# Multimodal Gesture Recognition with Spatio-Temporal

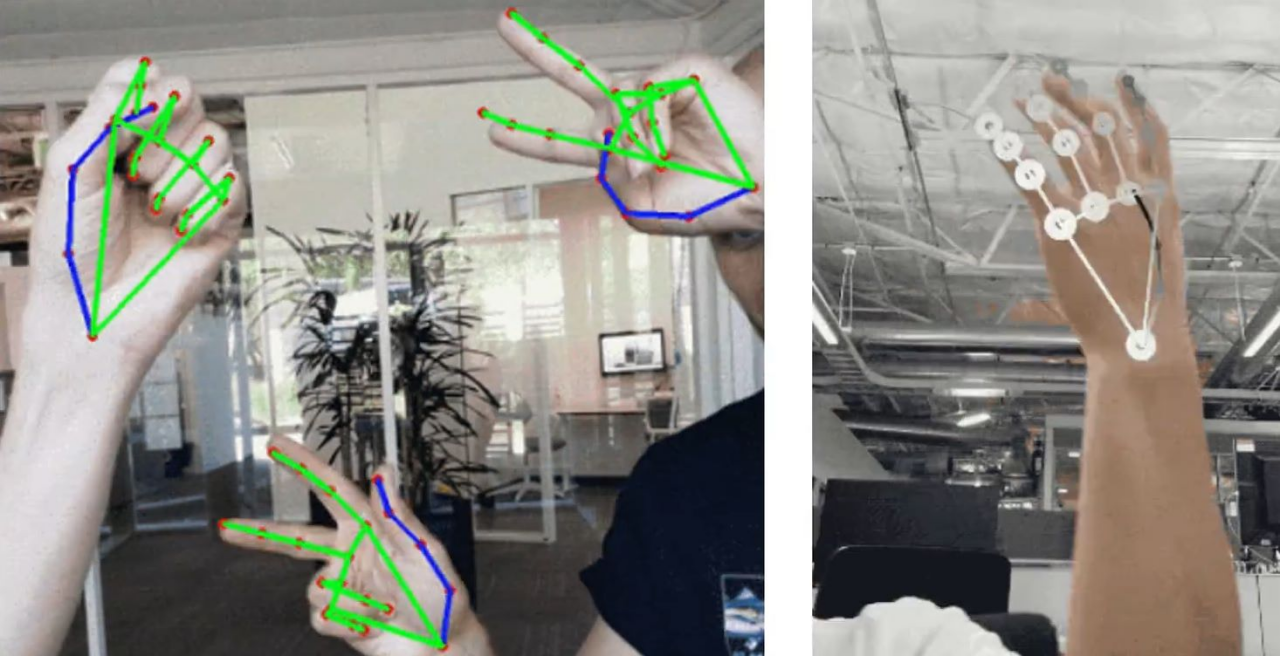
# Features Fusion Based on YOLOv5 and MediaPipe

*Hebei GEO University, Shijiazhuang 050031, P. R. China*

### Abstract (300 word limit)

As a natural, intuitive and easy-to-learn mode of interaction, gesture plays an important role in communication. Hand detection, containing multimodal information, includes static and dynamic detection and involves intricate spatial relationship problems such as different hand sizes, complex joints, occlusion, and self-occlusion. This study focused on a multimodal hand gesture recognition system based on YOLOv5 and MediaPipe with fused spatio-temporal features. Firstly, the Mediapipe and OpenCV library were employed to implement hand keypoint detection. Subsequently, the human-computer interaction of volume control was realized by identifying the distance between thumb and index. Finally, model training was conducted based on YOLOv5 algorithm, and the recognition of different gesture categories was realized. The performance was evaluated and compared through YOLOv5s, YOLOv5m, and YOLOv5l. The gesture recognition system interface visualization was achieved through pyqt5. Experiments show that the average detection accuracy of the model is 99.4% and the recognition speed is around 0.2 seconds.

### Image

****

**Figure. 2D (Left) and 3D (Right) Hand Keypoint Detection.**

**Recent Publications (minimum 5)**

1. Cao, W., Lu, P., & Cao, W. (2024). Multimodal Gesture Recognition with Spatio-Temporal Features Fusion Based on YOLOv5 and MediaPipe. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 38(08), 2455007 (2024). Doi: 10.1142/S0218001424550073

**Photograph**

Biography

WENYI CAO is a postgraduate student majored in Master of Applied Statistics at Hebei GEO University, China. The main research interests are in the fields of big data mining and analysis, machine learning, deep learning, data visualization, and time series analysis. Wenyi Cao has actively participated in numerous scientific research projects and academic skill competitions about big data and mathematical modeling, including 2 national-level innovation and entrepreneurship projects, 1 provincial-level project, and several university-level projects. The representative achievements include: (1) Multimodal Gesture Recognition with Spatio-Temporal Features Fusion Based on YOLOv5 and MediaPipe. International Journal of Pattern Recognition and Artificial Intelligence. (SCI & EI); (2) Study on the Relationship between Hang Seng Index and FTSE CHI Index in Hong Kong: Evidence from VAR Model. Proceedings of the 2nd International Conference on Business and Policy Studies Series: Advances in Economics, Management and Political Sciences. (CPCI).

Email: habaekcao0120@gmail.com

**Notes/Comments:**