一种基于深度学习的视惯融合 SLAM 方法

张嘉翔，叶平

（北京邮电大学人工智能学院，北京 100876）

摘要：近年来，随着高科技领域相关技术的快速发展，移动机器人愈发成为社会生产生活中不可或缺的得力助手。由于 SLAM 技术与移动机器人之间存在非常紧密的联系，因而受到广泛关注。目前，以视觉 SLAM 为主，融合 IMU（惯性测量单元）完成定位与建图工作的视惯融合 SLAM 已经被证明具备较强的鲁棒性和实用性，同时成本低廉。因此，本文以一类较为经典且开源的视惯融合 SLAM 算法 VINS-Mono 为主要研究对象，首先使用深度学习领域的一类卷积神经网络 SuperPoint 提取图像特征点，并将 VINS-Mono 的闭环检测模块本身提供的 FAST 角点替换为 SuperPoint 特征点，有效提升了视觉点特征的鲁棒性和分布的均匀性。同时，考虑到 SuperPoint 特征描述符在可区分性方面依然具有提升的空间，因此，本文在 VINS-Mono 的闭环检测模块引入一种轻量型网络 FeatureBooster，相关实验结果证明，若使用经过 FeatureBooster 处理的局部描述符进行闭环帧间的特征匹配，则能够获取准确性更高的特征匹配结果，从而改善了视惯融合 SLAM 算法定位与建图的精度和鲁棒性，体现了 FeatureBooster 对于 SLAM 算法的重要意义。

关键词：控制科学与工程；视惯融合 SLAM；SuperPoint；视觉点特征；FeatureBooster；移动机器人

中图分类号：TP242.6

A Visual Inertial SLAM Method Based on Deep Learning

Jiaxu Zhang, Ping Ye

(School of Artificial Intelligence，Beijing University of Posts and Telecommunications，Beijing 100876)

Abstract: In recent years, with the rapid development of related technologies in the high-tech field, mobile robots have become an indispensable assistant in social production and life. Due to the close relationship between SLAM technology and mobile robots, it has received widespread attention. At present, the visual inertial fusion SLAM, which mainly relies on visual SLAM and integrates IMU (Inertial Measurement Unit) to complete positioning and mapping work, has been proven to have strong robustness and practicality, while also being low-cost. Therefore, this article focuses on a classic and open-source visual inertial fusion SLAM algorithm VINSMono as the main research object. Firstly, a type of convolutional neural network SuperPoint in the field of deep learning is used to extract image feature points, and the FAST corner points provided by VINSMono's loop-closed detection module are replaced with SuperPoint feature points, effectively improving the robustness and distribution uniformity of visual point features. At the same time, considering that SuperPoint feature descriptors still have room for improvement in distinguishability, this paper introduces a lightweight network FeatureBooster in the loop-closed detection module of VINSMono. The relevant experimental results show that if local descriptors processed by FeatureBooster are used for loop-closed inter frame feature matching, higher accuracy of feature matching results can be obtained, which improves the accuracy and robustness of the localization and mapping of the visual inertial fusion SLAM algorithm, thusreflecting the important significance of FeatureBooster for SLAM algorithms.

Keywords: Control Science and Engineering; visual-inertial fusion SLAM; SuperPoint; visual point features; FeatureBooster; mobile robot

作者简介：张嘉翔（1996-），男，硕士研究生，主要研究方向：智能机器人

通信联系人：叶平（1979-），男，副教授，博士，主要研究方向：机器人智能感知与控制、Visual SLAM、智能信号分析与处理. E-mail: yeping@bupt.edu.cn

