文章编号:1671-4598(2025)02-0161-07 DOI:10.16526/j.cnki.11-4762/tp.2025.02.021 中图分类号:TP391 文献标识码:A

基于改进 C3D 模型的料仓视频分类识别方法

曹庆园，朱建鸿
（江南大学 轻工过程先进控制教育部重点实验室，江苏 无锡 214122）

摘要：在自动上料控制系统中，针对传统电感式传感器容易受到外界复杂环境干扰，且需要进行繁琐校准工作等问题，提出了一种基于改进 C3D 模型的料仓视频视觉分类识别方法；基于实验需求，设计了合作标靶和建立了料仓识别视频数据集；将初始 C3D 模型作为主干网络进行改进，将该模型第 3、4、5 层卷积层进行精简，使得模型参数量大幅降低，有利于加快推理速度；在轻量化后的 C3D 模型上融合 SE 注意力机制，C3D 模型从时空两个维度中提取特征，SE 注意力机制可以有效在复杂场景视频帧中找出标靶显著区域，在兼顾时序信息的同时能够高效提取特征进而提高识别准确率；实验结果表明，SE-C3D 识别模型准确率达到 99.61%，与初始 C3D 模型相比，准确率提高 2.48%，与其他典型三维卷积模型对比，各项性能指标也均有明显提升，对未来智能化上料系统的发展具有重要意义。

关键词：上料系统；3D 卷积神经网络；视频分类；SE 注意力；模型轻量化

Video Recognition Method for Silo Target Based on Improved C3D Model

CAO Qingyuan, ZHU Jianhong
(Key Laboratory of Advanced Process Control for Light Industry, Jiangnan University, Wuxi 214122, China)

Abstract: In automatic feeding control systems, in order to solve the problems that traditional inductive sensors are easy to be disturbed by external complex environments and need to be frequently calibrated, a visual classification and recognition method for silo video based on improved C3D model is proposed. Based on experimental requirements, a cooperative target is designed, and a video dataset for silo identification is established. The initial C3D model is improved as the backbone network, and the 3rd, 4th, and 5th convolutional layers of the model are simplified, which greatly reduces the model parameters, and it is beneficial for speeding up the inference speed. A SE attention mechanism can effectively be fused on the lightweight C3D model, the features of the C3D model are extracted from the dimensions of time and space. The SE attention mechanism can efficiently find out the area of the target in the complex scene video frames, while taking into account the time series information to improve the recognition accuracy. Experimental results show that the accuracy of the SE-C3D recognition model reaches 99.61%, which is 2.48% higher than that of the initial C3D model. Compared with other typical 3D convolution models, this model also significantly improves the performance indicators, which is of great significance for the development of intelligent feeding systems in the future.

Keywords: feeding system; 3D convolutional neural network; video classification; SE attention mechanism; model lightweighting

0 引言

工业制造 4.0 正成为世界工业发展的趋势，智能生产会成为现代工业的常规配置[1]。上料控制系统是部分生产线的首要环节，也是制造业生产线进行智能化、自动化的第一个落脚点，快速、准确上料是保证工业生产过程环节正常运作的基础[2]，自动化控制系统在工业制造 4.0 中扮演着至关重要的角色，它是实现制造业智能化、灵活化和高效化的关键技术之一。混凝土搅拌站在工业制造中具有重要的地位和作用，对基础设施建设、工业生产、环境保护等方面都起到关键性的支持作用。

随着深度学习技术的快速发展，国内外研究人员提出多种用于视频识别领域的卷积神经网络，其帮助制造业向自动化、智能化的生产方式转变[3]。由工业相机拍

收稿日期:2023-11-17； **修回日期**:2024-03-11。
基金项目:国家自然科学基金项目(61973139)。
作者简介:曹庆园(1998-),男,硕士研究生。
通讯作者:朱建鸿(1964-),男,大学本科,副教授,硕士生导师。
引用格式:曹庆园,朱建鸿.基于改进 C3D 模型的料仓视频分类识别方法[J].计算机测量与控制,2025,33(2):161-167,183.

摄的视频由大量的视频帧图像所构成。相比于单一图像，视频中增加了时间维度的信息，物体在先后帧中出现的顺序、状态等信息都非常关键。所以在视频任务中，在兼顾时序信息的同时高效提取特征是至关重要的问题[4]。目前有许多视频分类深度网络取得了较好的效果，在二维卷积领域，张传雷等人[5]针对视频分析中时间域和空间域上特征提取存在的易混乱以及运算成本高的问题，提出了 ResNetLSTM-Attention 网络模型结构，该网络结构将空间域和时间域的特征分开提取，该模型不仅有较高的识别准确率，而且训练用时和运算成本也大大降低。Jiang 等人[6]提出一种在用于视频分类的混合深度学习框架中建模多模态线索，提高视频识别的准确率。Wang 等人[7]提出了利用具有空间注意机制的卷积神经网络（CNN，convolutional neural networks）和具有时间注意机制的深度双向长短时记忆网络对视频数据进行处理，提出了一种新颖的视频分类方法。

