Pure Mathematics 理论数学, 2023, 13(3), 381-394
Published Online March 2023 in Hans. https://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2023.133042

Hans汉斯

基于C3D-APReLU的工业过程视频故障诊断

宋启哲，田　颖*，李嘉乐

上海理工大学，上海

收稿日期：2023年1月24日；录用日期：2023年2月23日；发布日期：2023年3月2日

摘　要

近些年来，人们对生产安全的要求越来越高。随着图像采集设备在工业过程监控中的普及，基于视频的深度学习故障诊断技术得到了快速的发展。然而使用传统激活函数的深度学习方法只能提供相同的非线性映射，这不利于模型对输入信号特征的学习和分类。针对这个问题，本文提出了一种用于视频分类模型的、可以自适应调整参数的激活函数APReLU-3D。该激活函数内嵌了一个可以对输入信号进行学习从而对坡度自动做出相应调整的子网络，使得每个输入信号都可以有自己的非线性映射。本文将APReLU-3D应用于视频分类模型C3D中，提出了C3D-APReLU模型。采用PRONTO工业数据集中的视频数据对该方法进行对比实验，结果表明，C3D-APReLU实现了比使用ReLU激活函数的C3D更好的故障诊断性能，其平均精度为0.978。

关键词

视频故障诊断，三维卷积网络，自适应参数激活函数，非线性映射

Video Fault Diagnosis of Industrial Process Based on C3D-APReLU

Qizhe Song, Ying Tian*, Jiale Li

University of Shanghai for Science and Technology, Shanghai

Received: Jan. 24th, 2023; accepted: Feb. 23rd, 2023; published: Mar. 2nd, 2023

Abstract

In recent years, people have higher and higher requirements for production safety. With the popularity of image acquisition equipment in industrial process monitoring, video-based deep learning fault diagnosis technology has been developed rapidly. However, the deep learning method

*通讯作者。

文章引用: 宋启哲, 田颖, 李嘉乐. 基于 C3D-APReLU 的工业过程视频故障诊断[J]. 理论数学, 2023, 13(3): 381-394.
DOI: 10.12677/pm.2023.133042

using the traditional activation function can only provide the same nonlinear mapping, which is not conducive to the learning and classification of input signal features. To solve this problem, this paper proposes a new activation function APReLU-3D which can adjust parameters adaptively for video classification model. The activation function is embedded with a subnetwork that can learn from the input signal and automatically adjust the slope accordingly, so that each input signal can have its own nonlinear mapping. In this paper, APReLU-3D is applied to video classification model C3D, and a model of C3D-APReLU is proposed. The method was compared using video data from PRONTO industrial dataset. The results show that C3D-APReLU achieves better fault diagnosis performance than C3D using ReLU activation function, with an average accuracy of 0.978.

Keywords

Video Fault Diagnosis, Three-Dimensional Convolutional Network, Adaptive Parameter Activation Function, Nonlinear Mapping

1. 引言

随着过程工业的不断发展，生产过程的不断优化，以及产品质量的不断提高，工业生产过程日益自动化、复杂化，且用于生产的设备成本也越来越高。一旦发生故障往往会导致严重的经济损失和社会危害。因此工业生产过程的安全性受到了极大的关注，维护生产安全的故障诊断技术成为了过程工业的研究重点之一。

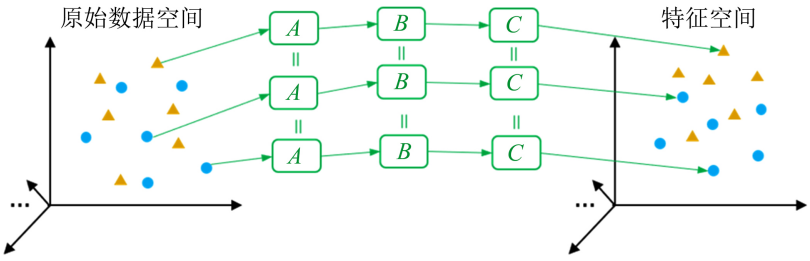
随着传感器和计算机技术的高速发展，工业过程中海量的反映过程运行机理和运行状态的关键数据得到了准确的采集。如何利用这些数据来满足日益提高的系统安全性要求已成为亟待解决的问题，这催化了基于数据驱动的故障诊断技术的发展。基于数据驱动的故障诊断技术不需要建立具体的数学模型，也不必具备准确的先验知识，以采集过程数据为基础，利用数据处理与分析方法提炼出数据中的特征信息，形成知识，最后训练出可以自主决策的模型，在复杂工业过程控制中有着极为广泛的应用[1]。

视频数据是最常见的数据类型之一，既包含着空间信息，又包含着时间信息，因此基于视频的故障诊断技术得到了快速发展。视频故障诊断技术常用来对故障场景进行判别和分类。视频分类算法最初基于传统的手工特征，主要是提取视频及图像的纹理、颜色、整体、显著区域等特征进行场景分类。Itti 等[2]提出先使用 Gabor 滤波法对提前划分好的图像在颜色、方向和密度三个通道上提取特征，然后使用 PCA/ICA 技术对特征进行降维，最后通过神经网络进行分类。贾澎涛等[3]提出多特征的视频场景分类，首先提取视频的平均关键帧，再将关键帧划分成感兴趣区域与不感兴趣区域，之后再分别提取场景特征，最后将特征进行融合并利用特征阈值对场景进行分类。然而，如何为不同的任务选择合适的手工特征成为了困扰学者们的难题。随着深度学习技术的发展，视频分类算法进入到了新的发展阶段。

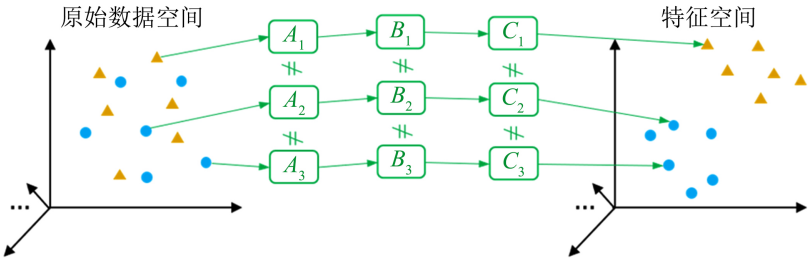
基于二维卷积神经网络(2 Dimensional Convolutional Neural Network, 2D CNN)的图像分类模型[4] [5]利用二维卷积(2D Convolution)提取特征和使用端到端的训练方式实现了自动地从图像中学习一系列有辨别性的高级语义特征，将图像分类的准确度提升到了新的高度。Andrej [6]在 2D CNN 的基础上考虑了将视频的时间特征与二维卷积提取到的空间特征在特征层面进行融合，实现了视频分类。双流网络[7]使用两个独立的 2D CNN 分别提取空间特征和时间特征，最后在分类层进行融合，得出最终的视频分类结果。

