2023 IEEE International Conference on Mechatronics and Automation:

Development of a Lightweight Real-Time Application for Dynamic Hand Gesture Recognition

Oluwaleke Yusuf, Maki Habib

Robotics, Control and Smart Systems (RCSS) Program,
Department of Mechanical Engineering,
The American University in Cairo (AUC), Cairo, Egypt.

Presentation Outline

- Background
- Literature Review & SOTA
- Proposed Framework & Application
- Evaluation Results & Comparison with SOTA
- Lightweight Application & Performance
- Conclusion & Future Work

Dynamic Hand Gesture Recognition

> Computer Vision
>> Perceptual Computing
>>> **Dynamic Hand Gesture Recognition (HGR)**

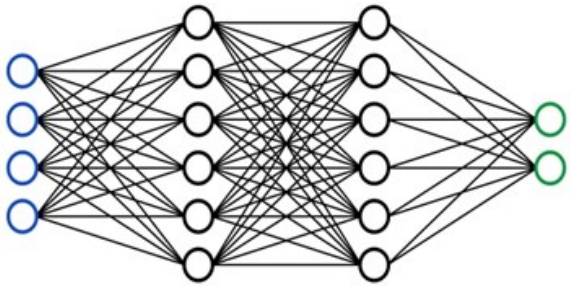
* **HGR Applications:**
  - Human-Machine Interactions
  - Human Behavior Analysis
  - Active & Assisted Living
  - Virtual & Augmented Reality
* **HGR Complexity:**
  - Human Hand
  - User Environment
  - Sensor Types

Existing HGR Frameworks

Input Modality

HGR Framework

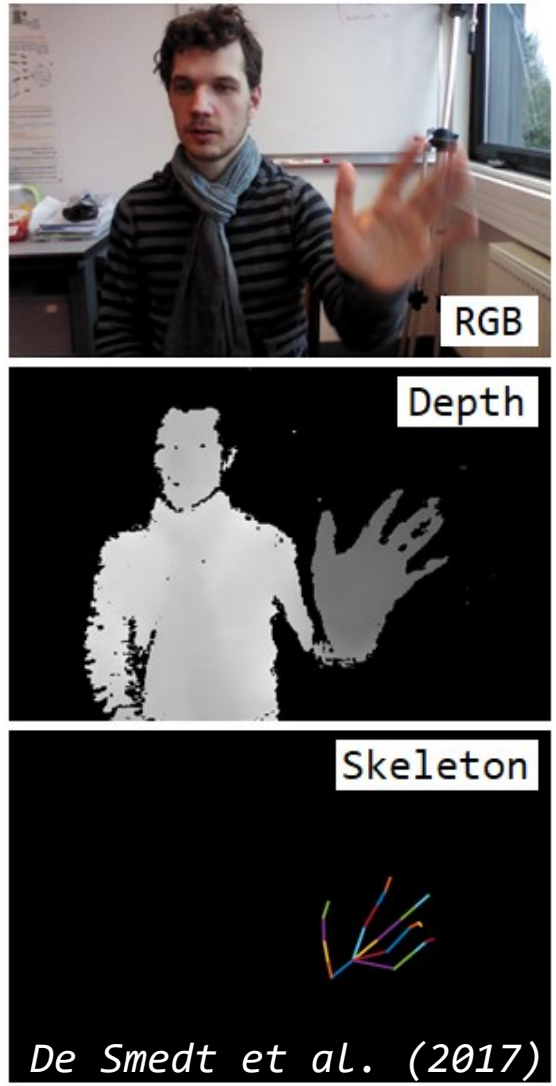
Gesture Classification



Network Architecture

State-Of-The-Art HGR Frameworks

- **Gesture Input Modalities:**
  - RGB [*Narayana et al. (2017), Köpüklü et al. (2019)*]
  - Depth [*Narayana et al. (2017), Köpüklü et al. (2019)*]
  - Skeleton [*Li et al. (2021), Shi et al. (2020), Sabater et al. (2021)*]
  - Optical Flow (RGB & Depth) [*Narayana et al. (2017)*]
- **Neural Network Architectures:**
  - Multi-Stream Fusion [*Li et al. (2021), Narayana et al. (2017)*]
  - Recurrent Neural Network [*Li et al. (2021)*]
  - (2D & 3D) Convolutional Neural Network [*Narayana et al. (2017), Köpüklü et al. (2019)*]
  - Graph Convolutional Network [*Li et al. (2021)*]
  - Attention Network [*Li et al. (2021), Shi et al. (2020)*]
  - Temporal Convolutional Network [*Sabater et al. (2021)*]