在三维卷积领域，2015 年 Tran 等人[8]提出了 C3D 网络（Convolutional 3D），C3D 网络使用 3D 卷积和 3D 池化直接处理视频，3D 卷积核可以同时在时间和空间双维度进行操作，可以有效处理视频序列中的时间信息，最终在 UCF101 数据集上达到 52.8% 的准确率。王粉花等人[9]针对人机交互中对动态手势识别准确率的要求，提出了一种融合双流三维卷积神经网络（I3D）和注意力机制（CBAM）的动态手势识别方法 CBAM-I3D。吴晓雨等人[10]提出了一种基于音视频多模态特征融合与多任务学习的特种视频识别方法，提取视频信息随时空变化特征，构建具有语义保持的共享特征子空间，以实现音视频多种模态特征的融合，实现端到端的特种视频智能识别。张瑗涵等人[11]提出了融合深度学习技术的双流程短视频分类方法，通过滑动窗口机制与级联分类器融合的方式对其进行分类检测，进一步提高短视频分类准确性。冯宇等人[12]提出了一种基于改进型 C3D 的注意力残差网络模型，针对原始 C3D 卷积神经网络的层数较少、参数量较大和难以关注关键帧而导致的人体行为识别准确率较低的问题，通过采用卷积核合并与拆分操作，采用全预激活式残差网络结构来增加构建的非对称卷积层，并且在残差块中增加时空通道注意力模块，得到实验结果优于其他流行算法。Qiao 等人[13]针对高度相似动作视频难以识别的问题，提出了一种深度学习框架来监测和分类奶牛行为，并将其与 C3D 网络和卷积长短时记忆网络智能结合，使用 C3D 网络从视频帧中提取 3D CNN 特征，利用卷积长短时记忆网络进一步提取时空特征，大量实验表明，将 C3D 和卷积长短时记忆网络结合在一起，可以利用时空特征显著提高基于视频的行为分类精度。上述三维卷积的模型改进中，针对不同目标的分类识别，准确率和模型轻量化仍是需要进一步进行提升。

本文基于混凝土搅拌站自动上料控制系统，在上料过程中，旋转喂料机需要准确寻找料仓，通常使用电感式传感器来获取料仓位置信息。然而，这种识别方式需要使用多个传感器，并且容易受到外界环境的干扰，如温度、湿度、电磁场等，以及这种方式还需要进行繁琐校准工作。针对此类问题，首先，设计了一种合作标靶用于料仓识别和建立了实验数据集。然后采用机器视觉和深度学习方法进行识别，在视频识别的任务中，需要兼顾时序信息的同时能够高效提取特征从而提高识别准确率，但现有基于深度学习的视频目标分类方法中，普遍存在背景图像干扰、无法高效利用前后帧时序信息等问题导致识别效果不佳。对此本文提出了一种基于深度学习的 SE-C3D 料仓口视频识别模型，将 SE 注意力机制融合入轻量化 C3D 网络中，该网络模型相较于原 C3D 网络参数量更低且拥有更高的识别精度。

1 数据采集与数据集制作

1.1 实验材料与环境

旋转喂料机是混凝土搅拌站上料控制系统中对各种原材料进行分料的装置，受电动机驱动，由下方的电感传感器获取位置信息，从而将各种原料分别传输到对应各料仓中。项目实地的旋转喂料机如图 1 所示。

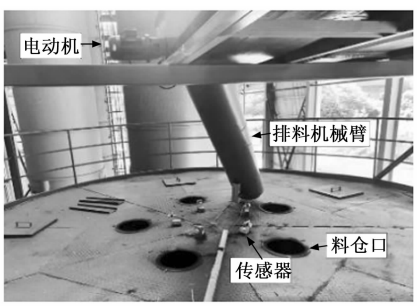


图 1 旋转喂料机结构图

因项目实地采集数据集困难，故实验数据采集于江南大学物联网工程学院实验室内搭建的小型模拟实验平台。实验采用的工业相机与机械臂的合作方式是“Eye-in-Hand”[14]型，工业相机是识别的唯一器件。在实验采集视频的过程中，将工业相机固定在机械臂末端，再由电动机带动机械臂使工业相机高速运动采集视频。这种方式对工业相机要求更高，故实验中采用海康机器人公司的 MV-CA004-10UC 进阶型面阵相机，相机采用 Sony 的 IMX287 CMOS 芯片，图像质量优异，该工业相机的分辨率为 720 像素×540 像素，最高帧率可达 526.5 fps。

为实现对料仓口精准识别，首先设计一套与各料仓口相对应的合作标靶。如图 2 所示：将工业相机固定在排料机械臂的下方，在料仓口对应的半径上固定着各个料仓合作标靶，标靶号与各料仓相对应。这种设计的优