0 引言

近年来，随着深度学习领域相关技术的快速发展，使用卷积神经网络实现视觉点特征提取的方案逐渐被越来越多 SLAM[1]领域的科研人员所重视，主要是因为在光照强度变化明显、相机快速大幅度运动等复杂环境条件下，使用深度学习特征点的视惯融合 SLAM 算法表现更佳。因此，本文首先在 VINS-Mono 算法[2]的闭环检测模块使用卷积神经网络 SuperPoint[3]提取图像特征点，并选取一定数量的 SuperPoint 特征点替换 FAST[4]角点实现闭环帧间的特征匹配，从而形成一类称作 SuperPoint-SLAM 的视惯融合 SLAM 算法；同时，本文引入深度学习领域的一种轻量级网络 FeatureBooster[5]，其可以在满足计算代价尽可能小的条件下，进一步提升 SuperPoint 特征描述符的可区分性，因而在大视角变化、弱纹理、重复纹理或光照强度变化明显等充满挑战性的场景中，将 FeatureBooster 输出的局部描述符应用于闭环帧间的特征匹配时能够获取准确性更高的匹配结果，从而实现闭环帧间更鲁棒的 3D-2D 位姿估计[6]，进而可以构建更高精度的闭环约束关系，优化了视惯融合 SLAM 算法在这类场景中的性能。

综上所述，本文在 VINS-Mono 算法的闭环检测模块开展相关工作，提出了基于 FeatureBooster 的闭环帧间特征匹配算法。相较于传统的闭环帧间特征匹配算法，使用本文设计的算法能够在弱纹理、重复纹理或者光照强度显著变化等复杂环境条件下更好地修正移动机器人的全局位姿，避免了由于构建的闭环约束的准确性不足而导致视惯融合 SLAM 算法定位误差的累积。

1 基于 FeatureBooster 的闭环帧间特征匹配算法研究

1.1 SuperPoint 特征点目前存在的问题

一般认为，若移动机器人处在光照强度显著变化、相机大幅度快速运动等复杂环境条件下，相较于使用传统特征点，使用基于深度学习的 SuperPoint 特征点可以实现一个更高精度的定位与建图过程。然而，若移动机器人实际的工作环境中包含弱纹理或重复纹理的区域，此时使用 K 最近邻（K Nearest Neighbor，KNN）[7]算法进行闭环帧间的 SuperPoint 特征匹配时容易产生较多的误匹配结果，从而给闭环约束的构建过程带来巨大挑战，导致视惯融合 SLAM 算法定位误差的累积。因此，如何增强这类场景下 SuperPoint 特征描述符的性能成为现阶段需要重点解决的问题。

1.2 FeatureBooster 的工作原理

轻量型网络 FeatureBooster 的结构示意图如图 1 所示，该网络以原始的局部描述符和特征点的几何属性作为输入，通过一个轻量型 Transformer 将所有特征关键点的视觉信息和几何信息集成到单个局部描述符中，使得局部描述符包含全局信息，从而有效提升了局部描述符在特定场景下的性能。

FeatureBooster 包含自增强和交叉增强两个阶段。在自增强阶段，使用一个轻量的多层感知器（MLP）[8]网络将原始的局部描述符投影到一个新的空间中。假设将投影函数记作 MLP_{desc}，特征点 p_i 的局部描述符为 d_i，转换后的局部描述符为 d_i^{tr}，则投影表达式为：

$$MLP_{desc}\left(d_i\right)\rightarrow d_i^{tr} \tag{1-1}$$

同时，在该阶段使用另一个 MLP（记作 MLP_v）将特征点的 2D 像素坐标 $\left(x_i,y_i\right)$、检测得分 c_i、方向 θ_i 以及尺度 s_i 等几何信息嵌入 d_i^{tr} 中，目的是进一步提升局部描述符的性能。该过程的表达式为：

$$d_i^{tr}+MLP_v\left(p_i\right)\rightarrow d_i^{tr} \tag{1-2}$$

其中 $p_i=\left(x_i,y_i,c_i,\theta_i,s_i\right)$ 表示上述所有可用的几何信息。

然而，由于在自增强阶段并未考虑单帧图像不同特征点之间可能存在的关联，从而导致获取的特性增强的局部描述符在重复纹理或者弱纹理等场景下依然表现不佳。因此，针对来源于自增强阶段的特性增强的局部描述符，需要在交叉增强阶段使用一个轻量且高效的 Transformer[9]对其作进一步处理。相比自增强阶段，在交叉增强阶段能够同时处理单帧图像的所有特征点，并将特征点的信息聚合成为全局信息集成到单个局部描述符中。若将 Transformer 表示为 $Trans$，则交叉增强阶段的投影表达式为：

$$Trans\left(d_1^{tr},d_2^{tr},\cdots d_N^{tr}\right)\rightarrow\left(d_1^{tr},d_2^{tr},\cdots d_N^{tr}\right) \tag{1-3}$$