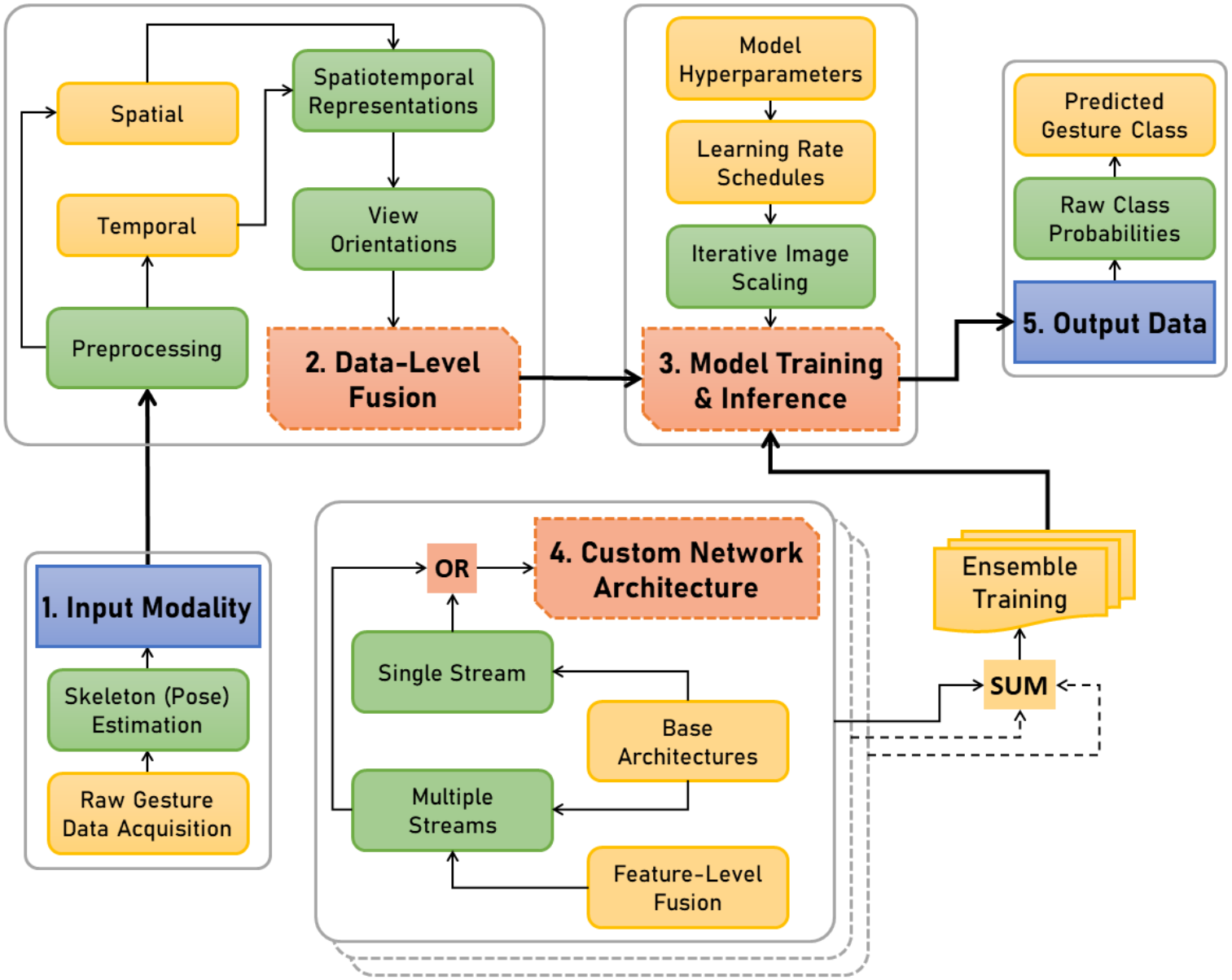
Constraints::

