基于组约束深度神经网络的航运监控事件识别

明道睿[1,2]，张 鸿[1,2]

（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：针对传统的机器学习算法以及常规的深度学习模型对于大数据量的航运监控视频识别效果不佳的问题，设计一种组约束深度神经网络模型（GCDNN）对实时航运监控视频进行识别。模型主要由结合 Inception 结构的 VGG-16 组件和优化 LSTM 单元的深层双向循环神经网络 DBO-LSTM 组件构成，充分提取视频帧序列的时空特征，使用稀疏组套索正则化算法进行网络稀疏处理，使用随机森林算法输出分类结果。实验结果表明，所提模型可以较好提升大数据量下的视频识别准确率，对于受恶劣天气影响的数据具有较强的识别能力，验证了所提模型的有效性。

关键词：航运监控；视频识别；神经网络；稀疏组套索；随机森林

中图法分类号：TP391.4 **文献标识号**：A **文章编号**：1000-7024（2020）10-2949-09

doi：10.16208/j.issn1000-7024.2020.10.041

Shipping hull behavior recognition based on group constrained deep neural network

MING Dao-rui[1,2], ZHANG Hong[1,2]

(1. College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China;
2. Hubei Province Key Laboratory of Intelligent Processing and Real-Time Industrial Systems, Wuhan University of Science and Technology, Wuhan 430065, China)

Abstract: Aiming at the problem that the traditional machine learning algorithm and the conventional deep learning model have poor effects on large-volume shipping surveillance video recognition, a group-constrained deep neural network model (GCDNN) was designed to identify real-time shipping surveillance video. The model was mainly composed of the VGG-16 component combined with the Inception structure and the deep bidirectional cyclic neural network DBO-LSTM component of the optimized LSTM unit. After fully extracting the spatio-temporal features of the video frame sequence, the sparse group lasso regularization algorithm was used for network sparse processing. The classification result was outputted using a random forest algorithm. Experimental results show that the proposed model can improve the video recognition accuracy under large data volume and has strong recognition ability for the data affected by bad weather, which verifies the validity of the proposed model.

Key words: shipping monitoring; video recognition; neural network; sparse group lasso; random forest

0 引 言

航运监控是智能视频监控[1]系统的一个重要应用，即是将船只上安装的摄像头所拍摄到的视频流实时传回到后台控制系统的识别模型当中，经过视频数据的预处理提取出连续的视频帧输入到识别模型当中，从而得到该视频流所对应的事件内容。智能视频监控从根本上解决了人力监控工作所可能导致的疏漏，实现了全天候对监控区域的实时监控，能够对视频内容进行及时准确的分析并且报告异常事件，这极大地提升了视频监控的安全级别的同时也降低了人力成本，从而成为当前的一个极具挑战性的前沿课题[2]。智能视频监控的核心就是能够对视频内容进行准确识别，所设计的模型需要从视频资源中提取数据的时空特征，并且还会因为大风、大雾、摄像头抖动等意外因素影

收稿日期：2019-07-15；**修订日期**：2020-07-20

基金项目：国家自然科学基金项目（61373109）

作者简介：明道睿（1995-），男，湖北十堰人，硕士研究生，研究方向为神经网络、深度学习；张鸿（1979-），女，湖北襄阳人，博士，教授，CCF 专业会员，研究方向为跨媒体检索、数据挖掘、机器学习。E-mail：1135025935@qq.com

响数据质量，这使得智能视频监控成为一项具有相当挑战性的工作。

近几年深度学习[3-5]技术获得了长足发展，计算机视觉的各个领域均开始引入深度学习的方法，尤其对于一些结构复杂、训练量巨大的神经网络模型，有着传统算法无法相比的优势。目前的图像分类、识别以及目标检测等领域已经获得了长足的技术发展，但它们只能用来识别静态性质的图片数据集，相同类别的图片之间也无法挖掘出时空序列的相关性，而近年来热门的 AlexNet[6]、VGG[7]、GoogleNet[8]都无法解决此类问题，它们并不能用于处理视频数据，因为视频帧的预测需要提取视频帧数据的时空特征，二维的卷积神经网络无法做到，视频数据的识别问题因此成为计算机视觉领域多年来的一个技术难题。如今视频识别领域的研究学者们以深度学习思想为核心开始设计出一些契合视频数据特点的新模型，比较有代表性的有三维卷积神经网络[9]以及双通道网络[10]等。本文充分调研了该领域最新技术成果的架构思想，依据视频数据时空特征的特殊性设计一种端到端神经网络模型，在实际的航运视频数据集中得到了理想的效果。

1 相关工作

随着近年来深度学习技术的推广，众多学者们在视频识别的领域提出一系列新的算法或者较之以前有所改进的算法，推动了视频识别技术的高速发展，它们以是否将深度学习作为核心策略被区分开来。

1.1 传统方法

传统的机器学习方法需要经历数据的预处理、特征提取、特征选择等过程，最后使用机器学习领域的相关分类方法对事件内容进行分类得到最终的结果。此类方法对于解决视频问题通常的切入点是检测数据的时空兴趣点(STIP)[11-13]，常用的算法比如 DT[14]算法，它是通过构建视频数据的光流场来获取运动目标的轨迹序列，接着使用 Fisher Vector[15]算法以矢量量化的策略构建视觉词典，将之前提取到的视频数据的 HOF 等 4 种特征进行编码，最后使用 SVM[16]分类器输出最终结果。IDT[17]算法是对 DT 算法的一个优化策略，主要是通过消除相机运动对于算法过程的负面影响来提升算法效率。

