

基于残差网络和注意力机制的夜间航运识别

段家家¹ 张鸿²

¹(武汉科技大学计算机科学与技术学院 湖北 武汉 430065)

²(武汉科技大学智能信息处理与实时工业系统湖北省重点实验室 湖北 武汉 430065)

摘要 针对常规的深度学习模型对夜间监控视频进行识别的效果不佳,提出一种基于残差网络和注意力机制的夜间航运事件的识别方法。增强夜间监控视频生成的暗光图像中的光照,采用SE-R2(2+1)模型对增强图像组合成的视频进行识别。该识别模型基于R(2+1)D模型,通过改进模型的激活结构,提升模型的泛化能力。嵌入SENet网络来提高模型的表征能力。实验结果表明,在增强后形成的数据集下,该方法识别准确率达到了88.2%,验证了模型的有效性。

关键词 低光照 夜间视频 图像增强 注意力机制 行为分类

中图分类号 TP391.4

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.12.020

NIGHT SHIPPING IDENTIFICATION BASED ON RESIDUAL NETWORK AND ATTENTION MECHANISM

Duan Jiajia¹ Zhang Hong²

¹(College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

²(Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

Abstract Aimed at the poor performance of conventional deep learning models in identifying night surveillance videos, a method for identifying night shipping events based on residual network and attention mechanism is proposed. The illumination in the dark image generated by the night surveillance video was enhanced, and the SE-R2(2+1) model was used to identify the video combined by the enhanced image. The recognition model was based on the R(2+1)D model. By improving the activation structure of the model, the generalization ability of the model was improved. At the same time, the SENet network was embedded to improve the characterization ability of the model. Experimental results show that under the enhanced dataset, the recognition accuracy of the proposed method reaches 88.2%, which verifies the effectiveness of the model.

Keywords Low-light Night video Image enhancement Attention mechanism Behavior recognition

0 引言

随着深度学习技术的发展,人们将深度学习与低光照增强技术相结合,使低光照图像增强技术飞速进步。它可以在增强图像光照和色彩的同时,优秀地保留图像的结构和细节,因此基于深度学习的低光照增强技术被广泛地应用到夜间识别的领域中,如夜间交通摄像头对车辆和行人的识别。

航运事件识别是通过船上的摄像头来监控船上货物,防止船家非法盗取货物^[1],同时对货物在航运时的遭遇的异常状态进行报警(如下雨时货物没有盖雨布或雨布没有盖好),这样可以有效地降低货主在运输途中受到的损失,本文将研究夜间环境下的航运事件识别。夜间行驶时,船只在江上处于一种低光照状态,所以生成的监控视频光照偏低,质量比较差,直接采用深度学习算法提取视频特征会得到较差的效果,因此需要低光照图像增强算法对拍摄的视频中截取的图像

进行光照增强。

目前基于深度学习的低光照增强模型绝大多数采用的是配对数据集(航运数据集是非配对数据集)进行模型训练,导致这些模型的泛化性较差,应用到夜间航运图像中进行增强会有较严重的色彩或细节失真。因此我们采用较热门的无监督低光照增强模型-EnlightenGAN^[2],使用非配对数据集进行训练,泛化性较强,它可以有效地增强夜间航运视频中截取的图像的光照。本文针对夜间航运监控图像进行增强后形成的视频噪音较大的关键问题,受到 $R(2+1)D$ ^[3] 方法的启发,提出 SE-R2(2+1)D 模型,以解决夜间场景下航运事件识别的问题。本文方法的创新点是提高了 $R(2+1)D$ 模型的代表能力,从而降低增强后图像重组成的视频中的噪音对航运事件识别的影响,同时增强模型的泛化能力来应对航运中出现新的货物对航运事件识别造成的影响。

1 相关工作

EnlightenGAN 是目前较热门的低光照增强模型之一。本文提出的 SE-R2(2+1)D 模型属于一种视频识别模型,参考了视频识别领域的经典 C3D 模型^[4]。

1.1 主流低光照增强方法

近年来,低光照增强方法大致分为四种类型:直方图均衡算法、暗通道先验算法、基于 Retinex 理论^[5] 的算法,以及目前比较热门的基于深度学习的低光照增强算法。文献[6]提出的 LLNET 模型较早地将深度学习算法用于低光照增强。它证明了基于合成数据训练的堆叠稀疏去噪自编码器能够对低光照噪声图像进行增强和去噪。文献[7]基于 Retinex 理论,采用先降解后增强两个步骤来增强光照,降解网络(Decom-Net)能够对图像进行解耦,得到光照图和反射图,增强网络(Enhance-Net)对前面得到的光照图进行增强,增强后的光照图和原来的反射图相乘可以得到增强结果。文献[8]是一个多分支低光照图像增强网络,它的核心思想是在不同的特征等级中提取出丰富的图像特征,可以使图片能够获得更加自然的增强效果,增强之后的细节更加清晰,但是在光照不是很均匀的情况下,该算法输出的结果有点过曝光的效果。文献[9]在文献[7]的思路基础上更进一步将网络分为三部分:图像分解、反射图恢复、光照图调整。它可以灵活地调整光照等级,在提升图像亮度的同时,极大降低了噪音和颜色失真之类的退化。文献[2]第一次将非配对训练样本技术引入到了低光照图像增强中,解决了上述算法在

