一种基于 DWConvLSTM 与局部敏感哈希注意力的视频摘要方案

朱 頔[1,2] 滕晓宇[1,2] 王 刚[1,2] 许文丽[1,2] 樊懿雯[1,2] 何 勇[3] 张 清[1,2]
ZHU Di TENG Xiaoyu WANG Gang XU Wenli FAN Yiwen HE Yong ZHANG Qing

摘 要 针对现有视频摘要算法存在特征敏感性不足、特征提取不够细腻、算法复杂性高等问题，文章基于深度可分离卷积（DWConv）、卷积长短期记忆网络（ConvLSTM）、多头注意力（Multi-Attention）与局部敏感哈希（LSH），共同设计一种轻量级视频摘要方案（DWCH-Attention）。在这一方案中，为降低算法复杂性，通过改进 ConvLSTM 并结合 DWConv 搭建 DWConvLSTM，再将其与注意力机制结合以提取全局特征；为提升特征提取细腻程度及算法对特征的敏感程度，借助局部敏感哈希（LSH）与交叉注意力机制设计局部特征提取方法；为进一步强化特征，设计以查询为驱动的特征融合方法，实现全局与局部特征的融合。为验证方案的有效性与可行性，将方案在 TVSum 与 SumMe 两个数据集上开展实验验证，结果表明，该方案在交叉验证、消融实验及对比分析中均表现出较好的性能。

关键词 DWConv；ConvLSTM；Multi-Attention；局部敏感哈希；视频摘要

doi：10.3969/j.issn.1672-9528.2025.09.015

0 引言

随着 5G 网络在下沉市场降价提速，6G 网络研究加速，以及智能终端技术的突破性发展，诱发了视频数据量的新一轮爆炸式增长的同时刺激了以应用、社交为主的短视频网站、手机 APP 的快速发展，国外有 NetFlix、Vimeo、Hulu、TubiTV、Twitch 与 YouTube 等，国内有抖音、快手、西瓜视频、bilibili、微视、好看、秒拍、全民小视频、小红书与美拍等。以短视频为主的微短剧逐渐成为娱乐的主流，也标志着用户对视频内容短而精的需求越来越大。但随着视频创作逐渐趋向无门槛化，创作群体趋向全面化，视频数据量不断增加，视频内容也不断丰富，但是不变的是用户对信息的接受能力，另外，随着生活节奏的不断加快，用户阅读视频时间逐渐趋向于碎片化，阅读方式逐渐趋向多元化，这就造成了用户所能接收的信息量与海量视频数据之间的极度不平衡，数据量增加在一定程度上制约了用户对有效信息的管理、检索能力。

1. 西安财经大学信息学院 陕西西安 710100
2. 智财协同可信计算陕西省高等学校重点实验室 陕西西安 710100
3. 中国电建集团西北勘探设计研究院有限公司 陕西西安 710065

[基金项目] 陕西省自然科学基础研究计划项目（2023-JC-ZD-38）；陕西省社会基金项目（2020F002）；陕西省提升公众科学素质研究计划项目（2021PSL09）

视频摘要算法作为一种简化视频数据的算法，已经成为抽取视频部分重要信息，解决上述问题的主要方式之一。现有针对视频摘要的研究已有不少成果，并广泛应用于教育、影视、监控、工业等领域中。其算法框架主要由预处理、特征提取、关键帧 / 镜头的筛选、摘要生成四部分组成，在基于深度学习的方法中，主要有 Auto-Encode[1]、深度卷积神经网络 [2]（deep convolutional neural networks, DCNN）、长短时记忆网络 [3]（long short-term memory, LSTM）、图神经网络 [4]、注意力机制 [5] 等，但随着创作视频类型与风格多元化，视频内容冗余量加剧，致使视频摘要算法面临着如下挑战：