1.2 深度学习方法

深度学习的基本思想在于对目标对象能够构建出不同的契合于原理、特征或相关概念的表达层次，层次本身有着较高和较低的区别，较低层次可以作为基础架构推导出较高层次的定义，相对地，较高层次也能够分解为不同的较低层次概念。深度学习最关键的两个方面：①由多层或多个阶段的非线性信息处理组成的模型；②拥有更高更抽象层次表达的特征能够更好地应用于自适应类型的学习方式。由于深度学习能够直接利用原始数据，可以为识别行为提供更高效的特征表达，因而在视频识别领域，基于深度学习的方法也取得了一定的进步，例如基于单帧识别的方法以及构建三维卷积的识别方法。基于单帧识别的方法采用直接对从视频流中截取的单帧图像进行特征学习的策略，但这种方法完全无法利用视频帧的时序信息，因而效率较低。该方法可以做进一步的优化，比如采用间隔取帧的策略，每跳过一定数量的视频帧再进行取帧，最后将所有学习到的图像特征送往全连接层进行特征融合[18]。为了解决二维卷积无法有效提取视频数据的时序信息的问题，提高训练效率，Heskes 将二维卷积扩展到三维，利用高维去表示低维的计算特征，采用第三个维度去表达视频数据的时序特征，有着非常不错的识别能力。

Clark 提出了一种时空双流神经网络（two-stream neural network，TSNN），该模型由两组并列的卷积神经网络构成，两组网络分别以视频数据的光流图片和等间隔抽样视频帧作为输入，提取视频数据的时空特征信息，最后将时序特征和空间特征进行融合。虽然该模型分离了时序特征和空间特征的识别，但特征提取的充分性仍然不够，准确率并未达到理想的程度。

为了解决上述问题，本文设计一种组约束深度神经网络模型（group-constrained deep neural network，GCDNN）进行时空特征的提取，首先是采用与 Inception 模型相结合的 VGG-16 模型提取视频帧序列的空间特征，之后采用优化 LSTM 单元的深层双向循环神经网络（DBO-LSTM）提取视频帧序列的时序特征，以端到端的方式将两个模型连接，并利用稀疏组套索算法（sparse group lasso，SGL）[19]实现网络中变量组级的稀疏化达到网络的修剪效果，充分训练之后使用随机森林算法[20]实现分类输出。本文的 GCDNN 模型基于时空特征的特殊性采取对应的网络结构进行提取，取得了对比模型中最高的识别精度，验证了所提模型的有效性。

2 航运监控事件识别

本文基于实际的航运监控项目撰写，是针对航运过程中船舶的各类事件进行识别，基于项目的推进不断地优化模型。由于船舶体积较为庞大，摄像头所拍摄的短时间内的连续视频帧无法提取到有效的反映船舶运动轨迹的时序信息，因而本文将提取间隔设置为 5 min，以连续的 6 帧视频帧作为模型输入数据，实现监控事件的识别过程。

2.1 模型整体架构设计

本文所设计的 GCDNN 模型包括以端到端的方式连接的处理视频数据空间类型特征的经过改进的 VGG-16、处理视频数据时序特征的优化 LSTM 单元的深层双向循环神经网络（DBO-LSTM），以及最后的网络修剪层和输出层。模型的整体架构如图 1 所示。

模型的输入视频帧序列〈T_1，T_2，T_3，…，T_k〉经过

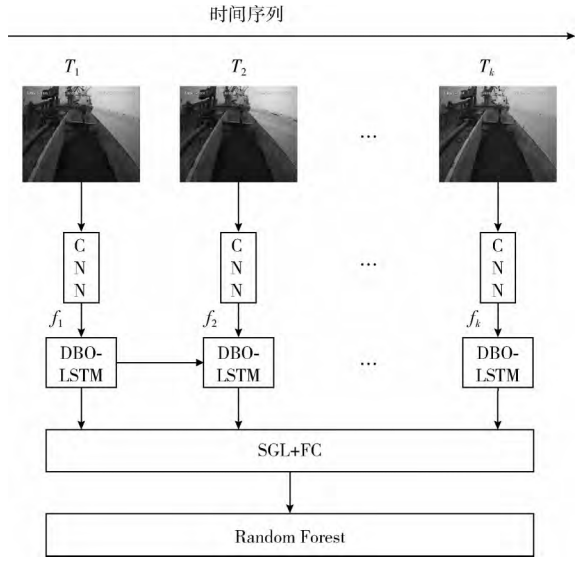

图 1　GCDNN 整体架构

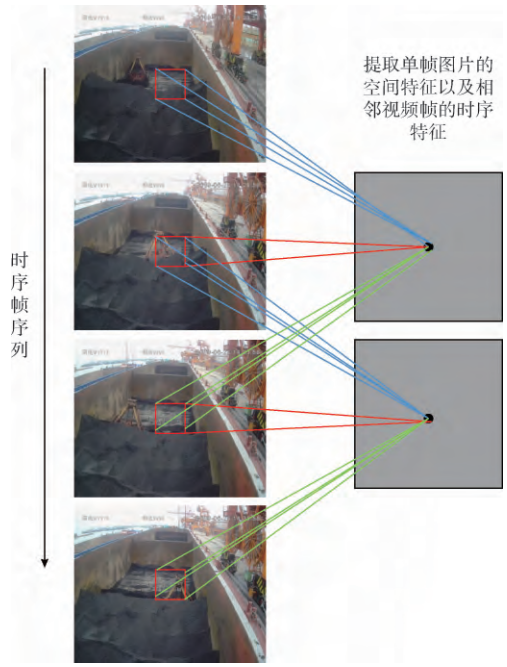

图 2　GCDNN 的时空特征提取建模

GCDNN 的卷积神经网络部分提取每一帧的空间特征〈f_1，f_2，f_3，…，f_k〉依次输入到 DBO-LSTM 中，然后将 DBO-LSTM 中各个时刻的输出值依次输入到具有稀疏组套索（SGL）正则化的全连接层中实现网络的修剪，最后通过随机森林算法输出分类结果。