构建配对数据集中产生的巨额开销问题。

1.2 主流视频识别方法

随着深度学习技术的发展,学者们发现卷积神经网络不仅可以应用到二维的图像识别,还可以更进一步地应用到视频识别领域。目前主流的视频行为识别网络为双流神经网络和 3D 卷积神经网络。

文献[10]提出了双流(two-stream)结构的 CNN (Convolutional Neural Network),由空间和时间两个维度的网络组成。一个是普通的单帧的卷积神经网络,另一个卷积网络使用多帧的密集光流场作为训练输入,用来提取动作信息,然后利用多任务训练的方法把两个数据集联合起来,把两个 Softmax 层的输出融合,但是不能在空间和时间特征之间学习像素级的对应关系。文献[11]在上面双流神经网络的基础上,采用了两个独立的卷积神经网络来进行时间和空间的融合,将单帧的图像信息和帧与帧之间的变化信息进行融合,单帧的图像可以形成对空间的描述,而通过光流法等方法可以形成对时间的描述,从而达到时间和空间互补的目的。文献[12]提出了一种新颖的 CNN 体系结构,隐式捕获相邻帧之间的运动信息,它仅将原始视频帧作为输入并直接预测动作类别,而无须明确计算光流。它通过给定的行为网络,将其与时间流 CNN 连接起来,该 CNN 将估计的光流映射到目标动作标签。然后以端到端的方式微调堆叠的时间流 CNN,以此来预测输入帧的动作类别,这样就可以隐式生成用于动作识别的运动信息。它的训练速度达到了传统双流神经网络的十倍,但在时序信息上会有一定的丢失。

文献[4]提出的 C3D 卷积网络是 3D 卷积网络的里程碑,采用 $3 \times 3 \times 3$ 的三维卷积核来提取视频特征,相比二维卷积神经网络可以更好地保留时序信息。文献[3]提出的 $R(2+1)D$ 卷积神经网络是文献[4]作者在 C3D 神经网络改进而来的,将 3DCNN (Three-Dimensional Convolutional Neural Network) 替换为 3D ResNet (Residual Network),并且将 3D 卷积分解成了二维的空间卷积和一维的时间卷积。

以上可以看出,文献[2]提出的 EnlightenGAN 模型可以满足航运数据集的要求,适合增强船舶夜间图像。由于船舶运输的货物可能发生变化,需要提高模型的泛化能力,并且该模型主要识别的是低光照图像增强后形成的视频,某些区域会有一些噪音,需要提升该模型提取主要特征的能力。因此本文提出 SE-R2(2+1)D 模型,它通过将 $R(2+1)D$ 模型中的 ResNet^[13] 结构的后激活优化为 ResNet V2^[14] 的完全预激活,以此来提升模型泛化能力,然后结合 SENet^[15] 神

经网络来提取船舶主要特征抑制不重要特征,对夜间航运事件识别取得了不错的准确率。

2 方法介绍

由于船舶在较短时间内并不会有明显的变化,如果采用识别连续帧的方法难以获取到有效的时序信息,对夜间航运事件识别准确率影响较大。因此每五分钟从监控视频中截取一帧,这样形成的视频中相邻帧的变化会比较明显,将截取的帧图像通过航运光照数据集训练过的 EnlightenGAN 模型来进行低光照增强,然后以六帧时序帧序列作为识别模型的输入,以 $R(2+1)D$ 网络作为基础框架,使用完全预激活的残差结构,并结合 SENet 网络的 SE-R2(2+1)D 模型来训练数据集,最后用该模型对夜间航运事件进行识别并预警。

2.1 基于夜间航运事件识别的 $R(2+1)D$ 方法

$R(2+1)D$ 神经网络的特点是时空卷积滤波器,时空卷积滤波器的特点是在 ResNet 的架构上将三维卷积滤波器分解成了二维的空间卷积滤波器和一维的时间卷积滤波器,让空间卷积与时间卷积交替进行。这样分解相对于传统的三维卷积滤波器在模型性能有两个较大的提升:1) 两个子卷积之间多出来一个非线性操作,与原来同样参数的三维卷积相比翻倍了非线性操作,从而使模型能够表示更复杂的函数。2) 让模型训练时优化的过程也分解开来,普通的三维时空卷积会在对输入的数据进行卷积时将其中空间信息和动态信息僵硬地结合在一起,导致模型不易优化。分解为二维的空间卷积和一维的时空卷积会使模型更容易优化,训练的损失更低。本文的夜间航运事件识别中对光照增强图像形成的视频进行时空卷积来提取特征图的方法如图 1 所示。