然而基于 2D CNN 的视频分类模型对于视频空间和时间特征是单独提取的，难以高效学习到更有用的时空特征(Spatiotemporal Features)。为克服在提取时空特征上的劣势，三维卷积(3D Convolution)被提出[8]。与二维卷积不同之处在于，三维卷积的卷积核同时考虑了视频中连续帧之间的空间和时间特征，更适合时空特征的学习。三维卷积被提出后，被广泛应用于视频理解任务的时空特征提取中[9] [10]，其中 C3D (Convolutional 3D)模型[11]是视频分类领域最具代表性的工作之一，被广泛应用[12] [13]。[12]提出了基于 C3D-ConvLSTM 的基于视频数据的奶牛行为分类方法，该模型可用于对动物在不同生长阶段的行为进行分类。[14]基于 C3D 提出了一种区域卷积模型 RC3D，该模型可以对视频流进行编码，生成包含活动的候选时间区域，解决了连续的、未处理的视频流中的活动检测问题。C3D 在视频分类中的优秀表现已在众多研究中得到了验证，因此本文以 C3D 为基准，进一步完善其在故障诊断领域中的表现。

然而，在包含 C3D 在内的经典神经网络模型中都使用的是传统的激活函数，比如：sigmoid, tanh, ReLU [16]，LReLU (leaky ReLUs) [17]，和 PReLU (parametric ReLUs) [18]。这会造成一个问题，那就是对于不同的输入数据模型都会应用相同的非线性映射，这样不利于模型学习有用的特征从而将类内信号投射到同一区域，将类间信号投射到不同区域实现高精度的分类，如图 1(a)所示。针对这个问题，赵明航等[15]提出了一个新的激活函数，即自适应参数线性修正单元 APReLU (adaptively parametric ReLUs)。APReLU 基于 PReLU 改进而来，但与之不同的是其中的坡度参数可使用一个子网络对输入数据进行学习做出相应调整，从而允许其提出的 ResNet-APReLU 模型对输入数据的每个通道进行自适应的非线性映射，达到提高振动故障诊断精度的目的，如图 1(b)所示。



(a) 使用传统激活函数的神经网络非线性映射示意图



(b) 使用APReLU激活函数的神经网络非线性映射示意图

Figure 1. Neural network nonlinear mapping diagram with different activation functions, where A, B and C represent nonlinear mapping at different stages
图 1. 使用不同激活函数的神经网络非线性映射示意图，其中 A，B 和 C 表示不同阶段的非线性映射

综上所述，本研究提出了一种使用 APReLU 作为激活函数的 C3D-APReLU 视频分类模型。该方法很好地将 C3D 和 APReLU 结合在了一起，通过 APReLU 引入的自适应的非线性映射，提取到了工业监控

视频中具有高辨别性的时空特征，从而实现了优秀的视频故障诊断效果。值得注意的是，由于ResNet-APReLU 中提出的 APReLU 处理的是经过一维卷积操作后的数据，本文为使其适用于三维卷积，进行了相对应的改造，提出了 APReLU-3D，这是本研究主要的创新之一。

2. 相关理论

2.1. C3D

C3D 模型是 2015 年由 Tran 等[11]提出的一个简单却有效的使用三维卷积对视频提取时空特征并分类的方法。C3D 的模型架构如图 2 所示。其中 Conv 模块为三维卷积层，Pool 模块为三维最大池化下采样层，Fc 为全连接层，Softmax 为输出归一化层。所涉及的组成成分会在下面一一讲解。

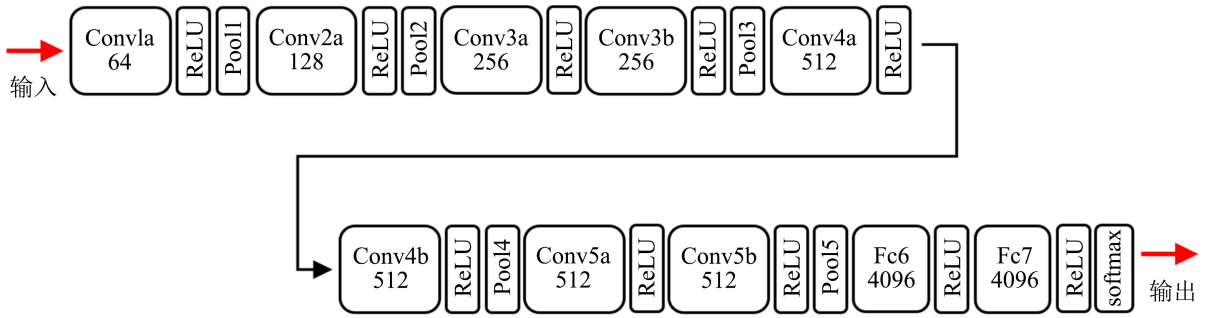
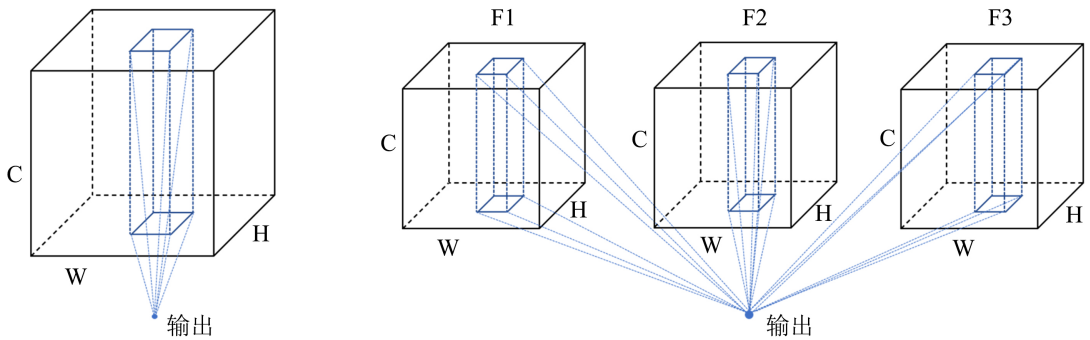