2.2　拓展的 VGG-16 提取视频帧的空间特征

卷积神经网络对于提取数据的空间特征有着最好的效果，因此本文的 GCDNN 模型将其作为第一部分，结合第二部分的 DBO-LSTM 共同提取视频帧数据的时空特征。图 2展示了航运中船舶卸货过程的特征提取过程。

VGG 模型为了加深增宽模型的架构，并减轻计算压力，采用小规格的池化核和卷积核，分别用以减小模型的宽高尺度以及增加网络的通道数量，但因此限制了网络的整体深度。为了提高网络的特征提取能力，同时避免网络结构过于复杂，提升训练速度，本文将 InceptionV4 中的第 3 个模块 Inception-C 加入到 VGG-16 的卷积层之后。Inception 是 Google 在 2014 年的 ILSVRC 比赛中取得优异表现的 CNN 模型，从 InceptionV1 发展到 InceptionV4。Inception 模块是做一个卷积分解，采用两个结构简单的一维卷积构成原有的多维卷积，例如一个 5 * 5 的卷积块可以由 5 * 1 和 1 * 5 两部分构成。这种拆分方式可以通过拓展模型非线性的表达能力来降低过拟合现象的发生几率，并且可以较大程度地减少网络参数，提升模型效率。因此这种非对称的卷积拆分策略可以取得更加优秀的性能。但为了避免网络结构过深出现梯度弥散而导致模型性能下降，本文将 V4 模型中表现最佳的一个模块 Inception-C 添加到 VGG 模型中。

本文所拓展的 VGG 模型包括 5 个卷积层（convolution）、5 个最大池化层（max pooling）、ELU（exponential linear unit）激活层、LRN（local response normalization）以及 Inception-C 等结构，由于高维特征在经过池化操作之后会产生特征的分离，原有的时序类信息将会丢失而只留存下数据的空间特征，因此模型去掉最后的全连接层以及分类层以避免这种现象的发生，卷积层的输出将作为 DBO-LSTM 的输入。前两层卷积层均连续进行 2 次卷积，后三层则均连续进行 3 次卷积，卷积层之后都会跟上相应的最大池化层。网络结构如图 3 所示。

模型的输入是VGG的标准输入224×224，模型中每

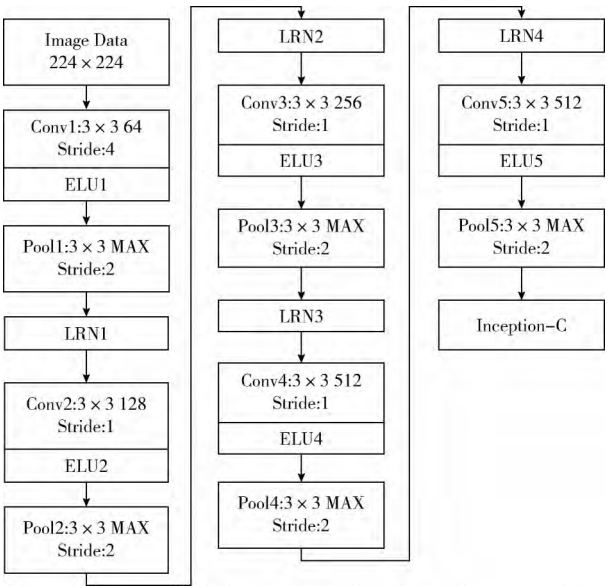

图 3　拓展的 VGG-16 网络模型

个卷积块均进行 3×3 的卷积操作，前 3 个卷积层的卷积核分别设置为 64、128、256 个，后两个均设置为 512 个，顶层的移动步长设为 4 个像素，后 4 个卷积层的移动步长均设置为 1 个像素，模型的激活层采用 ELU（指数线性单元）激活函数，该函数左右两侧具备不同的饱和性，左侧的软饱和特性可以让函数的抗干扰能力得到较大提高，并且可以促使经过函数之后的输出均值接近零，因而具有更快的收敛速度。在前 4 个池化层的每一层后面均增加局部响应归一化（local response normalization，LRN）[21]层，提升模型的泛化能力，最后的 Inception-C 模块则用于增加网络深度和宽度，提升网络的性能，卷积的结果将输入到 DBO-LSTM 中进行下一部分的训练。

2.3 DBO-LSTM 提取视频帧的时序特征

序列数据通常是指沿着时间轨迹所提取的数据，而用以操作序列数据的模型则是 RNN（recurrent neural network），此类数据反映了某些现象、事物等随时间的变化程度或状态，RNN 的网络架构通过模块的循环完成信息从上到下的逐层传输，信息流从输入单元依次流向隐藏单元以及输出单元，网络模块的隐含层每个时刻的输出都依赖于以往时刻的信息，在提取数据时序特征方面，RNN 有着比卷积神经网络更好的效果，因此本文设计一种优化 LSTM 单元的深层双向 RNN 来提取数据的时序特征。

