

# Gesture Vocabulary in OPTIMAI

---



**CERTH**  
CENTRE FOR  
RESEARCH & TECHNOLOGY  
HELLAS



This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 958264

The material presented and views expressed here are the responsibility of the author(s) only.  
The EU Commission takes no responsibility for any use made of the information set out.

**OPTIMAI**

# Contents

- 1 Introduction and Functionality
- 2 Real-time Gesture Recognition Module
- 3 EgoGesture Dataset
- 4 Indicative Outputs



A woman wearing a white hard hat and safety glasses is looking at a tablet in a construction site. The background is a blurred construction site with other workers.

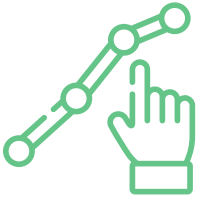
# Introduction

1

# What is Gesture Recognition?



Gesture recognition is the AI method deployed in the OPTIMAI project for understanding the operator's motives in the factory floor using the AR glasses.

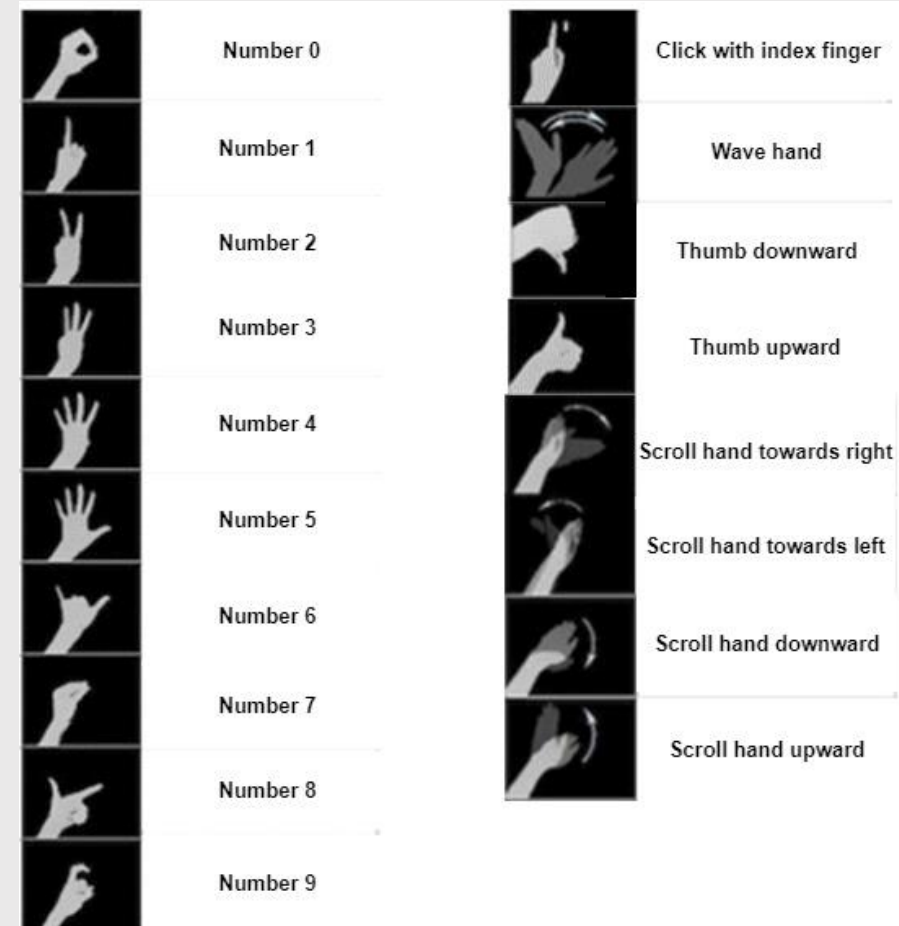


The AR functionalities allow the rapid adjustment of significant parameters and the ability to start/stop the inspection process of defected products, using a vocabulary of hand gestures. Thus, hand gesture recognition will be performed in the feed from the AR glasses aiming to understand the operator's intentions.