（1）在算法复杂性上，现有算法为了充分提取高冗余的视频特征，搭建复杂网络牺牲时间空间为代价；

（2）在特征提取上，题材多样致使特征分布的多样性，致使现有算法在特征敏感性上存在不足。

基于上述不足，本文设计了一种轻量级视频摘要方案，在该方案中，通过借助 DWConv 与 ConvLSTM 设计 DW-ConvLSTM 降低算法时间复杂度，通过借助将局部敏感哈希（locality-sensitive hashing, LSH）思想嵌入到交叉多注意力筛选中，设计局部差异化注意力提取，通过结合全局注意力增加算法对视频特征的敏感性的同时，进一步强化关键帧，达到提升特征提取力度的目的。结合全局与局部特征提取，不仅可以从时间和空间两个层面更好地提取满足生成摘要的特征提高算法性能的同时，还可以降低时间复杂度。

1 DWH-Attention

本方案主要分为基于 DWConvLSTM 与多头注意力机制的全局特征提取，基于 ConvLSTM、局部敏感哈希、多头注意力的局部特征提取以及基于查询的特征融合，3 个模块组成，具体如图 1 所示。

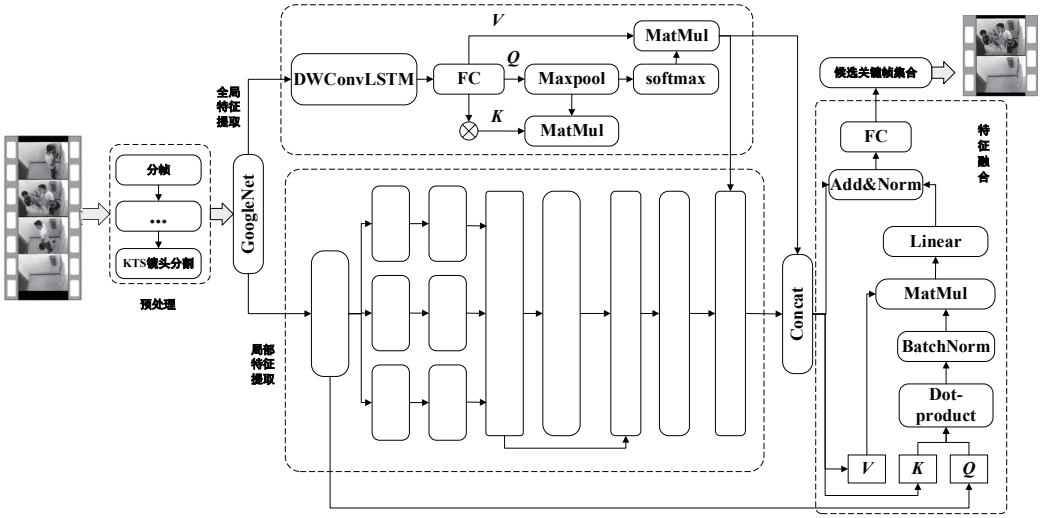

图 1 基于 DWH-Attention 视频摘要模型图

1.1 DWConvLSTM+Multi-Attention 全局特征提取

该部分主要完成从全局角度上提取特征，为降低模型的时间复杂度借助深度可分离卷积 DWConv[6] 改进 ConvLSTM[7]，并结合多头注意力共同提取特征。深度可分离卷积 DWConv 由逐深度卷积和逐点卷积组成，深度卷积用于提取空间特征，逐点卷积用于提取通道特征。深度可分离卷积在特征维度上分组卷积，对每个通道进行独立的逐深度卷积（depthwise convolution），并在输出前使用一个 1×1 卷积（pointwise convolution）将所有通道进行聚合。其中逐深度卷积（depthwise convolution,DWConv）与标准卷积的区别在于，深度卷积的卷积核为单通道模式，需要对输入的每一个通道进行卷积，这样就会得到和输入特征图通道数一致的输出特征图。DWConv 与传统的卷积 Conv 相比，优势在于在参数上，可减少输入通道数量，从而有效地减少卷积层所需的参数；运行速度要比传统卷积快，计算量更小，更易于实现和部署在不同的平台上，另外还能够精简计算模型，从而在较小的设备上实现高精度的运算。

