华南师範大學

《人工智能导论》课程项目

开 题 报 告

项　目　题　目：基于 LRNet 的 AI 换脸识别

所　在　学　院：计算机学院

项　目　组　长：梁玥

小　组　成　员：詹梦纯、何育霏

开　题　时　间：2024 年 9 月 30 日

一、选题背景

近年来，AI 换脸逐渐走入了人们的视野，为公众所了解。该项技术目前已经被广泛应用于影视剧等领域，然而，该技术也存在一些风险和问题。由于**自动编码器和生成对抗网络（GAN）**[1]的改进，人们能够轻易地在现有的图像与视频在将一个人的面孔替换为另一个人，制造出难以辨识真伪的图像与视频，对社会和个人造成了各种各样的负面影响。其中 **Deepfake**[2]（即 AI 深度换脸模型）最为臭名昭著。今年 9 月，韩国“深度伪造”(deepfake)AI 换脸性犯罪事件曝光，以女性为主的近千人受到侵害，200 多所学校牵涉其中，初高中生未成年人作案比例惊人。犯罪者用深度伪造技术(Deepfake)合成色情照片和视频，并在通信软件 Telegram 的群聊中传播，参与者多达 22 万人，引发韩国民众恐慌。

因此，在当今伪造的图像与视频充斥着网络空间的大背景下，辨别相关伪造信息迫在眉睫。开发高准确率与高精确率的高效 AI 伪造人脸识别模型不仅符合市场需求，同时也有利于维护网络空间与社会的公序良俗。

二、相关研究综述

现阶段相关领域的从业者引入了多个用于辨别 Deepfake 模型制造的伪造人脸的方法，其中最著名的两种分别是**帧级检测法（Frame-level detection）**以及**视频级检测法（Video-level detection)**[3,4]，但这两种方法均存在着成本高、缺乏稳健性以及难以复制的问题[5,6,7]。近年来，随着深度学习技术的发展，**卷积神经网络(Convolutional Neural Networks, CNN)**被面部表情识别领域的学者用于从输入数据中提取高级特征[8]。然而，CNN 虽然适合用于提取深度空间特征，却于提取时序特征[9] 表现不佳，这给动态表情识别带来了挑战。以 CNN 网络模型在地标检测中的特点为例，虽然 CNN 在学习局部判别信息方面表现出了不错的性能，检测速度相对较快[10]，但它无法学习相对空间特征，并且由于受限的感受野而丢失了重要信息，在地标检测帧级操作时无法达到高精确度的标准[11]。

针对这些问题，研究者们提出基于几何学的特征点检测的 **LRNet 模型**[12]，极大地提高了辨识图像或视频信息的准确度（Accuracy）和精确度（Precision）[13]。这类模型使得面部图像及视频中的几何特征序列得到了更多重视。同时相比与 CNN 网络模型，

LRNet 模型拥有特定的校准模块来提高地表检测结果[14]。

现阶段的 LRNet 的 LRNet 模型由四个部分组成：**面部预处理模块、校准模块、特征嵌入程序和 RNN 分类程序**。本项目在此重点对前两个模块的国内外研究现状进行分析。