* **Hardware Requirements:**
  - Inbuilt PC Webcam
* **Computational Complexity:**
  - Minimize CPU & RAM Utilization

Requirements::

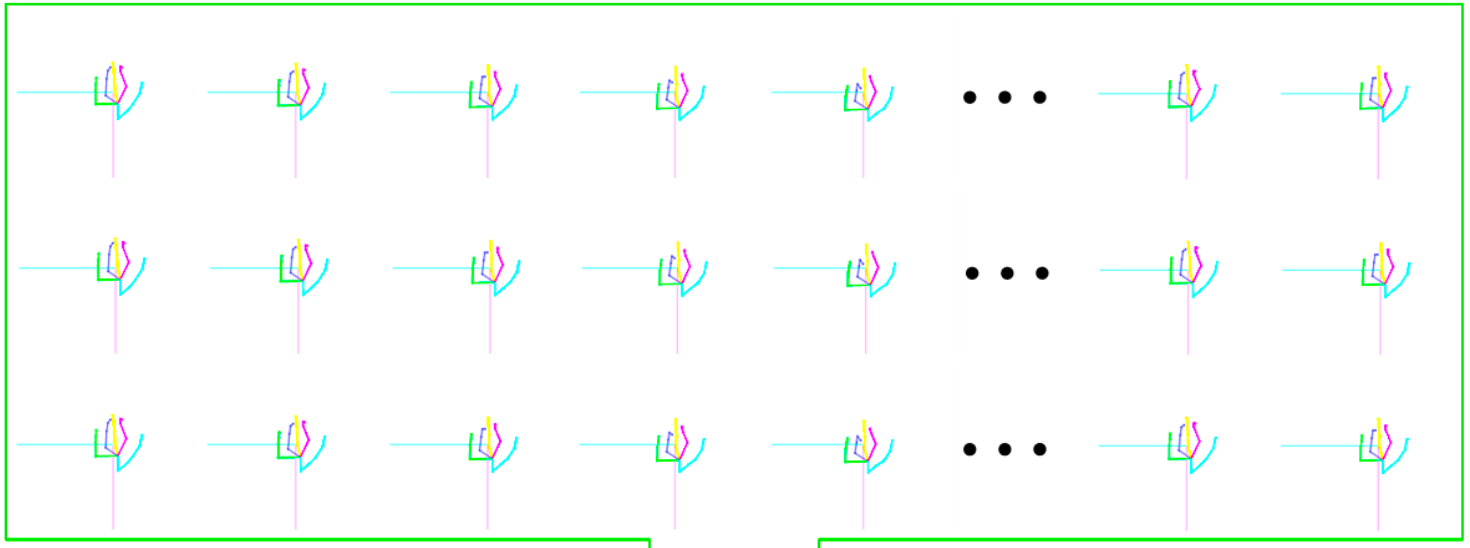
* **Classification Accuracy:**
  - Maintain SOTA Evaluation Performance
* **Real-Time Performance:**
  - Maximize FPS
  - Minimize Latency

