IARIA

PANEL #3

PORTO
July 2024

IARIA Congress 2024 & DigiTech 2024

Theme: Advances on Data Processing and Computation Power Optimization

IARIA Congress 2024 & DigiTech 2024

Theme: Advances on Data Processing and Computation Power Optimization

Panel #3 Advances on Data Processing and Computation Power Optimization

Porto
July 2024

Moderator

Prof. Dr. Constantine Kotropoulos, Aristotle University of Thessaloniki, Greece

Panelists

Emeritus Prof. Dr. Fritz Laux, Reutlingen University, Germany
Prof. Dr. Tatiana Kalganova, Brunel University London. UK
Prof. Dr. Matthias Harter, Rhein Main University of Applied Sciences, Germany
Prof. Dr. Constantine Kotropoulos, Aristotle University of Thessaloniki, Greece

Panel #3 Advances on Data Processing and Computation Power Optimization

Porto
July 2024

Emeritus Prof. Dr. Fritz Laux

- **Limitations in processor clock speed and parallel processing**
- **Need for powerful algorithms**

Prof. Dr. Tatiana Kalganova

- **Algorithm structure adaption against available hardware: Accuracy first, efficiency second!**
- **Event-based problems in AI systems**
- **Understanding useful data in the context of AI model**

Panel #3 Advances on Data Processing and Computation Power Optimization

Porto
July 2024

Prof. Dr. Matthias Harter

- Power consumption for AI training and inferencing

Prof. Dr. Constantine Kotropoulos

- Objective in multimedia forensics: maximizing classification, detection, and estimation accuracy
- Camera model identification depends on RESNET models; needs graphics card NVIDIA 48 GB RTX A6000, previous NVIDIA RTX 2080TI 11G becomes obsolete
- Optimizers for deep learning circumventing saddle points in a non-convex setting

Panelist Position: Algorithms!

PORTO
July 2024

Fritz Laux
Reutlingen
University

- **My Position**
  - **Processor clock speed stagnates**
    - **Moore's law not any more valid for clock speed**
    - **Only 50% increase in the last 10 years.**
  - **Parallel processing has limitations**
    - **Amdahl's law: Sequential part limits speedup by parallel processing**
    - **It is unclear if all algorithmic problems can be solved by a parallel algorithm**

Panelist Position: Algorithms!

PORTO
July 2024

Fritz Laux
Reutlingen
University

- **My Position**
  - Algorithms are more important than Processing power
    - **Using better algorithms can improve performance and power consumption by orders of magnitudes compared to hardware improvements**
  - **Examples:**
    - **using hash vs indexing for searching,**
      - ✓ hash is 20 times faster in searching in 10^6 items than a binary index
    - **using asymptotic expansion vs polynomial approximation for functions,**
      - ✓ An asymptotic expansion can be 100 times faster than a power series approximating sin(x) with the same accuracy
    - **using decision trees vs unsupervised mining.**
      - ✓ Certain cluster algorithms may not converge at all.

Panelist Position: Green AI

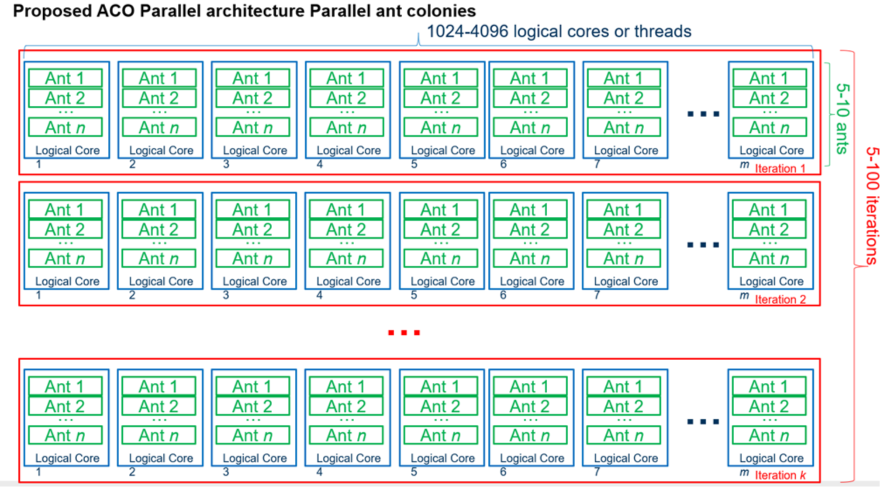
Porto
July 2024

- **Panel #3 Advances on Data Processing and Computation Power Optimization**
  - Algorithm structure adaption against available hardware: Accuracy first, efficiency second!