与此同时，ConvLSTM 是一种结合卷积神经网络（CNN）和长短时记忆网络（LSTM）的架构，专门用于处理时序数据，可以充分提取视频的时空特性。因此利用 DWConv 替换 ConvLSTM 中传统卷积部分，有利于提升精度、模型可迁移性的同时，降低算法的时间复杂度。用公式表示为：

$$\boldsymbol{f}_t = \sigma_g\left(\boldsymbol{W}_{\mathrm{xf}} * \boldsymbol{X}_t + \boldsymbol{W}_{\mathrm{hf}} * \boldsymbol{H}_{t-1} + \boldsymbol{W}_{\mathrm{cf}} \circ \boldsymbol{C}_{t-1} + \boldsymbol{b}_f\right) \tag{1}$$

$$\boldsymbol{i}_t = \sigma_g\left(\boldsymbol{W}_{\mathrm{xi}} * \boldsymbol{X}_t + \boldsymbol{W}_{\mathrm{hi}} * \boldsymbol{H}_{t-1} + \boldsymbol{W}_{\mathrm{ci}} \circ \boldsymbol{C}_{t-1} + \boldsymbol{b}_i\right) \tag{2}$$

$$\boldsymbol{C}_t = \boldsymbol{f}_t \circ \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \circ \tanh_g\left(\boldsymbol{W}_{xc} * \boldsymbol{X}_t + \boldsymbol{W}_{\mathrm{hc}} * \boldsymbol{H}_{t-1} + \boldsymbol{b}_c\right) \tag{3}$$

$$\boldsymbol{o}_t = \sigma_g\left(\boldsymbol{W}_{\mathrm{xo}} * \boldsymbol{X}_t + \boldsymbol{W}_{\mathrm{ho}} * \boldsymbol{H}_{t-1} + \boldsymbol{W}_{\mathrm{co}} \circ C_t + \boldsymbol{b}_o\right) \tag{4}$$

$$\boldsymbol{H}_t = \boldsymbol{o}_t \circ \tanh_g\left(\boldsymbol{C}_t\right) \tag{5}$$

式中：σ_g 为 sigmoid 激活函数；$\tanh_g$ 为双曲正切激活函数；$\boldsymbol{H}_{t-1}$ 为隐藏状态；$\boldsymbol{C}_{t-1}$ 为记忆单元；* 为卷积操作，在 DWConvLSTM 中，该操作替换成 DWConv；$\boldsymbol{f}_t$、$\boldsymbol{i}_t$、$\boldsymbol{C}_t$、$\boldsymbol{o}_t$ 分别表示遗忘门、输入门、记忆单元和输出门。

另外，注意力机制因其模拟人类筛选关键帧的方法，从而增强感兴趣的特定目标区域同时弱化不相关的背景区域，将注意力机制应用到特征提取中，能够更深层次的提取特征，另一方面，视频摘要主要依据为关键帧的打分，关键帧的打分依据为特征的权重，因此本方案将通过 DWConvLSTM 得到的时空特征，在此输入到多头注意力机制，进而可得全局特征矩阵。用公式表示为：

$$\boldsymbol{Y}_i = \mathrm{FC}\left(\mathrm{ConvLSTM}\left(\boldsymbol{X}_i\right)\right) = \boldsymbol{V} = \boldsymbol{Q} \tag{6}$$

$$\boldsymbol{K} = \boldsymbol{Y}_i \oplus \boldsymbol{F}_w \tag{7}$$

$$\boldsymbol{G} = (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{Soft\,max}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{8}$$

式中：$\oplus$为融合操作；$\boldsymbol{F}_w$ 为反馈模块的特征矩阵；$\boldsymbol{X}_i$ 为经过预处理操作后的特征矩阵。