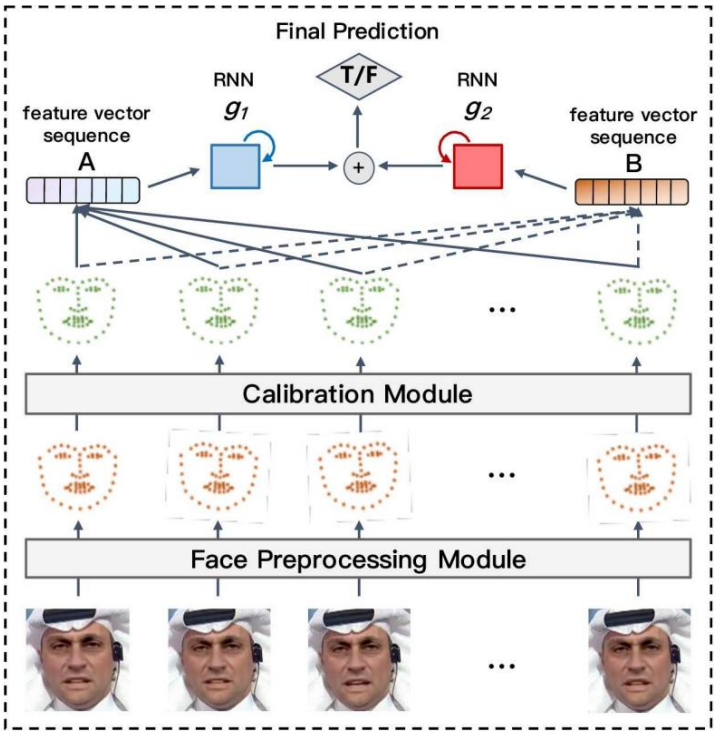

图 1：LRNet 模型流程图

1. 面部预处理模块方面的研究及其局限性

人脸识别领域的面部预处理模块是非常关键的一部分，其主要目的是对输入的人脸图像进行**去噪**[15]、**对齐**[16]、**归一化**[17]等处理操作，从而提高后续人脸特征提取和匹配的准确度和稳定性。当前在人脸预处理方面较为常见的技术包括以下几种：

① **人工标记关键点**。这种方法需要人工标注人脸关键点，在这些关键点上计算变换矩阵来完成对齐[18]。这种方法精度较高，但需要大量人工工作。

② **基于几何变换**[19]**的方法**。这种方法利用旋转、平移、缩放等基本变换来实现人脸对齐，具有较好的效果。但对于光照、角度等变化比较大的情况，效果可能不理想。

③ **基于联合建模[20]的方法**。这种方法利用多个训练好的模型，分别对人脸进行检测、对齐和归一化等处理，在整个流程中保证了精度和速度。

④ **基于深度学习的方法**。这种方法近几年被广泛应用，通过训练深度神经网络来实现人脸关键点的检测和对齐，目前在人脸预处理和人脸特征提取方面都取得了不错的效果。

然而，这些方法也存在着一些局限性[20,21]：

①需要大规模标注数据集。许多预处理方法需要大量标注数据集来训练并优化算法，但人工标注数据集往往耗时、耗力，并且容易出现质量不高的情况。

②受限于光照、角度等因素。对于照片中的人脸位置、光线、角度等变化比较大的情况，各种预处理方法的效果都可能大打折扣。

③计算量大。基于深度学习的方法通常需要大量的计算资源和时间来完成网络训练和预测，在实际应用中可能存在一定的难度。

④特征点数量不足。面部预处理模块使用预训练模型在人脸图像上提取出 68 个特征点进行检测，凭借这些特征点构建起人脸的标志性轮廓再送往特征点校准模块处理[22]。但由于人脸的复杂程度，68 个特征点显然远远不足够于以极高的精确率提取完整的人脸信息。

⑤算法局限性。现阶段 LRNet 模型的特征点校准模块所选用的 Lucas-Kanade 算法[23]是高效的特征点追踪算法。如果像素点的运动幅度过大，运动到了局部窗口之外，那么算法将无法找到对应的像素点，导致像素点丢失。

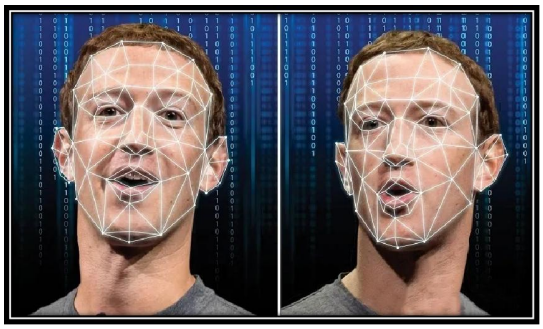

图 2：运用 LRNet 模型提取面部特征点

2.校准模块方面的研究及其局限性

LRNet 模型的校准模块是指在人脸图像预处理中，为了提高识别精度，对输入的人脸图像进行标准化和校准的过程。该过程主要包括两个步骤：人脸检测和对齐[24]。其中，人脸检测是指确定人脸在图像中的位置和大小，对齐是指将不同大小、姿态和光照条件下的人脸图像转化为一个统一的标准形式，以便于后续的特征提取和识别。