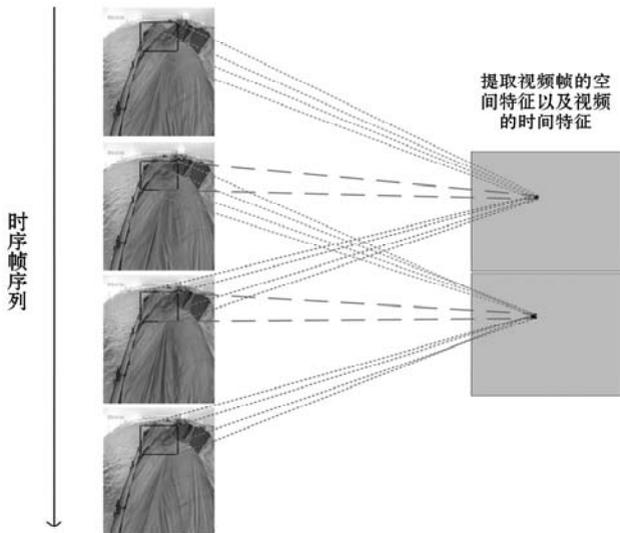


图1 夜间航运增强视频的时空卷积建模

可以看出,时空卷积既提取了视频单帧的空间特征也提取视频的时间特征。输入的视频会先经过空间卷积提取空间特征,再经过时间卷积提取时间特征。空间卷积和时间卷积之间通过中间通道结合起来,决定中间通道 M_i 的式子如下:

$$M_i = \frac{td^2 N_{i-1} N_i}{d^2 N_{i-1} + t N_i} \quad (1)$$

式中: t 表示卷积滤波器的时间长度; d 表示卷积空间的长和宽; N_{i-1} 表示输入通道; N_i 表示输出通道。

但在夜间航运事件识别过程中,船舶运输的货物可能会发生变化,因此需要提高模型的泛化能力来提高识别准确率。

2.2 基于 $R(2+1)D$ 神经网络的 ResNet V2 结构

最初,He 等^[13]对传统的卷积网络进行了优化,提出了残差网络(ResNet)结构,主要特点是通过捷径连接和恒等映射解决了梯度消失或者梯度爆炸的问题,提高了神经网络的性能。从最开始的 ResNet 不断地优化改进,学者们陆续地提出了 Wide ResNet^[16]、ResNeXt^[17]、ResNet V2^[14]。

本文受到 ResNet V2 中预激活的启发,改变了 $R(2+1)D$ 网络中激活单元的顺序,并且对模型中参数做出了微调来适应本文的 $R(2+1)D$ 网络模型。该改进将激活函数(BN, ReLU)调整到时空卷积之前,作为时空卷积的预激活,改进后的 $R2(2+1)D$ 模型的残差块结构如图 2 所示。

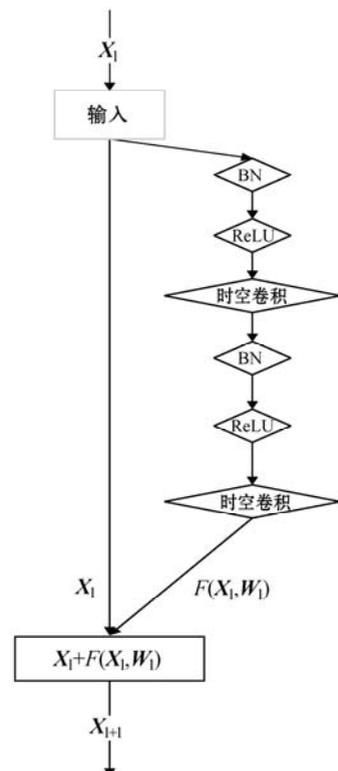


图2 改进后的 $R(2+1)D$ 模型残差块

图 2 中 BN (Batch Normalization)^[18] 表示批量归一化,用来解决在训练过程中中间层数据分布发生改变的问题,ReLU (Rectified Linear Unit)^[19] 表示线性整流函数,是一种神经网络中常用的激活函数。图 2 右侧部分被称作残差分支,左侧部分被称作恒等分支。由图 2 可以看出,输入的浅层网络单元会在右侧残差分支进行批量归一化,ReLU 函数激活,接着进行时空卷积提取特征,然后重复一次,最后与左侧恒等分支中没有发生变化浅层网络单元相加输出深层网络单元。该结构将原始结构中的 ReLU 激活函数都移到了图 2 的右侧,使得信息传递通道更加干净,让信息在模型中的前向传播和反向传播更加平滑,模型的收敛速度加快,训练损失降低。原始结构中 BN 将信息进行归一化之后马上就与左侧恒等分支进行相加,导致融合后的信息并不是归一化的,该信息会当作下一层的输入。改进后将 BN 放在时空卷积之前,使得所有权重层的输入都经过了归一化,能加快模型收敛速度,增强模型的抗过拟合能力。这两处改进都提升了模型的泛化能力。

原始残差网络的结构如式(2)和式(3)所示。

$$F(\mathbf{x}_l, \mathbf{w}_l) = \mathbf{w}_2 \times f(\mathbf{w}_1 \times \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2 \quad (2)$$

$$\mathbf{x}_{l+1} = f(\mathbf{x}_l + F(\mathbf{x}_l, \mathbf{w}_l)) \quad (3)$$

式中: \mathbf{x}_{l+1} 表示深层网络单元; \mathbf{w}_1 和 \mathbf{w}_2 表示进行一次时空卷积操作; \mathbf{x}_l 表示浅层网络单元; \mathbf{w}_l 表示时空卷积; \mathbf{b}_1 和 \mathbf{b}_2 表示偏差因子; F 表示残差函数; f 表示 ReLU 激活函数。修改后的残差网络结构公式为:

$$F(\mathbf{x}_l, \mathbf{w}_l) = f(\mathbf{w}_2 \times f(\mathbf{w}_1 \times \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) \quad (4)$$

$$\mathbf{x}_{l+1} = \mathbf{x}_l + F(\mathbf{x}_l, \mathbf{w}_l) \quad (5)$$

对式(5)进行递归计算可以得到式(6)。

$$\mathbf{x}_L = \mathbf{x}_l + \sum_{i=1}^{L-1} F(\mathbf{x}_i, \mathbf{w}_i) \quad (6)$$

