Human-centric Artificial Intelligence enabled Digital Images and Videos Forensic Triage

Shancang Li, Yifan Liu
School of Computer Science and Informatics
Cardiff University
Cardiff, UK.

***Abstract*—Digital forensics and incident response (DFIR) involve huge volume of data collected from digital systems that requires investigators to quickly sift through and prioritise relevant evidence. The forensics investigator team faces many challenges when analysing the processes to keep them on target and improve them. A lack of a systematic approach to data analysis can lead to slower decision-making. This work aims to enhance the effectiveness and efficiency in forensics analysis using human-centric artificial intelligence (HAI) enabled data triage. Specifically, an image and video triage was proposed that can significantly speed up the investigation and identify highly related evidence items from huge volume of images and videos.**

***Index Terms*—Digital forensics, Artificial intelligence, Image recognition and classification, forensics triage**

I. Introduction

The emerging techniques, such as Internet of Things (IoT) and Artificial Intelligence (AI), 5G, smart technologies, are generating huge volume of data. It could be text, image, video, or a mix of all. In digital forensics, the text, images, and video make up for a significant share of evidence sources. Image analysis or forensics image analysis in digital forensics mainly focus on image contents and authenticity analysis, which usually is performed in forensics lab.

Specifically, technology evolution makes social networks platforms and media sites, such as *TikTok, X (the former Twitter), Linkedin, etc.*, contains personal and private information or sensitive information related to specific cases. The social networks have been the valuable source evidence items in digital forensics, in which huge volume of images, videos, profiles, and user information can be extracted. The AI algorithms shows great potential in data classification, categorisation, and relevance analysis in an investigation. AI based forensics tools have been used to analyse both contents and metadata of a file to prioritise files for closer crutiny. Some deep learning algorithms can help identify unseen relevant connections between evidence items. Using AI techniques, forensics investigators can quickly identify and focus on more relevant evidence, leading to faster and more effective examination.

The AI algorithms shows great potential in helping digital forensics investigators efficiently analysis the vast amount of unstructured images and video data we obtain from cameras and smart devices. Using pre-trained AI models, we can analyse images with results that for specific tasks already surpass human-level accuracy. Specifically, the machine learning models allows digital forensics investigation close to the source of data and make it possible to overcome the limitation of privacy, real-time performance, efficacy, robustness, and more.

Human-centric machine learning [1], which brings the human behavioural analysis into the image and video triage of. It is extremely complex and evidence-based methods render findings in this field unreliable, which needs to understand the images or video before conduct triage. On the other hand, it is critical bring human's insights in digital forensics.

As the new technologies, such as cloud computing, the Internet of things (IoT), artificial intelligence (AI), 5G, etc., are growing more sophisticated, so must the field of modern forensics. Modern forensics techniques mainly focus on three areas: (1) data analysis; (2) network forensics, and (3) reverse engineering, which involves inspecting malware samples, traces, network traffic, and log files. The key challenges that the modern forensics face includes:

1) Move devices involved. Smart devices, including IoT sensors, mobile devices, tablets, etc., contain many different type of information that makes it extremely challenges to reconstruct physical events and evidence-driven court-proof framework;
2) Exponential growth in the volume of images. Forensic examiners may have huage volume of images and videos in almost every investigation.
3) In digital forensic investigation, it is very important to categorize and prioritise relevant content from the seized devices.

To address above challenges, this work focus on a data triage in digital forensics framework and approaches, the main contribution of this work are summarised as:

(1) This work focus on HXAI enabled digital forensic triage, addressing key challenges that digital forensic investigators facing.

(2) A HXAI enabled image and video classification approach was proposed based on the features defined by the human expertise, allowing practitioners to rapidly search while prioritising specific relevant features;

(3) A case study is presented to illustrate the proposed solution can empower investigators to meet the challenges outlied above.

Overall, AI in digital forensics offers the potential to expedite investigations, improve accuracy, uncover hidden patterns, and enhance decision-making processes, ultimately assisting law enforcement agencies in combating cyber crime and ensuring a fair judicial system.