Prof Tatiana Kalganova

Brunel University London

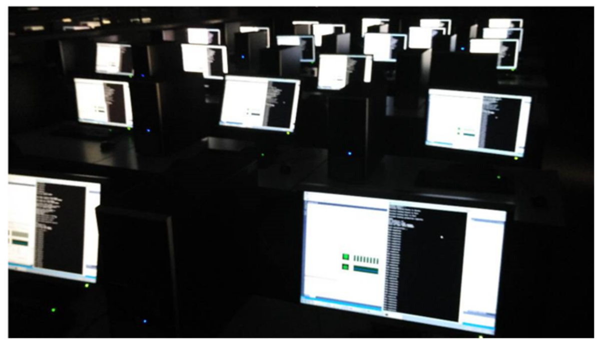
Proposed ACO Parallel architecture Parallel ant colonies



Parallel ACO architectures

Speed… Speed… Speed…

- Speed performance improved from **1.5 months** on HPC to **24 min** on PC, **without** loosing the quality of the solution.
- The algorithm has been modified to accommodate the distribution features of available hardware.
- The developed algorithm provides **consistently** better results in comparison with the base line provided.
- Total experimental computation time for one case study is **803 hours** that is equal to **33.5 days** (on Intel Core i7-4790) having completed **2 715** individual simulations.