Figure 2. Structure diagram of C3D
图 2. C3D 结构示意图

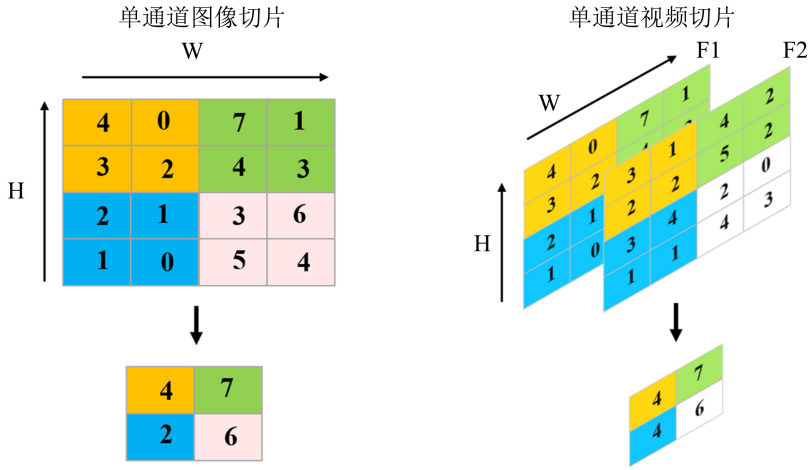
1) 三维卷积(3D Convolution)。三维卷积常用来提取视频数据中的时空特征，其输入输出数据形状为(N_{batch} , F, C, W, H)，其中 N_{batch} 表示在训练或测试时一个小批量中样本的数量，F 和 C 分别表示一个视频样本中的帧数(Frames)和图像通道数(Channels)，W 和 H 分别表示每帧图像的宽(Width)和高(Height)。二维卷积和三维卷积的示意图见图 3，其中蓝色部分为卷积核。图 3(a)中，无论是对一张图片还是多张图片执行二维卷积，输出均为一张图片(多张图片会被视作不同的通道)，这使得视频数据在通过二维卷积之后会丢失原始的时间信息。但图 3(b)中的三维卷积考虑了视频的时间维度，这样的机制可以使得视频片段经三维卷积后输出的还是视频数据形状，有效的保留了时间特征。因此，三维卷积拥有更好的提取时空特征的能力。



(a) 二维卷积示意图　　(b) 三维卷积示意图

Figure 3. Diagram of 2D convolution and 3D convolution
图 3. 二维卷积与三维卷积示意图

2) 三维最大池化下采样层(3D max pooling layer)。池化层以降低特征的尺寸为目的，对输入特征进行下采样。二维和三维最大池化层如图 4 所示，其中图 4(a)中的单通道图片在高和宽维度上被执行了尺寸为 2 的二维最大池化，被池化区域仅输出其中的最大值，输出图像的尺寸缩减为原图的一半。而图 4(b)中的单通道视频切片被执行了尺寸为 2 的三维最大池化，相比二维池化增加了时间维度上的下采样，这样可以缩减视频的帧数，进一步提取时间维度上的特征。



(a) 二维最大池化　　(b) 三维最大池化

Figure 4. Diagram of 2D max pooling and 3D max pooling
图 4. 二维最大池化与三维最大池化示意图

3) ReLU 激活函数。ReLU 是深度学习方法中最流行的激活函数之一，相较于 sigmoid 函数和 tanh 函数，ReLU 更有效地防止了梯度消失问题。ReLU 函数被表示如下

$$y = \max(x, 0) \tag{1}$$

其中，x 和 y 分别是输入和输出特征。

4) 全连接层(fully connected layer)。全连接层是一排神经元，它的每一个结点都与上一层的所有结点连接。在 C3D 中，Conv5b 输出的特征会展开与 Fc6 全部连接，将之前由三维卷积提取到的特征综合起来。

5) Softmax 函数。最后的输出层使用 Softmax 函数对输出向量进行归一化操作，该函数将输出特征转化到(0, 1)范围内，使用公式可以表示为：

$$y_j = \frac{e^{x_j}}{\sum_{i=1}^{N_{class}} e^{x_i}} \tag{2}$$

其中 x 和 y 分别是输入和输出，N_{class} 是分类类别数量，j 表示第 j 个类别。y_j 表示相应样本预测为第 j 个类别的概率，因此 $\sum_{i=1}^{N_{class}} y_j = 1$，即概率之和为 1。

在 C3D 中，所有的三维卷积核的尺寸都是 3 × 3 × 3，步长为 1 × 1 × 1，卷积核的个数标注在每个卷积模块上。除了 Pool1 池化核尺寸为 1 × 2 × 2，步幅为 1 × 2 × 2，其他所有三维池化层池化核均为 2 × 2 × 2，步长为 2 × 2 × 2。每个全连接层有 4096 个神经元。C3D 的输入输出数据大小为(N_{batch} , 16, 3, 112, 112)，

N_{batch} 视计算机能力与数据集大小而定，一般为 2^n，n 为大于 0 的整数。

2.2. ReLUs 的改进版本

ReLU 是深度学习方法中最流行的激活函数之一，为充分发挥其性能，已有多种 ReLU 的变体得到了研究并被开发，例如 LReLU 和 PReLU。LReLU 与传统的 ReLU 不同之处在于 LReLU 对小于零的特征应用了一个小的、非零的乘法系数，即坡度(slope)，而不是强制它们为零。LReLU 表示为(当坡度为 0.1 时)

$$y = \max(x,0) + 0.1 \cdot \min(x,0) \tag{3}$$

PReLU 是 LReLUs 的变种。如上所述，LReLUs 中的坡度是一个预设的常数。但在 PReLU 中允许使用梯度反向传播来训练坡度。PReLU 表示为

$$y = \max(x,0) + \alpha \cdot \min(x,0) \tag{4}$$

其中 α 是可训练的坡度。需要注意的是，PReLU 中的 α 在训练过程中是可调整的，但在测试过程中固定为一个常数，不能根据每次具体的测试信号进行调整。

3. 模型构建

3.1. APReLU-3D

APReLU 集成了一个特殊设计的子网络作为一个嵌入式模块，用于对具体的输入信号自适应地估计在非线性变换中使用的坡度 α，这是其与 PReLU 的不同之处，也是 APReLU 的创新所在。如图 5 所示，本节用一个实际输入输出案例来展示 APReLU-3D 的工作原理。