本文的 DBO-LSTM 是一个深层双向的 RNN，将每一步的输出与后续的序列联系起来，每个序列向前和向后呈现到两个单独的 LSTM，最后的输出结果将为正向反向过程输出的串联向量。由于实际航运视频监控项目的数据量较大，需要网络有更强大的表达与学习能力，因此 DBO-LSTM 将每一个 RNN 的隐含层增加至三层，加深网络结构，使其能够在高层更抽象地表达特征，较好提升网络性能。DBO-LSTM 的网络结构如图 4 所示。

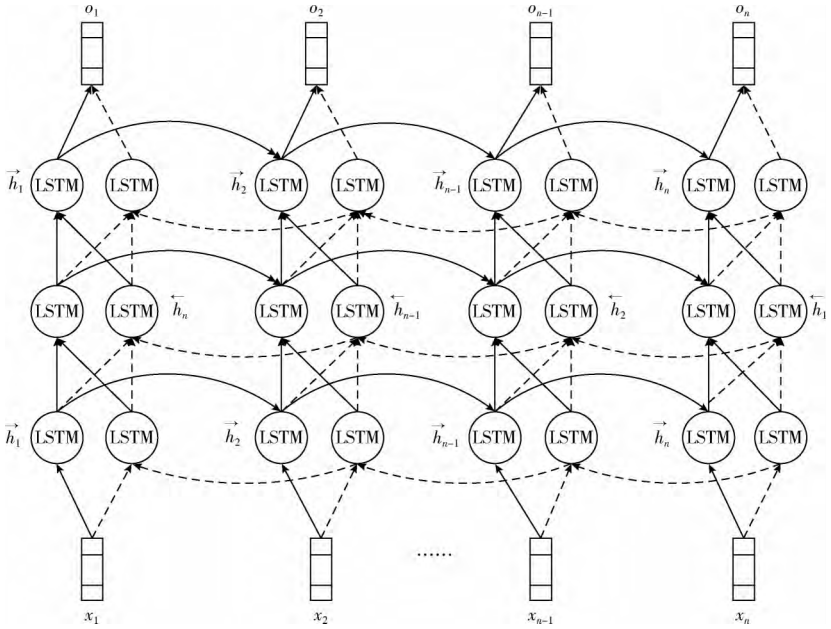

图 4 DBO-LSTM 网络结构

图 4 中，输入集标记为 $\{x_1, x_2, \cdots, x_{n-1}, x_n\}$，输出集标记为 $\{o_1, o_2, \cdots, o_{n-1}, o_n\}$，$\overrightarrow{h}_t$ 为前向传播状态信息序列，$\overleftarrow{h}_t$ 为后向传播状态信息序列，完整的特征信息将由前后状态信息序列融合得到。

LSTM 是 RNN 模型的一种特殊结构形式，对于具有长期依赖特征的问题具备较好的解决能力，针对连续时间轨迹点之间的延迟和间隔都较长的事件类型，LSTM 使用软门这一独特结构来处理，软门本身通过激活 RNN 来建模，共分为输入门、忘记门、输出门 3 种，软门通常被用来调整模型中的细胞状态，通过其相应的门激活与相应动作有关的激活之间的乘积来进行状态的调节。

本文对 LSTM 单元做出优化，将 LSTM 单元的忘记门和输入门进行耦合。模型在任一时刻状态中需要丢弃的信息由忘记门决定，会读取 h_{t-1} 和 x_t，h_{t-1} 表示上一个细胞的输出，x_t 表示的是当前细胞的输入，忘记门通过输出一个 0 到 1 之间的数值给每个在细胞状态 C_{t-1} 中的数字来决定信息的取舍，其中，1 和 0 分别代表着保留数据和舍弃数据的含义，具体输入到细胞状态中的信息数量由输入门决定，常规的单元结构是分开确定丢弃和添加信息的时刻，将两种软门耦合之后，新单元将会统一决定信息的流通，优化后的 LSTM 单元仅会当将要输入在当前位置时忘记，实时更新细胞中信息的存在状态，对于时序类信息的处理会更加高效。优化后的 LSTM 单元结构如图 5 所示。

DBO-LSTM 构建了前向传播和后向传播两个过程，最

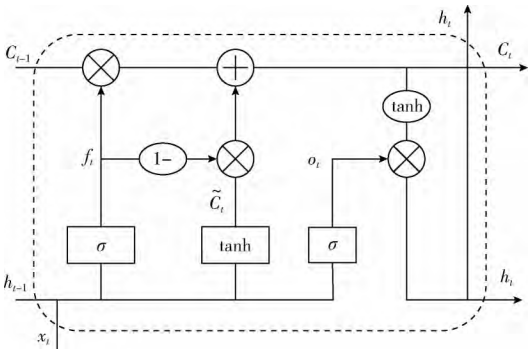

图 5　优化的 LSTM 单元结构

终得到的前向传播输出序列 $\vec{h}_t$ 的运算过程如下

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \sigma(W_C[h_{t-1}, x_t] + b_C) \tag{3}$$