# EgoGesture Dataset




- > All of the gestures required by the OPTIMAI necessities are included in the open-source **EgoGesture dataset**.
- > The defined gesture vocabulary contains 18 gestures described as:
  - > Numbers from 0-9
  - > Click with index finger
  - > Thumbs up/down
  - > Wave
  - > Swipe up/down/left right



^ (1) EgoGesture dataset

# Functionality of the Gestures

- > The initial OPTIMAI gestures vocabulary is finalized for MTCL and TVES use case aiming to support the operators efficiently to interact with the system at the shop-floor.
- > Specifically, the users can utilize the implemented AR applications through gestures to activate functions related to the outputs of defect analysis provided by the OPTIMAI system.

Gesture	Function
	Accept initiation of Inspection process
	Select respective menu item
	Stop all process and return to home page

- > (2) Gestures vocabulary for MTCL use case

Gesture	Function
	Accept initiation of Inspection process or Play video & Pause Video
	Select respective menu item
	Rewind & Fast Forward video by seconds
	Stop Production or stop all processes and return to home page

- > (3) Gestures vocabulary for TVES use case

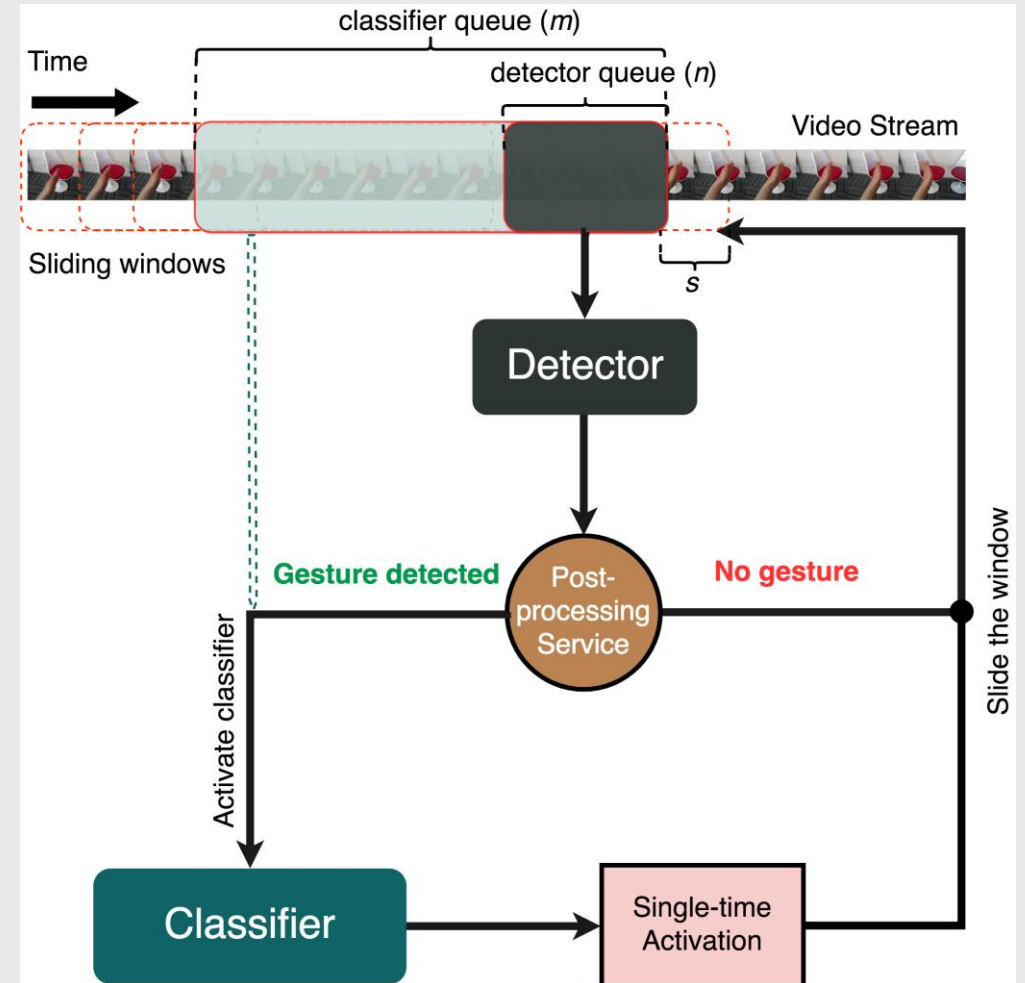


# Real-time Gesture Recognition Module

2

# Real-time Gesture Recognition Module

- › Regarding the gesture recognition task with real-time performance, the authors propose using the architecture that is presented in [1], [2].
- › The authors propose a workflow with a sliding window approach that activates the classifier upon detecting a gesture, using a detector as a switch. The proposed architecture is illustrated here.
- › Additionally, the workflow includes post-processing and single-time activation services