1.2 LSH+Multi-Attention 局部特征提取

为了增加算法对视频特征变化的敏感性的同时突出关键信息，受到稀疏注意力机制的启发，借助局部敏感哈希与多头注意力机制共同设计局部特征提取模块。局部敏感哈希能够在一定程度上聚合相似度较高的特征值，相当于对特征值进行分类，该分类可以用于多注意力机制中的降维操作。另外，为更加细腻特征与结合全局与局部特征提取，采用查询为驱动的交叉多头注意力，进一步细化特征，其实现过程如图 2 所示。用 $\boldsymbol{C}$ 代表全局特征提取结果，具体过程为：

$$\boldsymbol{C} = \boldsymbol{Q} \tag{9}$$

$$\boldsymbol{K} = \mathrm{LSH}\left(\mathrm{ConvLSTM}\left(\boldsymbol{X}_i\right)\right) \tag{10}$$

$$\boldsymbol{V} = \mathrm{ConvLSTM}\left(\boldsymbol{X}_i\right) \tag{11}$$

$$\boldsymbol{Y} = (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{Soft\,max}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{12}$$

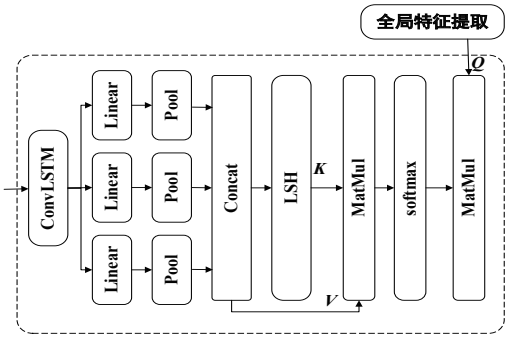

图 2 局部特征提取

1.3 基于交叉注意力的特征融合

为更好地融合特征，本文采用以查询为驱动的交叉注意力机制融合全局与局部特征：

$$\boldsymbol{Q} = \mathrm{ConvLSTM}(\boldsymbol{X}_i) \tag{13}$$

$$\boldsymbol{V} = \boldsymbol{K} = \mathrm{Concat}(\boldsymbol{C}, \boldsymbol{L}) \tag{14}$$

$$\boldsymbol{Z} = (\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{Soft\,max}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{15}$$

式中：$\boldsymbol{C}$ 表示全局特征提取后的结果；$\boldsymbol{L}$ 表示局部特征提取后的特征矩阵；$\boldsymbol{Z}$ 表示融合结果。

2 实验分析验证

针对所提出的 DWCH-Attention 视频摘要方案进行验证与分析，为充分验证方案的可行性与有效性，本文主要从方案自身验证以及与目前领域内相对较为先进的算法对比验证两个部分进行展开。

2.1 实验设置

2.1.1 实验数据集

本研究选用了视频摘要领域内最为常用的两个公开数据集 SumMe[8]，此两个数据库的具体介绍如表 1 所示。

表 1 实验数据集说明

数据集	描述
TVSum	TVSum 中共有 50 个视频，它将视频内容分为了新闻、操作指南、纪录片等 10 个类别，每类有 5 个。每个视频由 20 人标注，标注的类型为多组关键片段
SumMe	SumMe 中共有 25 个视频，内容大致包含假期，新闻和运动等。视频时长都在 1~6 min 之间。每个视频由 15~18 人标注，标注的类型为多组关键片段，这些摘要长度为原始视频长度的 5%~15% 之间

TVSum[9] 与 SumMe 两个数据集合通常用在检测监督学习下的视频摘要算法性能。另外在训练过程中涉及到 OVP（Open Video Project）[10] 和 YouTube 两个数据集，OVP 和 YouTube 均包含 50 个视频，视频注释是由 5 个用户生成的一组关键帧，其中 OVP 的视频时长为 1~4 min，YouTube 的视频时长为 1~10 min。