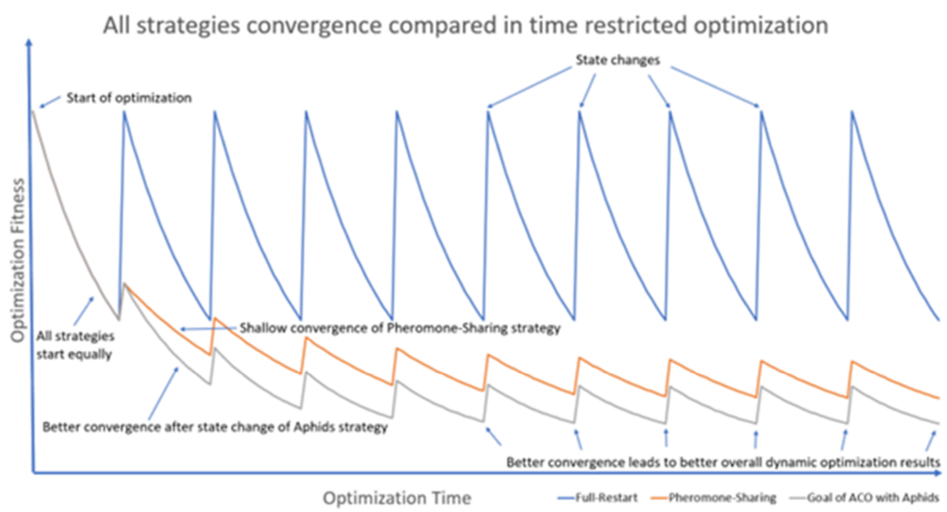


Impact of parallelisation

Panelist Position: Green AI

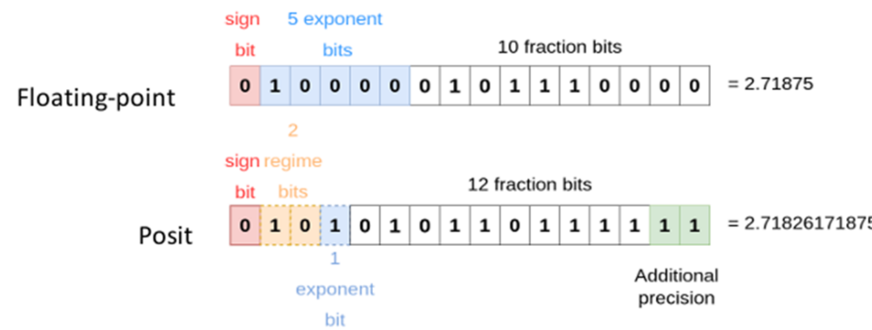
Porto
July 2024

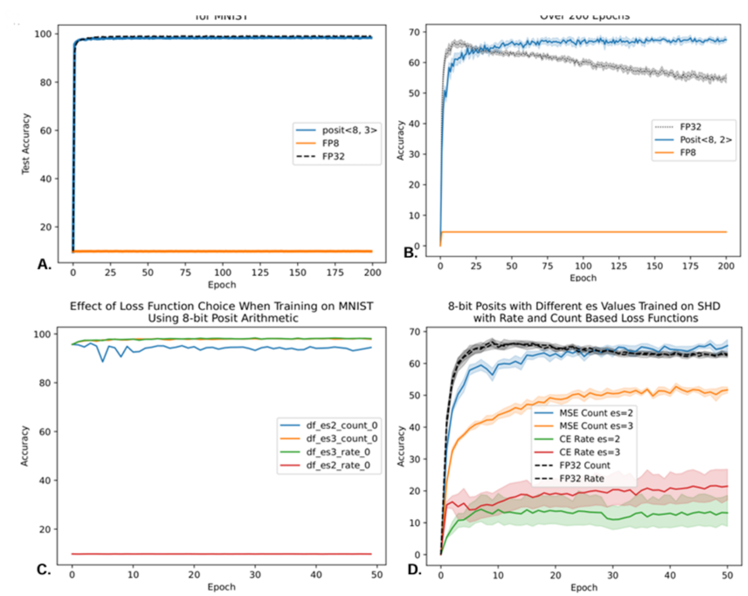
- Event-based problems in AI systems



All strategies convergence compared in time-restricted optimization. The chart displays a minimization problem's convergence for ten dynamic states changing frequently. Aphids' strategy combines great optimization convergence observed in the Full-Restart Strategy with low state change fitness penalty observed in the Pheromone-Sharing strategy. The combination of these strengths allows for better interstate convergence.

Never start from beginning:
Dynamic optimisation





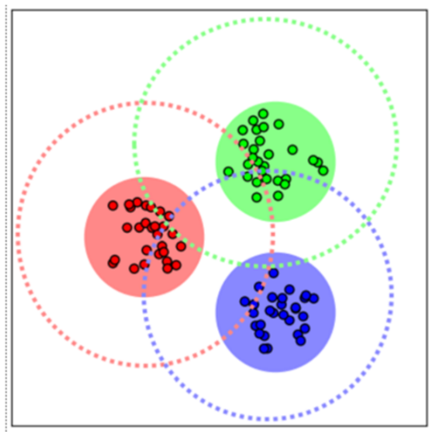
A & C show the results for MNIST, C & D show the results for SHD. The top row (A & B) show the training results and the bottom row (C & D) the loss function test results.

Spiking neural networks using posits:
event-based neuromorphic approach

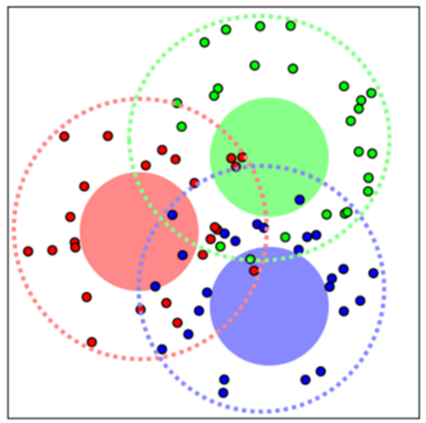
Panelist Position: Green AI

Porto
July 2024

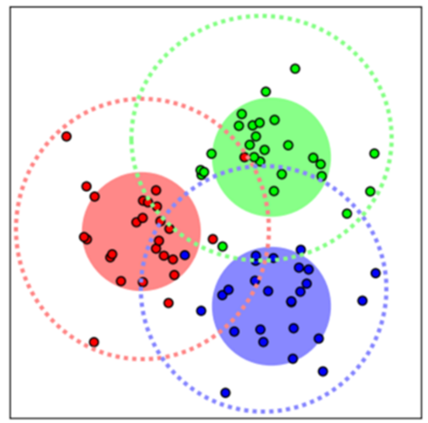
- Understanding useful data in the context of AI model