II. Related Works

In digital forensics, the AI technologies have been used in data triage and prioritisation. Helping prioritising data sources and files for digital forensics analysis based on potential relevance, importance, or unusual activities. Pirrung *et al.* developed an interactive AI based image triage tool, Sharkzor [2], can perform image triage, organiation, and automate.

In healthcare industry, the images and video triage is used for ambulance service patients and cliniciants [1]. Using video consulation AcuRX, clinicians determine acceptability and impact on the assessment. The method shows great potential in enhancing data collection.

Aiming at reduce redundant pictures, Chang *et al.* developed automatic photo triage using a relative quality measure, which can augment photos taken from the same scene [3]. The machine learning shows great potential in images and video classification, the pre-trained machine learning classification model can receive video frames as input and outputs the probability of each class being represented in the video [1].

Kaur and Jindal reviewed the image and video forensics analysis methods and approaches, specifically focus on multimedia materials extracted from social media platforms such as Whatsapp, Facebook, X, etc [4]. Ferreira *et al.* proposed a SVM based manipulation detection for photos and videos in cyber crime investigation in [5], which processes DFT features of a large dataset containing manipulated photos. Jafar *et al.* investigated photos taken by mobile phone and developed an deep learning AI platforms (Deepfake) that can detect fake photos and videos using isolating, analysing, and verifying lip/mouth movement [6]. Mittal *et al.* developed AI algorithms to distinguish manipulated content in images and videos [7].

The modern digital image analysis is more complicated than it used to be. Existing forensics tools, such as *ExifExtractor, CameraForensics, etc.*, are used to analyse images for data, focusing on: (1) highlighting key intelligence; (2) displaying areas with an identifier; (3) identifying areas with modifications. Recent advances in image classification focus on image classification using Convolutional Neural Networks (CNNs), in which pre-trained models and data augmentation techniques can be used to enhance image classification accuracy. In video classification spatiotemporal models (e.g., two-stream CNNs) and Long Short-Term Memory (LSTM) are used for sequential video analysis. It is very challenging to handle large-scale datasets and data imbalance and provide robustness images and videos classification. Specifically, it is very difficult to deal with real-time video classification and low-latency applications.

[1]https://www.tensorflow.org/lite/examples/video_classification/overview

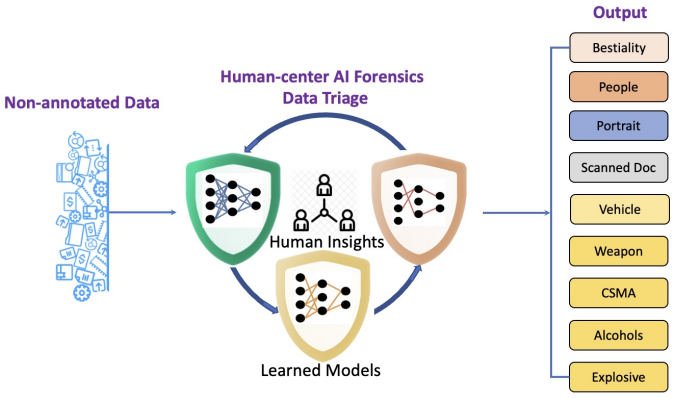

Fig. 1. AI assisted human-centric data triage

III. Human-centric Digital Forensics Triage

Using digital forensic triage can significantly reduce the extracting and analysing process time, which can also address the imbalance between growing demand for analysis and availability of resources, ensuring that specialist forensic capability focus on the most impact investigation.

Human-centric AI Data Triage (HDT) is an efficient digital forensics analysis approach prioritising the human factor in the process of selecting, analysing, and interpreting digital evidence during an investigation. HDT acknowledge the importance of human expertise of investigators alongside technological forensics tools. The HDT integrates human-centric approach with advanced analysis technology to address complex cases and ensuring justice and accuracy in digital forensics investigation.

Fig. 1 presents the key components in a HDT, in which the forensic analysts and investigators play a central role in decision-making (human insights). Machine learning algorithms (Learned models) can identify subtle nuances and patterns that automated tools (e.g., Autopsy, Encase, etc.) may miss. The HDT is able to enhance the quality of evidence and reducing false positives/negatives.