2.1.2 评价指标

为了能够与现有的算法进行合理化比较，本章采用最为常见的评价指标 Precision、Recall 和 F-score。设 S 为摘要算法生成的数据，G 为数据库中用户标注的背景数据，则 Precision、Recall 和 F-score[11] 的定义为：

$$\mathrm{Precision} = \frac{|S \cup G|}{|S|} \tag{16}$$

$$\mathrm{Recall} = \frac{|S \cup G|}{|G|} \tag{17}$$

$$F\text{-score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{18}$$

但 Fontes[12] 发现用随机算法来做视频摘要任务，其效果可以目前最好的算法提并论，即单纯以 F-score 为衡量可信度下降。本文所设置的交叉验证实验如表 2 所示。

表 2 交叉实验设置

数据集	设置	训练集	测试集
TVSum	C	80% TVSum	剩余的 20%TVSum
	A	80%TVSum + SumMe + OVP + YouTube	剩余的 20% TVSum
	T	SumMe + OVP + YouTube	TVSum
SumMe	C	80% SumMe	剩余的 20% SumMe
	A	TVSum + 80%SumMe + OVP + YouTube	剩余的 20% SumMe
	T	TVSum + OVP + YouTube	SumMe

注：C—Canonical；A—Augmented；T—Transfer

2.1.3 实验环境

本文实验操作平台为 PyTorch，ConvLSTM 维度为 256，参数学习率初始化为 5，batch 大小为 5，kernel-size 设置为 (5,1)，在训练过程中最大训练轮次 Epoch 设置为 100。

2.2 实验验证

为充分检验模型的有效性与可用性，本文首先对模型自身进行实验验证，然后与领域内其他算法进行对比分析。

2.2.1 模型自身验证

该部分首先进行数据集内独立验证，然后再进行不同数据集内交叉验证，采用 F-score 作为衡量标准，表 3 为本文所提出的摘要算法在 TVSum 与 SumMe 两个数据集上的验证结果。

表 3 方案在 TVSum 与 SumMe 两个数据集上性能分析

单位：%

	TVSum（F-score）	SumMe F-score
Max	62.13	58.6
Min	58.2	41.53
Average	61.65	53.2

在此验证过程中，将每个数据集拆成 3 部分，60% 作为训练集，20% 作为测试集，剩余 20% 作为调整。从表 3 中最大值、最小值与平均值分析可知，本文所提出的视频摘要算法在两个数据集上，性能表现优越，为进一步检验算法的性能，采用表 2 的实验设置对算法进行交叉验证，验证结果如表 4 所示。

表 4 方案交叉验证分析

单位：%

	TVSum（*F*-score）			SumMe（*F*-score）		
	C	A	T	C	A	T
Max	64.40	67.14	63.97	57.05	52.8	53.02
Min	58.21	53.06	51.20	45.60	46.41	47.00

从表 4 中可知，方案在 TVSum 数据集上，Canonical、Augmented、Transfer 三种交叉验证 *F*-score 的最大、最小值都在 50% 以上，在 SumMe 数据集上，3 种交叉验证，*F*-score 的最大最小都在 45% 以上，由此可知本文所设计的方案是可行的，所生成的摘要是有效的。

2.2.2 对比分析验证

对特征进行细化角度上，基于 LSTM 的较为经典的算法有 vsLSTM[13] 与 dppLSTM，在 RNN 上，较为经典的算法为 H-RNN[14]，另外，文献 [13] 给出了一种借助深度金字塔网络，提取多尺度特征的方法。文献 [15] 提出了一种新颖的注意力引导多粒度融合摘要算法，有效模拟上下文信息，感知内容；涵盖时空特征的文献 [16]。在监督环境下，实验结果对比如表 6 所示。