在 LRNet 模型中，对齐过程[25]主要通过估计眼睛、嘴巴等人脸关键点位置来实现。具体来说，利用已有的检测器检测到人脸区域后，根据预先训练的神经网络模型，可以自动估计出图像中的人脸关键点信息（如眼睛中心、嘴巴中心等）。接着，通过这些关键点信息，使用仿射变换或透视变换等技术对人脸图像进行变换和对齐，使得不同人脸图像在姿态和大小上具有统一性[26]。这样就可以保证后续的特征提取和识别过程能够得到更稳定、鲁棒的特征表示，提高模型的识别精度。其具有以下特点：

① **自适应性**

LRNet 模型的校准模块可以自动检测人脸图像中的关键点，并根据这些点的位置信息自适应地对不同大小、姿态和光照条件下的人脸进行对齐。因此，该模块能够适应更多的场景和情况。

② **鲁棒性**

LRNet 模型的校准模块采用多个关键点来对齐人脸图像，因此即使某些关键点检测出现误差，仍然能够保持相对较好的对齐效果。这样可以增强模型的鲁棒性，使其在面对噪声、干扰等问题时能够保持较高的识别率。

③ **稳定性**

LRNet 模型的校准模块通过将不同大小、姿态和光照条件下的人脸转化为统一的标准形式，可以保证后续的特征提取和识别过程得到更稳定、鲁棒的表示。这样可以减少由于变化引起的特征漂移和误差，提高模型的识别精度和泛化能力。

④ **可扩展性**

LRNet 模型的校准模块采用基于关键点的对齐方法，因此可以通过增加或调整

关键点的数量和位置，来扩展模型的适用范围和对不同场景、任务的适应性。

然而，校准模块也存在以下方面的缺陷：

① 精度受限

LRNet 模型的校准模块是通过关键点检测和对齐来实现的，因此其精度受到关键点检测算法和对齐方法的影响。在某些情况下，由于人脸图像中的遮挡、姿态变化、低质量图像等原因，关键点检测可能会出现误差，从而导致对齐效果不佳。

②计算复杂度高。

LRNet 模型的校准模块需要对每个输入的人脸图像进行关键点检测和对齐操作，这些操作需要耗费大量的计算资源。在处理大规模数据集时，可能会受到计算资源限制，导致无法实现较高的处理速度和效率。

③对不同特征的适应性有限。

LRNet 模型的校准模块是针对人脸图像预处理和识别任务设计的，其关键点和对齐方式仅适用于人脸表征学习任务，对于其他类型的图像或特征表示任务可能需要重新设计适合的关键点和对齐方式。

最初的 LRNet 模型通过估计眼睛、嘴巴等特征点来实现对齐，但该方法在遇到头部倾斜和人脸部分遮挡等情况时效果不佳。因此，一些研究工作尝试通过改进校准模块来提高 LRNet 模型的性能。例如，“Coupled Cascaded Pose Regression”（CCPR）模型[27]采用联合级联回归的方式来估计人脸的姿态参数和关键点，从而实现更精确的对齐。该模型通过多个子级联回归网络共同完成任务，使得模型的精度和泛化能力都得到了提升。另外，“Deep Descriptors for Face Verification and Image Retrieval”（DD）模型[28]则采用了一种基于稠密网格的对齐方法，该方法能够更好地适应人脸的形状变化和遮挡情况，提高了模型的泛化能力。DD 模型在多个数据集和评测中都取得了不错的结果。除了针对校准模块改进的研究外，一些研究工作也探索了 LRNet 模型在实际应用中的局限性问题。例如，“On Face Recognition with Occlusions”（Zhu et al.，2019）研究发现，在人脸遮挡情况下，原始的 LRNet 模型的识别率会显著下降，需要在模型中加入特殊的处理机制来解决这个问题[29,30]。此外，还有一些研究工作