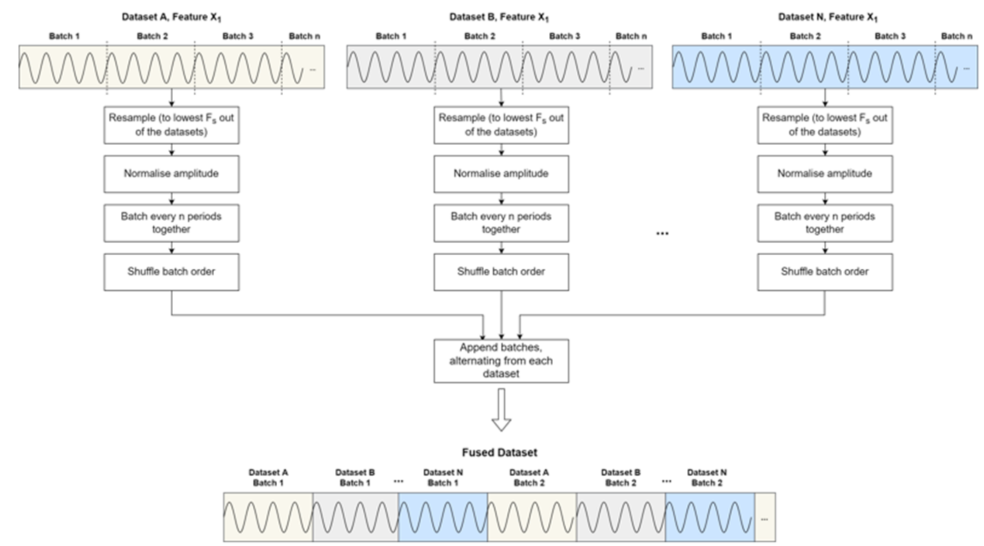
Visualization of 2-Dimensional Data Generated for 3 Relatively Well Separated Classes (*Lateral Exclusion* Visualized)



Visualization of 2-Dimensional Data Generated for 3 Relatively Well Separated Classes (*Central Exclusion* Visualized)



Visualization of 2-Dimensional Data Generated for 3 Relatively Well Separated Classes (*Random Exclusion* Visualized)



Data exclusion strategy influences the AI ML model performance

Dataset fusion together with generalization techniques lead to not only 90% savings in data used, but minimizes efforts on data collection

Panelist Position: Green AI

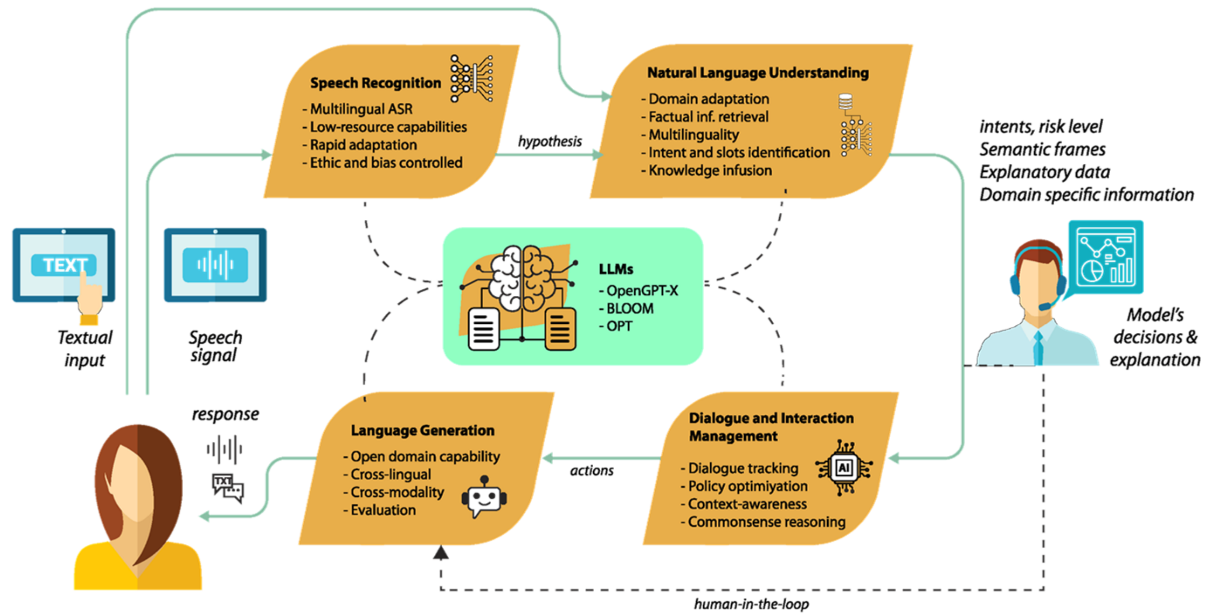
Porto
July 2024

The EU ELOQUENCE project aims to develop a Large Language Model (LLM) that is a skilled conversational agent, with minimised bias, strong EU values and high-levels of trustworthiness. Intelligent data use and **computational minimisation** techniques are vital to this project in ensuring compliance with European values, and the promotion of **green AI**.

The LLM will perform this role across both speech and text input, all EU member languages, and with utilisation of external knowledge.