点是通过标靶实现了机械臂排料口和各料仓口间接识别，相较于工业相机直接拍摄料仓口进行识别，这种识别方式更加精准，能达到与直接识别方式相同的效果。

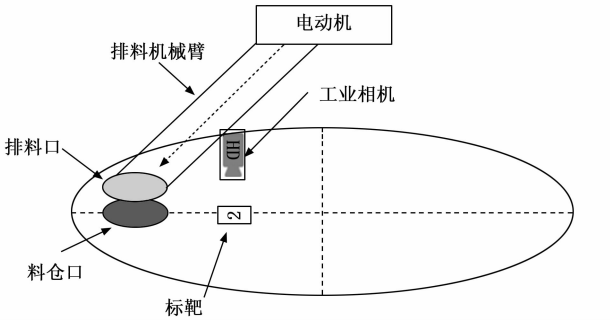


图 2　视频采集示意图

考虑到项目实际要求，实验装置共设计了 9 个料仓号。实验中视频采集帧率为 100 fps，视频帧图像的原始尺寸是 720 像素×540 像素。工业相机采集不同时间段，不同光照条件下各个料仓标靶的运动视频。实验平台中工业相机旋转一周采集视频需要的时间大约为 15 s，为了采集到丰富的数据集，机械臂需要旋转多次进行采集。

1.2　数据集的制作

将实验室中利用小型机械臂模型原始采集的数据经过处理按照 1～9 号料仓以及非料仓共 10 种类别划分，考虑到后续控制工作实时性，以 3 帧图像作为一个视频样本。数据集样本如图 3 所示。以 1 号料仓为例，以每个视频样本中最后一帧第一次出现完整的标靶，为有效料仓视频类别，其他的则都判定为非料仓类别。

在实际生产环境中，除了正常视频样本外，还会出现光线过亮、光线过暗以及机械臂在运动的过程中因抖动导致相机拍摄视频模糊等复杂的识别场景，这都会给识别带来巨大的困难。因此本实验在数据集制作过程中，添加了识别困难的复杂样本，如图 4 所示。

本实验共采集 37 000 个视频样本。其中约 10％是

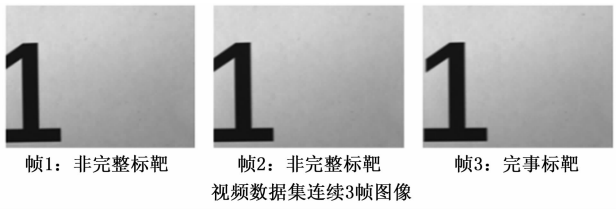


图 3　有效类别示意图

图 4　复杂样本展示

由强光照（复杂样本 1）和弱光照（复杂样本 2）下采集，40％是由模糊样本（复杂样本 3）构成。每类对应仓口的样本各为 3 500 个，处于非料仓口的样本 5 500 个。训练集、测试集按照 7∶3 的比例划分，训练集共有 25 900 个视频样本，测试集共有 11 100 个视频样本，具体划分如表 1 所示。

表 1　数据集构成信息

拍摄场景	正常	亮光	暗光	抖动	总计
训练集/个	13 000	1 370	1 330	10 200	25 900
测试集/个	5 590	520	540	4 450	11 100

2　改进 C3D 网络模型

2.1　C3D 网络模型轻量化

为保证料仓合作标靶分类任务的高准确率和实时性，实验中采用 C3D 模型提取视频特征，C3D 具有大量的参数和计算需求，使用初始 C3D 模型可能会面临性能瓶颈和运行速度缓慢的问题。故选择将模型轻量化，轻量化过程如图 5 所示。初始 C3D 模块有 5 个块，