$$\vec{C}_t = f_t * \vec{C}_{t-1} + (1 - f_t) * \tilde{C}_t \tag{4}$$

$$o_t = \sigma(W_o[\vec{h}_{t-1}, x_t] + b_o) \tag{5}$$

$$\vec{h}_t = o_t * \tanh(\vec{C}_t) \tag{6}$$

其中，$\vec{h}_{t-1}$ 属于前一个前向 LSTM 单元输出状态信息，σ 为 sigmod 函数，W 是权值，b 是偏置。采用忘记门计算前一个前向单元所传递的细胞记忆信息 $\vec{C}_{t-1}$ 与细胞丢弃的信息的概率 f_t，在输入门中 sigmod 决定那些需要更新的信息值 i_t，使用 tanh 所需要添加的记忆信息 $\tilde{C}_t$ 到更新的记忆信息 $\vec{C}_t$，最后在细胞的输出门中确定 $\vec{h}_{t-1}$ 所要输出的一部分状态信息 o_t 与 tanh（$\vec{C}_t$）得到 LSTM 最后所要求的前向传播输出序列 $\vec{h}_t$。模型后向传播流程中，序列信息从后往前传递，数据计算过程与前向保持一致，因此可以得出后向传播信息 $\overleftarrow{h}_t$，最后融合前向传播特征信息 $\vec{h}_t$ 与后向传播特征信息 $\overleftarrow{h}_t$ 而获得完整的特征信息 h_t，它的计算式为

$$h_t = [\vec{h}_t + \overleftarrow{h}_t] \tag{7}$$

2.4　网络的修剪和分类输出

在处理高维数据时，需要构建出输入数据当中拥有最多有效信息的最小子集，以此完成特征选取这一关键步骤。而网络中不断输入的特征数据通常需要增加隐层神经元数量支撑，这就需要进行网络的修剪，在这方面目前常见的是 l_1 和 l_2 正则化，但在应用到深度神经网络中时，移除神经元必须当它的所有输入和输出权值一致为零时才可以办到，因而网络的修剪效果并不太理想。本文利用稀疏组套索（SGL）正则化算法，将一个神经元的所有输出权值视为一组，在组套索项的约束下同组的变量同时为零或同时不为零，进而可以将它们同时移除，不同位置的神经元被移除将会有着不同的作用：输入层神经元被移除相当于一次特征选择过程，隐层神经元被移除可以起到简化网络的作用。

SGL 的核心策略是采取组级稀疏的思想将每一个神经元的全部输出交互统一置为零或非零，详细来说就是，SGL 中不同的组级稀疏效果分别由 3 个不同的变量组实现：

（1）输入组 G_{in}：组中一个元素 $g_i \in G_{in}$，$i = 1, \cdots, d$，它是网络中第 i 个输入神经元所有输出连接构成的向量，即矩阵 W_1 转置的第一行；

（2）隐层组 G_h：组中一个元素 $g \in G_h$ 是隐层中任一神经元所有输出连接构成的向量，即矩阵 W_k 转置中的一行（$k > 1$）。对应于网络内部直到输出层的神经元，一共有 $\sum_{K=2}^{H+1} N_k$ 个组，H 为隐层的层数，N 为训练样本的总数；

（3）偏移组 G_b：这是一个对应网络中偏移的一维组，共有 $\sum_{K=1}^{H+1} N_k$ 个组。每组偏移对应于向量 $\{b_1, \cdots, b_{H+1}\}$ 中的一个元素。

综上，一共有 $G = \sum_{K=1}^{H+1} N_k$ 个变量组，分别对应 SGL 中的 3 种不同稀疏效果。如果需要完成对输入数据的特征选择，忽略其预测阶段的无效特征，则可以将输入组变量统一置为零。其次，如果需要精简网络结构，有效减少所在网络模型中的神经元数量，则可以将隐层变量统一置为零。最后，如果将偏移组变量统一置为零的话，可以为相应的神经元移除偏移，以一种权值分组的策略完成正则化过程。

定义所有的组为

$$G = G_{in} \cup G_h \cup G_b \tag{8}$$

组稀疏正则化可被写为

$$R_{l_{2,1}}(w) \triangleq \sum\nolimits_{g \in G} \sqrt{|g|}\, \|g\|_2 \tag{9}$$

式中：$|g|$ 为向量 g 的维度，它确保每个组都得到统一的加权，操作符 $\|\ \|_p$ 为欧几里得标准空间 l_p 范数，当 $p = 2$ 为欧几里得范数，$p = 1$ 为曼哈顿范数。式（9）存在一定的不足，在删除了一些组之后，就会失去对剩下连接的稀疏性的保障，因此引入“稀疏组套索”惩罚

$$R_{SGL}(w) \triangleq R_{l_{2,1}}(w) + R_{l_1}(w) \tag{10}$$

利用此 SGL 公式的任意一项就能获得最优解，达到最佳的修剪效果可以使 GCDNN 获得最高的组稀疏性，并得到一个非常紧凑和高效的神经网络。

网络的最后输出层，采用随机森林（random forest）算法完成输出过程。随机森林算法由 LeoBreiman 提出，利用随机策略构建出一个由众多决策树[22]组成的森林结构，决策树（decision tree）也是树结构，它的叶子结点和非叶子节点分别存储目标数据的类别属性和相应属性测试的结果，而对于该属性在所属范围内的输出存储则是由决策树的分支来完成。决策树之间不会有任何相关性，每当森林中进来了一个输入样本，所有的决策树都会对其类别进行判断，每一个判断过程都是独立的。最终被选择最多的类别，就被预测为样本的分类结果。本文采用随机森林算法

的原因在于，RF 在处理高维数据方面有着优秀的表现，它不用做特征选择，对数据集的适应能力很强，并且由于随机性的引入，使用 RF 作为分类器不容易发生过拟合现象，抗噪声能力优秀。