其中，Transformer 的输入为经过自增强阶段处理而得到的局部描述符，输出为单帧图像的 N 个特性增强的实值型或二进制整型局部描述符。

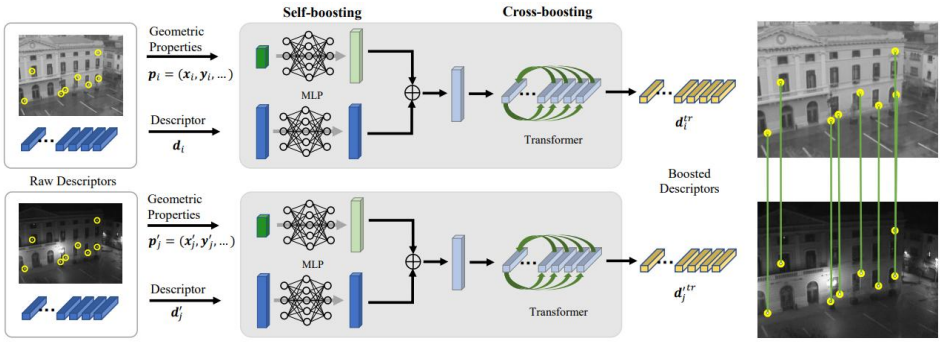

图 1 轻量型网络 FeatureBooster 的结构示意图

Fig. 1 Thestructural diagram of the lightweight network FeatureBooster

1.3 基于 FeatureBooster 的闭环帧间特征匹配算法设计与实现

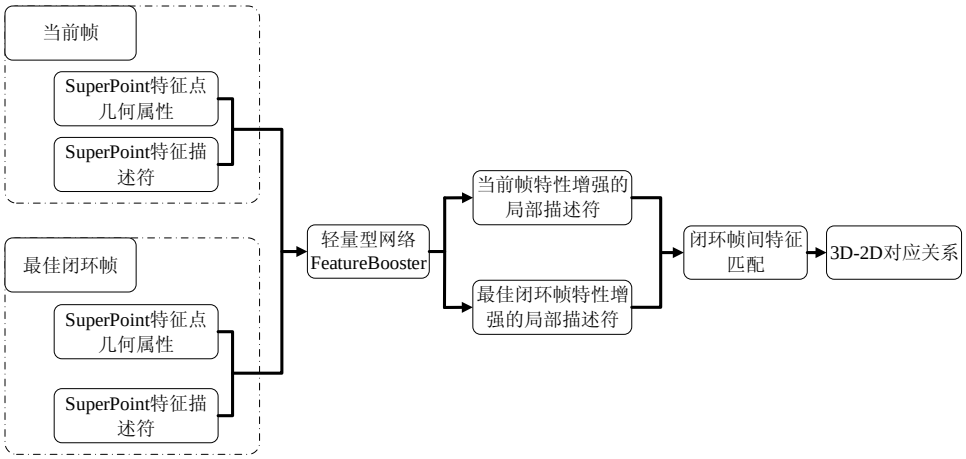

图 2 基于 FeatureBooster 的闭环帧间特征匹配算法流程图

Fig. 2The flow chart of the loop-closed inter frame feature matching algorithm based on FeatureBooster

本文设计并实现的基于 FeatureBooster 的闭环帧间特征匹配算法的流程图如图 2 所示。若当前帧满足执行闭环检测的条件，并且在关键帧数据库中找到一个满足要求的最佳闭环帧，此时将当前帧以及最佳闭环帧的 SuperPoint 特征点和对应的局部描述符分别作为轻量型网络 FeatureBooster 的输入，可以获取两帧的特性增强的局部描述符，并从中选取一定数量满足要求的局部描述符替换未经过 FeatureBooster 网络处理的 SuperPoint 特征描述符完成闭环帧间的特征匹配，最终获取闭环帧间的 3D-2D 对应关系。

