University POLITEHNICA of Bucharest
Faculty of Electronics, Telecommunications and Information Technology

Demo application for real-time person identification

Thesis advisor

Conf. dr. ing. Horia CUCU

Student

Gabriel Octavian GALER

Contents

- Motivation
- State of the project
- Application overview
- Application flowcharts
- Contributions
- Conclusions
- Further development

Motivation

- CDep meetings broadcast (http://www.cdep.ro/)
- Human intervention for banners
- Room for improvement using deep learning:
  - Identity prediction (showcased in this project)
  - TTS (text-to-speech) addition

State of the project

2019

- Acquisition of data
- Speech recording web service [1]

2020

- Face identification system
- Voice identification system
- Multimodal system [2]

2021

- Video processing application
- Plenum / non-plenum dataset
- Video player optimization

[1] Cristian MANOLACHE - Speech Recording Web Service and Application

[2] Gabriel SANDU - Multimodal Person Identification using Deep Neural Networks

Application overview

- An application which adds identity banners for persons speaking in plenum, given an input video
- Optimization for video player
- Plenum/non-plenum dataset and classifier

Application overview

presidium

plenum

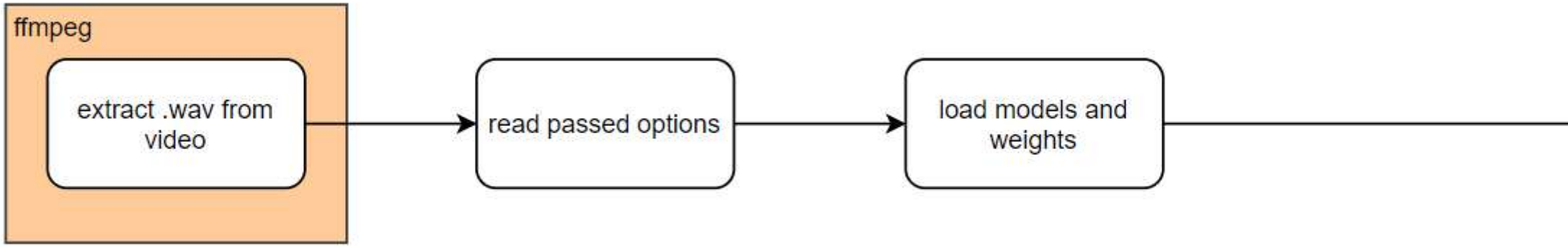
amphitheater

Examples of captures from a CDep recording

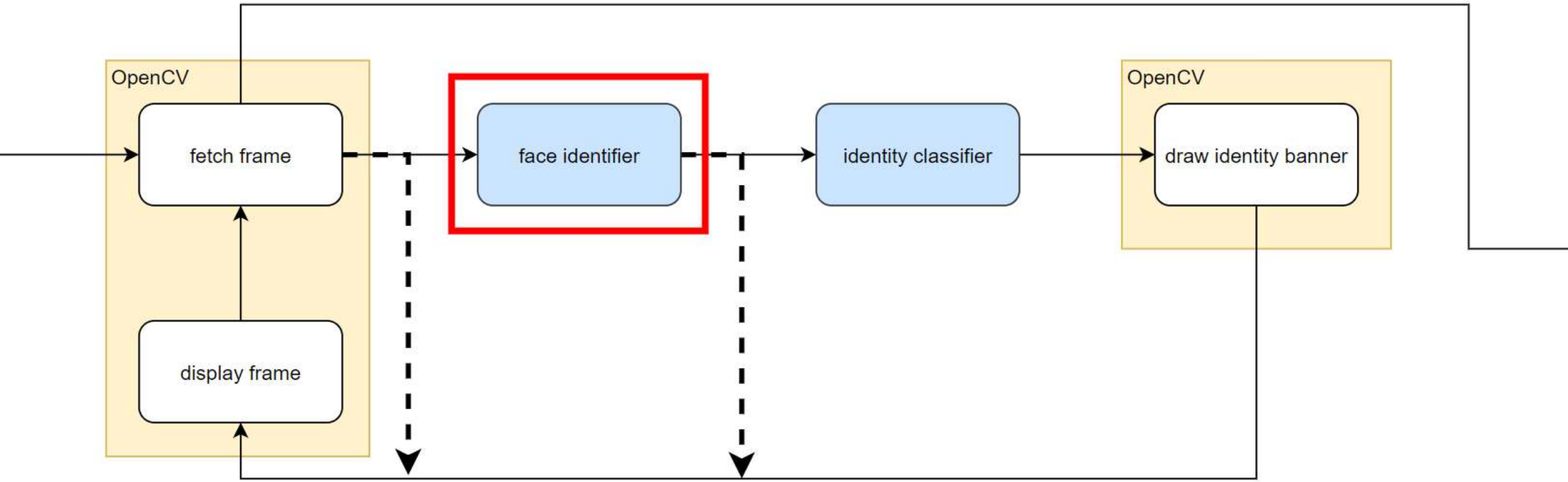
Flowcharts overview

- Video preprocessing
- Main loop
- Merging and logs generation

Flowcharts – Video preprocessing



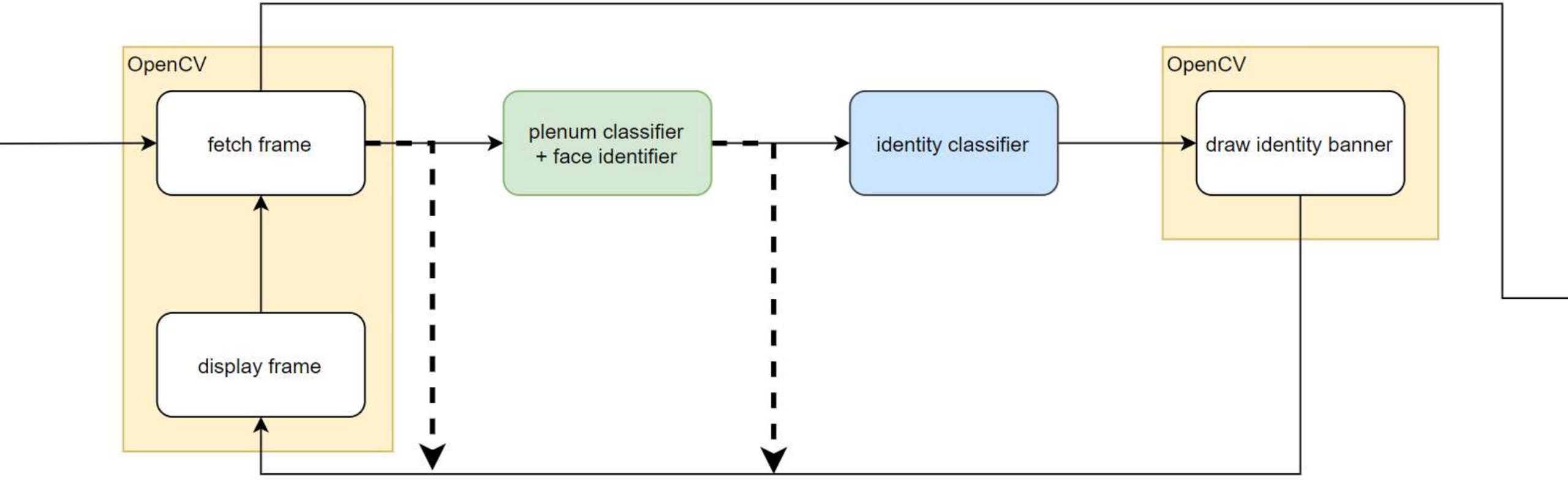
Flowcharts – Main loop



Identified problems

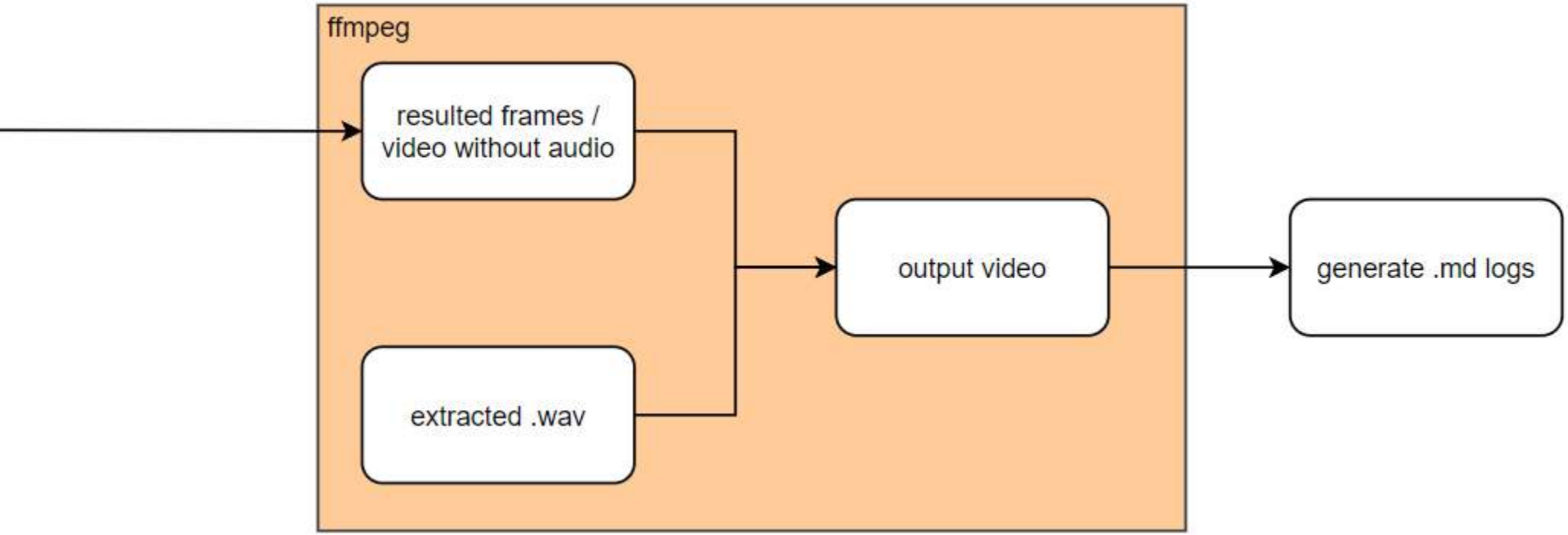
- Face detection in other camera angles other than plenum
- The need for plenum classifier