Our current roles for this project:

- Dataset sourcing and fusion, suitable for conversational dialogues
- Evaluation of trustworthiness across freeform, unlabelled text
- Production of external knowledge databases for trustworthy evaluation:
  - RAG and Knowledge graphs
- Reduction of computational requirements through data quality identification



ELOQUENCE introduces four pilots aimed at advancing conversational AI in privacy, bias detection, empathetic virtual agents, and AI-enhanced call centers.

Pilot 1: Privacy-preserving language model learning through decentralised training in smart homes

Pilot 2: Social context-aware language model detecting biases

Pilot 3: Retrieval-augmented LLMs as virtual agents, capable of understanding the user's goals, making API calls and respond with empathy

Pilot 4: Upgrading support call centres through AI-based supervision of multimodal dialogues

Safety critical applications, requiring advanced technologies to ensure privacy throughout

LLM hybridisation with external knowledge to enhance explainability of AI decisions

Conversational evaluation framework for continued improvements to the system

Validate all ELOQUENCE technologies across the range of pilots stated above

Trustworthy and unbiased dialogues for the conversational agent

Validate technologies across EU cultural values, as well as ethical, legal and societal requirements

eloquence

Panelist Position

PORTO
July 2024

THE WALL STREET JOURNAL.

SIGN IN SUBSCRIBE

Cooling towers at Plant Vogtle, a nuclear power station in Waynesboro, Ga. KENDRICK BRINSON FOR THE WALL STREET JOURNAL

BUSINESS | ENERGY & OIL

There's Not Enough Power for America's High-Tech Ambitions

Georgia is a magnet for data centers and other cutting-edge industries, but vast electricity demands are clashing with the newcomers' green-energy goals

By *David Uberti* Follow
Updated May 12, 2024 12:16 pm ET

THE WALL STREET JOURNAL.

SUBSCRIBE SIGN IN

BUSINESS | ENERGY & OIL

Big Tech's Latest Obsession Is Finding Enough Energy

The AI boom is fueling an insatiable appetite for electricity, which is creating risks to the grid and the transition to cleaner energy sources

By Katherine Blunt Follow *and Jennifer Hiller* Follow
March 24, 2024 7:00 am ET

Share Resize 384 Listen (2 min)

Bill Gates said at the 2024 CERAWeek in Houston last week that electricity is the key input for deciding whether a data center will be profitable. PHOTO: F. CARTER SMITH/BLOOMBERG NEWS

HOUSTON—Every March, thousands of executives take over a downtown hotel here to reach oil and gas deals and haggle over plans to tackle climate change. This year, the dominant theme of the energy industry's flagship conference was a new one: artificial intelligence.

Matthias Harter
Hochschule RheinMain

Panelist Position

PORTO
July 2024

- **Power consumption for AI training and inferencing: Some alarming statements by Casey Handmer’s blog**
  https://caseyhandmer.wordpress.com/2024/03/12/how-to-feed-the-ais/
  - **“... AGI will create an irresistibly strong economic forcing function to pave the entire world with solar panels – including the oceans. ... ”**
  - **“At current rates of progress, we have about 20 years before paving is complete.”**
  - **“Is AI a solar panel maximizer, rather than a paperclip maximizer?”**