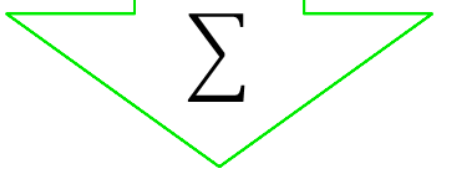
Application Constraints & Requirements

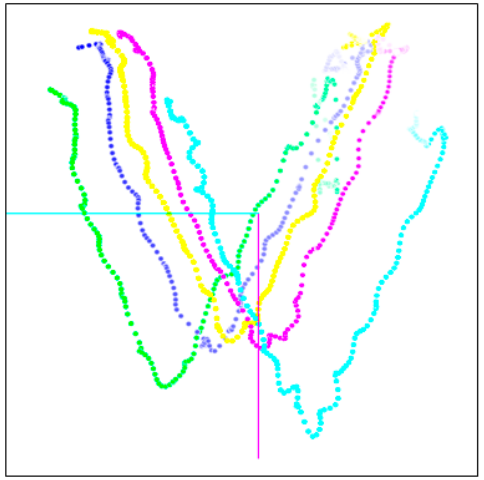


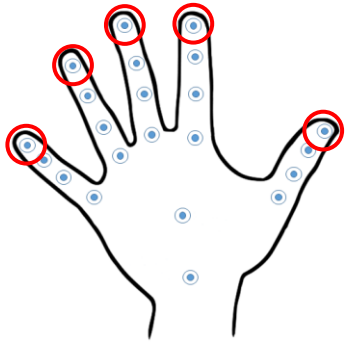
Proposed Dynamic HGR Framework

Temporal Trails Data-Level Fusion



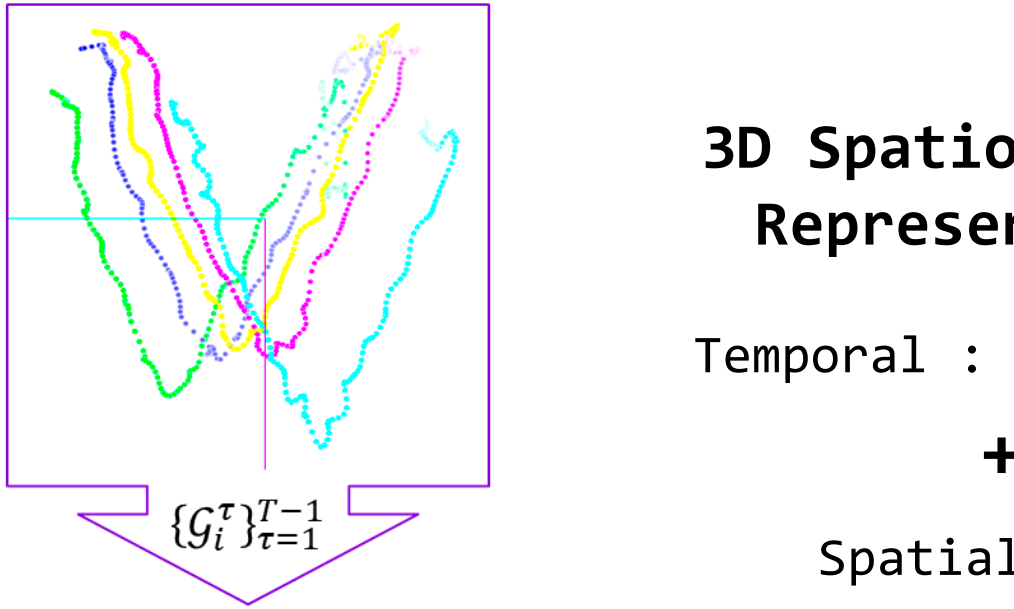






Temporal Information Condensation::

- $\sum\{\mathcal{G}_i^{\tau}\}_{\tau=1}^{T-1}$
  - Five Fingertips
  - Dynamic Gesture, i
  - Temporal Window, T

