

面部特征点定位性能在手术环境中的评估

Ines Frajtag¹, Marko Švaco¹, and Filip Šuligoj¹

Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Croatia

ines.frajtag@fsb.unizg.hr

Abstract. 机器人、计算机视觉及其应用在包括医学在内的各个领域中的使用变得越来越普遍。许多面部检测算法已在神经外科、眼科和整形外科中获得应用。使用这些算法的一个常见挑战是可变的光照条件，以及检测位置的灵活性以识别和准确定位患者。提议的实验在一个受控环境中测试 MediaPipe 算法，用于检测面部特征点，使用一个能够自动调整位置的机械臂，同时手术灯和模型保持在固定位置。研究结果表明，在手术照明下改进的面部特征点检测精度显著提高了在较大偏航和俯仰角度下的检测性能。标准差/离散度的增加是由于选定面部特征点检测不准确造成的。这项分析允许讨论 MediaPipe 算法在医疗程序中潜在的整合。

1 引言

随着技术和自动化的发展，医学实践已经显著进步，为提高医疗程序的精确性、效率 and 安全性提供了新的解决方案。传统的外科手术方法往往依赖于外科医生的经验 [1] 和视觉技术，这可能容易出现人为错误。现代智能解决方案如机器人系统 [2]、计算机视觉算法和精确规划 [3] 被引入到外科环境中。

关键的计算机视觉技术之一是面部标志检测，这使得识别诸如眼睛、眉毛、鼻子和嘴巴等特定的解剖特征成为可能。在医学应用中，特别是整形外科 [4]、神经外科 [3] 或眼科 [5]，跟踪标志位置对于确定精确的手术部位至关重要。大多数人脸检测算法，如 OpenPose [6]、MediaPipe [7]、Dlib 库 [8] 和 FaceNet [9]，作为开源工具广泛用于情感识别 [10]、监控和识别等任务。然而，这些算法在手术室中的应用仍未被深入探索。在手术室中实施面部识别算法和面部标志检测，可以顺畅地识别病人的手术目标区域。这不仅加快了手术过程，还通过在整个手术过程中允许对面部及其特征的连续跟踪，提高了手术的安全性。本研究需要解决的一个具体研究问题是：诸如光强度和角度这样的光照条件如何影响手术室中面部标志检测的准确性和重复性？

MediaPipe 是一个基于 BlazeFace 方法的开源框架，[11]，本研究的目的是评估其算法在实际手术室条件下检测面部标志的能力。实验涉及在控制的医疗条件下使用带有 3D 相机的机械臂和类人模型在手术室内测试算法。此外，本研究还探讨了 MediaPipe 能否在手术过程的计划和执行阶段用于估计患者头部的姿态。目标是评估该算法在不同拍摄角度下检测相应面部标志的一致性。本研究的结果将有助于更好地理解该算法在医疗场景中的局限性和可靠性，可能推动其在手术室中的应用。

2 相关研究

与光照相关的问题在面部识别领域是一个显著的挑战。诸如弱光、强光或光照不均匀等条件使得算法难以检测面部的几何特征，这可能导致检测失败或重复性降低。在研究 [12] 中，作者结合了多种人脸检测方法，以在不同的光照、肤色和背景条件下改善结果。为了解决光照问题，他们建议优化算法对外部因素的敏感性较低，使用在多样面部光照数据集上训练的神经网络以实现更好的面部检测 [13]。另一种方法是多重曝光概念。在 [14] 中，作者展示了他们的循环曝光生成 (REG) 模块能够有效地与各种人脸检测算法相结合，而不需要在低或正常光照下拍摄多张图像。在 [15] 中，对 MediaPipe 的人脸标志识别和头部姿态预测进行了研究，显示其能够可靠地使用深度神经网络方法估算头部姿态角度。然而，光照条件和头部姿态角度被认为是影响面部标志检测准确性的关键挑战。根据 [15] 的结论，我们的研究重点是光照和面部标志检测在现实临床条件下的影响。在 [16] 中，采用运动捕捉系统作为金标准，比较了 MediaPipe、OpenPose 2.0 和 3DDFA-V2。基于研究结果，3DDFA-V2 被推荐为最适合患者在多方向移动的临床应用。[16] 与我们研究的主要区别是整合了一个带有 3D 摄像头的机械臂和一个固定的仿体，模拟神经外科手术设备。光照方面有所不同 (我们的研究使用手术灯)，并且在检测角度设置上有限制，监测 7 个特定面部标志的变化。采用 2D 和 3D 坐标观察预测面部标志的位置，以展示其分散性和定位的可重复性。在他们的研究 [17] 中，Abdullah 和 Ali 比较了几种基于深度学习的人脸检测器。在 MTCNN、Dlib 库和 MediaPipe 中，结果显示 MediaPipe 以 99.3 % [17] 的准确率达到了最高。用于确定准确性的标准包括检测中的错误类型和数量、精度和执行时间。MediaPipe 可以确定 468 个相互依赖的点，即使是最微小的面部运动，也会导致这些点的位置变化。由于其对变化的敏感性以及易于集成到各种系统中，因此被选用于此研究。此外，该算法经常应用于监控系统和其他医学领域 (康复 [18]，诊断 [19])，主要用于研究目的。该算法用于医学与工程学的跨学科领域 [20]，其中在驾驶过程中眨眼次数与医学监测方法 (EOG 和 EEG) 进行比较。在研究中 [20]，MediaPipe 在眨眼检测中实现了 99.87 % 的准确率，并记录的假眨眼次数是 Dlib 库的一半。这一章概述了 MediaPipe 算法在人脸特征点检测中的应用。主要重点在于评估特征点识别的再现性，并分析算法在位置和光线变化情况下的性能。这些评估旨在确定该算法的鲁棒性及其在临床环境中的潜在适用性。