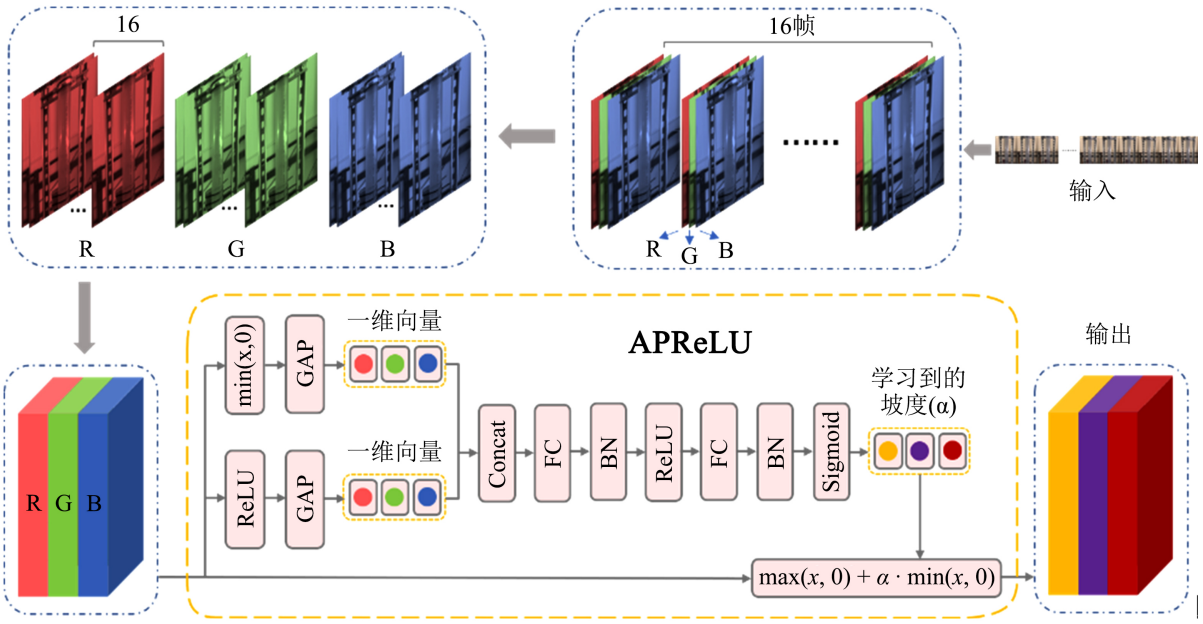


Figure 5. Schematic diagram of APReLU-3D
图 5. APReLU-3D 原理示意图

APReLU-3D 的输入数据同 C3D 相同，是一段形状为(N_{batch}, F, C, W, H)的视频数据。图中的输入数据的形状为(32, 16, 3, 112, 112) (为方便展示，图中只画出一个样本)，其中的 3 个通道分别用 R、G、B 表示。

在被送入 APReLU 之前，数据形状会被重构为(32, 3, 16, 112, 112)，即所有视频帧中相同通道的特征图会被抽出放在一起。之后这样排列的数据就会被送入 APReLU 中，首先被传播到一个 ReLU 和一个三维全局平均池化下采样层(GAP)中，计算一个表示正数特征全局信息的一维向量。同时，将输入特征图传播到一个 min(x, 0)函数和一个 GAP 层来计算另一个一维向量来表示负数特征的全局信息。其中三维全局平均池化下采样 GAP 与三维最大池化下采样类似，均以降低特征的维度为目的，不同之处在于全局平均池化操作求取的是池化区域的平均值。然后将两个一维向量串联并传播到一个计算路径(即 FC → BN → ReLU → FC → BN → sigmoid)后计算得到各通道对应的坡度 α，最后，将学习到的坡度 α 应用于式(4)中对原输入信号进行非线性映射，得到输出特征图。在该计算路径中，每个 FC 层的神经元数等于 APReLU 输入信号的通道数，Sigmoid 函数将坡度映射到(0, 1)范围内的浮点数，这是为了防止为坡度分配的值过大。计算路径中的 BN 表示批归一化(Batch Normalization) [19]，其被经常应用于深度神经网络的中间层，为中间层数据进行归一化处理，以减小各层数据的差异性，加速网络的收敛。BN 的公式化表示如下：

$$\mu = \frac{1}{N_{batch}} \sum_{s=1}^{N_{batch}} x_s \tag{5}$$

$$\sigma^2 = \frac{1}{N_{batch}} \sum_{s=1}^{N_{batch}} \left(x_s - \mu\right)^2 \tag{6}$$

$$\hat{x}_s = \frac{x_s - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{7}$$

$$y_s = \gamma \hat{x}_s + \beta \tag{8}$$

其中 X_s 和 Y_s 分别是一个 batch 中第 s 个样本的输入和输出。μ 是批内样本均值，σ^2 为批内样本方差，ϵ 是为了防止分母为 0 而引入的超参数。γ 和 β 分别是为数据缩放和偏移而要学习的参数。最后，将学习到的坡度 α 应用于式(4)中进行非线性变换，得到输出数据。

值得注意的是，APReLU 自适应地调整非线性变换是以视频数据中图像通道为单位的，这是因为在 APReLU 的上游三维卷积中，输出数据的相同通道的特征图都是由同一个卷积核进行三维卷积操作而得来，又因为同一个卷积核关注的是相同的特征，所以 APReLU 对相同的特征施以相同的非线性映射。

3.2. C3D-APReLU

C3D-APReLU 的模型结构很简单，如图 6 所示，只需在 C3D 模型结构的基础上，将其中在卷积层后使用的 ReLU 函数替换为 APReLU-3D 即可。因为 APReLU-3D 的输入与输出数据与 C3D 具有相同的形状和类型，因此 APReLU 可以很容易地插入到 ReLU 出现的位置，而无需进行任何其他修改。