Flowcharts – New main loop



Flowcharts – merging and logs generation

- On end (if in video writer mode)
- The frames obtained from the main loop are piped into a video file

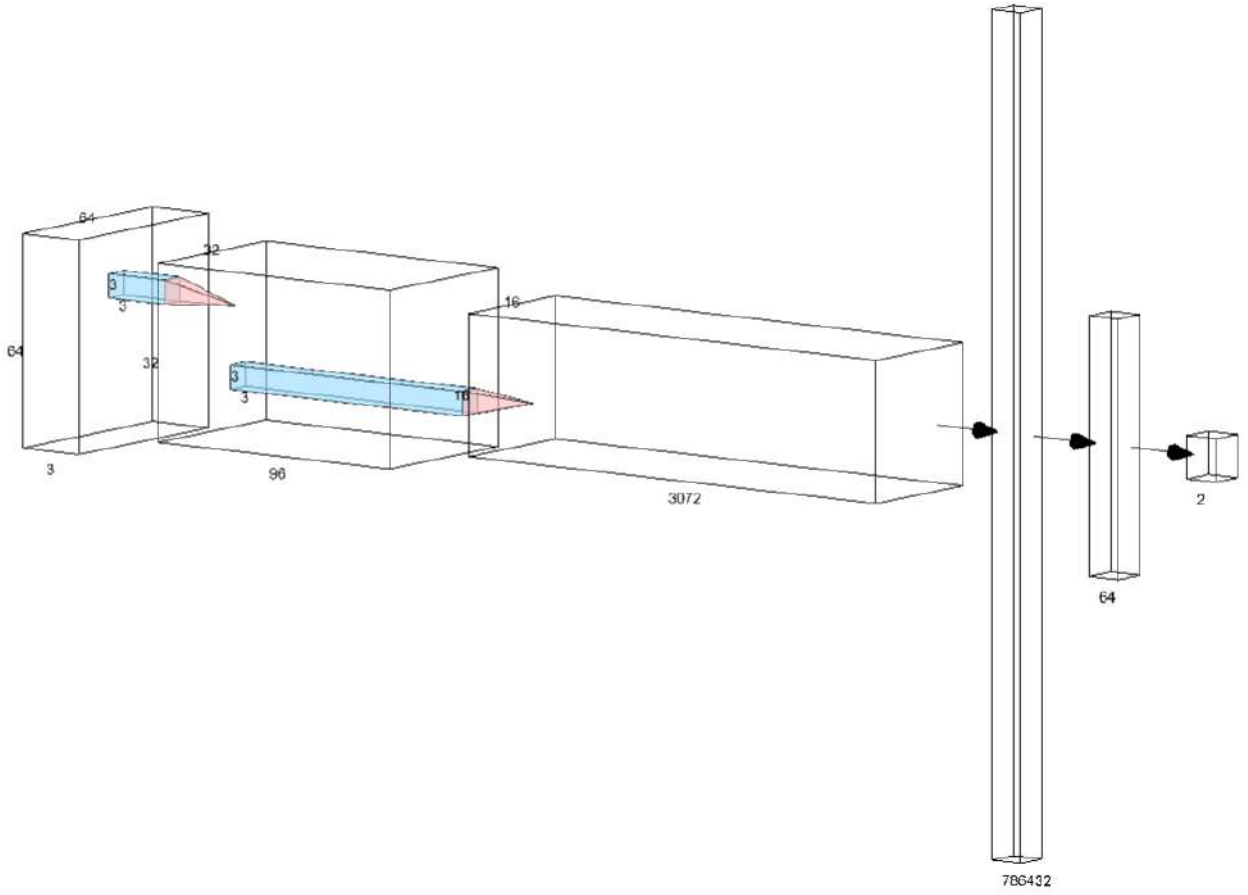


Contributions – Plenum/non-plenum dataset

- Created a plenum/non-plenum dataset
- ~7k images (~3k for plenum, ~4k for non-plenum)