实验是在萨格勒布大学机械工程和海军建筑学院的医学机器人实验室进行的，该实验室被设计成一个模拟手术室。使用了 Figure 1 (第 2 部分) 的设备来进行实验。模型被牢固地固定在连接到 Hillrom PST 300 手术台的 DORO QR3 多功能颅骨钳上。模型几何形状是从患者的 CT 图像重建中得到的，并使用 3D 打印机由 PLA 材料制成。模型的开发使用了患者数据，并得到了患者的知情同意和伦理委员会关于在研究中使用 CT 图像的批准。在模型上方，有一盏 Trumpf TruLight 5000 手术灯，而配备 3D 摄像头 Ensensio N35-606-16-IR 的 KUKA LBR iiwa 14 R820 机器人则执行预定义位置的运动。位置计算、指令执行和 MediaPipe 算法的启动是在运行 Ubuntu 22.04 的研究工作站上进行的，工作站配备了 64 位英特尔 Core i5 处理器 (2.80 GHz)。初步验证和摄像头定位基于模型鼻尖面部特征标记的坐标 (在 Figure 2 (第 2 部分) 中标记为 T5) 确定。在 10 次重复捕获中，使用 MediaPipe FaceMesh 算法在 2D 图像上识别出面部特征标记坐标。3D 摄像头捕获的视差图数据结合 2D 点坐标用于计算 3D