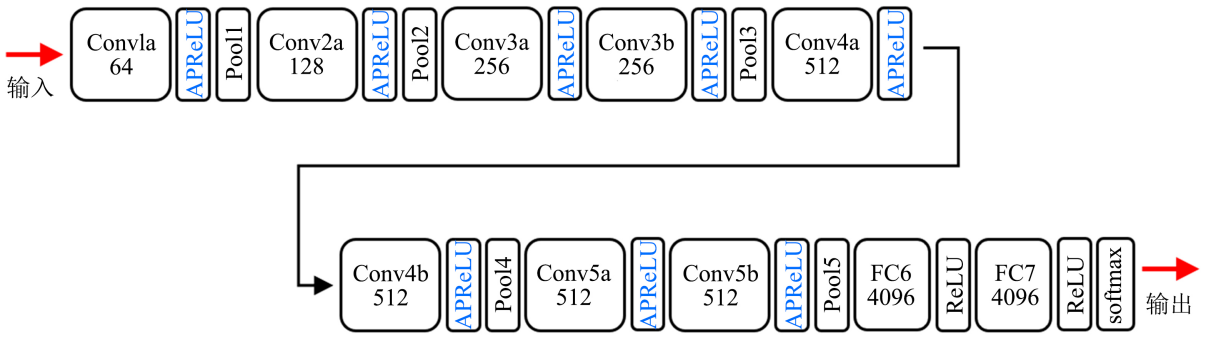


Figure 6. Structure diagram of C3D-APReLU
图 6. C3D-APReLU 结构示意图

C3D-APReLU 的优化通过最小化由式 8 所计算得到交叉熵损失实现，其中 y_j 是式 3 的输出，t_j 是第 j 个类别的真实标签

$$L = -\sum_{j=1}^{N_{class}} t_j \log\left(y_j\right) \tag{9}$$

4. 实验与分析

本章将本文提出的 C3D-APReLU 应用于某工业案例研究中工业级多相流设备的故障诊断中，数据集使用该案例中在多相流体监控区域收集的视频数据。设计了对比实验以比较 C3D-APReLU 和使用 sigmoid、tanh、ReLU、LReLU 和 PReLU 激活函数的 C3D 在该数据集上的故障诊断性能，以验证提出的 APReLU-3D 对帮助 C3D 性能提升的有效性。实验还对比了 C3D-APReLU 与近年来较为流行的双流网络、ViViT 等视频分类模型的性能。

4.1. 数据描述

本研究使用的数据集是 PRONTO 异构基准数据集[20]中的视频监控数据，该数据集是从克兰菲尔德大学(Cranfield University)的工业级多相流设备中收集的。该设备是专门为研究由水、空气和油组成的多相流体在工业生产中的输送、测量和控制而设计的。该多相流设备流程图如图 7 所示。

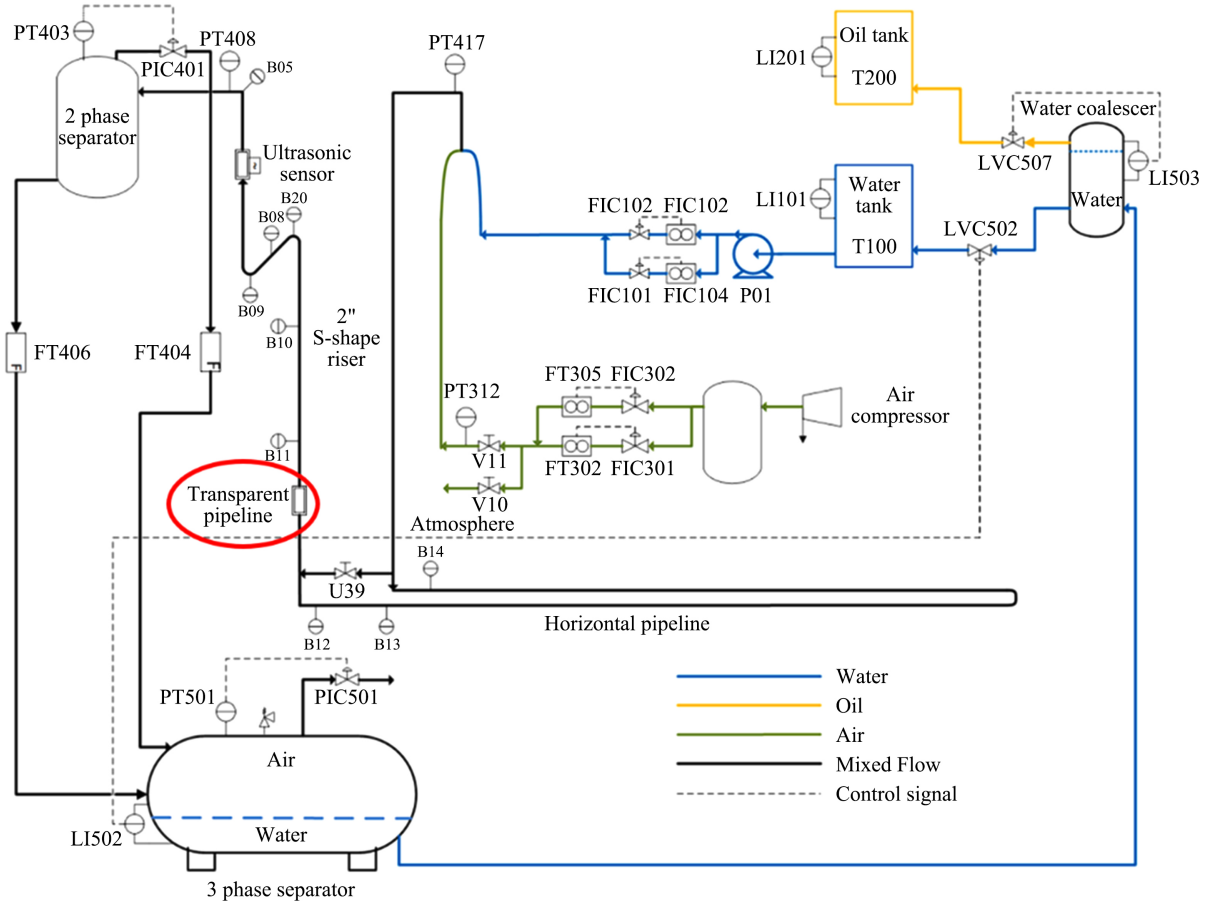
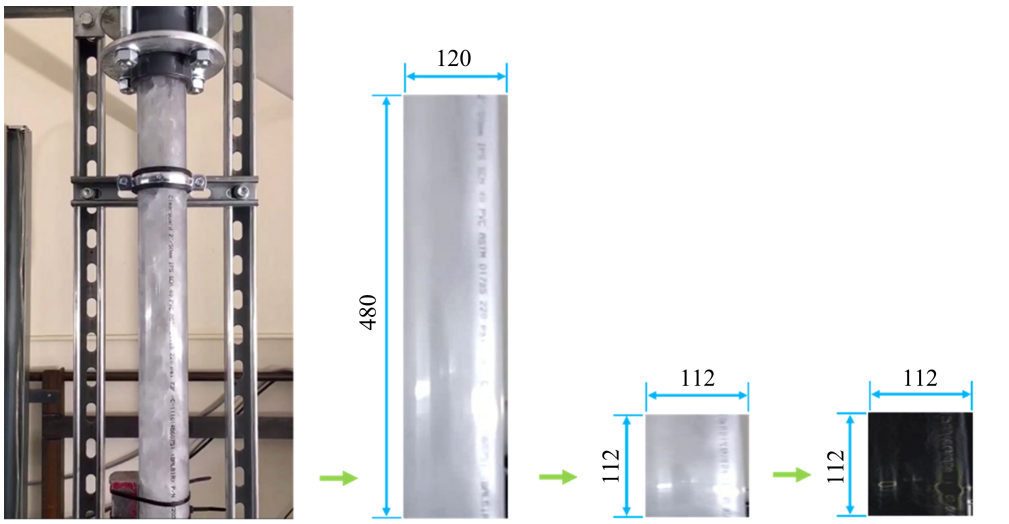