综上所述，本文设计并实现的基于 FeatureBooster 的闭环帧间特征匹配算法的效果图如图 3 所示。

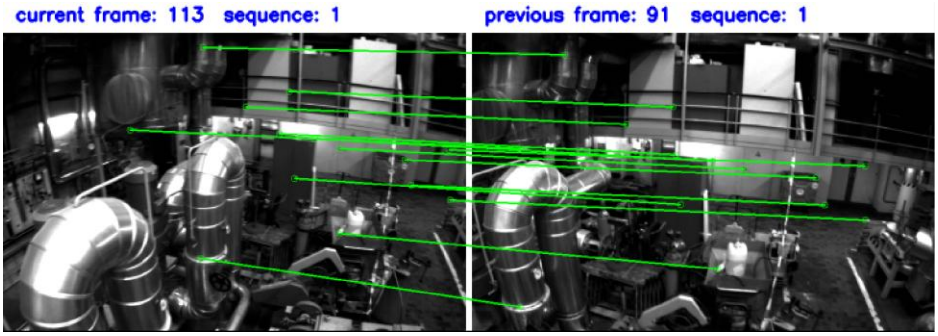

图 3 基于 FeatureBooster 的闭环帧间特征匹配算法效果图

Fig. 3 Therendering of the loop-closed inter frame feature matching algorithm based on FeatureBooster

由图 3 所示的效果图可知，即使实际环境中包含管道等重复纹理的物体，由于使用 FeatureBooster 网络增强了 SuperPoint 特征描述符的可区分性，因而能够有效提升这类场景中闭环帧间特征匹配的准确率，进而使用 PNP+RANSAC[10]算法可以求解得到更高精度的闭环帧间相对位姿，优化了视惯融合 SLAM 算法在定位与建图方面的精度和鲁棒性，这是仅使用原始的 SuperPoint 特征描述符所无法实现的结果，表明轻量型网络 FeatureBooster 对于视惯融合 SLAM 算法的闭环检测模块的重要意义。

2 实验结果与数据分析

本章内容主要由三部分组成，第一部分是闭环帧间特征匹配准确率对比实验，主要测试使用本文提出的算法和传统算法对应的闭环帧间特征匹配准确率；第二部分是闭环帧间相对位姿精度对比实验，主要测试使用本文提出的算法和传统算法进行特征匹配时对应的闭环帧间相对位姿精度；第三部分是视惯融合 SLAM 算法定位精度对比实验，通过公开数据集分别测试在闭环检测模块使用本文提出的算法和传统算法的视惯融合 SLAM 算法对应的定位与建图精度。本文的实验部分使用的笔记本电脑性能参数和相关环境配置如表 1 所示。

表 1 笔记本电脑性能参数和相关环境配置

Tab. 1 the performance parameters and related environmental configurations of the laptop

操作系统	64 位 ubuntu18.04+ROS melodic
内存大小	15.4GiB
CPU	12th Gen Intel® Core™ i7-12700H×20
GPU	RTX3050Ti

2.1 闭环帧间特征匹配准确率对比实验

针对本文设计并实现的基于 FeatureBooster 的闭环帧间特征匹配算法，为了证明其相较

于传统的闭环帧间特征匹配算法在性能方面的优势，因此使用开源的 EuRoC MAV[8]数据集包含的 5 个 Machine Hall 图像序列和 6 个 Vicon Room 图像序列设计了闭环帧间特征匹配准确率对比实验，实验结果如表 2 所示。

表 2 闭环帧间特征匹配准确率对比

Tab. 2 Comparison of the loop-closed inter-frame feature matching accuracy

算法类型 / 数据集	传统的闭环帧间特征匹配算法	基于 FeatureBooster 的闭环帧间特征匹配算法
MH_01_easy	96.59%	**97.98%**
MH_02_easy	98.34%	**99.09%**
MH_03_medium	93.66%	**94.06%**
MH_04_difficult	87.25%	**89.29%**
MH_05_difficult	89.97%	**91.50%**
V1_01_easy	98.00%	**99.28%**
V1_02_ medium	98.49%	**98.77%**
V1_03_difficult	95.38%	**96.09%**
V2_01_easy	98.85%	**99.01%**
V2_02_ medium	97.13%	**97.22%**
V2_03_difficult	91.72%	**92.58%**