As shown in Fig. 1, the HDT includes following components:

- Human insights driven decision making, which uses the experienced forensics analysts to determine data relevance;
- Combining human insights with automated tools, strategies for leveraging both human judgement and technologies;
- Documenting decision-making processes, providing transparency and well-documented procedures.
- AI enabled human-centric approach
- Collaboration between forensics experts, legal professionals, and law enforcement agencies.

A. Human-Centered AI in Digital Forensics

It is crucial in digital forensics to integrate human-centric AI, forensic experts, legal professionals, and law enforcement

to build an human-centric AI investigation approaches [8]. Human-Centered AI considers human insights in digital forensic investigation. Investigators need to think what data can do, what data science can do, and what AI can do. AI has shows great potential in digital forensics, specifically, in automated log analysis, malware detection, image and video analysis, natural language processing, network traffic analysis, and forensic triage [2]. The human-centric AI delivers intelligence to support and enhance forensic analysts' capabilities in digital forensics investigation tasks ranging from analyse log files to video recognition, to ensuring compliance with legal or policy requirements.

B. Expert Companion in Forensics Investigation

Digital forensics analysis is a specialized field that requires a combination of technical, investigative, and legal expertise. Professionals in this field should possess a range of skills and knowledge to effectively collect, preserve, analyze, and present digital evidence. Unlike general AI reasoning problems, digital forensics investigation significantly relies on expertise of examiners. For example, examiners usually need to be trained and gain practical experiences in specific area, e.g., computer forensics, mobile forensics, etc. Then, within the framework of HDT, digital forensics knowledge that associated to specific cases needs to be introduced.

$$I \leftarrow (\mathcal{E}, \mathcal{A}, \mathcal{R}) \tag{1}$$

in which I denotes expert insights, $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ denotes evidence items, $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$ denotes attributes of evidence items, and $\mathcal{R} = \{r_1, r_2, \ldots, r_k\}$ denotes relationships between evidence items and attributes.

IV. Human-centric AI Digital Forensics Triage for Image and Video analysis

In digital forensics analysis, hundreds of thousands of images on digital devices can be available for examination. Existing techniques, such as hash sets and file metadata (*size, name, created time, accessed time, changed time, etc.*), can be used to conduct analysis. However, it is insufficient for sheer volume of data examiners. Pre-trained AI models can be used to discover relevant evidence quickly rather than waiting for a full examination over seized devices. This section focuses on an images and videos analysis using HAI empowered digital forensic triage.

A. Digital Forensics Triage for Images and Videos

Images and videos classification are time-consuming and the existing forensic tools, such as Encase, Autopsy, etc., are often used to quickly qualify exhibits on scene or in the lab. It is important for data examiners to quickly identify files contain relevant evidence that can then be prioritised by investigators. The pre-trained AI models are able to categorise images and videos based on specific features, like *drugs, guns, knives, pornography, vehicles, etc.*. When combining image categorisation with previous prioritisation techniques, investigators can more quickly triage a device, this will allow investigators to look at the best candidates that may be relevant to their case.

HAI empowered images and videos classifier, classifying all images and videos from file system for further categorisation, analysis, and investigation process. HAI can help investigator to make decision using AI recommendation that can detect abnormally and misinformation through large volumes of images and videos. The undergoing AI use cases (*e.g., LLM, ChatGPT, etc.*) can be very useful in data triage of digital forensics investigation. The HAI enabled data triage will play a key role in modern digital forensics and incident response, including investigate malicious activity, reverse engineering malware, obtain threat intelligence, and assist with incident recovery.