Figure 7. Flow chart of multiphase flow equipment
图 7. 多相流设备流程图

本次实验用到的多相流体的流动监控视频记录于图 7 中的红色框位置区域，该位置处有一段透明管道，可以方便的对流体流动状态进行观察和记录。在该工业案例研究中，一共记录了 3 种典型的多相流流动状态的视频数据，这 3 种流动状态来自于模拟的 3 种不同的设备真实运行场景，其中 2 个场景分别对应不同的设备故障情况，即管道的空气泄漏(Air leakage)和分流(Diverted flow)，另一个则对应设备正常运行的情况(Normal)。该数据集中 Air leakage 和 Diverted flow 两种类别的视频分辨率为 960 × 544，Normal 类别视频的分辨率为 1280 × 720。所有视频按照每秒 30 帧的标准进行录制。在该案例中，三个类别收集的视频时长分别为：Air leakage 121s，Diverted flow 287s，Normal 119s。图 8 为 Diverted flow 故障类别下，某段视频中连续的 10 帧图像。

Figure 8. 10 consecutive frames of images in a video under the diverted flow fault category
图 8. Diverted flow 故障类别下某段视频中连续的 10 帧图像

为了防止视频分辨率、背景和色温对分类的影响，需要对数据进行预处理，如图 9 所示。第一步使用 Adobe Premiere 软件对视频图像中的管道部分进行截取并缩放，将所有视频片段均处理为仅包含管道部分、分辨率为 480 × 120 大小的视频，如图 9(b)所示；第二步需要将经第一步处理得到的视频数据缩放至 C3D-APReLU 可以接受输入的数据尺寸，即高和宽均为 112，如图 9(c)所示；第三步对数据集进行数值归一化处理，首先计算出训练数据中所有视频的各个通道的特征图数值均值，然后训练数据和测试数据的每个通道都减去相应数值。如图 9(d)所示，图像数据归一化的目的是为了去除视频图像色温对分类的影响。



(a) 原视频图像　(b) 管道部分图像　(c) 管道缩放图像　(d) 对图像归一化

Figure 9. Video image preprocessing
图 9. 视频图像预处理

因为 C3D 模型处理的数据为连续 16 帧图像的视频片段，所以根据每类的视频时长和帧率可以截取得到每类的样本数分别为 Air leakage 类别 227 个，Diverted flow 类别 539 个，Normal 类别 224。每个样本的形状为(16, 3, 112, 112)。本文将所有数据按照 7:3 的比例随机划分为训练集和测试集，并随机划分 5 次进行多组重复实验。至此，该视频数据集预处理部分介绍完毕，所有信息整理如表 1 所示。

Table 1. PRONTO video dataset
表 1. PRONTO 视频数据集

类别	Normal	Air-leakage	Diverted-flow
视频时长	119s	121s	287s
视频帧率		30 帧/s	
样本尺寸		16 × 3 × 112 × 112	
样本数量	224	227	539
训练/测试样本量	157/67	159/68	377/162

4.2. 训练设置

完成数据预处理和划分之后，就可以对模型进行训练和测试。首先进行模型训练。运行本实验的硬件环境的 CPU 型号为 Intel(R) Xeon(R) Gold 5220R，GPU 型号为 Nvidia RTX3090。在训练所有模型时，使用随机梯度下降(SGD)算法对模型的参数进行更新，为了加快模型的训练，使用到了动量加速(Momentum)的训练技术。在本实验中，动量比按照 Deep Learning [21]文中所述被设为 0.9。为了抑制训练过程中出现的过拟合现象，应用了权重衰减，其通过在损失函数中增加惩罚项可以有效地将权重推向零。在本实验中，权重衰减系数设置为 0.0005。训练时每个小批量中的样本个数为 32 个。本实验采用可变学习率进行训练，初始学习率为 0.00001，之后随着训练迭代数增加而变化，其变化示意图如图 10 所示。

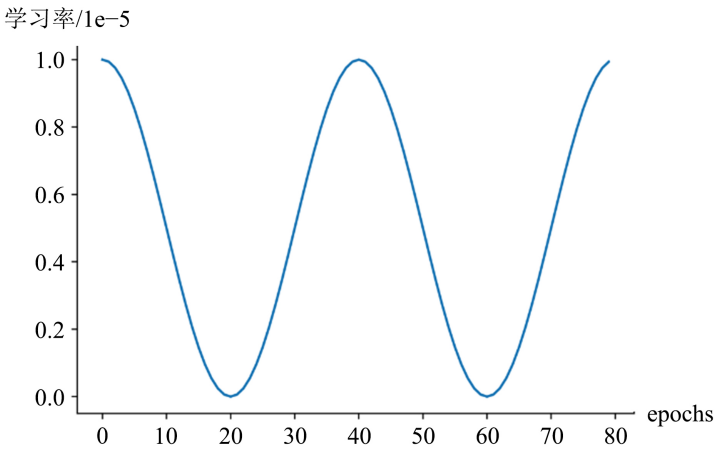


Figure 10. Schematic diagram of learning rate changes
图 10. 学习率变化示意图

4.3. 实验结果与分析

模型在训练集上训练完毕之后，进行其在测试集上的性能测试。本研究用到的性能度量指标包括精

度(accuracy)、查准率(precision)、查全率(recall)和 F1 度量(F1-score)。本研究所设置的实验主要是为证明本文提出的 C3D-APReLU 在 PRONTO 工业案例视频故障诊断场景的有效性及 APReLU-3D 在 C3D 中对比传统激活函数的优越性。

4.3.1. APReLU 与经典激活函数的对比

本节对比了 C3D-APReLU 和使用传统激活函数的 C3D 模型在以上数据集上的性能表现。每种方法模型参数均调至最优后对其进行 5 次重复实验，下表 2 展示了各项性能指标的平均值，其中加粗数字为相应指标下的最大值。

Table 2. The performance of C3D using different activation functions
表 2. 使用不同激活函数的 C3D 的性能表现

激活函数	性能指标			
	精度	查准率	召回率	F1-score
APReLU	**0.978**	**0.980**	**0.978**	**0.979**
ReLU	0.953	0.962	0.958	0.956
Sigmoid	0.421	0.435	0.406	0.420
Tanh	0.864	0.873	0.843	0.858
LPeLU	0.934	0.942	0.934	0.929
PReLU	0.967	0.969	0.963	0.966