3　实验与分析

3.1　数据集

本文实验所采用的数据集是在整个项目进程中所积累下来的数据，船只上安装的摄像头将所拍摄的监控视频上传到服务器，对视频进行截帧处理后完成数据集的分类下载。视频数据本身是全天时段的，因此需要从中删除掉夜晚时段的数据。由于监控视频的拍摄可能会受到各种干扰因素的影响，比如摄像头晃动、硬件设施故障等因素，所以在数据集的筛选过程当中，需要去除掉那些质量不佳的图像，对数据进行统一的归类收集，其中对于受到比如大雾天气影响的模糊数据则保存下来并统一收集，用于进行相关的测试。

截止到本文撰写阶段，所采集的数据包含 25 500 个视频片段，截帧处理后相当于 153 000 张图像，一共分为 8 个事件类别，分别为装船、卸船、空仓、正常行驶、雨布吹飞、停泊、未盖布行驶、摄像头遮挡，按照 9∶1的比例建立训练集与测试集。图 6 为数据集事件样例展示。

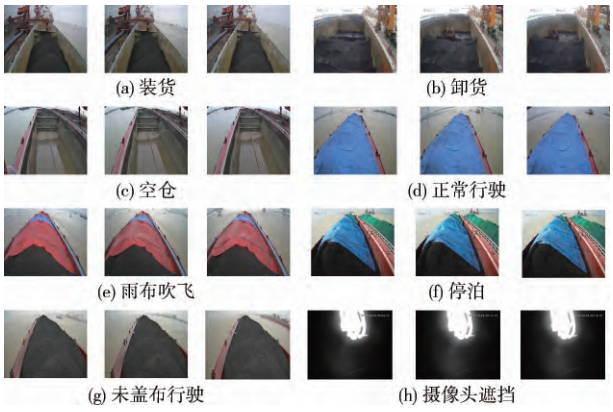

图 6　数据集样例展示

数据集中的 8 个类别是应项目需求所划分，需要模型实时准确地给出每一个事件的识别结果并由识别系统的前台将事件类型划分，对于其中的异常事件将发出预警交由管理人员给出解决方案。其中雨布吹飞、空仓、未盖布行驶等事件容易互相识别错误，它们的识别准确率彼此之间也有着不小的差距，原因在于不同事件中时间轨迹特征的区分难度不同，这些事件当中，雨布吹飞和未盖布行驶属于识别事件当中的异常事件，因此非常需要模型拥有更高更全面的特征提取能力，根据视频帧的背景移动状况以及船面的空间特征变化来准确地做出识别。

3.2　模型训练与实验结果分析

本文实验的硬件平台为：Intel（R）Core（TM）i7-8700k CPU，NVIDIA GeForce GTX 1070GPU，在此平台上将 GCDNN 模型与其它用于视频识别的主流模型作相关对比。首先是双流神经网络模型（TSNN），该模型输入数据的标准格式为 256 * 256，由于模型是基于双通道识别思想的，所以需要先提取数据的空间类型特征，之后利用另一个相同规格的卷积神经网络提取时序特征，需要以光流信息作为输入，所以首先用 OpenCV 获取每两帧之间提取的所有点的光流信息输入到第二个卷积神经网络中，得出识别结果，最后使用 SVM 融合两者的结果得到最终结果。实验中的学习率设置为 0.001，丢失率设为 0.9，batch _ size 设为 32，训练 10 000 次。之后本文将 TSNN 的卷积神经网络结构更换成更深的 VGG-11，训练次数设置为 10 000 次进行相同的对比实验。

对于 3D 卷积神经网络，由于基础的网络结构较为简单识别率不甚理想，难以达到实际需求的标准，因此本文首先对基于 VGG-11 网络结构的三维卷积模型进行训练，由于输入数据的格式不匹配，所以需要将输入的连续视频帧数据缩放为 224 * 224 以契合 VGG-11 的输入数据标准格式[23]，实验中的学习率设置为 0.0001，丢失率设为 0.8，batch _ size 设为 8，训练 10 000 次，接着将 VGG-11 网络替换为 VGG-16 网络，训练次数设置为 12 000 次。实验部分将统一展示每一个对比模型在 8 个事件类别上的识别准确率，并且添加了平均准确率作为参考，并添加了每个类别单张图片的识别速度作为比对参考。

对于本文的 GCDNN 模型，实验将会对两部分网络结构的优化效果进行测试，分别将卷积神经网络部分的 VGG-16 拓展前后的识别能力做出对比，以及循环神经网络部分，采用 LSTM 与优化后的 DBO-LSTM 做出对比，丢失率设置为 0.8，batch _ size 为 8，学习率设置为 0.0001，训练次数为 10 000 次。而对于网络的修剪以及分类输出性能，实验将本文所用稀疏组套索正则化（SGL）与 L1 和 L2 正则化之间的训练效果进行对比，设置相同的正则项参数，利用 softmax 和 RF 进行分类输出，并对于不同的正则化算法和分类算法的组合进行了对比。表 1 为各模型在各事件上的测试结果汇总。

从表 1 中可以看出，采用 TSNN 模型在应对大数据量的识别任务时效果并不理想，识别的平均准确率在所有对比模型中最低，并且识别速度也相对较慢，将 TSNN 的卷积模型更换成 VGG-11 之后，各个类别的识别准确率均有了明显的提高，平均准确率从 0.839 提高到 0.875，提高了 3.6%，但仍未达到实际项目需求的标准，且处理速度更慢，因此需要寻找更有效的模型。