HAI enabled digital forensics triage can rapidly scan through vast amounts of digital contents, pinpointing individuals of interest and significantly reducing the manual effort required. The HAI enabled digital forensics expedites the identification process, enabling investigators to focus their efforts on the most relevant leads and accelerate the progress of the investigation. Key features could be used by the HAI to conduct examination:

1) *Relevant Features.* Deep learning models (*e.g., CNN, RNN, etc.*) could be used to extract higher-order features (*e.g., faces, objects, text, etc.*) that could be used in classification, manipulation, and fake detection;
2) *Prioritising Contents.* Using HAI to well understand and identify contents of images and videos, such as *bestiality child abuse, pornography, portraits, people, scanned documents, vehicles, weapons, etc.*;
3) *Image or Video Fingerprint*. Image or video fingerprints are generated using various algorithms, typical fingerprint is a sequence of number that represents the content of the image. It presents the compactness, uniqueness and local features of an image.
4) *Prioritising devices.* Diverse sources from which images and videos can be obtained [9], including {*Storage devices, Mobile devices/tablets, Digital cameras, Social media and online platforms, Email, Surveillance cameras, External media (CDs, DVDs), IoT devices, In-car systems, etc.*}
5) *Case based features. E.g., criminal cases, civil cases, incident response, compliance and regulatory cases, etc.*

Images or video triage in digital forensics is the task of categorising and assigning labels to group of features or rules. Unlike traditional image classification, in this work we bring human insights and previously classified reference samples to train classifier for new, unknown data (images and video). As shown in Fig. 2, the HDT uses deep learning algorithms to extract features of input data (an image in this example), and then use LSTM encoder and decoder to produce captions for the input data and predict the features. In this example, it

[2]https://www.darkreading.com/dr-tech/6-ways-ai-can-revolutionize-digital-forensics

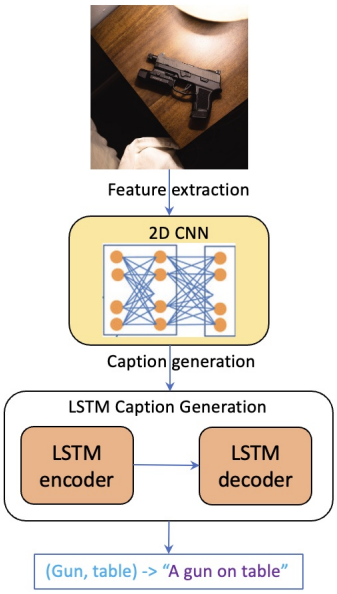

Fig. 2. Caption generation for a specific image or video frame

identified a 'gun' and a 'table', and the create a relationship between 'gun' and 'table' as "A gun on table".

B. Deep learning based photos and video classification

It is an essential task to analyse vast collections of multimedia content (photos and videos) in digital forensics. The deep learning has emerged as a powerful tool for automating the classification of photos and videos. The CNNs have been widely used in photos classification due to their capability to learn the spatial hierarchies of features (edges, textures, shapes, etc., which are critical for object recognition in images) automatically.

1) Forensic Features Recognition: Images and videos classification with CNN-RNN architecture includes following procedures:

Step 1: Take images or extract frames from given video;
Step 2: Use feature extractors (*e.g., CNNs, average algorithms, etc.*) to extract features from images or from all the frames;
Step 3: Classify the image or every frame based on these extracted features;

Fig. 3 shows an example to use AI models to extract key objects from given images or videos, in which the HDT model can use pre-defined features, e.g, based on the contents like "car, bus, motocycle, truck", to recognise the objects in the image and video frames.

2) Noise pattern based image analysis: Using Photo-Response Non-Uniformity (PRNU), examiners can identify and analyse source devices and manipulation of images. The key idea includes following steps

Step 1: Estimate reference PRNU using average multiple images or frames of video captured by the same camera, as Eq. (2)

$$\bar{I}_k = F(I_k), W_k = I_k - \bar{I}_k \tag{2}$$

in which F is denoising filter, I_k denotes the k^{th} frame of the video and $\bar{I}_k$ is the denonised frame; W_k is the noise residual for the k_{th} frame.

$$N_{ref} = \frac{1}{N}\sum_{k=1}^{N} W_k \tag{3}$$

in which N_{ref} is the reference PRNU and can be estimated from a set of N frames using Eq. (3).