式(6)表达了任何第 L 层网络单元(深层)与第 l 层网络单元(浅层)之间关系;通过设损失函数为 loss,则其反向传播公式如式(7)所示。

$$\frac{\partial \text{loss}}{\partial \mathbf{x}_l} = \frac{\partial \text{loss}}{\partial \mathbf{x}_L} \left(1 + \frac{\partial}{\partial \mathbf{x}_l} \sum_{i=1}^{L-1} F(\mathbf{x}_i, \mathbf{w}_i) \right) \quad (7)$$

可以看出,图 2 进行的结构调整可以使反向传播基本符合假设,信息的传递不会受到阻碍,简化优化过程,并且让 BN 起到了正则化作用。让模型整体上更易于训练,泛化能力更强。解决了船舶运输的货物可能发生变化导致识别错误的情况,从而提高模型的识别准确率。

由于本文主要识别的是夜间图像进行低光照增强后形成的视频,因此视频中可能存在一些噪音来影响视频的识别,我们需要通过增强主要特征的提取能力

来抑制这些噪音,从而提高模型的识别准确率。

2.3 基于改进 R(2+1)D 和 SENet 网络的 SE-R2(2+1)D 网络模型

本文在 R(2+1)D 神经网络的基础上,选用泛化能力更强的改进后的 R(2+1) 作为网络的基本结构,加入隐式注意力机制^[20-21] 的 SENet 网络来提升模型特征提取能力。SENet 网络由一个全局平均池化层、两个全连接层、Sigmoid 激活函数、ReLU 激活函数组成。

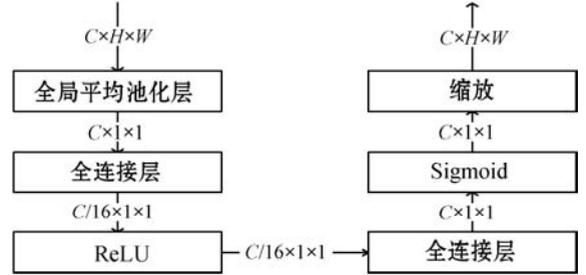


图 3 SENet 网络的特征提取流程

从图 3 可以看出,当经过两次时空卷积得到特征图的 H (高) \times W (宽) \times C (通道数)时,通过 SENet 网络中的全局平均池化层来降维特征图,将其降维为 $1 \times 1 \times C$ 的一维特征图,该操作将得到通道级的全局特征,如式(8)所示,然后通过一个全连接层来减少通道个数从而降低模型的计算量,接着采用 ReLU 激活函数激活,后通过第二个全连接层将全局特征恢复到原来的通道数,这个过程可以降低模型复杂度以及提升泛化能力,然后通过一个 Sigmoid 的门获得 0~1 之间归一化的权重,如式(9)所示。最后通过一个缩放的操作来将归一化后的权重加权到每个通道的特征上,得到原来的特征维度。

$$\mathbf{Z}_c = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H \mathbf{u}_c(i, j) \quad (8)$$

$$s = \sigma(\mathbf{W}_2 \times \delta(\mathbf{W}_1 \times \mathbf{Z}_c)) \quad (9)$$

式中: W 和 H 表示特征图的宽和高; \mathbf{u}_c 表示特征图中第 C 个二维矩阵; \mathbf{Z}_c 表示得到的一维特征图; \mathbf{W}_1 表示第一个全连接操作, \mathbf{W}_1 的维度是 $C/r \times C$, r (本文设为 16) 是一个缩放参数; δ 表示 ReLU 激活函数; \mathbf{W}_2 是第二个全连接层; \mathbf{W}_2 的维度是 $C \times C/r$; σ 表示 Sigmoid 激活函数; s 表示通道权重系数。

图像进行光照增强后会在部分区域形成一定程度的噪音,某些时段运输的货物是煤,在光照较低的环境中容易与周围环境混淆,这种情形下形成的噪音将会比较严重,本文通过 SENet 网络对可能影响主体船舶识别的失真区域进行抑制,从而降低噪音的影响,提高模型的准确率。本文提出的 SE-R2(2+1)D 模型的整体结构如图 4 所示。