3D 卷积神经网络在测试中拥有着较好的表现，基于 VGG-11的3D卷积模型在每个类别的识别准确率上均有了

表 1　不同模型在不同类别数据集上的测试结果

模型名称	正常行驶	雨布吹飞	摄像头遮挡	未盖布行驶	停泊	空仓	装货	卸货	平均准确率	处理速度/s
TSNN	0.876	0.844	0.906	0.877	0.858	0.814	0.765	0.773	0.839	0.832
TSNN(VGG-11)	0.890	0.867	0.926	0.887	0.882	0.836	0.774	0.785	0.875	0.851
3DCNN(VGG-11)	0.914	0.877	0.962	0.931	0.894	0.877	0.844	0.832	0.891	0.411
3DCNN(VGG-16)	0.942	0.884	0.974	0.965	0.944	0.874	0.863	0.859	0.913	0.448
GCDNN(VGG-16)	0.947	0.892	0.977	0.965	0.952	0.893	0.868	0.865	0.920	1.126
GCDNN(LSTM)	0.932	0.883	0.971	0.964	0.946	0.873	0.862	0.858	0.911	1.391
GCDNN	0.974	0.935	0.996	0.971	0.972	0.920	0.878	0.876	0.940	1.472

较大幅度的提升，平均识别准确率较之 TSNN 提高了 5.2%，比 TSNN（VGG-11）提高了 1.6%，初步达到了项目需求的准确率水平，且处理速度只有很短的 0.411 s，主要是因为 3D 卷积神经网络模型的复杂度大大减小，且能较好地在卷积操作之后保存时序特征，因而整个网络的性能得到了较大提升。在将 VGG-11 更换为 VGG-16 之后，3D 卷积神经网络的识别准确率得到了进一步提升，从 0.891 提升到了 0.913，且处理速度只增加了 0.037 s，取得了不错的效果。

表 1 的数据显示，未改动 VGG 的 GCDNN 识别准确率已经达到 0.920，相比 VGG-16 的 3D 卷积神经网络提高了 0.7%，但处理速度由于模型的复杂性降低了 0.678 s。而循环神经网络部分则是将常规的 LSTM 网络改进为 DBO-LSTM，网络的性能得到了较大提高，平均识别准确率从 0.911 提高到 0.940，取得了对比模型中最好的效果，满足了项目的实际业务需求，但处理速度不可避免地降低到 1.472 s，相当于以识别速度的牺牲换取更高的识别精度，在实际业务中，可以根据不同的需求来对模型做出调整，在识别速度和识别精度两方面做出权衡。

本文对于常用的 L1 和 L2 正则化进行了测试实验，并与本文的 SGL 进行了比对，图 7 展示了 3 种稀疏算法的稀疏性对比，横轴为正则项参数，纵轴为稀疏度，稀疏度通过零权重相对于连接总数的百分比来计算。可以看出 L2 正则化的效果非常的差，无法满足复杂网络的稀疏要求，而 L1 和 SGL 则达到了不错的效果，SGL 更优。

表 2 展示了对于 GCDNN 模型，3 种正则化算法和分类算法即 softmax 与 RF 的识别效率对比，不同的正则化算法和分类算法的组合有着显著的网络修剪性能差异。

表 2　修剪输出能力对比

模型名称	平均准确率	处理速度/s
L1+Softmax	0.927	1.506
L2+Softmax	0.914	1.594
SGL+ Softmax	0.935	1.477
SGL+RF	0.940	1.472

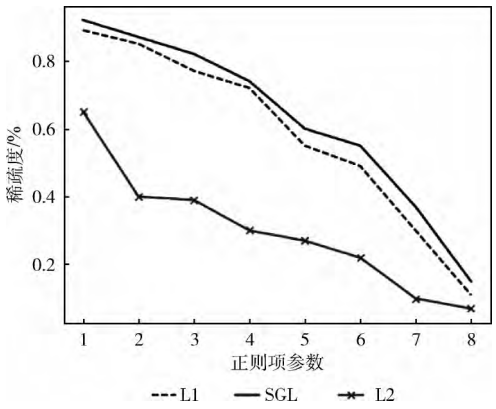

图 7　稀疏性比较

从表 2 中可以看出，SGL 在 3 种算法中有着最优的输出性能，以 softmax 为分类器时，平均准确率已经达到 0.935，相比 L1 和 L2 分别提升了 1.3%和 2.1%，处理速度也较快，而与 RF 的组合则是达到了最优结果。

本文在进行模型训练以及对比实验时，去除掉了一些严重受到天气影响的数据，是为了最大程度保证训练效果，但在实际应用中，除开硬件设备出现问题导致的图像质量差之外，天气原因也是一个不可避免的因素，但它具有一定的研究价值，本文实验收集了一部分受到大雾天气影响导致的停泊事件类别图像模糊数据集，样例如图 8 所示。

图 8　停泊事件模糊数据样例

模糊数据更考验模型对于图像特征的提取能力，本文实验将 GCDNN 与其它实验相关模型在此事件数据集上进行了比对实验，结果见表 3。

项目的实际运作中，遇到特殊天气时会增加人力因素的投入，以保证航行的顺利进行，但仍需要智能识别系统的数据作为参考，实验部分的对比模型在特殊数据集下识

表 3 模糊数据集上的模型识别结果

模型名称	识别准确率	处理速度/s
TSNN	0.692	0.837
TSNN(VGG-11)	0.714	0.862
3DCNN(VGG-11)	0.726	0.411
3DCNN(VGG-16)	0.747	0.450
GCDNN(VGG-16)	0.753	1.128
GCDNN(LSTM)	0.751	1.392
GCDNN	0.795	1.479