致力于改善 LRNet 模型的计算效率和内存开销等问题。例如，“Lightweight Face Recognition System Based on Compressed Convolution and Multi-Scale Feature Fusion”（Dong et al.，2020）在保证精度的前提下，采用了压缩卷积和多尺度特征融合等技术优化了模型的计算和内存开销[31]。

在提高预处理模块提供的逐帧检测特征点的精确度（Precision）方面，校准模块使用了 Lucas-Kanade 光流计算算法来预测特征点的后期位置，将其中准确的预测与去检测结果通过降噪的卡尔曼滤波器再次提高精确度。经过追踪与去噪的特征点序列被嵌入到两种类型的特征向量序列中，然后输入到双流 RNN[32]进行虚假视频分类。

然而由于去噪目的而采用的卡尔曼滤波器[33]根据建模方法不同而有诸多改进版本：基础卡尔曼滤波器、扩展卡尔曼滤波器到互补卡尔曼滤波器……根据所使用的系统不同，不同版本的卡尔曼滤波器所产生的效果也不同，导致校准模块的普适性较差[34]。

三、拟解决的问题和研究内容

本项目的研究目标是通过结合 LRNet 模型和几何特征，改进人脸预处理模块和校准模块，从而实现对人脸识别中面部伪造的有效识别。具体方法包括：深入研究 LRNet 模型的预处理模块、结合几何特征和机器学习算法设计新的校准方法、大量实验验证方法可行性和性能表现。突破目前面部预处理模块以及校准模块研究领域上的局限性，实现 AI 伪造人脸的识别方法。

1.基于 DLIB 库基础上开发面部预处理模块

本项目的目标是在继承对 Dlib 库的选择的前提下改进或重新开发预处理模型，将现阶段预处理模型所能提取的特征点最大数从 68 往上提升，设计新的算法或修改现有算法以提取更多的特征点，并将该算法代码转化为 Python 语言版本以满足实际需求。

2.基于 LK 金字塔光流算法优化校准模块

本项目将以 LK 算法的改进版——金字塔-LK 算法为切入点，测试出将金字塔-LK 算法的最佳改进版本并将其引入现阶段的 LRNet 模型。LK 算法是一种普遍应用于计算

机视觉领域的光流算法，其主要目的是在连续帧序列上估量出物体或场景的运动信息。结合使用金字塔-LK 算法和图像金字塔技术，可以更加准确地获取光流估计结果，并从而提高模型的稳定性和鲁棒性，减少图像像素点丢失的概率，有效地适应不同尺度的数据处理需求。

3.选择最优版本的卡尔曼滤波器提高模型鲁棒性和可靠性

本项目将深入了解卡尔曼滤波器，同时选择最优版本的卡尔曼滤波器，以此最大限度提高特征点去噪效果，并根据模型的使用环境不同选用，得出不同版本的组件选用与调整方案，减少在不同环境下产生的误差，以提高 LRNet 模型的泛用性、稳定性以及鲁棒性。

四、可行性分析

1. 技术可行性

目前，对于假人脸识别有着成熟的理论基础。且多个研究表明 deepfakes 的 LRNet 模型是可行的，且准确率比其他模型高。同时其具有以下优点：

①鲁棒性强与高准确率

LRNet 模型在训练时使用了大量真实人脸数据和 ai 伪造人脸数据，因此它对于各种不同类型的 ai 伪造人脸都能够有很好的鲁棒性，可以有效地区分真实人脸和伪造人脸。通过对多个数据集进行测试，LRNet 模型在人脸识别方面的准确率高达 99%以上，这意味着它能够高效地识别 ai 伪造人脸并且准确率将会非常高。

②开源易学

现如今已有丰富的关于伪造人脸鉴别的开源知识。这种开源特性极大地促进了技术的普及与学习。开发者可以直接进行深入研究与定制优化模型，且社区内还能充分进行资源共享与经验交流。