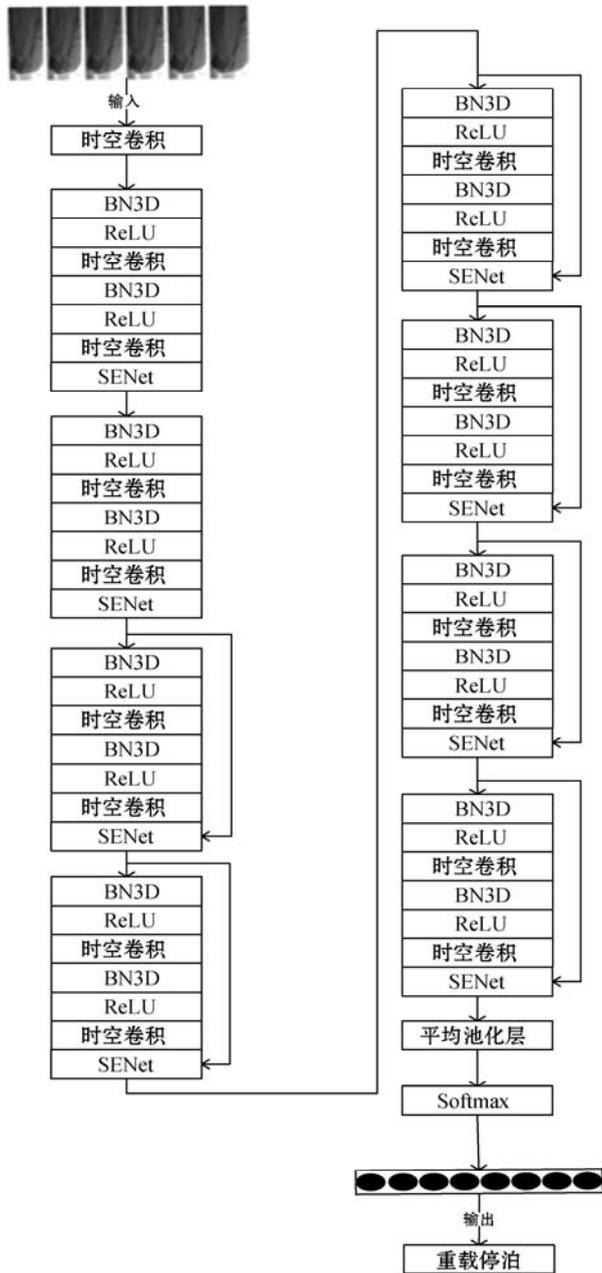


图4 SE-R2(2+1)D模型整体结构

图4中的模型由8个残差块组成,每个残差块进行两次时空卷积,共进行了17次时空卷积,整体结构类似于ResNet18的网络结构。该模型会将输入的六帧图像尺寸大小调整为 172×128 ,根据图像尺寸的大小,设置模型的第一个时空卷积的卷积核设为 $1 \times 7 \times 7$,之后的时空卷积核都设为 $3 \times 3 \times 3$ 来提取特征。SENet模块放在每个残差块的最后,对时空卷积提取的特征图进行加强,这种注意力机制让SE-R2(2+1)D模型可以更加关注信息量最大的船舶的通道特征,而抑制些不重要的通道特征。

3 实验

本文选用了3DCNN(VGG11)^[22]、R-C3D^[23](Region

Convolutional 3D Network)、R3D^[24]、R(2+1)D模型与本文提出的SE-R2(2+1)D模型来进行对比,并对SE-R2(2+1)D模型进行消融实验。实验环境的配置如下: Intel(R) Core(TM) i9-9900, NVIDIA GeForce RTX 1080Ti 16 GB显卡, 32 GB内存。操作系统是Windows 10, 编程语言是Python 3.6.5, 深度学习框架为PyTorch。

3.1 数据集描述

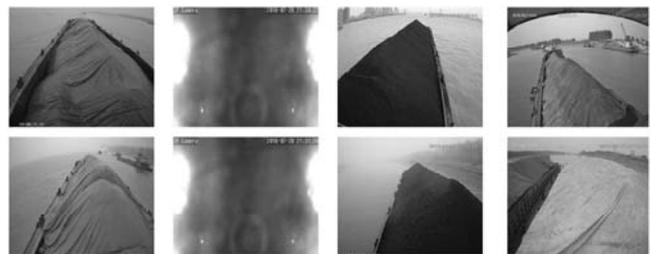
SE-R2(2+1)D模型的数据集源于被EnlightenGAN模型增强后的图像和未增强的原始图像数据集形成的视频,视频分辨率为 700×600 ,所占内存为17.4 GB,分为8类,分别是正常行驶、摄像头遮挡、未盖布行驶、停泊、空仓、装货、卸货、雨布吹飞。原始图像数据集由68 645幅图像构成,正常行驶9 311幅,摄像头遮挡7 584幅,未盖布行驶11 352幅,停泊8 648幅,空仓8 127幅,装货7 981幅,卸货8 354幅,雨布吹飞7 288幅。其中未盖布行驶种类数据集中包含三种运输货物,分别是煤、土、沙,其中:煤4 024幅,土3 357幅,沙3 971幅。训练集和测试集按7:3的比例分配。

3.2 参数设置

采用SE-R2(2+1)D卷积神经网络初始化特征并提取网络卷积层权重,采用随机梯度下降算法优化整个网络模型。将数据预处理后形成的视频每次截取六帧为输入,大小调整为 172×128 ,网络训练的初始学习率0.01,网络每经过10次迭代学习率降为原来的0.1,共迭代80次,批大小设置为8。