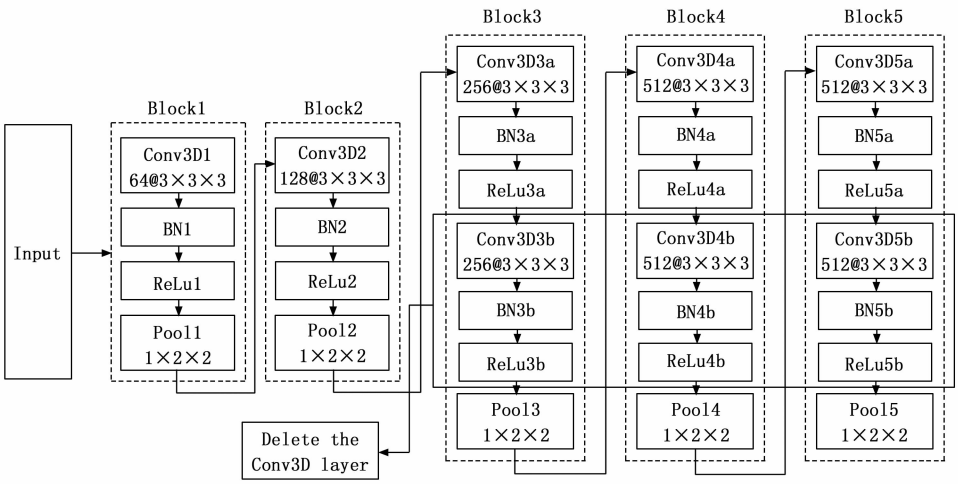


图 5　轻量化 C3D 网络结构图

前两个块分别有 3D 卷积层、ReLU 激活函数和 3D 池化层。从 Block3 到 Block5，原始 C3D 模型的每个 Block 均有两个 3D 卷积层，两个 ReLU 激活函数和一个 3D 池化层。轻量化操作是只保留 Block3 到 Block5 中的第一个 3D 卷积层、ReLU 激活函数和 3D 池化层，删除 Block3 到 Block5 中的第二个 3D 卷积层、ReLU 激活函数和 3D 池化层，该操作得到的轻量化 C3D 网络使得模型参数量更小，模型推理速度更快。

C3D 网络模型中的 3D 卷积、ReLU 激活和 3D MaxPool 池化的计算如下[15]：将三维卷积和池化核大小表示为，其中为核时间深度，为核空间大小。3D 卷积是通过将一个 3D 核卷积到由多个连续帧叠加在一起形成的立方体来实现的。这种结构使得卷积层中的特征映射与前一层中的多个连续帧相连接，从而有效地捕获运动信息。具体而言，第 i 层第 j 个特征图在位置的数值可以表示为：

$$V_{ij}^{xyz}=f\{b_{ij}+\sum_{k=1}^{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1}w_{ijk}^{pqr}v_{(i-1)k}^{(x+p)(y+q)(z+r)}\} \quad (1)$$

式中，P_i、Q_i、R_i 为三维卷积核的大小，m 为 $i-1$ 层特征图的个数；$v_{(i-1)k}^{(x+p)(y+q)(z+r)}$ 为第 $i-1$ 层第 k 个特征映射 $(x+p,\ y+q,\ z+r)$ 的值；w_{ijk}^{pqr} 是连接到第 $i-1$ 层的第 k 个特征映射的卷积核；$b_{i,j}$ 是偏集；$f(\cdot)$ 为 ReLU 激活函数。

三维池化操作使特征立方体在时间维度上具有一定的不变性，同时大大减少了计算量，从而提高了三维卷积神经网络在时间维度上的鲁棒性[16]，最大池化的公式如下：

$$m_{pool}^{xyz}=\max_{0\ll i\ll S_1,0\ll j\ll S_2,0\ll k\ll S_3}(n_{x\times s+i,y\times t+j,z\times r+k}) \quad (2)$$

式中，$\boldsymbol{u}$ 为三维输入向量，m 为三维池化运算后的输出，$(s,\ t,\ r)$ 为坐标 $(x,\ y,\ z)$ 沿 x，y，z 方向的上采样步长；$S_1\times S_2\times S_3$ 为采样面积大小。本文在卷积过程中采用步长为 3×3×3 小核尺寸，有助于从时空信息方面捕捉所有变化[17]，使用小核尺寸可以显著减少卷积层的参数数量，可以降低模型的复杂度，也可以获得更丰富、更复杂的特征表示，有助于提高模型的表达能力。小核尺寸的卷积层具有参数共享和局部连接的特点，有助于提取局部纹理、边缘等细节信息。

2.2 SE 注意力机制

通道域注意力机制的代表作是 Hu 等人[18]提出的基于图像分类的 SE 注意力机制。SE 注意力机制思想由 2 个关键步骤组成：挤压（Squeeze），激励（Excitation）。在挤压阶段，通过全局平均池化操作，将每个通道上的特征图压缩为一个标量值，以获得通道的全局信息。在激励阶段，将这些压缩后的特征输入到两个全连接层中，经过非线性激活函数后产生激励权重。最后，将这些激励权重与原始特征图进行点乘操作，对每个通道进行加权，得到加权后的特征图。其中，SE 注意力机制基本结构如图 6 所示：其中，Squeeze 是通过全局平均池化将每个通道的二维特征（$H\times W$）压缩得到 $1\times1\times C$ 的特征向量，计算过程如公式（3）所示：

$$Z_c=F_{sq}(U_c)=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}U_c(i,j) \quad (3)$$

式中，H 和 W 分别为每个特征图的高和宽，C 表示特征图的通道维度。

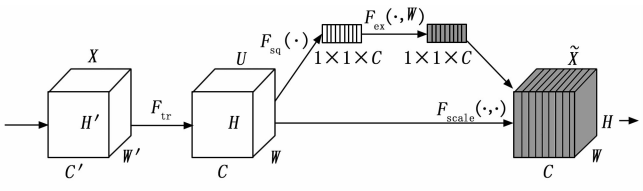


图 6 SE 注意力机制

Excitation 是两个全连接层形成一个瓶颈层结构以模拟通道之间的关联性，并输出与输入特征相同数量的权值，计算过程如公式（4）所示：

$$s=F_{ex}(z,W)=\sigma[g(z,W)]=\sigma[W_2\,\delta(W_1\,z)] \quad (4)$$

式中，$\delta(\cdot)$ 为 ReLU 激活函数，$\sigma(\cdot)$ 为 sigmoid 函数。W_1 和 W_2 代表两个不同的全连接操作，s 代表每张特征图的重要程度。