③轻量级结构

LRNet 模型的网络结构相对简单，参数数量比其他人脸识别模型少，因此它能够在

较短的时间内进行训练，并且在实际应用中运行速度非常快。

综上，针对 LRNet 模型的研究能够有效解决识别 AI 伪造人脸的课题，具有广阔的应用前景和使用价值。

2. 团队条件

本项目小组成员均来自华南师范大学计算机学院，对于学习一丝不苟，能够积极的进行深究探讨。本团队已有一年的合作，已形成相近的价值观念，是一个知识互补、分工互补、同方向、拥共识的强有力的团队。

梁玥，项目团队负责人，计算机科学与技术专业本科生。学习态度端正，有较强的自我学习能力。有良好的算法基础，能熟练运用 C++、Python 等编程语言。熟悉机器学习和人工智能等领域知识，尝试复现过人工智能算法与模型，并能运用于实际。对机器学习与人工智能领域颇感兴趣。具有良好的沟通交流以及团队协作能力。

詹梦纯，项目团队成员，计算机科学与技术专业本科生。擅于和他人沟通，组织负责过校级省级比赛，具有良好的团结合作精神。对 Linux、Unix 有一定了解，能较熟练操作 C++、Python、Java 等语言，同时对 SQL 也有一定程度上的认识。参与过游戏设计开发和实验室平台搭建，目前对深度学习有较大的兴趣，乐于学习更多的知识。

何育霏，项目团队成员，计算机科学与技术专业本科生。具备扎实变成基础，掌握 Python 和 C++编程语言，具备良好的逻辑思维能力和问题解决能力，热衷于计算机技术的学习与应用，对深度学习框架 PyTorch 有一定实践经验，积极参与团队合作，善于倾听和整合各方意见，乐于分享自己的想法与经验。

五、计划进度安排

研究时间：2024 年 9 月-2024 年 12 月

项目初期：2024 年 9 月-2024 年 10 月

1. 收集 LRNet 模型识别 AI 伪造人脸方向论文文献与有关深度人脸伪造的图像与视频的数据集

2. 整合论文材料根据项目需求进行方案比较。

3. 制定项目研究思路，开发计划及文本，及项目的总体架构。

项目中期：2024 年 10 月-2024 年 11 月

1. 设计模型并进行测试。

2. 测试模型性能，提高模型泛化能力。

3. 整合子模块，检查系统错误。

4. 逐步优化迭代，整理实验数据并撰写结题报告。

项目后期：2024 年 11 月-2024 年 12 月

1. 修改并完善项目。

2. 总结开发过程，准备结题材料。

六、参考文献

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ..., Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672-2680.

[2] Deepfakes.Deepfakes.github[EB/OL].[2022-09-14].https://github.com/Deepfakes/faceswap.

[3] Siwei Lyu and Hany Farid. "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection." IEEE Access 7 (2019): 128611-128636.

[4] Yuezun Li, Ming-Ching Chang, Xiaoming Liu, and Siwei Lyu. "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking." Proceedings of the IEEE International Conference on Computer Vision. 2019.

[5] 曹申豪,刘晓辉,毛秀青,邹勤. 人脸伪造及检测技术综述[D].武汉大学,2022.

[6] Xinyu Liu, Weiming Zhang, Jing Shao, Xiaoguang Tu, and Jianhua Li. "Detecting Deepfake Videos from Face Detection and Recognition Perspectives: A Review." arXiv preprint arXiv:2101.06860 (2021).

[7] Yuezun Li, Ming-Ching Chang, Xiaoming Liu, and Siwei Lyu. "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking." Proceedings of the IEEE International Conference on Computer Vision. 2019.

[8] 谢天，于灵云，罗常伟,谢洪涛,张勇东.深度人脸伪造与检测技术综述.安徽大学,2023.

[9] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[10] Amirhossein Tavanaei and Mohammad Reza Karami. "Towards a unified system for facial expression recognition: A deep learning approach." IEEE Transactions on Affective Computing 11, no. 2 (2020): 199-215.