3.3 实验结果分析

航运事件识别方法用两个数据集来进行测试:1) 夜间航运事件生成的原生图像数据集;2) 用EnlightenGAN模型对其进行增强后形成的增强数据集。将数据集按7:3的比例分配训练集和测试集。最后统计各类别准确率以及平均准确率,并且将平均准确率作为不同方法优劣的客观评价指标。识别的事件类别共被分为8类,这8类是根据船舶在江上运输最常出现的状况以及对运出货物会造成重要影响的事件得出的分类,分别是正常行驶、摄像头遮挡、未盖布行驶、停泊、空仓、装货、卸货、雨布吹飞。类别示例如图5所示。



(a) 正常行驶 (b) 摄像头遮挡 (c) 未盖布行驶 (d) 停泊

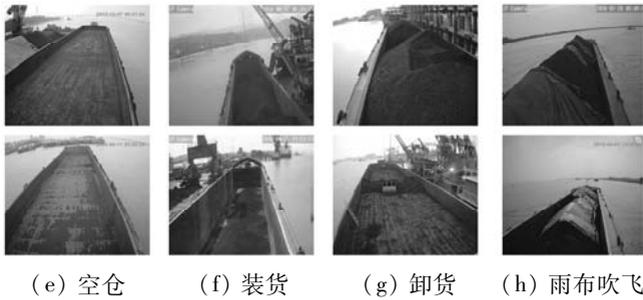


图 5 类别实例

表 1 显示了本文划分出的航运事件在不同的识别算法中的准确率,可以看出,当我们不采用 EnlightenGAN 模型来增强数据集时,各个模型的识别准确率都处于比较低的状态,其中文献[22]提出的 3DCNN(VGG-11)模型的航运事件识别准确率只达到了 63.3%,处于较低水平。文献[23]提出的 R-C3D 达到了 66.7% 的准确率,文献[24]提出的 R3D 模型达到了 63.9% 的准确率,在 R3D 基础上改进而来的 R(2+1)D 的准确率达到了 67.2%,相比 R-C3D 模型的准确率有所提高。而本文提出 SE-R2(2+1)D 模型对航运事件的识别准确率更进一步,达到了 69.5%,但结果还是处于较低水平,无法直接应用到夜间识别的实际环境中。

表 1 不同模型在不同类别的原始数据集下的准确率

类别	3DCNN VGG11	R-C3D	R3D	R(2+1)D	本文方法
停泊	0.657	0.708	0.671	0.728	0.747
空仓	0.661	0.676	0.652	0.654	0.676
装货	0.631	0.629	0.603	0.652	0.669
卸货	0.591	0.688	0.652	0.673	0.690
雨布吹飞	0.526	0.551	0.529	0.613	0.629
正常行驶	0.584	0.604	0.588	0.631	0.656
未盖布行驶	0.646	0.679	0.636	0.698	0.714
摄像头遮挡	0.622	0.643	0.627	0.675	0.693
平均准确率	0.633	0.667	0.639	0.672	0.695

从表 2 可以看出对原始数据集用 Enlighten 模型进行低光照增强后,再用同样的识别模型进行视频识别时识别准确率有了较大幅度的提升。3DCNN(VGG-11)神经网络模型有了 0.184 的准确率提升,达到了 81.7%,而 R-C3D 模型的准确率达到 84.2%,有了 0.175 的提升。而 R(2+1)D 神经网络模型则提升了 0.186,达到了 85.8% 的准确率。本文提出的 SE-R2(2+1)D 模型提升了 0.187 的准确率,达到了 88.2% 的准确率,呈现了一个比较优秀的结果。同时与表 2 中的其他模型相比较,本文提出的 SE-R2(2+1)D 模型的准确率也是最高的。

表 2 不同模型在不同类别的增强数据集下的准确率

类别	3DCNN VGG11	R-C3D	R3D	R(2+1)D	本文方法
停泊	0.798	0.847	0.806	0.852	0.864
空仓	0.784	0.790	0.784	0.841	0.872
装货	0.769	0.795	0.782	0.840	0.861
卸货	0.702	0.802	0.766	0.894	0.895
正常行驶	0.808	0.834	0.823	0.848	0.857
雨布吹飞	0.784	0.801	0.786	0.835	0.866
摄像头遮挡	0.836	0.843	0.837	0.862	0.893
未盖布行驶	0.821	0.859	0.824	0.871	0.912
平均准确率	0.817	0.842	0.823	0.858	0.882

表 3 和图 6 展示了对本文模型进行消融实验的结果,其中:用 R2(2+1)D 来表示改进 R(2+1)D 模型中激活单元顺序后形成的网络模型,SE-R(2+1)D 表示将 SENet 网络结合到 R(2+1)D 模型后形成的网络模型。从表 3 中可以看出 R2(2+1)D 模型以及 SE-R(2+1)D 模型相比 R(2+1)D 模型的识别准确率均有所提高。从整体上可以看出本文方法呈现的效果最好,而且模型的收敛速度也相对比较稳定。结合表 2 和表 3 可以看出,本文提出的 SE-R2(2+1)D 模型拥有对夜间图像增强后组成的视频中的噪音较好的抑制能力。