Step 2: Estimation of Test PRNU, based on the PRNU of video, one can calculate it's PRNU using Eq. (3);
Step 3: Calculation of correlation between reference and $PRNU_{test}$ using Eq. (4)

$$C_r(N_R, N_T) = \frac{(N_T - \bar{N_T}) \cdot (N_R - \bar{N_R})}{\|N_T - \bar{N_T}\| \cdot \|N_R - \bar{N_R}\|} \tag{4}$$

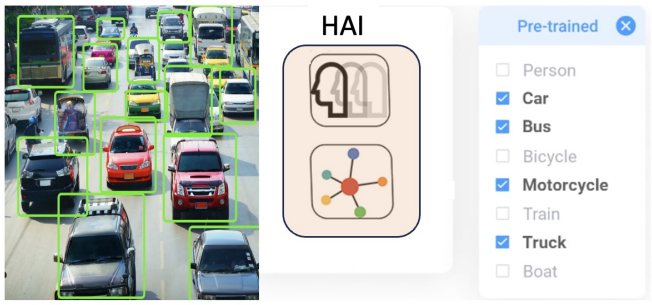

Fig. 3. Features based HAI based Data Triage

Fig. 4 shows an example of PRNU extraction from images using above methods. Specifically, the HDT module can extract PRNU fingerprint of mobile phone (e.g., Realme 7 in this example for Photo 1), and then when input an image or a video clip, it will analyse the PRNU feature of the image or clip and then match with pre-extracted PRNU feature, by doing this, the HDT can classify images or videos based on their source camera and mobile.

3) Classification Algorithm: Deep learning algorithms can be used in image and video classification, such as *CNNs, RNNs, Long short-term memory (LSTM), etc.*, have been developed. The videos can be treated as a collection of frame, using pre-trained deep learning models (such as *AlexNet, GoogleNet, ResNet, VGGNet, etc.*) we can extract frame features. These features can be averaged into video representation as input of classifiers, such as *support vectors machine (SVMs), decision-tree, k-nearest neighbors (KNN), etc.*, we can category images and videos based on the observations.

In digital forensics, it is important to implement image and video labelling or bookmark, which can describing an image (frame) with a label or a natural sentence, where the spatial relationships between objects or object and action are further described (e.g., "a gun on table"). Using a pre-trained LSTM model, appropriate caption can be generated to caption