Contributions – Plenum classifier

- CNN architecture, using 2 CONV+POOL layers, followed by 2 FC layers
- Trained for 10 epochs on the dataset
- Training accuracy 0.99, test accuracy 0.97

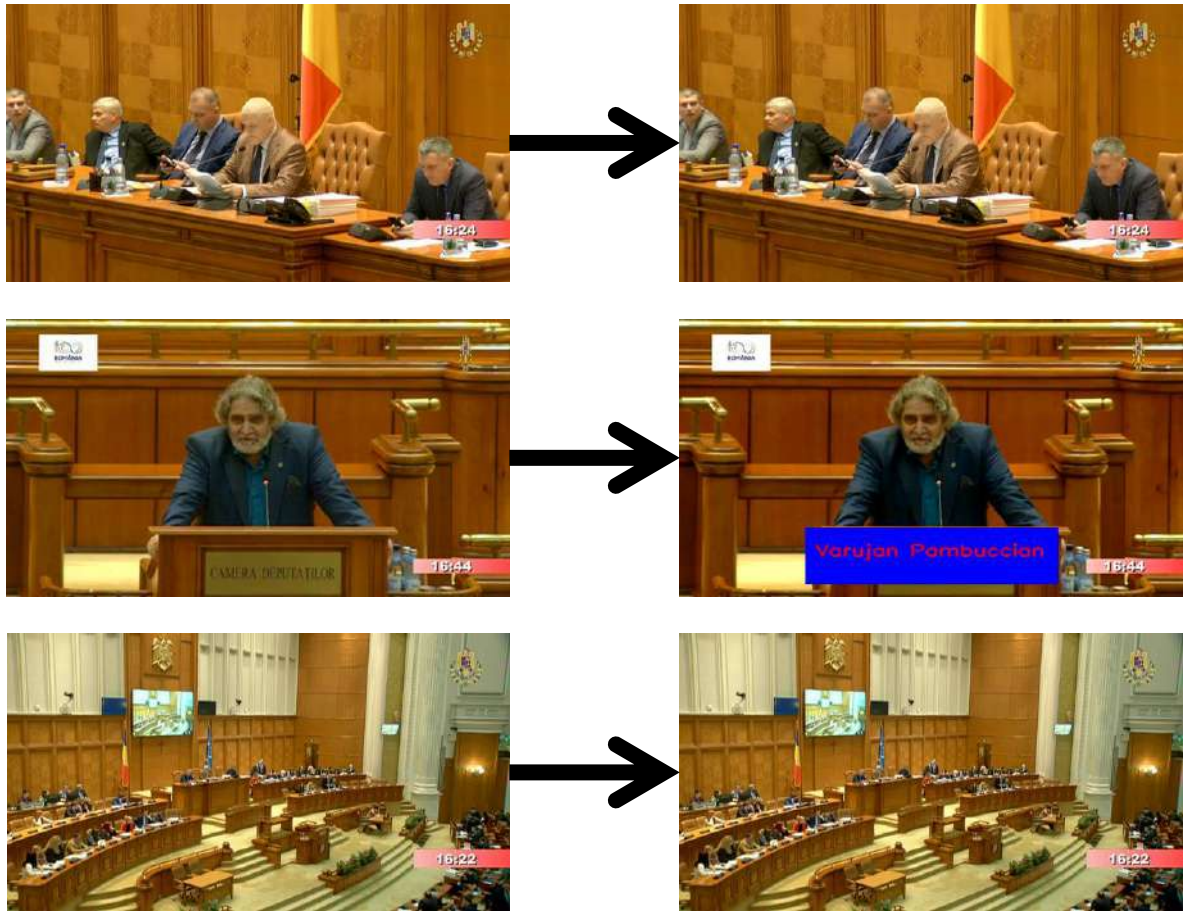


Contributions

- Main contributions:
  - An application which adds identity banners for persons speaking in plenum
  - Created a plenum/non-plenum dataset (~7k photos)
  - Created a CNN plenum classifier
- Additional contributions
  - Created a script for extracting the last 3 seconds of audio from a given frame index
  - Created handling functions for frames and audio which ensures compatibility with the network
  - Created classes for files cleanup and logging
  - Implemented a state-of-the-art NN for face detection (YOLO, pre-trained on WIDER)

Conclusions

- Successful first version of the application
- Further development is necessary for production



Further development

- Stream support
- Extend the identity classifier
- Testing with GPU
- Further investigations on face detection
- Tackle mask problem during the pandemic

Thank you for listening!