 WGD 和 WGRD2 能够获得中等的正确率。

*2) 学习率设置为 0.0001:* 由于在 0.01 学习率下除了失败的 WGRD1 模型外其他模型最终效果差异太小，所以我们还做了学习率为 0.0001 的实验，这个实验我们摒弃了失败的 WGRD1 模型，每个模型只训练了 10 轮，观察每个模型随着训练次数增加损失下降快慢和正确率上升的快慢，即学习收敛的快慢。实验结果如图 2 所示。我们看到损失下降最快同时正确率上升最快的是宽卷积神经网络，与此相反的是残差网络，我们设计的 WGD 和 WGRD2 位于中间。我们猜想可能是网络越宽，则损失下降最快同时正确率上升最快，即网络越宽收敛越快。

表 I: 学习率 0.01 模型实验 Top 1/5 正确率

| 模型（名称-层数） | Top 1 正确率（%） | Top 5 正确率（%） |
|---|---|---|
| CNN-16 | 99.3223 | 99.9197 |
| Wide CNN-12 | 99.3580 | 99.9643 |
| Group CNN-25 | 99.2153 | 99.9554 |
| ResNet-16 | **99.4560** | **99.9910** |
| DenseNet-88 | 99.3669 | 99.9465 |
| WGD-88 | 99.2688 | 99.9375 |
| WGRD1-88[1] | 9.8528 | 50.6286 |
| WGRD2-29 | 99.2331 | 99.9286 |

[1] WGRD1-88 是个失败的模型。

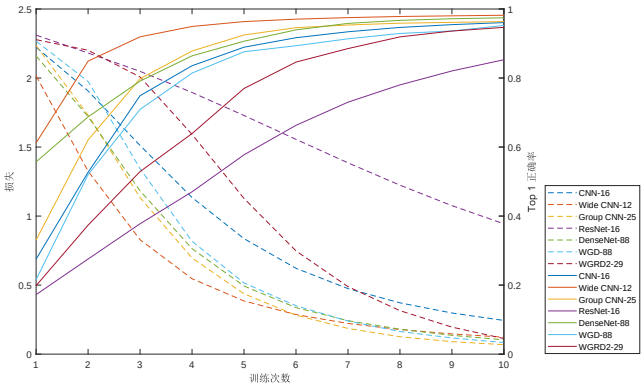


图 2: 学习率 0.0001 模型实验损失与 Top 1 正确率

## V. 结论

在本文中我们设计了三个融合深度模型 WGD、WGRD1 和 WGRD2。这三个模型基于稠密网络模型，并加入了其他卓越模型的特征。基于 Kaggle 竞赛数据集的实验表明了：1）使用深度模型可以很好地解决非正常驾驶检测问题；2）我们设计的 WGD 和 WGRD2 模型能够在实验中获得平均水平的正确率；3）我们设计的 WGRD1 模型的失败说明其数据操作方式“先按元素相加再并联”过于复杂；4）深度模型的宽度正相关与学习收敛速度。我们相信未来的研究会使用更复杂的深度模型并取得更好的非正常驾驶检测效果。

## 致谢

本文作者感谢徐子晨老师提供的指导及超算资源。

## 参考文献


[1] R. L. Olson, R. J. Hanowski, J. S. Hickman, and J. L. Bocanegra, "Driver distraction in commercial vehicle operations," tech. rep., 2009.

[2] A. Acıoğlu and E. Erçelebi, "Real time eye detection algorithm for perclos calculation," in *Signal Processing and Communication Application Conference (SIU), 2016 24th*, pp. 1641–1644, IEEE, 2016.

[3] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[9] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993*, 2016.

[10] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.

[11] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pp. 5987–5995, IEEE, 2017.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning.," in *AAAI*, pp. 4278–4284, 2017.

[13] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," *arXiv preprint arXiv:1505.00387*, 2015.