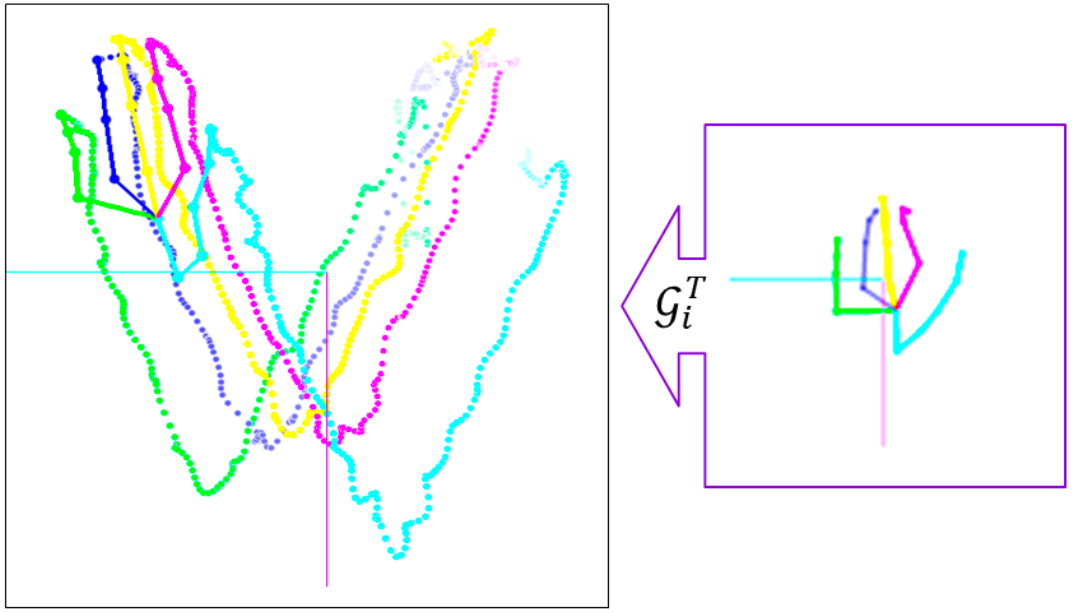


3D Spatiotemporal Representation

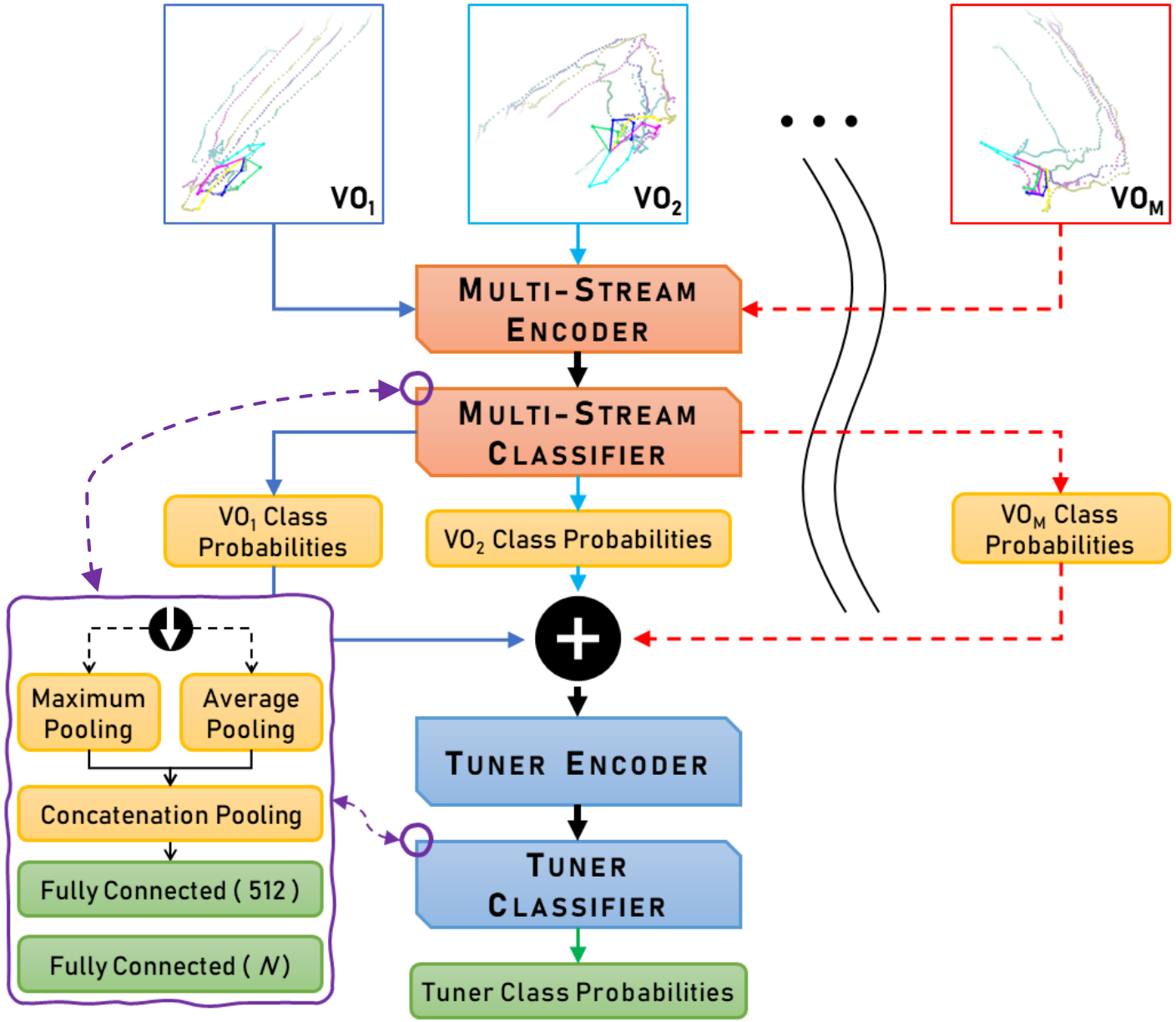
Temporal : $\sum\{\mathcal{G}_i^{\tau}\}_{\tau=1}^{T-1}$