坐标。摄像头被放置在距离面部特征标记 T5 为 $d = 270 \text{ mm}$ 的位置，随着机器人和摄像头移动到不同位置，该距离保持不变。

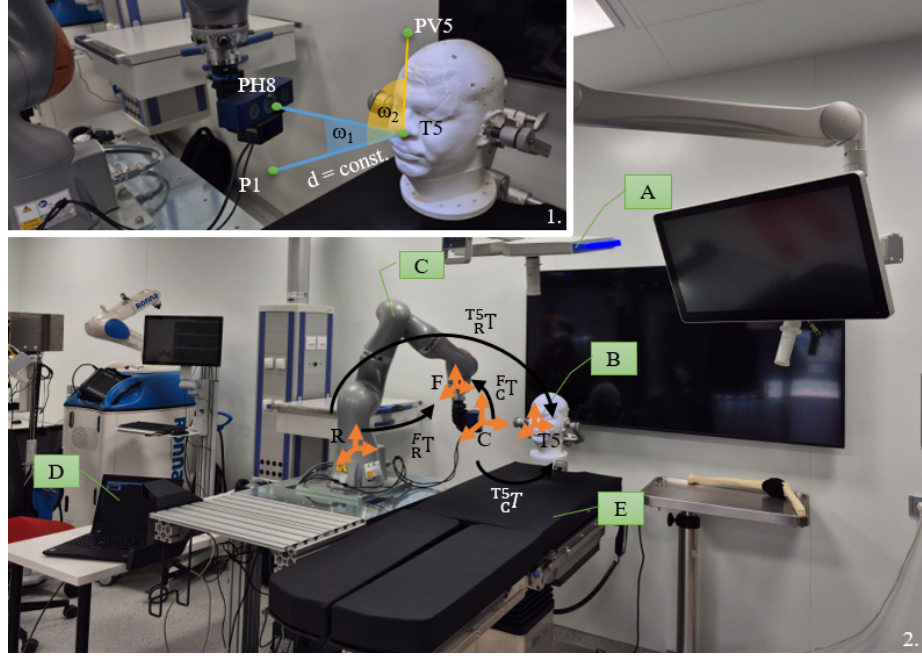


Fig. 1. 1. 一般实验室设置包括手术灯 (A)、仿体 (B)、安装了 3D 摄像头的机器人 (C)、运行 MediaPipe 的电脑 (D) 和手术台 (E)。主要参考坐标系有：机器人基座 (R)、机器人法兰 (F)、摄像头 (C) 和图像坐标系 (T5)。任意两个坐标系 (y 和 x) 之间的齐次变换矩阵表示为 xT_y [21]。为了实现机器人手臂在特定角度的旋转位置，我们使用 R_x 作为偏航的旋转矩阵和 R_y 作为俯仰的旋转矩阵。2. 从 P1 到 PH8 的位置是使用方程 ?? 和值 $\omega_1 = \{0^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 50^\circ, 60^\circ, 70^\circ\}$ 生成的。同样地，从 P1 到 PV5 的位置是使用方程 ?? 和值 $\omega_2 = \{0^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ\}$ 生成的。

机器人位置 PH 和 PV 是基于头部旋转角度（航向角为 ω_1 ，俯仰角为 ω_2 ）计算得出的，使用了 eigen C++ 库计算变换矩阵，并使用相应的方程 ?? 或 ??。进行实验时，手术灯设置为最大亮度 (160 000 勒克斯)，色温为 4500 K，手术灯与仿真人体模型之间的角度为 5° 。对面部分析的关注，MediaPipe FaceMesh 模型采用几何方法预测 468 个面部特征点，其中在研究中跟踪了七个关键点。

数据收集过程如下：机器人将摄像头移动到所需位置，等待 5 秒钟，捕获两张图像，并使用 MediaPipe 算法对其进行处理。输出包括旋转角度、检测到的人脸标志、标志的二维坐标及其对应的三维坐标。

3 结果

面部标志点检测性能在两种不同的光照条件下进行评估：手术灯和 3D 相机上的内置红外前灯。使用机械臂将相机定位在预设角度，实现图像捕捉过程的自动化。

为了确定在使用手术灯（相机的前置灯关闭）的情况下的检测限，机械臂以 10° 的增量进行旋转。在偏航超过 80° 和俯仰超过 40° 的旋转后检测失败后，以 1° 的增量进行了测试。这些测试确定算法在 73° 和 40° 时停止检测。在偏航 73° 时，面部标志检测的重复性是 10 次尝试中成功了 40 %。相机在偏航和俯仰方向分别移到了 7 个和 5 个位置，在预定角度（偏航为 ω_1 ，俯仰为 ω_2 ）上进行测试，详情见 Figure 1（第 1 部分）。每个位置均进行了 10 次测试。结果表明，所有位置的检测均成功，包括偏航 70° 和俯仰 40° 。使用红外投影仪照亮场景的相机前置灯，面部检测的最大角度显著降低。测试表明，算法在仅有 30 % 的情况下在偏航 30° 时检测到了人脸。然而，在 24° 时，面部地标的预测是 100% 成功的。由于 MediaPipe 是先检测到面部后再预测点，并不是所有地标位置都能精确识别到位。为此，我们对四个地标的坐标变化进行了监测：T1、T2、T5 和 T7，这些地标在所有位置都成功被检测到。在第一阶段，使用可变的手术照明和 3D 相机的集成前灯测试了面部检测的最大可实现旋转角度。结果显示在 Figure 2（第 2 部分）。在第二阶段，跟踪了一直被准确检测到的地标坐标值的变化。根据结果，在恒定光照条件下，旋转角度的增加会导致特定地标坐标的更大离散度，这可以通过标准差来表示。2D 图像中的坐标和 3D 坐标系中的标准差在图表上显示（见 Figure 3），作为欧氏值。坐标 i 和 j 表示像素值的 2D 图像中的位置，其中 i 代表水平坐标 (x)， j 代表垂直坐标 (y)。同样地，在 3D 空间中（在本例中为相机坐标系）， x 代表垂直坐标， y 代表水平分量， z 代表地标的距离。