Scale 将前面两个函数得到的归一化权重加权到每个通道的特征上，并权重系数逐个通道相乘，以完成通道维度上注意力机制的引入，如公式（5）所示：

$$\widetilde{x_c}=F_{\text{scale}}(u_c,s_c)=s_c\,u_c \quad (5)$$

式中，$F_{\text{scale}}(u_c,\ s_c)$ 为特征图 u_c 和标量 s 对应通道的乘积。

SE 注意力机制旨在通过使网络能够执行动态通道特征重新校准来提高网络的表示能力。因此选择在 C3D 的基础上增加 SE 注意力机制，可以对原始数据的特征向量序列生成注意力权重，从而提高视频识别的准确率。

2.3 融入注意力机制的 C3D 改进模型

鉴于 SE-C3D 模型的对视频片段可以同时学习图特征和相邻帧之间复杂的时序特征进而提高识别的准确性，以及 SE 注意力机制通过网络的不断学习来更新每个特征通道的权重值，以提升有效的通道响应。针对传统电感式传感器容易受到外界复杂环境干扰，且需要进行繁琐校准工作等问题，本文提出了改进 SE-C3D 视频识别分类模型。模型的整体结构如图 7 所示：在初始 C3D 模型的第 3、4、5 层卷积层进行轻量化的基础上，在轻量化 C3D 模型后融合了 SE 注意力机制，模型的输入层是为帧数为 3、大小为 32 像素×32 像素的样本，先用改进的 C3D 网络提取输入数据的局部时空特征，然后将提取的视频特征序列放入三维通道注意力模块中，用于凸显分类相关特征或削弱无关特征，最后输出识别结果。识别模型采用交叉熵损失函数进行训

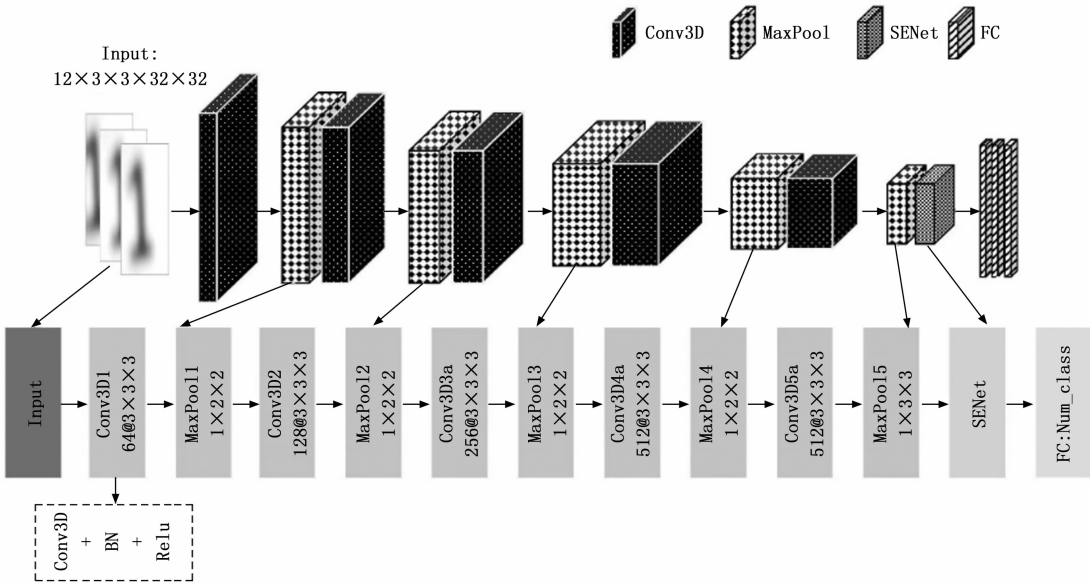


图 7　SE-C3D 模型结构图

练，即：

$$L = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \tag{6}$$

式中，M 为类别的数量，y_{ic} 符号函数（0 或 1），p_{ic} 是观测样本 i 为类别 c 的预测概率。

2.4　评价指标

本算法的性能评估采用了 4 个评价度量，包括综合准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F_1 值（F_1-score），计算得到结果高值意味着模型的预测能力好。

综合准确率计算的方式是将样本的正确预测数除以全部样本数。计算公式如下：

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

精确率表示某一类别的样本中被预测真正属于该类别的比例：

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

召回率表示在属于某一类别的所有样本中，准确地被预测为该类别的比例：

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

F_1 值则同时考虑了精确率和召回率，将精确率和召回率的乘积除以精确率与召回率的和的两倍，得到一个在 0 到 1 之间的比例：

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

式中，TP（True Positive）表示真阳性，TN（True Negative）表示真阴性，FP（False Positive）表示假阳性，FN（False Negative）表示假阴性。