别率均有了较大程度的下滑，但本文的 GCDNN 模型仍取得了最高的识别率，达到了 0.795，且处理速度未有较大变化，可以作为航运监控数据的有效参考。

4 结束语

本文针对现有深度学习模型面对大数据量的航运监控视频数据时识别准确率不高的问题，提出一种组约束深度神经网络模型，利用拓展的 VGG-16 模型和 DBO-LSTM 分别提取视频帧数据的空间特征和时序特征，最后利用稀疏组套索算法进行网络修剪并采用随机森林算法实现分类输出。在实际的航运监控数据集上的对比实验结果表明，本文所提模型取得了最佳的识别准确率，验证了模型的有效性。但在识别速度上，由于模型的复杂性，识别速度较低，有较大的提升空间，并且模型本身的网络结构也有得到更佳优化的可能性。视频识别是深度学习的一个前端领域，有着广阔的应用前景，未来将在现有工作的基础上，进一步调整网络结构，结合领域前沿的网络模型取得更强的识别能力。

参考文献：

[1] Huang W，Ding H，Chen G. A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance [J]. Signal Processing，2018，142 (6)：104-113.

[2] Wang H，Oneata D，Verbeek J，et al. A robust and efficient video representation for action recognition [J]. International Journal of Computer Vision，2016，119 (3)：219-238.

[3] Rui Z，Yan R，Chen Z，et al. Deep learning and its applications to based machine health monitoring [J]. Mechanical Systems & Signal Processing，2019，115 (2)：213-237.

[4] Zhang Q，Yang L T，Chen Z，et al. A survey on deep learning for big data [J]. Information Fusion，2018，42 (2)：146-157.

[5] Sun X，Wu P，Hoi S C H. Face detection using deep learning：An improved Faster RCNN approach [J]. Neurocomputing，2018，299：42-50.

[6] Liu Z，Dou Y，Jiang J，et al. Throughput-optimized FPGA accelerator action for deep convolutional neural networks [J]. ACM Transactions on Reconfigurable Technology & Systems，2017，10 (3)：17-28.

[7] Day M J，Horzinek M C，Schultz R D. Guidelines for the vaccination of dogs and cats. Compiled by the vaccination guidelines group (VGG) of the world small animal veterinary association (WSAVA) [J]. Journal of Small Animal Practice，2017，48 (9)：528-541.

[8] Ouyang W，Zeng X，Wang K，et al. DeepID-Net：Object detection with deformable part based convolutional neural networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2017，39 (7)：1320-1334.

[9] Ghafoorian M，Karssemeijer N，Heskes T，et al. Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin [J]. Neuroimage Clinical，2017，14 (C)：391-399.

[10] Muschietti L，Lembge Bertrand. Two-stream instabilities from the lower-hybrid frequency to the electron cyclotron frequency：Application to the front of quasi-perpendicular shocks [J]. Annales Geophysicae，2017，35 (5)：1093-1112.

[11] Gao S，Janowicz K，Couclelis H. Extracting urban functional regions from points of interest and human activities on location-based social networks [J]. Transactions in Gis，2017，21 (3)：446-467.

[12] Lim K H，Chan J，Leckie C，et al. Personalized trip recommendation for tourists based on user interests，points of interest visit durations and visit recency [J]. Knowledge and Information Systems，2018，54 (2)：375-406.

[13] DasDawn D，Shaikh Soharab Hossain. A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector [J]. Visual Computer，2016，32 (3)：289-306.

[14] Jeong W C，Sajib S Z，Katoch N，et al. Anisotropic conductivity tensor imaging of in vivo canine brain using DT-MREIT [J]. IEEE Trans Med Imaging，2017，36 (1)：124-131.

[15] Boulkenafet Z，Komulainen J，Hadid A. Face antispoofing using speeded-up robust features and fisher vector encoding [J]. IEEE Signal Processing Letters，2017，24 (2)：141-145.

[16] Aburomman A A，Reaz M B I. A novel SVM-kNN-PSO ensemble method for intrusion detection system [J]. Applied Soft Computing，2016，38 (C)：360-372.

[17] Niazmardi S，Safari Abdolreza，Homayouni Saeid. Similarity-based multiple kernel learning algorithms for classification of remotely sensed images [J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing，2017，10 (5)：2012-2021.

[18] Barata M，Bernardino J，Furtado P. An overview of decision

support benchmarks：TPC-DS，TPC-H and SSB［J］. Advances in Intelligent Systems & Computing，2015，353（4）：619-628.

［19］Mashayekhi M，Gras R. Rule extraction from decision trees ensembles：New algorithms based on heuristic search and sparse group lasso methods［J］. International Journal of Information Technology & Decision Making，2017，16（6）：21-32.

［20］Eva S，Bentzen P，Bradbury I R，et al. Applications of random forest feature selection for fine-scale genetic population assignment［J］. Evolutionary Applications，2018，11（2）：153-165.

［21］Klamka J，Maurer H，Swierniak A. Local controllability and optimal control for a model of combined anticancer therapy with control delays［J］. Mathematical Biosciences & Engineering，2017，14（1）：195-212.

［22］Tayefi M，Tajfard M，Saffar S，et al. hs-CRP is strongly associated with coronary heart disease（CHD）：A data mining approach using decision tree algorithm［J］. Computer Methods and Programs in Biomedicine，2017，141：105-109.

［23］WANG Zhongjie，ZHANG Hong. Shipping monitoring event recognition based on 3D convolutional neural network［J］. Journal of Computer Applications，2019，39（12）：3697-3702（in Chinese）.［王中杰，张鸿. 基于三维卷积神经网络的航运监控事件识别［J］. 计算机应用，2019，39（12）：3697-3702.］