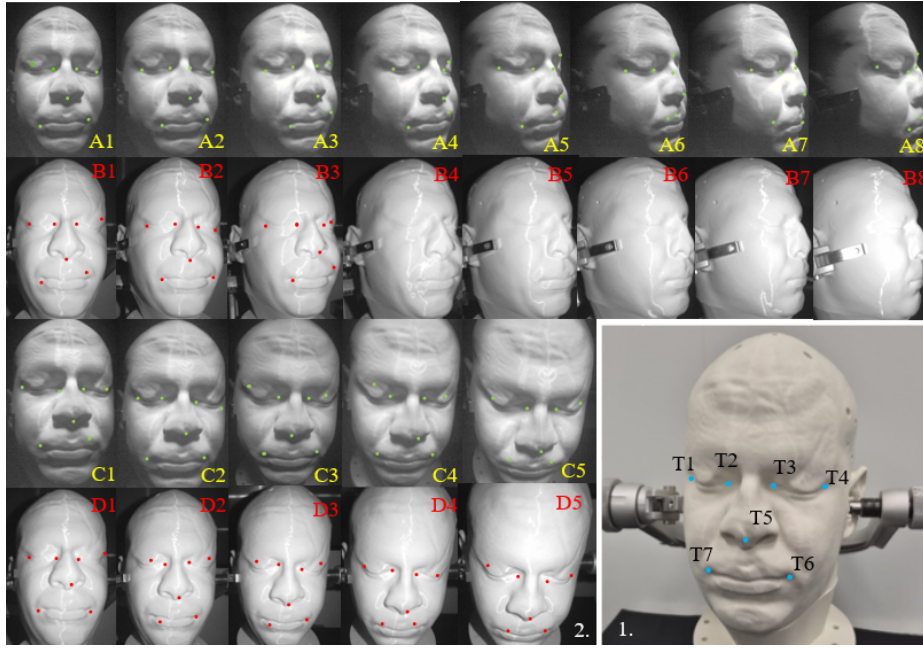


Fig. 2. 1. 在测试期间观察到的七个特定点是：T1（右眼外角）、T2（右眼内角）、T3（左眼内角）、T4（左眼外角）、T5（鼻尖）、T6（左嘴角）和 T7（右嘴角）。2. 头部位置从 P1 到 PH8 旋转一个角度 ω_1 ，当仅使用手术灯时，图片标记为 A1-A8 显示了关键点的检测或未检测。同样的位置在使用摄像机的综合前灯时显示，并标记为 B1-B8。同样，头部位置从 P1 到 PV5 旋转一个角度 ω_2 ，其中 C1-C5 表示仅使用手术灯的图像，D1-D5 表示使用摄像机前灯的图像。

为了展示标准差和角度之间的相关性，计算了每个地标在对应方向上的 Spearman 等级相关系数 (Equation 1)。在给定的 Equation 1 中， ρ 表示 Spearman 等级相关系数， d_i 是每个数据点上两个变量（这里是标准差和角度）等级之间的差异， n 是数据点的数量。

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

Spearman 等级相关系数的结果如下：对于地标 T1，为 0.9；对于地标 T2，为 0.7；对于地标 T5，为 0.2；对于地标 T7，为 0.8。系数值大于 0.9 表示变量（标准差和角度）之间有非常强的正相关。系数为 0.7 和 0.8 表明有很强的正相关，而系数为 0.2 表示正相关较弱。