+

Spatial : $\mathcal{G}_i^{T}$



Temporal Trails Data-Level Fusion

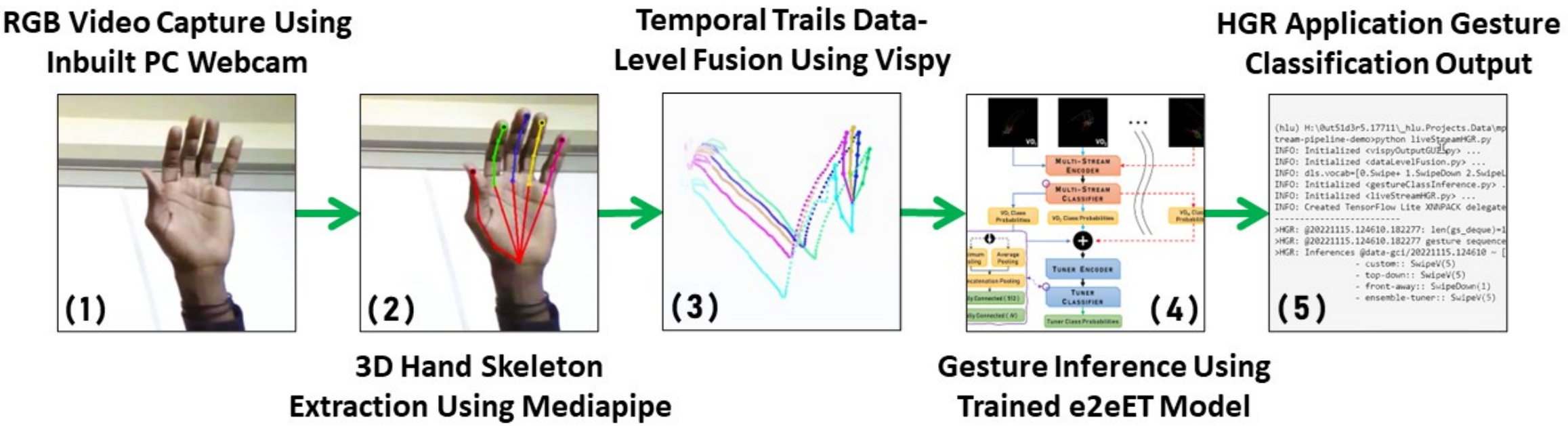


End-to-end Ensemble Tuner [e2eET] Multi-Stream CNN Architecture

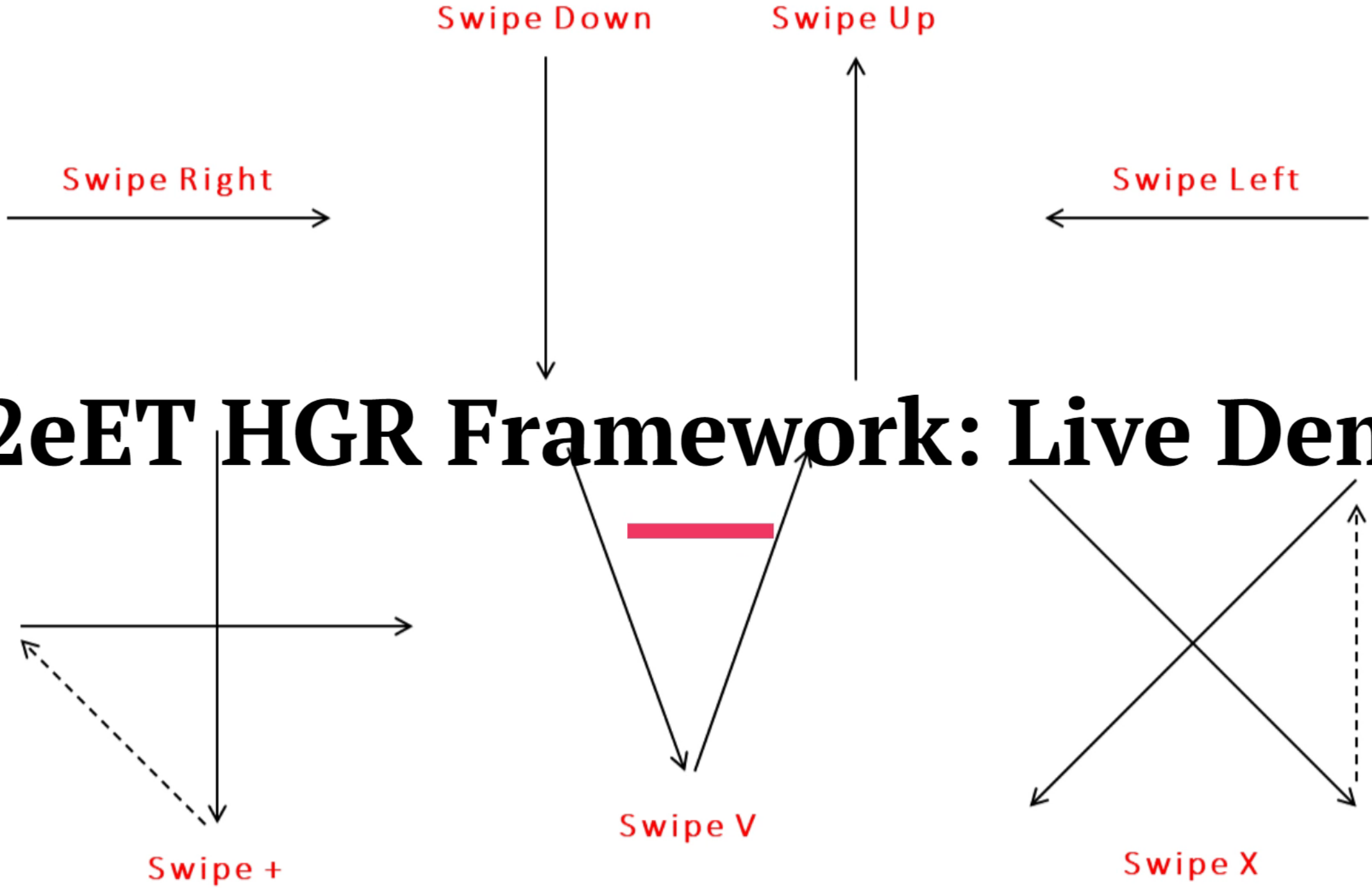
HGR Framework Accuracy Comparison w/SOTA

Datasets / SOTA Frameworks	Classification Accuracy	
	SOTA	Ours
1. Dynamic Hand Gesture 14/28 Dataset (DHG1428) – n=2800 ; c=14/28		
Li et al. (2021)	95.18%	94.11% [-1.07%]
2. 3D Hand Gesture Recognition....Dataset (SHREC2017) – n=2800 ; c=14/28		
Shi et al. (2020)	95.45%	96.61% [+1.16%]
3. Consiglio Nazionale delle Ricerche....Dataset (CNR) – n=1925 | c=16		
Lupinetti et al. (2020)	98.78%	97.05% [-1.73%]
4. Leap Motion Dynamic Hand Gesture Benchmark (LMDHG) – n=608 ; c=13		
SOTA: Lupinetti et al. (2020)	92.11%	98.97% [+6.86%]
5. First-Person Hand Action Benchmark (FPHA) – n=1175 ; c=45		
SOTA: Sabater et al. (2021)	92.93%	91.83% [-4.10%]

Lightweight Real-Time Application