表 3 不同改进的识别模型在增强数据集下的结果对比

模型	平均准确率
R(2+1)D	0.858
R2(2+1)D	0.865
SE-R(2+1)D	0.878
本文方法	0.882

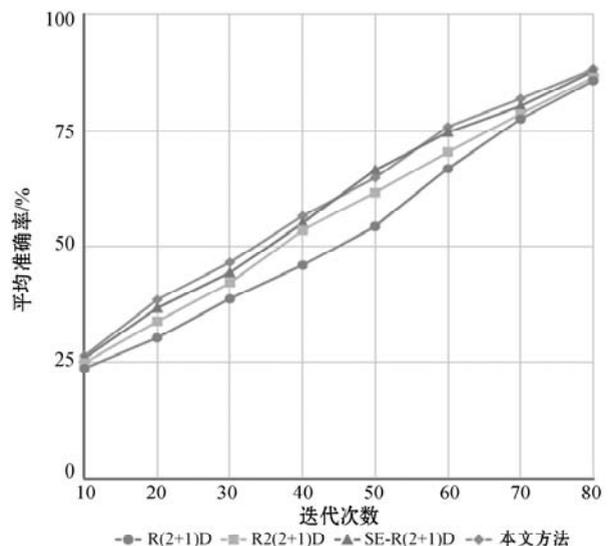


图 6 不同改进的识别模型在增强数据集下准确率变化情况

当货物发生变化时,它对未盖布行驶这一种类的识别影响是最大的,未盖布行驶时运输的货物如图 7 所示。

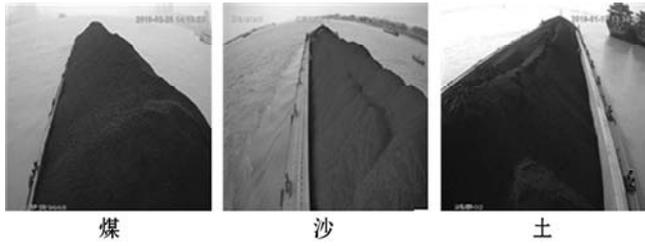


图 7 航运货物

图 7 中航运的货物分别是煤、沙和土。可以看出货物煤和土的颜色较为接近,但煤的外表面整体呈颗粒状,而土的外表面较为平滑,沙的颜色与其他两者有明显不同,外表面也看起来较为平滑。

由表 4 可以看出,当航运货物发生了变化时本文方法对货物的识别准确率均高于 90%,表现出了不错的泛化能力。

表 4 未盖布行驶时不同货物在增强数据集下的准确率

货物	3DCNN VGG11	R-C3D	R3D	R(2+1)D	本文方法
煤	0.811	0.853	0.807	0.885	0.908
土	0.835	0.866	0.848	0.867	0.921
沙	0.827	0.861	0.819	0.859	0.906

从表 5 中可以看出,随着模型的优化,结构也逐渐复杂了起来,导致模型对每段视频的识别所需时间也逐渐变长,从开始的 3DCNN(VGG-11)模型只需要 0.415 s,到本文提出的 SE-R2(2+1)D 模型需要 0.976 s,由于视频事件是半小时识别一次,所以 0.976 s 仍然可以满足夜间航运事件识别的时间需求。

表 5 增强数据集上不同模型平均识别时间 单位:s

模型	平均时间
3DCNN(VGG11)	0.415
R-C3D	0.372
R3D	0.513
R(2+1)D	0.826
本文方法	0.976

本文在进行模型训练和实验结果分析时,剔除了某些航行图中受到夜间强光照射到摄像头时生成的图像,这是为了可以最大程度上保证模型训练的效果。但是在船舶夜间航行时,会经常遇到强光照射到摄像头的情况,所以它也具有一定的研究价值。本文在实验中收集了一些受到强光照射时生成的夜间图像,如图 8 所示。

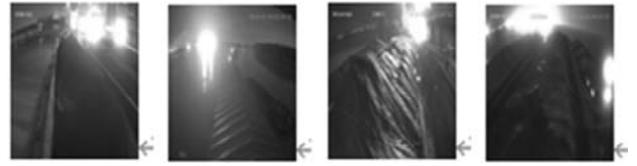


图 8 夜间强光图像样例

在强光照射环境下获取到的图像由于局部光照过于明亮,低光照增强后生成的图像失真会较为严重,导致模型的识别准确率下降。本文将夜间强光环境下的图像光照增强后用实验中事件识别模型进行对比实验,结果见表 6。

表 6 强光环境下增强的数据集识别结果

模型名称	识别准确率	处理时间/s
3DCNN(VGG11)	0.591	0.415
R-C3D	0.634	0.372
R3D	0.597	0.513
R(2+1)D	0.648	0.828
本文方法	0.714	0.976

可以看出,强光对模型识别的准确率影响很大,各模型对航运事件的识别平均准确率都有了较大幅度的下降,但本文提出的 SE-R2(2+1)D 模型凭借结合了 SENet 网络增加的特征提取能力,降低了失真对模型识别的干扰,同时改变激活函数的顺序来增加模型的泛化能力,使其获得了最高的准确率,达到了 71.4%,且处理时间并没有太大的变化,可以作为在夜间强光环境时进行事件识别的一个参考依据。