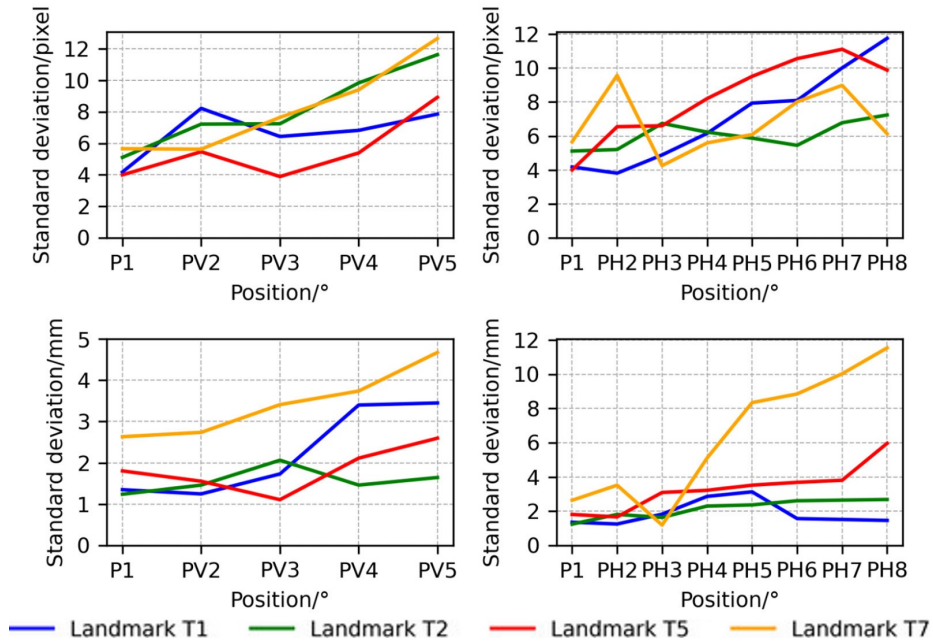


Fig. 3. 这些图展示了标准差与位置之间的关系。上面的图显示了像素中图像坐标的结果，而下面的图显示了 3D 坐标的结果。左侧的图是偏航方向的图，右侧则是俯仰方向的图。

跟踪面部标志在二维图像上的位置和坐标值及其三维坐标，观察到在手术灯下可以实现更高的旋转角度，从而使检测成为可能。手术灯在照亮面部解剖特征方面比相机上的正面灯光效果更好。在最大旋转角度（偏航 70° 和俯仰 40° ）下，旋转一侧的面部标志定位更加准确，而另一侧的标志逐渐变得分散，这一点进一步通过标志 T1、T2 和 T7 的 Spearman 等级相关系数大于 0.7（强正相关）得到了证实。

本文测试了 MediaPipe 算法在恒定手术照明下不同相机旋转角度下的人脸特征检测。研究表明，照明在最大化人脸特征检测的旋转角度时起到了重要作用，与没有手术照明相比，手术照明可以实现更大的旋转角度。精确定义光的方向和强度以最好地突出面部解剖特征的几何形状是至关重要的。此外，特定面部特征的坐标随着旋转角度和光照条件的变化而变化，旋转角度的增加导致了更大的分散，有些特征点在面部上散开。由于这种散开，T5 特征点不能在每个位置的一致深度图中被提取，从而导致缺失 3D 坐标数据。这就是为什么该特征点的相关性为 0.2，而对于相机旋转方向的面部侧面上的其他特征点，其相关性保持强烈。为了假设正态分布，需为每个点（T1、T2、T5、T7）的偏航和俯仰提供更多数据。在即将就此课题进行的研究中，计划使用更大的数据集以进行更详细的统计分析，包括 p 值和置信区间。未来研究将专注于开发一种能够在手术室内精确定位病人面部的的方法。系统应能在不同的照明条件下检测手术室内的物体，如手术台和照明设备，以便识别病人。无论患者处于何种姿势，系统都应在手术过程中定位他们的面部，并自动调整设备（机器人、摄像机）的位

置, 以确保面部检测和手术器械的可见性。通过这种方式, 手术室内的流程可以实现自动化, 从而降低手术过程中对患者的风险。

4 致谢

本研究由 INSPIRATION 项目资助——用于机器人神经外科的非侵入性患者注册 (资助编号: NPOO.C3.2.R2-I1.06.0153), 由欧盟通过国家复苏与韧性计划 (NPOO) 资助。