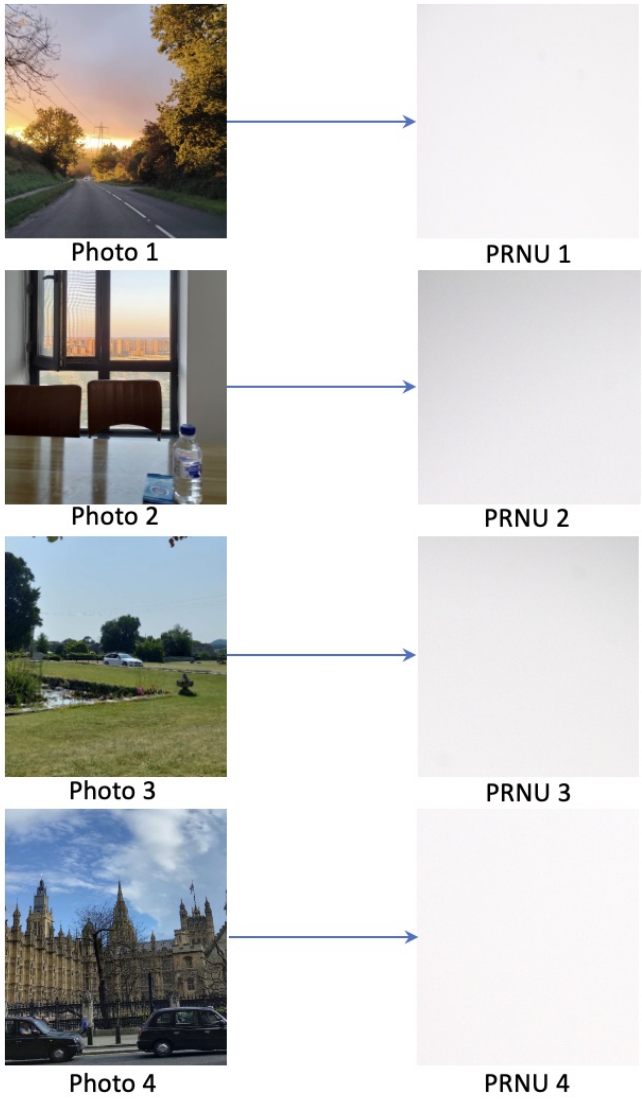

Fig. 4. PRNU features based image triage

the input image or frames, as shown in Fig. 2. In the past few years, a number of video caption methods have been developed, including LRCN [10], Two-Stream [11], LSTM [12], Image-Base [13], etc.

The accuracy can be obtained via

$$Acc = \frac{TP + TN}{Total}, Pre = \frac{TP}{TP + FP} \tag{5}$$

in which TP is true positive, TN is true negatives, $Total$ is total number observation. The recall is $\frac{TP}{TP+FN}$ and

V. Discussion and Conclusion

AI empowered digital triage can significantly speed up the critical evidence at early identification, saving time in extracting and analysing processes over large storage capacity. Specifically when multiple devices involved, digital forensics triage can quickly identify devices that contain relevant evidences.

This paper focus on machine learning based images and videos forensics classification and analysis by proposing a data triage, which provides explainable machine learning for images and video classification and can significantly enhance the analysis efficiency and accuracy.

References

[1] F. Bell, R. Pilbery, R. Connell, D. Fletcher, T. Leatherland, L. Cottrell, and P. Webster, "The acceptability and safety of video triage for ambulance service patients and clinicians during the covid-19 pandemic," *British paramedic journal*, vol. 6, no. 2, pp. 49–58, 2021.

[2] M. Pirrung, N. Hilliard, A. Yankov, N. O'Brien, P. Weidert, C. D. Corley, and N. O. Hodas, "Sharkzor: Interactive deep learning for image triage, sort and summary," *arXiv preprint arXiv:1802.05316*, 2018.

[3] H. Chang, F. Yu, J. Wang, D. Ashley, and A. Finkelstein, "Automatic triage for a photo series," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–10, 2016.

[4] H. Kaur and N. Jindal, "Image and video forensics: A critical survey," *Wireless Personal Communications*, vol. 112, pp. 1281–1302, 2020.

[5] S. Ferreira, M. Antunes, and M. E. Correia, "Forensic analysis of tampered digital photos," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 25th Iberoamerican Congress, CIARP 2021, Porto, Portugal, May 10–13, 2021, Revised Selected Papers 25*. Springer, 2021, pp. 461–470.

[6] M. T. Jafar, M. Ababneh, M. Al-Zoube, and A. Elhassan, "Forensics and analysis of deepfake videos," in *2020 11th International Conference on Information and Communication Systems (ICICS)*, 2020, pp. 053–058.

[7] T. Mittal, R. Sinha, V. Swaminathan, J. Collomosse, and D. Manocha, "Video manipulations beyond faces: A dataset with human-machine analysis," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 2023, pp. 643–652.

[8] E. Dietz, A. Kakas, and L. Michael, "Argumentation: A calculus for human-centric ai," *Frontiers in Artificial Intelligence*, vol. 5, p. 955579, 2022.

[9] Y. Akbari, S. Al-maadeed, O. Elharrouss, F. Khelifi, A. Lawgaly, and A. Bouridane, "Digital forensic analysis for source video identification: A survey," *Forensic Science International: Digital Investigation*, vol. 41, p. 301390, 2022.

[10] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.

[11] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *Advances in neural information processing systems*, vol. 27, 2014.

[12] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4694–4702.

[13] S. Zha, F. Luisier, W. Andrews, N. Srivastava, and R. Salakhutdinov, "Exploiting image-trained cnn architectures for unconstrained video classification," *arXiv preprint arXiv:1503.04144*, 2015.