2.5　模型训练与评估

本研究使用了以下硬件环境进行模型训练：Intel 酷睿 i5-11320H 处理器、16 GB 运行内存、3.2 GHz 的 CPU 频率，以及搭载 NVIDIA GeForce MX450 GPU。软件环境方面：电脑搭载 Win11 系统和 Anaconda 3 平台，在深度学习框架平台 Pytorch 上进行训练和测试网络模型。训练过程中，卷积神经网络模型的参数设置对识别结果有着密切联系。训练过程中将输入数据 batch _ size 设置为 128，一共训练迭代 50 轮，在数据输入网络模型之前，需要对图像进行预处理，包括归一化等操作，将样本图像统一调整为 32 像素×32 像素的大小。在优化器的选择中，是利用具有内存占用小、应用简单和计算效率高等优点的 Adam 优化器进行优化网络模型。学习率是深度学习训练中的一个关键超参数，控制着每次迭代更新权重的步长大小，其设置对训练过程有着显著影响，对于本文中料仓合作标靶的视频识别，设置为 0.000 1 是一个比较合适的学习率。为评估网络模型对料仓合作标靶的识别能力，最终通过计算精确率、召回率、特异度和 F_1 值 4 个评价度量，综合评价网络模型的性能。

3　结果与分析

3.1　网络训练结果分析

为了验证本文所提出的 SE-C3D 模型相对于初始 C3D 模型在料仓识别中的准确率改进，进行了对比实验。在相同的超参数和实验环境下，本实验分别使用 C3D 模型和 SE-C3D 模型对实验数据集进行训练，并观察它们在 50 轮迭代过程中准确率和损失值的变化情况。根据图 8 的结果可以看出，SE-C3D 模型在大多数情况

下都表现出比 C3D 模型更高的准确率。随着训练进行，SE-C3D 模型的准确率和损失值逐渐稳定并趋于收敛。这说明 SE-C3D 模型在训练过程中逐渐学习到了数据的特征，并取得了较好的性能。这表明引入 SE 注意力机制的 SE-C3D 模型能够提升视频分类任务的准确性。通过在网络中引入 SE 模块，模型可以自适应地学习到不同特征的重要性，并加强对关键信息的关注。这种注意力机制有助于提高模型的表达能力和泛化能力，从而改善料仓识别的准确率。

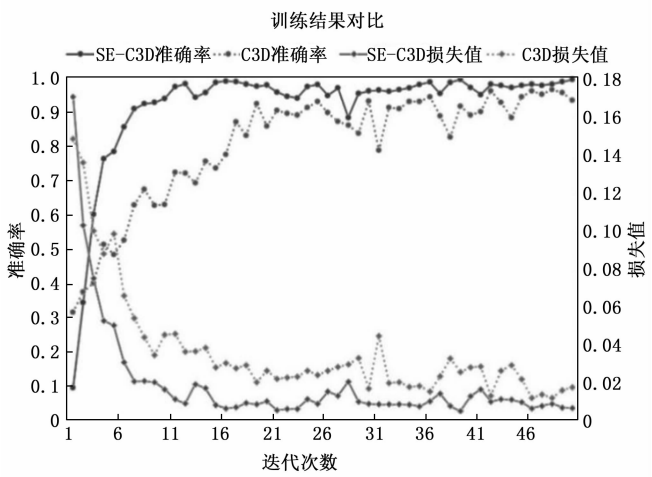


图 8 改进前后训练结果对比

在该料仓合作标靶视频分类任务中，SE-C3D 模型表现出更低的损失值，这意味着该模型能够更好地拟合训练数据。相比之下，C3D 模型损失值相对不稳定，可能是因为该模型没有引入 SE 模块，导致难以充分学习到数据的特征。SE-C3D 模型通过引入 SE 模块来增强模型的表示能力。SE 模块能够自适应地学习每个通道中不同特征的重要性，并将这些特征重新加权。这使得模型能够更好地捕捉数据的特征，并且在经过训练后，能够更好地适应新的数据。基于以上分析，可以得出 SE-C3D 模型相比于初始 C3D 模型，SE-C3D 模型能够提取更有效的特征并更好地捕捉视频数据中的关键信息，在该料仓视频分类任务上具有更高的准确率和更低的损失值，且更具稳定性。