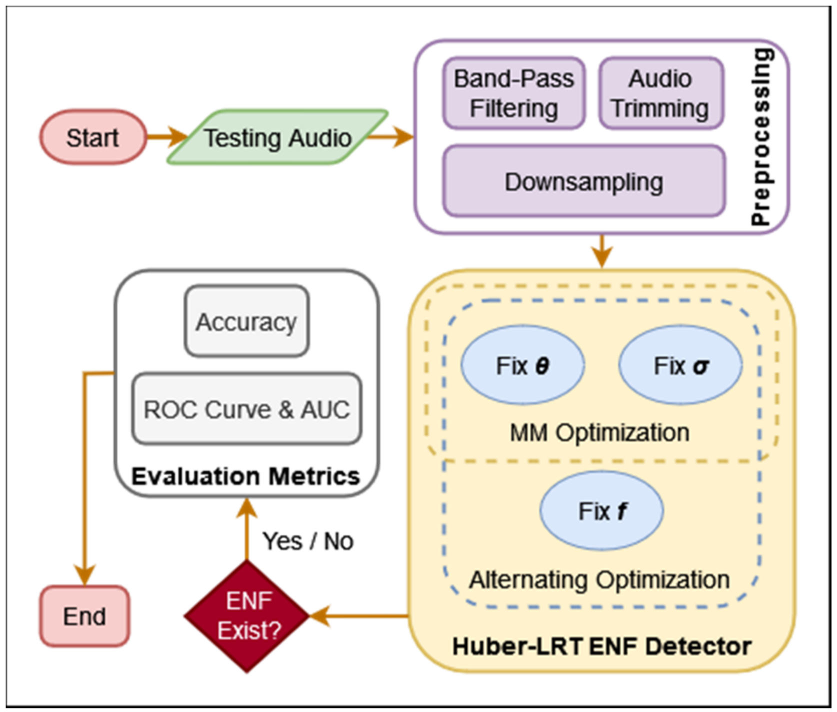
Matthias Harter
Hochschule
RheinMain

Panelist Position: Multimedia Forensics: quest for higher accuracy

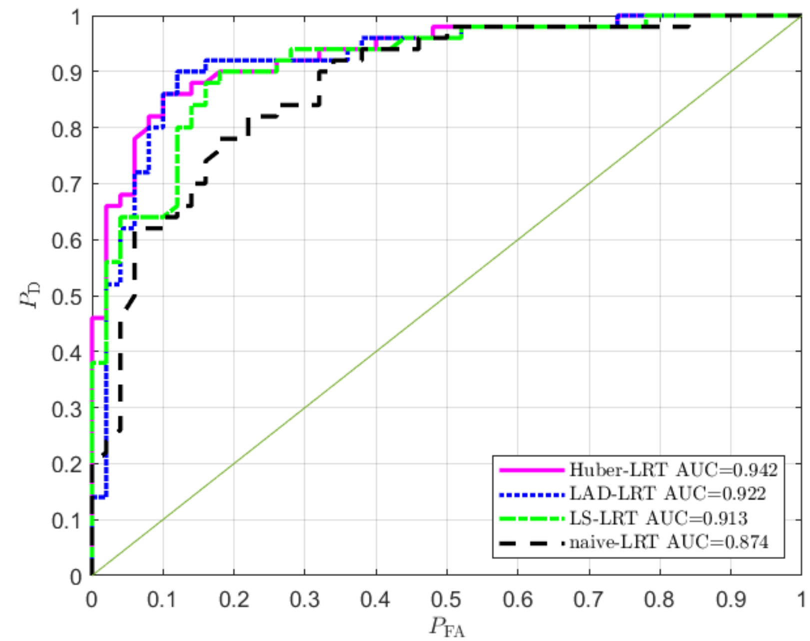
Porto
July 2024

- **Panel #3 Advances on Data Processing and Computation Power Optimization**
  - Detect Electric Network Frequency

Constantine
Kotropoulos
AUTH



Flowchart of the Huber-Likelihood Ratio Test (Huber-LRT) ENF Detector.

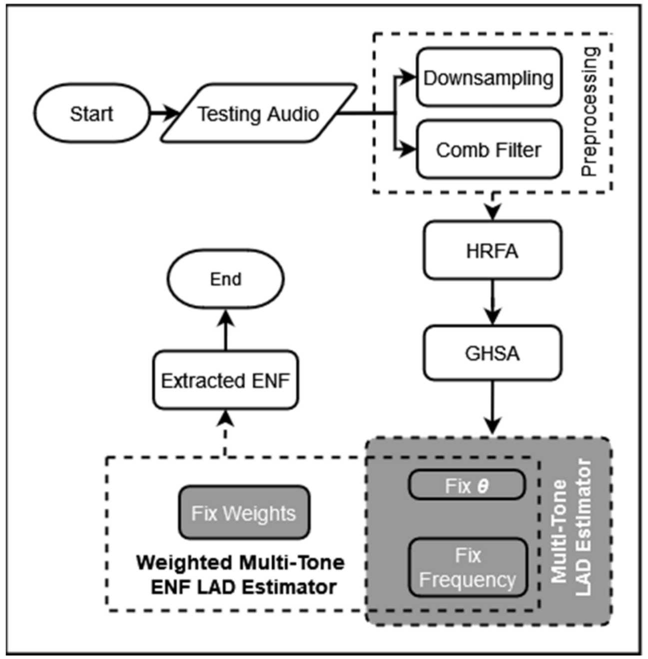


Receiver Operating Characteristic (ROC) curves and their Area Under the curve (AUC) for clips of 5-sec duration from ENF-WHU dataset.