References

1. Han, Jason J. et al.: Association Among Surgeon Experience, Patient Risk, and Outcomes in Coronary Artery Bypass Grafting. *The Annals of Thoracic Surgery* ; 111:86-93 (2021). doi:10.1016/j.athoracsur.2020.04.122
2. Haidegger, T. et al: Robot-Assisted Minimally Invasive Surgery-Surgical Robotics in the Data Age. *Proceedings of the IEEE*, vol. 110, no. 7 (2022). doi:10.1109/JPROC.2022.3180350
3. Jerbić, B. et al: RONNA G4-Robotic Neuronavigation: A Novel Robotic Navigation Device for Stereotactic Neurosurgery. In: Abedin-Nasab(eds.) *Handbook of Robotic and Image-Guided Surgery*, Elsevier Inc (2019). doi:10.1016/B978-0-12-814245-5.00035-9
4. Rehab M. Ibrahim, F. E. Z Abou-Chadi, A. S. Samra. Plastic Surgery Face Recognition: A comparative Study of Performance. *International Journal of Computer Science Issues*, Vol. 10, Issue 5, No 2. (2013). doi:10.13140/RG.2.2.21078.46404
5. Simsek, I. et al: Analysis of surgical outcome after upper eyelid surgery by computer vision algorithm using face and facial landmark detection. *Graefe' s Archive for Clinical and Experimental Ophthalmology* (259), 3119-3125 (2021). doi:10.1007/s00417-021-05219-8
6. Baltrusaitis, T. et al: OpenFace 2.0: Facial Behavior Analysis Toolkit. In: *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, Xi' an (2018). doi:10.1109/FG.2018.00019
7. Lugaresi, C. et al: MediaPipe: A Framework for Building Perception Pipelines. (arXiv:1906.08172). arXiv. (2019). doi:10.48550/arXiv.1906.08172
8. King, D.: Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research* (10), 1755-1758 (2009). doi:10.1145/1577069.1755843
9. Schroff, F. et al: FaceNet: A Unified Embedding for Face Recognition and Clustering. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Boston (2015). doi:10.1109/CVPR.2015.7298682
10. Jakhete, Sumitra A., Kulkarni, Nilima: A Comprehensive Survey and Evaluation of MediaPipe Face Mesh for Human Emotion Recognition. In: *8th International Conference, Communication, Control and Automation (ICCUBEA)*, IEEE, India (2024). doi:10.1109/ICCUBEA61740.2024.10775188
11. Z. Zhu and R. Jiao.: Real-time Facial Expression Recognition Research Based on Blazeface Face Detection and Resnet Emotion Classification, *2024 5th International Conference on Computer Vision. Image and Deep Learning (CVIDL)*, pp. 401-408. China (2024). doi:10.1109/CVIDL62147.2024.10603598
12. Kumar, A. et al: Face detection in still images under occlusion and non-uniform illumination. *Multimedia Tools and Applications* (80), 14565-14590 (2021). doi:10.1007/s11042-020-10457-9

13. Bao, Y., Dang, R.: Face Detection under Non-Uniform Low Light Based on Improved MTCNN. In: 2nd International Conference on Artificial Intelligence and Computer Engineering, IEEE, Hangzhou, China (2021). doi:10.1117/12.2539617
14. Liang, J. et al: Recurrent Exposure Generation for Low-Light Face Detection. IEEE Transactions on Multimedia, vol. 24 (2022). doi:10.48550/arXiv.2007.10963
15. Al-Nuaimi, Arqam M., Mohammed, Ghasan J.: Face Detection Estimation based on MediaPipe Landmark. In: 7th International Conference on Contemporary Information Technology and Mathematics, Mosul, Iraq (2021) doi:10.1109/ICCITM53167.2021.9677878
16. Hammadi, Y. et al: Evaluation of Various State of the Art Head Pose Estimation Algorithms for Clinical Scenarios. Sensors (22), 6850 (2022). doi:10.3390/s22186850
17. Abdullah T. M., Ali N. H. M. Facial DeepFake Performance Evaluation Based on Three Detection Tools: MTCNN, Dlib, and MediaPipe. Fifth international conference on applied sciences: ICAS2023. (2024). doi:10.1063/5.0213294
18. Dill S., Rösch A., Rohr M., Güney G., De Witte L., Schwartz E., Antink C. Accuracy Evaluation of 3D Pose Estimation with MediaPipe Pose for Physical Exercises. Current Directions in Biomedical Engineering 9(1);563-566. (2023). doi:10.1515/cdbme-2023-1141
19. Galanakis I., Soldatos F. R., Karanikolas N., Voulodimos A., Voyiatzis I., Samarakou M. A MediaPipe Holistic Behavior Classification Model as a Potential Model for Predicting Aggressive Behavior in Individuals with Dementia. Applied Sciences 14(22):10266. (2024). doi:10.3390/app142210266
20. Shanmugam K. S., Badruddin N., Asirvadam V. S. Comparative Study of State-of-the-art Face Landmark Detectors for Eye State Classification in Subjects with Face Occlusion. IEEE 4th Euroasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). (2022). doi:10.1109/ECBIOS54627.2022.9945018
21. F. Suligoj, C. M. Heunis, J. Sikorski and S. Misra. RobUSt—An Autonomous Robotic Ultrasound System for Medical Imaging. IEEE Access, vol. 9, pp. 67456-67465. (2021). doi:10.1109/ACCESS.2021.3077037