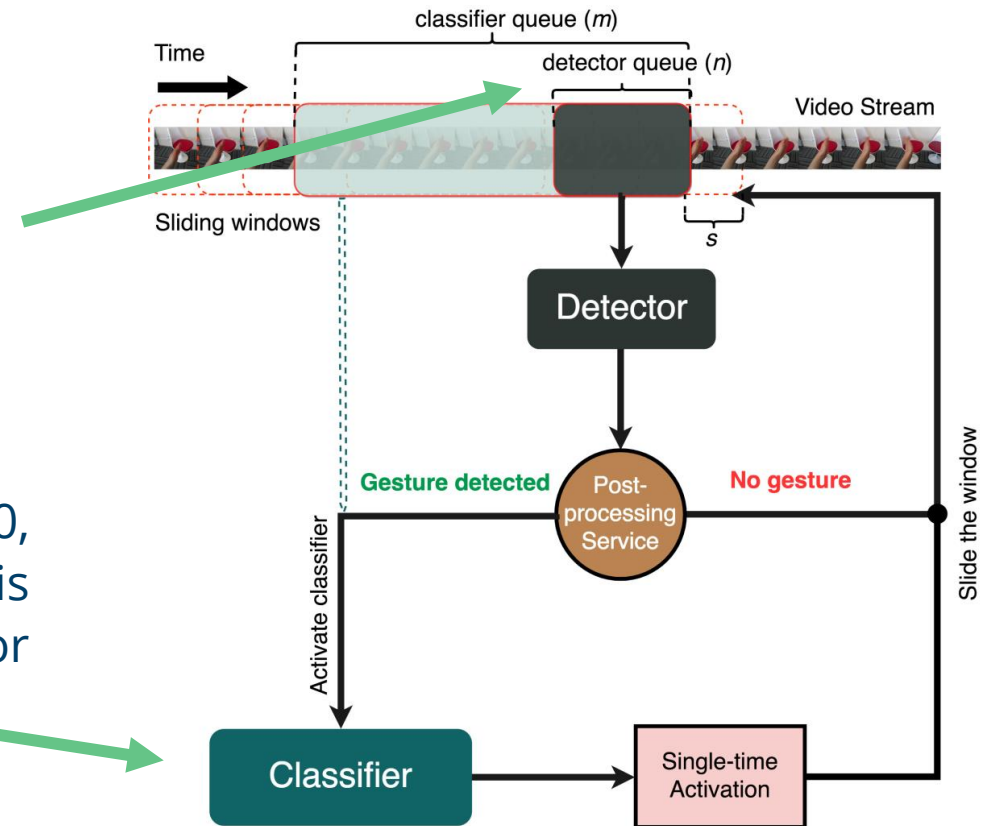




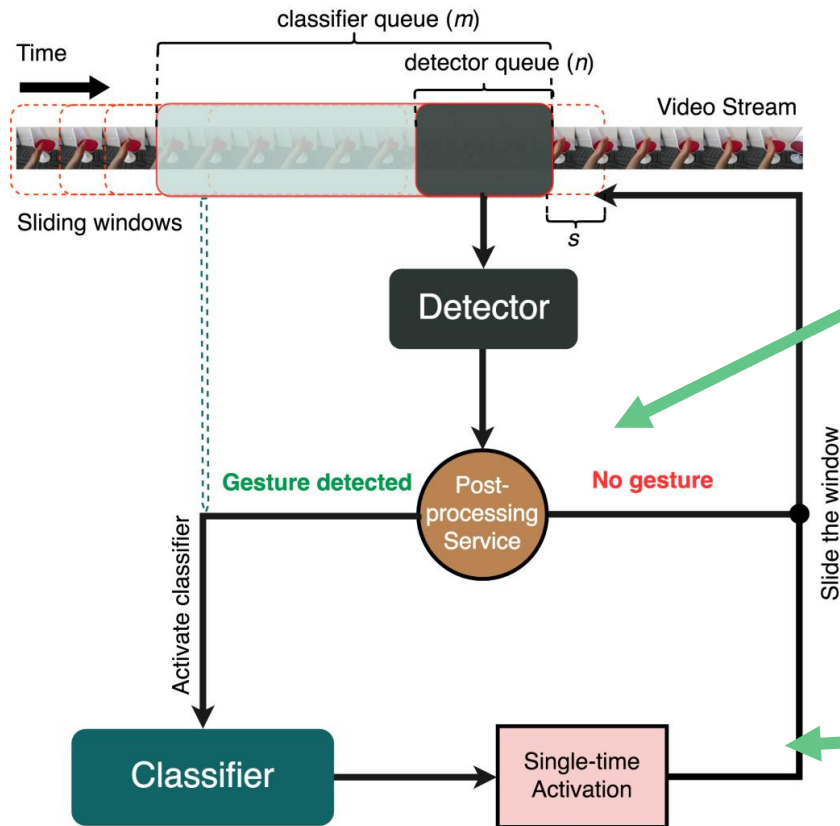
# Real-time Gesture Recognition Module

1 The **detector**'s main goal is to identify when a gesture takes place by analyzing a sequence of images, which is passed through the detector queue. The sole purpose of this component is to activate the classifier in case a gesture is detected.

2 In terms of a **classifier**, the authors use ResNext-10, although, as stated, their proposed architecture is independent of the actual model-type used for gesture classification purposes.



# Real-Time Gesture Recognition Module



3

**Post Processing:** A typical cause for misclassifications and overall decreased system performance, when it comes to dynamic hand gestures, is hands moving out of the camera's field of view previous detector predictions. To combat this issue, they put forward a technique that utilizes the raw softmax probabilities from.

4

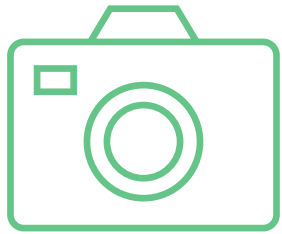
The proposed method involves a two-level control mechanism for achieving **single-time activation**. According to the authors, there are two possible scenarios for gesture detection: early-detection and late-detection. In the former case, the architecture detects a gesture when a threshold level is superseded by a confidence measure before the end of the gesture. In the latter case, when a deactivation of the classifier is triggered by the detector, there's a gesture prediction.