表 6 *F*-score 对比分析

单位：%

模型	TVSum（*F*-score）	SumMe（*F*-score）
VsLSTM[13]	54.2	41.6
DppLSTM	59.6	42.9
H-RNN[14]	61.9	43.8
文献 [15]	62.4	51.9
文献 [16]	61.0	51.8
文献 [4]	61.5	51.7
文献 [17]	59.1	48.0
本文（最大）	62.13	58.6
本文（平均）	61.65	53.2

由表 6 可知，本文所设计的方案相对于其他相似方案，虽然在平均值方面不低于 H-RNN 方案，但是在最大值以及 SumMe 数据集上，相对于 H-RNN 有所提升，表明了本文方案的有效性以及可用性。

综上所述，从方案的自身验证与对比分析验证结果可知，本文所提出的方案在摘要提取方面具有可用性与有效性，相对于其他类似方案，具有一定的优势。

3 总结

本文针对现有视频摘要算法存在算法复杂性高，特征敏感性不足，特征提取力度不够细腻等问题，基于 DWConv、ConvLSTM、Multi-Attention、局部敏感哈希共同设计一种轻量级的视频摘要算法。从架构上，该算法分为全局特征提取、局部特征提取与特征融合三部分。在全局特征提取中，通过结合 DWConv 与 ConvLSTM 构建 DWConvLSTM，并结合多头注意力共同完成；在局部特征提取中，为提升特征提取的细腻程度以及对算法对特征的敏感程度，借助局部敏感哈希（LSH）与交叉注意力机制设计局部特征提取方法；在特征融合模块，为进一步强化特征，设计以查询为驱动的特征融合方法。经过实验验证与分析可知，本文方案具有可用性与有效性。但是如何提升算法的泛化性仍旧是下一步面临的问题。

参考文献：

[1] MA M Y, MEI S H, WAN S, et al. Graph convolutional dictionary selection with $L_{2,p}$ norm for video summarization[J]. IEEE transactions on image processing, 2022, 31:1789-1804.

[2] JI Z, XIONG K L, PANG Y W, et al. Video summarization with attention-based encoder–decoder networks[J]. IEEE transactions on circuits and systems for video technology, 2019, 30(6): 1709-1717.

[3] APOSTOLIDIS E, ADAMANTIDOU E, METSAI A I, et al. AC-SUM-GAN: connecting actor-critic and generative adversarial networks for unsupervised video summarization[J]. IEEE transactions on circuits and systems for video technology, 2020, 31(8): 3278-3292.

[4] LIANG G Q, LÜ Y B, LI S C, et al. Video summarization with a dual-path attentive network[J]. Neurocomputing, 2022, 467: 1-9.

[5]TENG X Y, GUI X L, XU P, et al. A multi-flexible video summarization scheme using property-constraint decision tree[J]. Neurocomputing, 2022, 506: 406-417.

[6] DU Y, WU T, DAI Z F, et al. F-YOLOv7: fast and robust real-time UAV detection[J/OL].Computing, 2025[2025-07-02]. https://link.springer.com/article/10.1007/s00607-024-01406-7.

[7] 黄金贵，黄一举 . 基于注意力时空解耦 3D 卷积 LSTM 的视频预测 [J]. 微电子学与计算机 ,2022,39(9):63-72.

[8]LI Y, MERIALDO B. Multi-video summarization based on AV-MMR[C]//2010 International Workshop on Content Based Multimedia Indexing (CBMI). Piscataway:IEEE, 2010: 1-6.

[9]SONG Y L, VALLMITJANA J, STENT A, et al. TVSum: summarizing web videos using titles[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: IEEE,2015: 5179-5187.

[10]GYGLI M, GRABNER H, RIEMENSCHNEIDER H, et al. Creating summaries from user videos[C]//Computer Vision–ECCV 2014. Piscataway:IEEE,2014: 505-520.

[11]ZHAO B, LI X L, LU X Q. Property-constrained dual learning for video summarization[J]. IEEE transactions on neural networks and learning systems, 2019, 31(10): 3989-4000.