由表 2 所示的实验结果可以看出，由于在视惯融合 SLAM 算法的闭环检测模块使用轻量型网络 FeatureBooster 增强了 SuperPoint 特征描述符的可区分性，因而可以有效提升闭环帧间特征匹配的准确率。其中，运行复杂环境条件下采集的数据集 MH_04_difficult 时闭环帧间特征匹配准确率的增长幅度最大，为 2.04%。

2.2 闭环帧间相对位姿精度对比实验

本文的 2.1 小节设计的对比实验证明将一类轻量型网络 FeatureBooster 引入闭环检测模块能够有效提升闭环帧间特征匹配的准确率，为了进一步说明 FeatureBooster 的使用能够改善闭环帧间相对位姿的精度和鲁棒性，因此设计了闭环帧间相对位姿精度对比实验。本实验基于 EuRoCMAV 数据集中包含的复杂环境条件下的图像序列 MH_05_difficult 进行实现。同时，为了获取具有代表性的实验结果，本实验针对添加不同大小高斯噪声的情况执行 5 次平移精度和旋转精度的计算过程，最后将 5 次测试结果进行汇总并取平均值，实验结果分别如表 3 和表 4 所示。

表 3 闭环帧间相对位姿的平移部分精度对比（单位：m）

Tab. 3Comparison of accuracy in the translation part of relative pose between loop-closed frames(Unit:m)

方差 / 算法	0.2	0.4	0.6	0.8	1.0
传统的闭环帧间特征匹配算法	0.414	0.430	0.469	0.506	0.537
基于 FeatureBooster 的闭环帧间特征匹配算法	**0.359**	**0.373**	**0.411**	**0.458**	**0.475**

表 4 闭环帧间相对位姿的旋转部分精度对比（单位：度）

Tab. 4Comparison of rotational accuracy of relative pose between loop-closed frames(Unit:degree)

方差 / 算法	0.2	0.4	0.6	0.8	1.0
传统的闭环帧间特征匹配算法	0.611	0.643	0.657	0.694	0.719
基于 FeatureBooster 的闭环帧间特征匹配算法	**0.566**	**0.582**	**0.605**	**0.630**	**0.658**

由表 3 和表 4 所示的实验结果可以看出，相较于传统的闭环帧间特征匹配算法，本文提出的基于 FeatureBooster 的闭环帧间特征匹配算法具备更高的视觉点特征位置精度和旋转精度，优化了闭环帧间位姿估计的精度和鲁棒性，体现了轻量型网络 FeatureBooster 对于本文设计的一类视惯融合 SLAM 算法的闭环检测模块的重要意义。其中，当施加均值为 0，方差为 1.0 像素大小的高斯噪声时，闭环帧间相对位姿的平移部分精度提升最明显，为 6.2cm；当施加均值为 0，方差为 0.8 像素大小的高斯噪声时，闭环帧间相对位姿的旋转部分精度提升最明显，为 0.064 度。

2.3 视惯融合 SLAM 算法定位精度对比实验

本文提出的基于 FeatureBooster 的闭环帧间特征匹配算法能够更好地满足视惯融合 SLAM 算法对于建立闭环帧间准确性较高的特征匹配关系的要求。为了验证本文设计并实现的算法给视惯融合 SLAM 算法定位与建图的精度和鲁棒性方面带来的提升，因此设计了视惯融合 SLAM 算法定位精度对比实验，主要对比本文提出的 SuperPoint-SLAM 算法和在 SuperPoint-SLAM 算法的闭环检测模块引入轻量型网络 FeatureBooster 而形成的一类视惯融合 SLAM 算法（本文将后一类算法记作 Boost-SuperPoint-SLAM）。本实验基于开源的 EuRoC MAV 数据集包含的 11 个图像序列进行实现，通过使用轨迹评估工具 EVO[12]获取视惯融合 SLAM 算法的绝对轨迹误差（ATE）[13]，以此来反映 SLAM 算法在定位与建图方面的精度和鲁棒性。实验结果如表 5 所示。

表 5 视惯融合 SLAM 算法定位精度对比（单位：m）

Tab. 5 Comparison of positionalaccuracy of the visual inertial fusion SLAM algorithm(Unit: m)