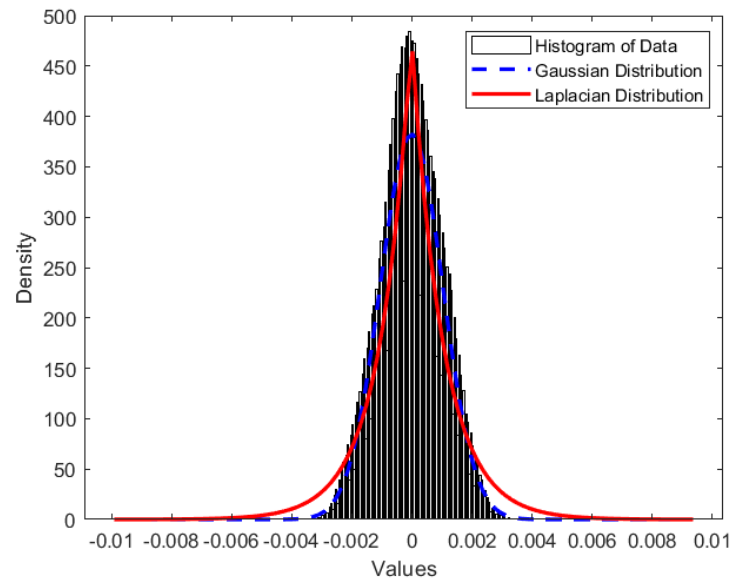
Panelist Position: Multimedia Forensics

Porto
July 2024

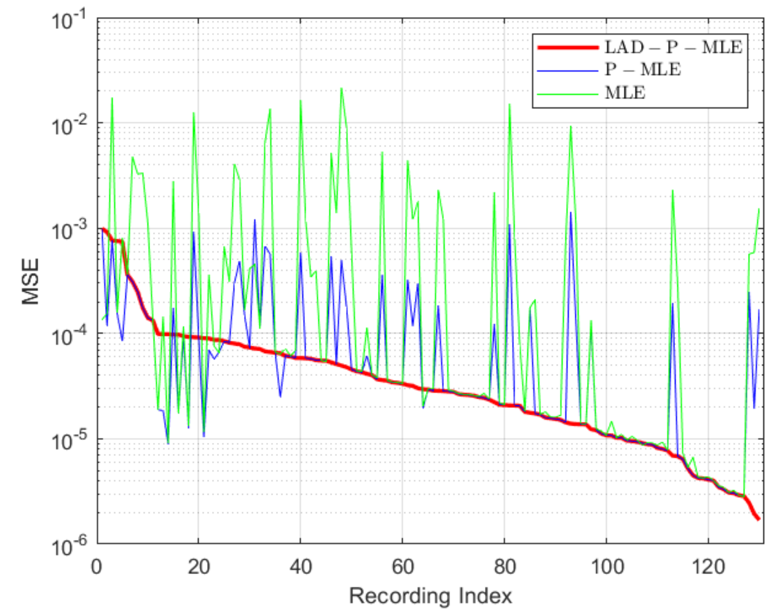
- Estimate the Electric Network Frequency



Flowchart of the proposed LAD-based ENF Estimation framework.



Histogram of the noise samples in audio recording #129.

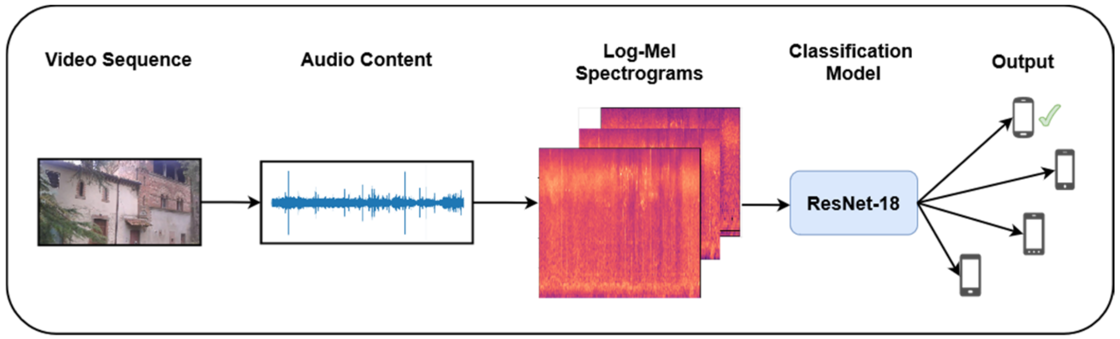


Comparison of MSE across the 130 audio recordings from ENF-WHU dataset for LAD-P-MLE P-MLE, and MLE estimation.