3.2 不同网络性能对比分析

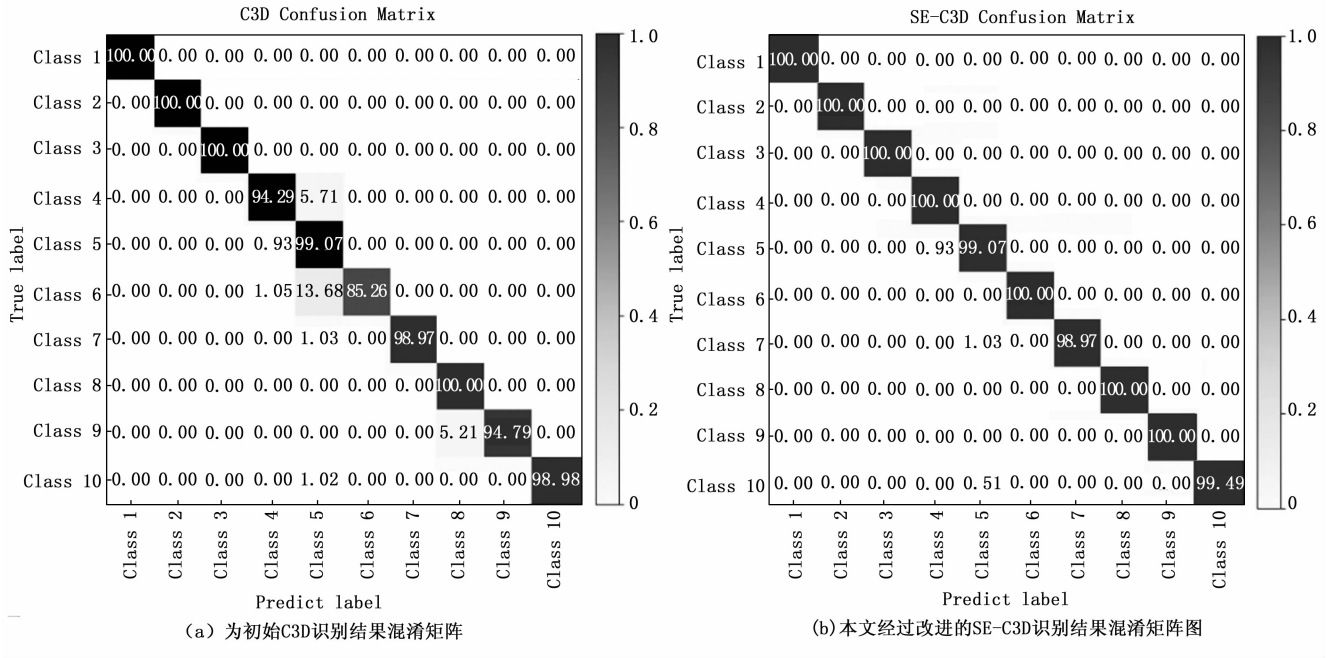
在模型性能对比中，不同网络会直接影响识别效果，其中典型的二维视频识别网络是长期卷积循环网络[19]（LRCN，long-term recurrent convolutional networks），该模型结合了传统卷积网络和长短期记忆网络的新型网络结构，具有处理视频时序信息的能力[20]。其融合了 LSTM[21]（Long Short Term Memory）。

LSTM 是一种特殊的循环神经网络，它能够识别和预测序列数据中的模式。与传统的循环神经网络不同，LSTM 具有多个门控单元，它们能够控制信息的流动和保留，从而捕捉到长期依赖关系，LSTM 的核心部分是记忆单元（Memory Cell），它在网络中负责存储和传递信息。LSTM 还包括 3 个门控单元：输入门（Input Gate）、遗忘门（Forget Gate）和输出门（Output Gate）。LSTM 通过这些门控单元的控制，可以选择性地读取、写入、遗忘和输出记忆单元中的信息，从而实现对序列数据中长期依赖关系的捕捉和预测。识别结果如表 2 所示：融合 LSTM 组成的 LRCN 网络准确率为 93.93%，此二维卷积模型由于使用逐帧识别方式，相对三维识别网络，输入模型信息最少，故导致准确率最低。

表 2 不同网络模型性能对比

网络类型	准确率/%	精确率/%	召回率/%	F_1/%	参数量
LRCN	93.93	93.24	93.93	93.89	5 361 170
3D_MobilenetV2	94.55	95.38	94.50	94.98	3 481 610
3D_SqueezeNet	96.13	96.34	95.78	95.99	1 837 002
3D_Resnet50	97.74	97.87	97.74	97.72	63 518 666
C3D	97.13	97.46	97.05	97.08	42 848 674
SE-C3D	99.61	99.57	99.63	99.60	10 997 538