算法类型 / 数据集	SuperPoint-SLAM	Boost-SuperPoint-SLAM
MH_01_easy	0.129	**0.109**
MH_02_easy	0.089	**0.066**
MH_03_medium	0.171	**0.168**
MH_04_difficult	0.302	**0.253**
MH_05_difficult	0.276	**0.234**
V1_01_easy	0.093	**0.063**
V1_02_ medium	0.087	**0.080**
V1_03_difficult	0.144	**0.142**
V2_01_easy	0.080	**0.072**
V2_02_ medium	0.112	**0.110**
V2_03_difficult	0.245	**0.203**

由表 5 所示的实验结果可以看出，无论移动机器人处在较为理想或是充满挑战性的工作

环境中，将轻量型网络 FeatureBooster 与视惯融合 SLAM 算法的闭环检测模块相结合均能够有效提升 SLAM 算法在定位方面的精度和鲁棒性。其中，运行复杂环境条件下采集的数据集 MH_04_difficult 时视惯融合 SLAM 算法定位精度的增长幅度最大，为 4.9cm。

3 结论

在基于 FeatureBooster 的闭环帧间特征匹配方面，本文首先阐述了使用 SuperPoint 特征描述符在实际应用过程中存在的不足；然后介绍了轻量型网络 FeatureBooster 内部的工作原理；接着阐述了将 FeatureBooster 与闭环帧间的特征匹配部分相结合的具体实现，并且通过展示实验效果图说明 FeatureBooster 的应用能够使得 SuperPoint 特征描述符具备更强的可区分性；最后通过实验验证了使用经过 FeatureBooster 处理的局部描述符可以获取闭环帧间准确率更高的特征匹配结果，提升了闭环帧间相对位姿的精度，优化了视惯融合 SLAM 算法在定位方面的性能，体现了 FeatureBooster 对于视惯融合 SLAM 算法的重要意义。本论文的主要贡献是：

1）在经典且开源的 VINS-Mono 算法的闭环检测模块，将用于闭环检测的特征点类型由传统的 FAST 角点替换为基于深度学习的 SuperPoint 特征点，有效提升了视觉点特征的鲁棒性和分布的均匀性。

2）提出了基于 FeatureBooster 的闭环帧间特征匹配算法，并且通过三类对比实验充分证明，相较于使用传统算法，使用本文提出的算法能够有效提升复杂环境条件下视惯融合 SLAM 算法的表现。

[参考文献] (References)

[1] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: part I[J]. IEEE robotics & automation magazine, 2006, 13(2): 99-110.
[2] Qin T, Li P, Shen S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on Robotics, 2018, 34(4):1004-1020.
[3] DeTone D, Malisiewicz T, Rabinovich A. Superpoint: Self-supervised interest point detection and description[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 224-236.
[4] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]// Computer Vision-ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer Berlin Heidelberg, 2006: 430-443.
[5] Wang X, Liu Z, Hu Y, et al. FeatureBooster: Boosting Feature Descriptors with a Lightweight Neural Network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7630-7639.
[6] 张国良，林志林，姚二亮，等. 考虑多位姿估计约束的双目视觉里程计[J]. 控制与决策，2018, 33(6): 1008-1016.
[7] 何坚，周明我，王晓懿. 基于卡尔曼滤波与 k-NN 算法的可穿戴跌倒检测技术研究[J]. 电子与信息学报, 2017, 39(11): 2627-2634.
[8] Taud H, Mas J F. Multilayer perceptron (MLP)[J]. Geomatic approaches for modeling land change scenarios, 2018: 451-455.
[9] Han K, Xiao A, Wu E, et al. Transformer in transformer[J]. Advances in neural information processing systems, 2021, 34: 15908-15919.
[10] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", Commun. ACM, vol. 24, no. 6, pp. 381-395, 1981
[11] Bradski G, Kaehler A. OpenCV[J]. Dr. Dobb's journal of software tools, 2000, 3(2).
[12] 周治国，曹江微，邸顺帆. 3D 激光雷达 SLAM 算法综述[J]. 仪器仪表学报, 2021, 42(9): 13-27.
[13] 陈翔，邹庆年，谢绍宇，等. 一种面向运动目标的关键帧自动选择算法[J]. 计算机与现代化，2020 (10): 81.