从表中可以看出，APReLU 带给 C3D 模型的性能提升是显而易见的，其精度在 0.953 的准确度基数上提升了 2 个百分点，达到了 0.978 的卓越的性能表现。而对于使用其他传统激活函数的改进版本，APReLU 也具有优势。可以看到除了 APReLU 之外，使用 PReLU 的 C3D 的效果最好，精度达到了 0.967，但也与 C3D-APReLU 的 0.979 具有一定的差距。针对 C3D-APReLU 和 C3D-ReLU，还画出了相应的测试结果混淆矩阵，如图 11 所示。

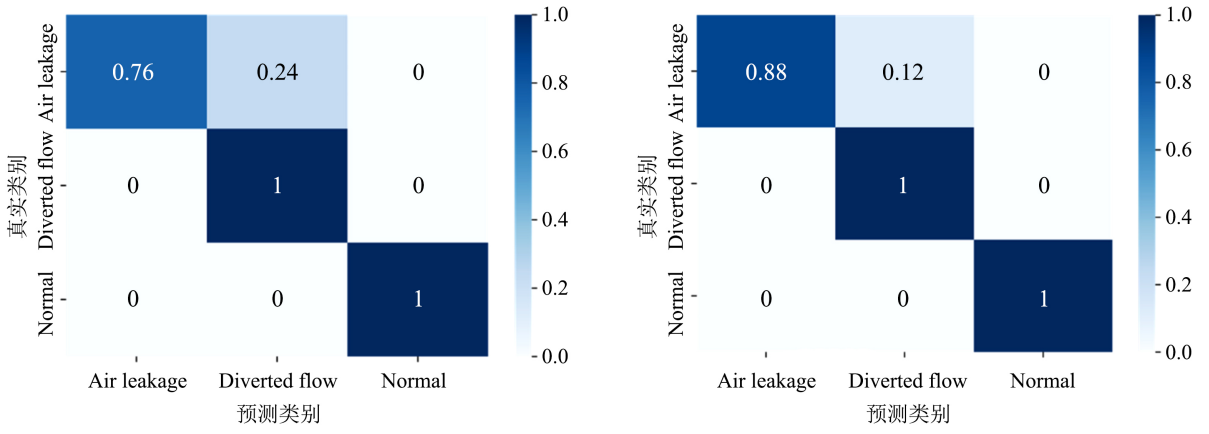
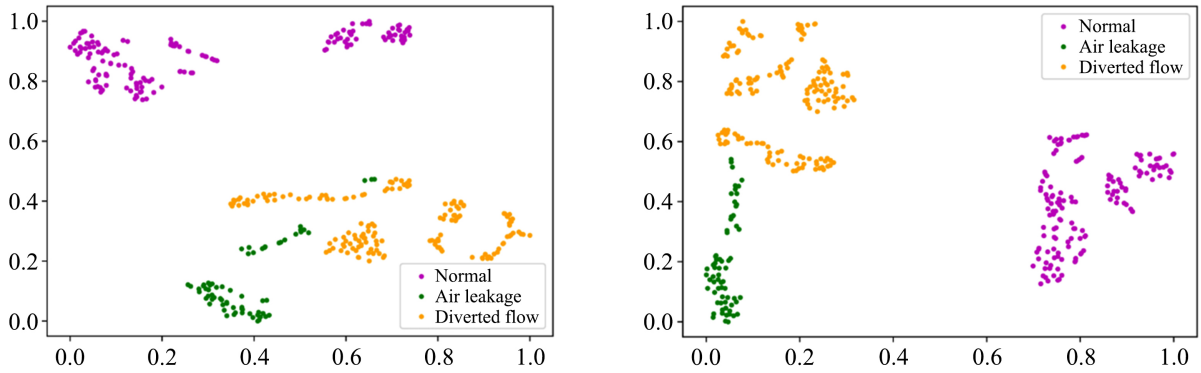


Figure 11. Test result confusion matrix of C3D using different activation functions
图 11. 使用不同激活函数的 C3D 测试结果混淆矩阵

从图中可以看到，Diverted flow 和 Normal 类别被分类的很好，能被全部分类正确。但 Air leakage 的分类精度不佳，会有一部分被分错至 Diverted flow 类别。其中 C3D-APReLU 对于 Air leakage 类别分类精度为 0.88，而 C3D-ReLU 则是 0.76。



(a) C3D-ReLU　　(b) C3D-APReLU

Figure 12. 2D visualization of output features of the final Fc layer in C3D using different activation functions
图 12. 使用不同激活函数的 C3D 最后 Fc 层输出特征的二维可视化

本研究还采用非线性降维方法 t 分布随机近邻嵌入(t-SNE) [22]将 Fc7 层的输出特征投影到二维空间以直观展示 APReLU 所带来的非线性映射效果。虽然低维特征在降维后通常会丢失大量信息，但二维可视化的目的是为了直观地了解学习到的特征，而不是将二维可视化用于故障分类。如图 12 所示，在 C3D-APReLU 中，各个类别的测试样本可以更为紧密的聚在一起，具有更好的可分性。而 C3D-ReLU 每个类别样本之间较为分散，且 Air leakage 和 Diverted flow 两种类别的样本出现了明显的交叠，可分性较差。以上可以证明 APReLU 所提出的自适应非线性映射是有效的。

4.3.2. C3D-APReLU 与其他方法的对比

本节比较了 C3D-APReLU 与其他 5 个视频分类模型在 PRONTO 视频数据集上的故障诊断效果。这 5 个方法的简单介绍如下所述：

- 双流网络：该方法对视频数据使用一个空间流卷积网络提取空间信息，同时使用一个时间流卷积网络提取时序信息，最后将二者融合作出分类决策；
- ResNet34LSTM：该方法是在 ConvLSTM [23]方法的基础上将普通的卷积神经网络替换成了经过 ImageNet 数据集预训练的 ResNet34 模型[4]；
- I3D [24]：该方法对经典的 2D 卷积图像分类模型进行改造，例如：AlexNet [5]、ResNet 等，将其中的 2D 卷积层替换为 3D 卷积层，实现了对视频数据的特征提取和分类；
- R(2 + 1)D [25]：该方法认为完整的 3D 卷积可以通过 2D 卷积和 1D 卷积来近似，于是将空间和时间建模分解为两个单独的步骤，由此设计了 R(2 + 1)D 方法；
- ViViT：该方法完全使用 Transformer 网络作为基础结构，对视频数据提取时空特征并分类。

