Internship proposal

Internship subject

Deep fuzzy extractors for document integrity check

Keywords

Deep learning, fuzzy extractors, integrity checking, document processing

Context of the study

The current health situation is forcing the authorities to use digitized copies of paper documents. Nevertheless, the widespread availability of professional image-editing tools, simple scanning devices and the accessibility of high-quality printing tools are increasing the number of document forgeries. Scanned copies can be easily forged using certain image editing tools (such as Photoshop or Gimp) or new approaches based on the use of deep learning. As a result, there is a great need for efficient, robust solutions for verifying the integrity of printed documents, which are then digitized. The aim is to extract a signature from an electronic document that can be used to verify the integrity of digitized documents.

Description of the subject

When an electronic document is printed and scanned several times, a slightly different image of the document - due to the optical characteristics of the capture devices - is obtained each time. A similar problem arises in biometrics. For example, when we capture the same fingerprint several times, it is not possible to obtain perfectly identical images, even if they are very close. The difficulty of developing a method for verifying the integrity of printed and scanned documents is similar to the difficulties encountered in biometrics: we want to record and then compare characteristics in order to deduce that two sets of data do indeed represent the same thing, despite the presence of noise.

To consider the uniqueness property of biometric data, fuzzy extractors [3] robust to sensor noise are used [1]. Unlike falsified biometric data, a falsified document does not differ significantly from its authentic version, making integrity verification more complex.

In previous work, an initial document integrity verification system was set up [1,2]. The features extracted are based on the analysis of intersections and bifurcations within alpha-numeric characters.

We are now interested in exploring the use of machine learning for fuzzy feature extraction, based on a method previously developed in biometrics [4].

The objectives of this internship are:

1. Explore the use of neural networks for fuzzy feature extraction [4].
2. Adapt method [4] to document integrity verification.
3. Compare the existing methods [1,2] with developed deep learning-based methods.

Required profile

- The candidate must currently be enrolled in a Master 2 program or in the final year of engineering school (that corresponds to Bac+5 in France) in Computer Science.
- Programming languages: Python.

- Libraries for image analysis and processing: OpenCV, scikit-image (Python).
- Machine learning frameworks: scikit-learn, Pytorch.
- Scientific knowledge: signal processing, image analysis, machine learning and deep learning.
- Knowledge in multimedia security will be considered a plus.
- Languages: French or English.

Place and allowance of internship

This research internship is part of the FuzzyDoc exploratory project funded by the GdR ISIS (Information, Signal, Image et ViSion). It will be supervised by Iuliia Tkachenko (Université Lyon 2, LIRIS, Lyon) and Pauline Puteaux (CNRS, CRIStAL, Lille).

Funding covers 5-6 months of internship starting from February 2024. The intern will be based at LIRIS (Laboratoire d'Informatique en Image et Systèmes d'information) on the campus of Lyon 2 University in Bron. Internship allowance is about 600 euros/month (3.90 euros per hour).

References

[1] P. Puteaux, I. Tkachenko, "Crossing number features: from biometrics to printed character matching", IWCDF@ICDAR 2021, September 2021, Lausanne, Switzerland.

[2] F. Yriarte, P. Puteaux, I. Tkachenko, "A Two-Step Method for Ensuring Printed Document Integrity using Crossing Number Distances", IEEE WIFS 2022, December 2022, online.

[3] K. Nandakumar, A. K. Jain, and S. Pankanti. "Fingerprint-based fuzzy vault : Implementation and performance." IEEE Transactions on Information Forensics and Security, 2(4) :744–757, 2007.

[4] C. Rathgeb, J. Merkle, J. Scholz, B. Tams, V. Nesterowicz, "Deep face fuzzy vault: Implementation and performance", Computers & Security, Volume 113, 2022.

[5] Python implementation of fuzzy extractor https://pypi.org/project/fuzzy-extractor/

Contact information

E-mail: iuliia.tkachenko@liris.cnrs.fr and pauline.puteaux@cnrs.fr

Please provide your CV, the motivation letter and the transcripts with your marks for the two years of Master's degree / the last two years of engineering school.