[11] Sun, Y., Wang, X., & Tang, X. (2013). Deep convolutional regression networks for facial landmark detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, 1385-1392.

[12] Changjie Fan, Guangxu Zhu, Yingjie Zhang, and Sitong Cheng. "Attention-Enhanced LRNet: A Lightweight Network for Facial Landmark Detection." Proceedings of the IEEE International Conference on Multimedia and Expo. 2020.

[13] Qi Fan, Yang Zhou, and Qian Zhang. "Landmark Detecting with an Improved LRNet." Proceedings of the Chinese Conference on Pattern Recognition. 2020.

[14] Basak, N., Dhole, A., Konar, D., Chakraborty, A., & Garai, A. (2020). LRNet: An Efficient Deep Learning Model for Facial Landmark Detection. IEEE Access, 8, 102687-102700.

[15] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2017). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 24(3), 328-332.

[16] Dong, X., Yu, S. I., Weng, X., & Wei, S. E. (2016). Facial landmark detection by deep multi-task learning. European Conference on Computer Vision (ECCV), 94-108.

[17] Li, H., Deng, J., & Du, X. (2019). Stack hourglass face detection with set-based classification for detecting small faces. IEEE Access, 7, 104332-104344.

[18] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499-1503.

[19] Zhu, X., Lei, Z., Liu, X., Shi, H., & Li, S. Z. (2016). Face alignment across large poses: A 3D solution. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 146-155.

[20] Xiong, X., & De la Torre, F. (2013). Supervised descent method and its applications to

face alignment. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 532-539.

[21]Buades, A., Coll, B., & Morel, J. M. (2005). A review of image denoising algorithms, with a new one. Multiscale modeling and simulation, 4(2), 490-530.

[22]Lee, S., Lee, J. H., & Lee, B. H. (2016). Multi-resolution facial landmark detection based on cascaded regression of shape and texture features. Pattern Recognition, 50, 94-103.

[23]B.D. Lucas and T. Kanade,(1981) "An Iterative Image Registration Technique with an Application to Stereo Vision," Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, pp. 674-679.

[24]S. Sun, M. Huang, and Z. Liu, "LRNet: A Low-Resolution Face Recognition Network Combined with a Calibration Module," International Journal of Pattern Recognition and Artificial Intelligence, vol. 34, no. 10, 2020.

[25]V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 1867-1874, June 2015.

[26]J. Deng, J. Guo, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 4690-4699, June 2018.

[27]C. Zou, Y. Lin, and H. Yang, "Coupled Cascaded Pose Regression for Face Alignment," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, pp. 5325-5334, June 2014.

[28]Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation by Joint Identification-Verification," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 3693-3702, June 2015.

[29]Chen, J., Shan, S., He, C., Zhao, G., & Pietikäinen, M. (2019). "Occlusion-aware face recognition: from holistic to parts-based approaches." IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(10), 2389-2405.

[30]Ge, H., Li, J., Liu, X., & Yang, X. (2018). "LR-VGG-HIS: An Occlusion-Aware Deep Network for Face Recognition." IEEE Transactions on Circuits and Systems for Video Technology, 29(10), 3046-3058.

[31]Dong, X., Zhang, S., Huang, Z., & Wang, W. (2020). "Lightweight Face Recognition System Based on Compressed Convolution and Multi-Scale Feature Fusion." IEEE Access, 8, 15259-15269.

[32]Fernando, B., Gavves, E., Oramas, J. M., Ghodrati, A., and Tuytelaars, T. "Bidirectional two-stream recurrent convolutional networks for multi-view action recognition." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 1, 2019, pp. 72-84.

[33]Julier, S. J., & Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems. Proceedings of the SPIE Aerospace Sensing Conference, 182-193.

[34]Pellny, T. K., & Rushton, S. K. (2015). An investigation into the robustness of the calibration of an inertial measurement unit. Journal of Navigation, 68(2), 211-223.