Code available at:: https://github.com/Outsiders17711/e2eET-Skeleton-Based-HGR-Using-Data-Level-Fusion



e2eET HGR Framework: Live Demo

HGR Application Performance Analysis

* **Test Platform::**
  - Lenovo PC @ Windows 10
  - Intel Core i7-9750H CPU
  - 16GB RAM
* **HGR Application Stats::**
  - 4 Python Modules
  - 3 UI Windows
  - ~2s Latency @ 15FPS
  - 93.46% Classification Accuracy

PC Software	RAM Utilized
HGR Application Modules	648.9 MB
RGB Video Capture + Mediapipe Skeleton Estimation	*90.0 MB*
Vispy Data-Level Fusion	*86.6 MB*
e2eET Model Inference	*358.8 MB*
Vispy Output GUI	*82.6 MB*
Terminal	*30.8 MB*
Google Chrome [9 Tabs]	821.7 MB
Adobe Acrobat DC [3 Documents]	228.7 MB
Microsoft Word [3 Documents]	199.8 MB
Windows Explorer [3 Windows]	166.7 MB

Conclusion

+ Demonstrated the viability of Data-Level Fusion in HGR domain.
+ Leveraged advances in deep, data-driven ML algorithms and architectures.
+ Using a custom end-to-end Ensemble Tuner Multi-stream CNN Architecture.