4 结 语

针对当前航运事件时识别中的传统模型学习能力不足、夜间监控视频中光照不足造成视频质量较差的问题,本文提出解决方法。针对低光照情况,首先研究了 EnlightenGAN 模型,它可以通过非配对图像的数据集进行训练,适合航运监控项目产生的数据集,增强图像的过程用时较短,不会影响到实时识别的要求,增强后的图像不论是色彩、亮度、语义细节的保留程度都不错。对于夜间航运监控视频中每六分钟截取的帧图像进行增强后形成的视频,本文提出 SE-R2(2+1)D 模型,在 R(2+1)D 模型基础上通过抑制图像增强后产生的视频中的噪音以及增加模型泛化能力使得识别准确率更进一步,达到了一个不错的效果。然而本文模型的参数数量较大,导致模型训练需要较多的计算资源,接下来的工作将研究如何对模型进行剪枝以及压缩来降低模型的参数数量。

参 考 文 献

- [1] 明道睿,张鸿. 基于约束深度神经网络的航运监控事件识别[J]. 计算机工程与设计,2020,41(10):2949-2957.
- [2] Jiang Y F, Gong X Y, Liu D, et al. EnlightenGAN: Deep light enhancement without paired supervision [J]. IEEE Transactions on Image Processing,2021,30:2340-2349.
- [3] Tran D, Wang H, Torresani L, et al. A closer look at spatiotemporal convolutions for action recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2018:6450-6459.
- [4] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks [C] // IEEE International Conference on Computer Vision, 2015:4489-4497.
- [5] Park S, Yu S, Kim M, et al. Dual autoencoder network for Retinex-based low-light image enhancement [J]. IEEE Access,2018,6:22084-22093.
- [6] Lore K G, Akintayo A, Sarkar S. LLNet: A deep autoencoder approach to natural low-light image enhancement [J]. Pattern Recognition,2017,61:650-662.
- [7] Wei C, Wang W J, Yang W, et al. Deep Retinex decomposition for low-light enhancement [EB]. arXiv:1808.04560, 2018.
- [8] Lv F, Lu F, Wu J H, et al. MBLLEN: Low-light Image/video enhancement using CNNs [C] // British Machine Vision Conference,2018:1-13.
- [9] Zhang Y H, Zhang J W, Guo X J. Kindling the darkness: A practical low-light image enhancer [C] // 27th ACM International Conference on Multimedia,2019:1632-1640.
- [10] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos [J]. Advances in Neural Information Processing Systems,2014,14(6):2199.
- [11] Feichtenhofer C, Pinz A, Zisserman A. Convolutional two-stream network fusion for video action recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016:1933-1941.
- [12] Zhu Y, Lan Z, Newsam S, et al. Hidden two-stream convolutional networks for action recognition [C] // Asian Conference on Computer Vision,2018:363-378.
- [13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition,2016:770-778.
- [14] He K M, Zhang X Y, Ren S Q, et al. Identity mappings in deep residual networks [C] // European Conference on Computer Vision,2016:630-645.
- [15] 黄扬. 基于 SEnet 的小样本 MRI 颈椎病变检测方法研究 [D]. 长春:吉林大学,2020.
- [16] Zerhouni E, Lanyi D, Viana M, et al. Wide residual networks for mitosis detection [C] // 14th International Symposium on Biomedical Imaging,2017:924-928.
- [17] Sharma A, Muttou S K. Spatial image steganalysis based on ResNeXt [C] // 18th International Conference on Communication Technology,2018:1213-1216.
- [18] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C] // International Conference on Machine Learning, 2015:448-456.
- [19] Hara K, Saito D, Shouno H. Analysis of function of rectified linear unit used in deep learning [C] // International Joint Conference on Neural Networks,2015:1-8.
- [20] Fukui H, Hirakawa T, Yamashita T, et al. Attention branch network: Learning of attention mechanism for visual explanation [C] // IEEE Conference on Computer Vision and Pattern Recognition,2019:10705-10714.
- [21] 郑顾平,李金华,曹锦纲. 基于注意力机制的单幅雨滴图像增强 [J]. 计算机应用与软件,2020,37(9):160-164.
- [22] Yang K, Qiao P, Li D S, et al. Exploring temporal preservation networks for precise temporal action localization [C] // 32nd AAAI Conference on Artificial Intelligence,2018:7477-7484.
- [23] Gao Y, Liang H L, Liu B D, et al. Action detection based on 3D convolution neural network with channel attention mechanism [C] // IEEE Symposium Series on Computational Intelligence,2020:602-606.
- [24] Chen J B, Mao J, Thiel C, et al. iWaste: Video-based medical waste detection and classification [C] // 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society,2020:5794-5797.

(上接第 62 页)

- [11] 房俊龙,甄景龙,马文川,等. 并联有源滤波器的改进重复控制策略研究 [J]. 电力系统及其自动化学报,2020,32(2):133-139.
- [12] 朱文. 基于改进重复控制的三相四桥臂 APF 研究 [D]. 镇江:江苏大学,2018.
- [13] 赵宇明,何新华,孟亨,等. 基于 ELM 优化 PI + 重复控制策略在 APF 中的应用研究 [J]. 计算机应用与软件,2021,38(3):46-50.
- [14] 王晓宇,王炜. 基于分数阶 PID 与重复控制逆变器优化方法研究 [J]. 计算机仿真,2019,36(3):147-152.
- [15] Lada M Y, Radzi M A M, Jasni J, et al. Performance of three-phase three-wire cascaded H-bridge multilevel inverter-based shunt active power filter [J]. International Journal of Power Electronics and Drive Systems (IJPEDS),2020,11(3):1430-1440.