对所有方法进行 5 次重复实验，结果记录于表 3 中。

从表中数据可以得知，C3D-APReLU 的性能仅次于双流网络。双流网络在此数据集上的性能表现最好，平均精度达到了 0.991。虽然双流网络的精度好于 C3D-APReLU，但由于在使用时要提取视频的光流特征，而提取光流特征是一件十分耗时的事情，当遇到大型数据集时，其时间成本会大大超出其精度优势，所以在实际应用场景中的实用性不如 C3D 等采用端到端训练方式的模型。

Table 3. The precision and mean of 5 repeated experiments of different methods
表 3. 不同方法 5 次重复实验的精度与均值

不同方法	精度					均值
	1	2	3	4	5	
C3D-APReLU	0.994	0.964	0.973	0.982	0.977	0.978
双流网络	0.988	0.996	0.992	0.987	0.992	**0.991**
ResNet34-LSTM	0.927	0.936	0.933	0.939	0.940	0.935
I3D	0.962	0.982	0.974	0.968	0.974	0.972
R(2 + 1)D	0.958	0.971	0.963	0.966	0.982	0.968
ViViT	0.945	0.956	0.946	0.937	0.926	0.942

5. 结论

本文提出了一种适用于三维卷积神经网络的、可以为输入信号做自适应非线性映射的激活函数APReLU-3D。该激活函数内嵌了一个可以对输入信号进行学习从而对坡度自动做出相应调整的子网络，使得每个输入信号都可以有自己的非线性变换。APReLU-3D 可以很方便的替换三维卷积后面的传统激活函数，而无需对原网络做任何修改。本文将 APReLU-3D 应用于 C3D 视频分类模型中，提出了 C3D-APReLU。对 PRONTO 工业数据集中的视频数据进行的故障诊断实验表明，APReLU-3D 的自适应非线性映射能够让 C3D-APReLU 对比使用传统激活函数的 C3D 实现更好的分类效果。但是，本文的不足之处在于，使用的视频数据集类别太少，缺乏一个大规模、多类别的数据集。这不仅是本文面临的问题，也是工业视频故障检测领域面临的问题之一。

基金项目

本工作受国家自然科学基金(61903251)资助。

参考文献

[1] 欧敬逸, 田颖, 向鑫, 宋启哲. 基于迁移 BN-CNN 框架的小样本工业过程故障诊断[J]. 电子科技, 2022.

[2] Itti, L., Koch, C. and Niebur, E. (1998) A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 1254-1259. https://doi.org/10.1109/34.730558

[3] 贾澎涛, 杨丽娜. 基于多特征的视频场景分类[J]. 计算机应用研究, 2018, 35(11): 3472-3475.

[4] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[5] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. https://doi.org/10.1145/3065386

[6] Karpathy, A., Toderici, G., Shetty, S., *et al*. (2014) Large-Scale Video Classification with Convolutional Neural Networks. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1725-1732. https://doi.org/10.1109/CVPR.2014.223

[7] Simonyan, K. and Zisserman, A. (2014) Two-Stream Convolutional Networks for Action Recognition in Videos. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. and Weinberger, K.Q., Eds., *Advances in Neural Information Processing Systems* 27 (*NIPS* 2014), Curran Associates, Inc., Red Hook.

[8] Ji, S., Xu, W., Yang, M. and Yu, K. (2012) 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 221-231. https://doi.org/10.1109/TPAMI.2012.59

[9] He, M., Li, B. and Chen, H. (2017) Multi-Scale 3D Deep Convolutional Neural Network for Hyperspectral Image Classification. 2017 *IEEE International Conference on Image Processing* (*ICIP*), Beijing, 17-20 September 2017, 3904-3908. https://doi.org/10.1109/ICIP.2017.8297014

[10] Li, Y., Zhang, H. and Shen, Q. (2017) Spectral-Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sensing*, **9**, Article No. 67. https://doi.org/10.3390/rs9010067

[11] Tran, D., Bourdev, L., Fergus, R., Torresani, L. and Paluri, M. (2015) Learning Spatiotemporal Features with 3D Convolutional Networks. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 4489-4497. https://doi.org/10.1109/ICCV.2015.510

[12] Qiao, Y., Guo, Y., Yu, K. and He, D. (2022) C3D-ConvLSTM Based Cow Behaviour Classification Using Video Data for Precision Livestock Farming. *Computers and Electronics in Agriculture*, **193**, Article ID: 106650. https://doi.org/10.1016/j.compag.2021.106650

[13] 李燕, 何敏. 基于 C3D 和 CBAM-ConvLSTM 的犯罪事件视频场景分类[J]. 刑事技术, 2022, 47(5): 448-457.

[14] Xu, H., Das, A. and Saenko, K. (2017) R-C3D: Region Convolutional 3D Network for Temporal Activity Detection. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 5794-5803. https://doi.org/10.1109/ICCV.2017.617

[15] Zhao, M., Zhong, S., Fu, X., *et al*. (2020) Deep Residual Networks with Adaptively Parametric Rectifier Linear Units for Fault Diagnosis. *IEEE Transactions on Industrial Electronics*, **68**, 2587-2597. https://doi.org/10.1109/TIE.2020.2972458

[16] Nair, V. and Hinton, G.E. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the* 27*th International Conference on International Conference on Machine Learning*, Haifa, 21-24 June 2010, 807-814.

[17] Maas, A.L., Hannun, A.Y. and Ng, A.Y. (2013) Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the* 30*th International Conference on Machine Learning*, 16-21 June 2013, Atlanta.

[18] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 7-13 December 2015, 1026-1034. https://doi.org/10.1109/ICCV.2015.123

[19] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proceedings of the* 32*nd International Conference on Machine Learning*, Lille, 6-11 July 2015.

[20] Stief, A., Tan, R., Cao, Y., *et al*. (2019) A Heterogeneous Benchmark Dataset for Data Analytics: Multiphase Flow Facility Case Study. *Journal of Process Control*, **79**, 41-55. https://doi.org/10.1016/j.jprocont.2019.04.009

[21] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. https://doi.org/10.1038/nature14539

[22] van der Maaten, L. and Hinton, G. (2008) Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, **9**, 2579-2605.

[23] Shi, X., Chen, Z., Wang, H. and Yeung, D.-Y. (2015) Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R., Eds., *Advances in Neural Information Processing Systems* 28 (*NIPS* 2015), Curran Associates, Inc., Red Hook.

[24] Carreira, J. and Zisserman, A. (2017) Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 4724-4733. https://doi.org/10.1109/CVPR.2017.502

[25] Tran, D., Wang, H., Torresani, L., *et al*. (2018) A Closer Look at Spatiotemporal Convolutions for Action Recognition. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6450-6459. https://doi.org/10.1109/CVPR.2018.00675.