↪ **HGR Framework & Application::**

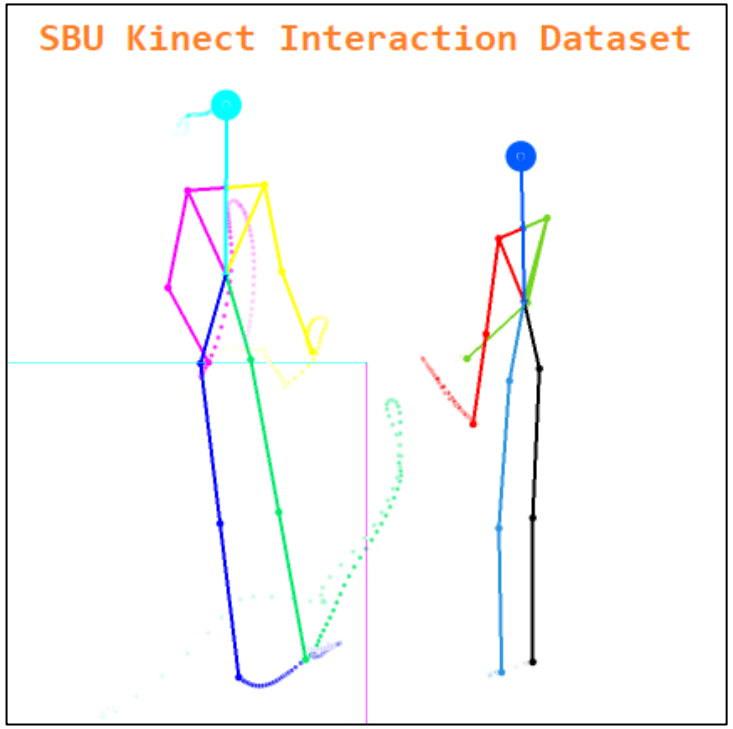
3D Hand Gesture Recognition ⟶ 2D Image Classification

☑ Obtained classification accuracies between -4.10% and +6.86% of SOTA.
☑ Minimized hardware requirements and computational complexity.

Future Work

Data-level Fusion::

- Gesture Visualization
- Image Generation



e2eET Architecture::

- Loss Function
- Ensemble Tuning Method

HGR Framework & Application::

- Unsegmented Data Streams
- Skeleton-based Human Action Recognition

References

[1] Bazil, *'Web Love GIF'*, GIFER, 2021. Available: https://gifer.com/en/2iv1.

[2] Q. De Smedt, H. Wannous, J.-P. Vandeborre, J. Guerry, B. Le Saux, and D. Filliat, *'SHREC'17 Track: 3D Hand Gesture Recognition Using a Depth and Skeletal Dataset'*, 2017.

[3] C. Li, S. Li, Y. Gao, X. Zhang, and W. Li, *'A Two-stream Neural Network for Pose-based Hand Gesture Recognition'*, 2021.

[4] P. Narayana, R. Beveridge, and B. A. Draper, *'Gesture Recognition: Focus on the Hands'*, 2018.

[5] L. Shi, Y. Zhang, J. Cheng, and H. Lu, *'Decoupled Spatial-Temporal Attention Network for Skeleton-Based Action Recognition'*, 2020.

[6] A. Sabater, I. Alonso, L. Montesano, and A. C. Murillo, *'Domain and View-point Agnostic Hand Action Recognition'*, 2021.

[7] O. Köpüklü, A. Gunduz, N. Kose, and G. Rigoll, *'Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks'*, 2019.

[8] K. Lupinetti, A. Ranieri, F. Giannini, and M. Monti, *'3D Dynamic Hand Gestures Recognition Using the Leap Motion Sensor and Convolutional Neural Networks'*, 2020.

Thank You!
Questions?