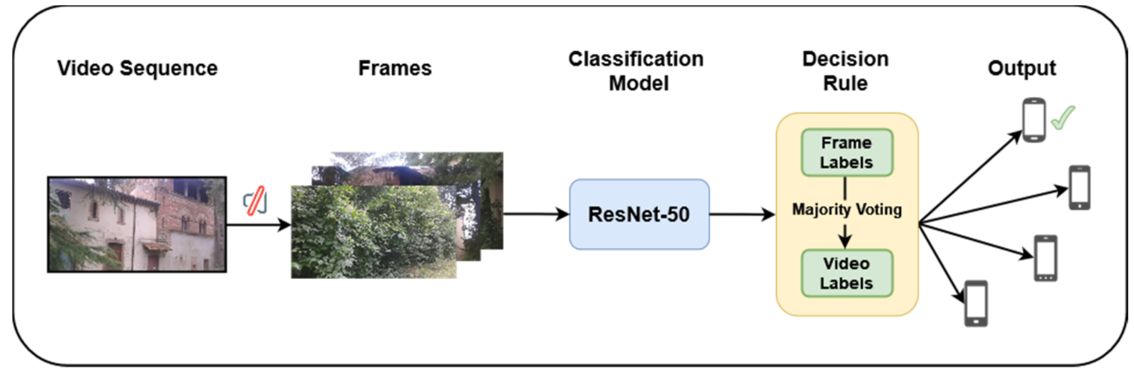
Panelist Position: Multimedia Forensics

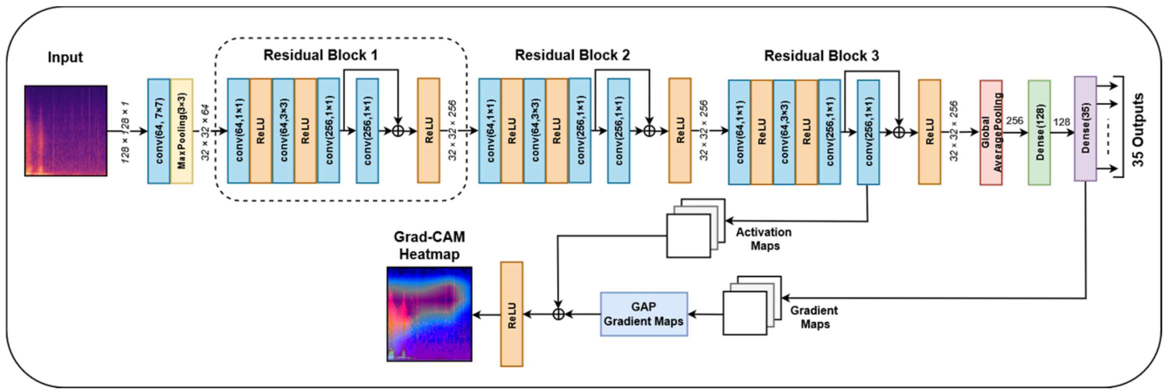
Porto
July 2024

- Camera Model Identification/Source Model Identification

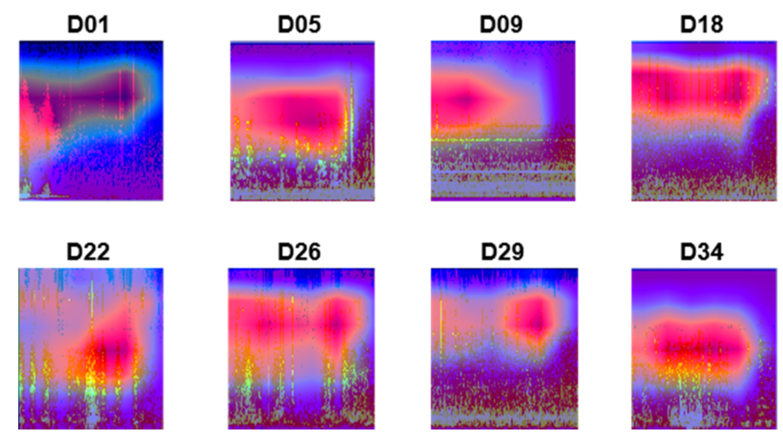


Vision dataset





xAI-CMI



The best classification accuracy on MOBIPHONE for model identification on all sets of augmentations

	Audio		Non-speech		UBM
	ML	SVM	ML	SVM	SVM
Baseline	99.2063%^{8}	96.4286%$^{1/MC}$	100%^{4}	99.2063%$^{1/M}$	98.8095%$^{4/M}$
All augmentations	93.739%^{64}	92.0635%$^{2/MC}$	89.9471%^{64}	94.0035%$^{1/M}$	93.2099%$^{2/M}$

STAGE IS YOURS