C3D 对比于基于二维卷积的 LRCN 网络，准确率提升了 3.20%，由此可知 C3D 网络带来的提升比较明显，这主要由于 C3D 在卷积层中引入了时间维度的输入，从而能够同时捕捉图像的空间特征和视频序列的时间特征，使得 C3D 能够更好地捕捉视频中的运动和动态变化，而基于传统二维 CNN 因在时序上只考虑单一帧从而无法做到这一点。本文提出的 SE-C3D 模型在测试集上的准确率达到 99.61%，精确率达到 99.57%，与初始 C3D 模型召回率提高 2.58%，F_1 值提高 2.52%，且与其他典型三维卷积模型进行对比（见表 2）：如 3D_MobilenetV2、3D_SqueezeNet、3D_Resnet50，在准确率、精确率、召回率、F_1-score 各项指标上也均有明显提升。在参数量方面，相较于二维 LRCN 模型，参数量有所增加，这是由于模型由二维卷积变为三维卷积的必然结果，同时由二维到三维卷积使模型性能也得到了明显提升；但相较于初始 C3D，两种模型均为三维卷积，表 2 可以得出改进后 SE-C3D 的参数量也大幅减少，说明删除 Block3 到 Block5 中的第二个 3D 卷积层、ReLU 激活函数和 3D 池化层之后，使得 SE-C3D 模型更加轻量化，三维模型往往包含大量的数据，这些数据会导致模型的处理时间和计算负荷增加，从而降低了计算效率。通过对三维模型进行轻量化，即减少其数据量，可以提高计算效率，使得实用性更强。SE-C3D 在参数量减少的情况下，由表 2 可知，在该料仓视频分类任务上，性能相较于初始 C3D 有所提升，该提升得益于融合的 SE 注意力机制提升了模型对 channel 特征的敏感性，增强特征图中的有用信息，但是对网络



(a) 为初始C3D识别结果混淆矩阵　　(b)本文经过改进的SE-C3D识别结果混淆矩阵图

图 9　混淆矩阵对比

模型参数量影响相对较少。

3.3　SE-C3D 模型结果分析

初始 C3D 网络仅仅依靠 3D 卷积神经网络进行视频特征的提取，虽然相较于 2D 卷积神经网络的逐帧识别方式准确率有所提升，但这对特征提取的能力提升依然有限。为了更清楚地展示模型在测试集中各个类别中的识别结果，经过 50 次迭代训练后，引入混淆矩阵对实验结果进行详细分析，如图 9 所示，初始 C3D 模型在类别 4、类别 5、类别 6、类别 7、类别 9、类别 10 中均出现错误识别，尤其类别 6 的错误率最高，达到了 14.74%，这是因为 C3D 模型对于输入样本本身模糊不清，C3D 的识别性能可能会受到限制，从而导致模型难以从中提取有用的特征，导致了识别准确性的降低。在 SE-C3D 模型识别结果的混淆矩阵图中，除了类别 5、类别 7、类别 10 出现个别错误识别外，其余 7 种类别都能得到 100%的准确识别。这由于 SE 注意力机制通过学习特征通道之间的关系，能够自适应地为每个通道分配不同的权重。这使得模型能够更加关注对于特定任务更有用的特征通道，从而增强了特征的表征能力。这有助于模型更好地区分不同类别之间的差异，提高了识别的准确性。使得 SE-C3D 模型能够整合特征通道的重要程度，可以有效地在复杂场景的视频帧中找出标靶显著区域，从而提高了识别的准确率。这表明本文算法有效提升 C3D 网络的性能。SE-C3D 具有更加强大的表达能力和学习能力，能够更好地处理具体有连续时序信息特征的分类识别问题。

4　结束语

经过将机器视觉和深度学习应用于混凝土搅拌站上料控制系统中旋转喂料机排料口寻找料仓的过程，可以用于取代原有的电感式传感器工作方式。为解决运动视频中难以精确识别运动料仓问题，本文设计了一种合作标靶，并建立了所需的实验数据集。针对目前基于深度学习的视频目标分类方法中存在的背景图像干扰、无法充分利用前后帧时序信息以及模型参数量大难以部署到移动小型设备等问题，本文提出了一种 SE-C3D 的料仓合作标靶视频识别模型。通过删除初始 C3D 部分卷积操作，以及添加 SE 注意力机制构成的 SE-C3D 模型，相较于初始 C3D 网络，该模型具有更低的参数量且拥有更高的识别精度。利用自行搭建的小型旋转机械臂模型进行实验并构建视频数据集来训练 SE-C3D 模型。最终，在测试集上，该模型的准确率达到了 99.61%，综合准确率、精确率、召回率、F_1 值 4 个评价度量相较于二维卷积网络以及其他典型的三维卷积网络对比均有提高，与传统二维 CNN 模型以及其他典型三维视频识别分类模型相比，SE-C3D 模型都展现出了更好的性能提升。该模型对于料仓的高效识别以及未来智能化上料系统的发展具有重要意义。为实现智能化上料系统提供了一种可行的解决方案，并为相关领域的进一步研究和开发提供了有价值的参考。

参考文献：

[1] 杨　涛，易新蕾，卢绍文，等. 工业人工智能驱动的流程工业智能制造 [J]. Engineering，2021，7 (9)：70－83.

[2] 郝伟凯，王亚敏. 自动上料系统优化改造研究 [J]. 科技资讯，2022，20 (13)：45－47.

（下转第 183 页）