# EgoGesture Dataset

3

# EgoGesture Dataset

- › For the OPTIMAI purposes, training, validation and testing of the network is performed using the **Egogesture dataset**. This is a large-scale dataset of hand gesture recognition in an egocentric viewpoint, which contains all the gestures of interest for the purposes of OPTIMAI. It includes:



*24,161 gesture samples using a wearable camera*



*With 83 gesture categories*



*Performed by 50 subjects*

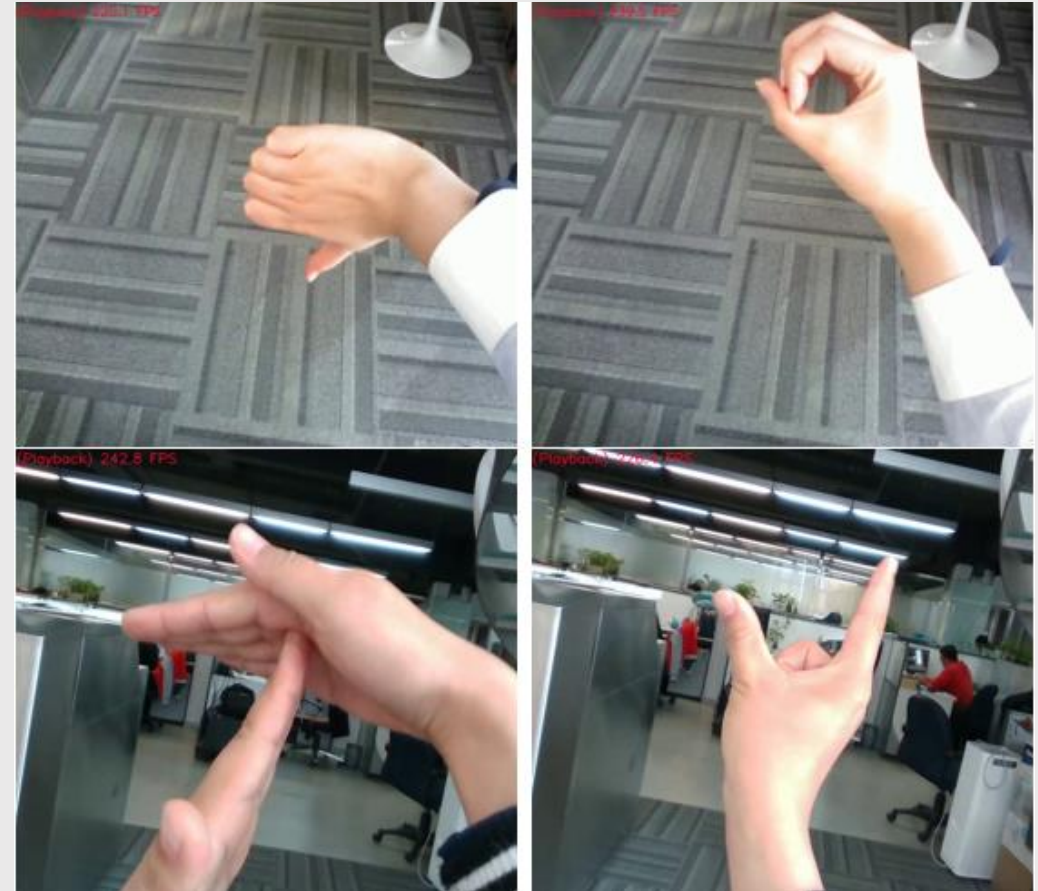
- › The dataset provides preprocessed skeleton data for each video sample, which can be used for gesture recognition and analysis. The dataset is freely available for research purposes.

# | Indicative Outputs

4

# Indicative Outputs

- › In this section, some sample outputs (frames) that are produced during online testing of the network in its current stage of development and improvement for the purposes of OPTIMAI are demonstrated.
- › These images serve as a demo for the gestures that are captured from the video. The sample images show four of the gestures in the vocabulary that are captured from video sources that are inputted to the network



# OPTIMAI

Thank you!



**CERTH**  
CENTRE FOR  
RESEARCH & TECHNOLOGY  
HELLAS



This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 958264

The material presented and views expressed here are the responsibility of the author(s) only. The EU Commission takes no responsibility for any use made of the information set out.