[12]FONTES DE AVILA S E, BRANDÃO LOPES A P, DA LUZ A, et al. VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method[J]. Pattern recognition letters, 2011, 32(1): 56-68.

[13]ZHANG K, CHAO W L, SHA F, et al. Video summarization with long short-term memory[C]//Computer Vision – ECCV 2016. Berlin:Springer,2016: 766-782.

[14]ZHAO B, LI X L, LU X Q. Hierarchical recurrent neural network for video summarization[C]//Proceedings of the 25th ACM international conference on Multimedia. NewYork: ACM, 2017: 863-871.

[15]KHAN H , HUSSAIN T , KHAN S U ,et al.Deep multi-scale pyramidal features network for supervised video summarization[J].Expert systems with application, 2024,237:121288.

[16]ZHONG R, WANG R, ZOU Y, et al. Graph attention networks adjusted Bi-LSTM for video summarization[J]. IEEE signal processing letters, 2021, 28: 663-667.

[17]WU J X, ZHONG S H, JIANG J M, et al. A novel clustering method for static video summarization[J].Mutimedia tools and applications, 2017, 76(7): 9625-9641.

【作者简介】

朱頔（2002—），男，陕西宝鸡人，硕士研究生，研究方向：工业视频摘要。

滕晓宇（1991—），女，山东临沂人，博士，讲师，研究方向：视频摘要、深度学习、工业信息化、可信计算。

王刚（1974—），男，江苏丰县人，博士，教授，研究方向：大数据分析、可信计算、深度学习。

许文丽（1970—），女，山东德州人，博士，副教授，研究方向：大数据分析、数据挖掘、机器学习。

樊懿雯（2003—），女，陕西渭南人，本科，研究方向：工业视频摘要。

何勇（1986—），男，甘肃会宁人，本科，高级工程师，研究方向：工程数字化、工程大数据应用。

张清（1998—），女，山西运城人，硕士，讲师，研究方向：隐私计算、深度学习。

（收稿日期：2025-03-31 修回日期：2025-09-04）

（上接第 63 页）

指导平台建设研究 [J]. 计算机应用与软件 ,2022,39(6):76-81.

[15] 许侃 , 吴鑫卓 , 林原 , 等 . 基于对抗型排序学习的混合推荐算法 [J]. 山西大学学报 (自然科学版),2024,47(3):481-493.

[16] GENUER R, POGGI J M, TULEAU-MALOT C.Variable selection using random forests[J].Pattern recognition letters, 2010, 31(14): 2225-2236.

[17] OTOK B W, MUSA M, PURHADI, et al.Propensity score stratification using bootstrap aggregating classification trees analysis[J]. Heliyon , 2020,6(7): e04288.

[18] MIJANUR RAHMAN M, OAISHI C R,MAHMUD J, et al. Elective course recommendation system using SVD algorithm[C]//Congress on Smart Computing Technologies. Berlin:Springer,2025:461–475.

[19] ABHINAV N, SUJATHA K. Content-based movie recommendation system using cosine similarity measure[C/OL]//Third International Conference on Advances in Physical Sciences and Materials: Icapsm 2022. New York: AIP Publishing, 2022[2025-05-13]. https://doi.org/10.1063/5.0178819.

[20] ELREEDY D, ATIYA A F, KAMALOV F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning[J].Machine learning, 2024, 113: 4903-4923.

[21] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. NewYork: ACM, 2016: 785–794.

【作者简介】

涂辉（1987—），男，江苏徐州人，硕士，工程师，研究方向：职业教育、机器学习。

许腾飞（1987—），男，江苏徐州人，硕士，讲师，研究方向：职业教育、人工智能。

丁中燕（1991—），女，江苏盐城人，硕士，助理研究员，研究方向：学生管理、职业教育。

张正金（1984—），男，安徽合肥人，博士，讲师，研究方向：人工智能、因果推理。

（收稿日期：2025-04-16 修回日期：2025-09-10）