International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2410117

Preserving Privacy in Multimedia : Text-Aware Sensitive Information Masking for Visual Data

Ardon Kotey, Tejan Gupta, Shivendra Bharuka, Abhishek Singh, Nikhil Ghugare, Lalith Samanthapuri

Department of Information Technology, Dwarkadas J. Sanghvi College of Engineering, Mumbai, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 01 Feb 2024

Published: 12 Feb 2024

Publication Issue

Volume 10, Issue 1

January-February-2024

Page Number

166-174

ABSTRACT

The unauthorised revelation of confidential data has become a source of concern due to the proliferation of multimedia content on the Internet. Unintentionally captured and exposed textual data, including but not limited to personally identifiable information (PII), financial details, and confidential documents, may be present in images and videos when they are being recorded or shared. This study presents an innovative method for concealing text-sensitive information in visual data with the intention of safeguarding privacy without compromising the context and integrity of multimedia content. By utilising cutting-edge text detection algorithms, our approach effectively discerns textual areas present in images and videos. Natural language processing (NLP) methods are subsequently utilised to categories the identified text into sensitive or non-sensitive categories according to predetermined standards. In the case of confidential information, we employ a context-aware concealing strategy that obscures only the pertinent segments while maintaining the visual indicators and encircling context. This approach diverges substantially from conventional pixel-level masking, which frequently obliterates crucial data and impedes interpretability. In pursuit of context-aware masking, we investigate a range of methodologies including semantic-based keyword masking, character-level redaction, and text-region-specific image inpainting. The efficacy of our methodology is assessed across a range of datasets comprising videos and images that contain text types and sensitivity levels that vary. The accuracy of text detection and classification, the impact on user comprehension, and the effectiveness of concealing in preserving privacy and visual quality are all evaluated. The findings of this study carry substantial ramifications for safeguarding privacy across a range of domains, encompassing social networking services, online media sharing platforms, and video surveillance systems. Our approach provides a valuable tool for protecting personal information and upholding privacy rights in the digital age by facilitating the selective masking of sensitive text while preserving the visual fidelity and context of multimedia content.

Keywords: Text Detection, NLP, ML, Masking, Image Detection, NER, Blurring

I. INTRODUCTION

The increased volume of online engagement that occurred amidst the COVID-19 pandemic served to underscore the critical nature of safeguarding sensitive data that is embedded within multimedia materials. As a greater proportion of daily activities—including education, employment, and even entertainment—migrated online, video blogging platforms (vlogs) experienced phenomenal growth. Regrettably, a distressing incident transpired during this time period as well: unintentional capture and exposure of personal financial information, including credit card details, during a live vlog. This severe illustration emphasised the susceptibility of textual information contained in visual media and the necessity for comprehensive measures to protect privacy in the era of digitalization. The proliferation of digital cameras and smartphones has precipitated a seismic increase in the acquisition and dissemination of visual data. Online interactions are replete with images and videos, which not only document scenes and events but also contain textual information. Although textual data, including captions, signs, and documents, enhances the multimedia experience by providing context, it also presents a substantial privacy risk. Inadvertent disclosure of sensitive information can occur through the capture or sharing of textual content. This includes financial details such as credit card numbers or bank statements, personally identifiable information (PII) including names, addresses, and phone numbers, and confidential documents containing private agreements or personal records.

Conventional methods for safeguarding privacy in multimedia frequently involve the implementation of pixel-level masking or obscuring techniques on specifically targeted sensitive areas. Nevertheless, these approaches frequently neglect granularity and context recognition, resulting in the obscurement of critical data encompassing the sensitive text. This may substantially impede the utilisation and interpretability of the concealed material, resulting in it becoming unintelligible or devoid of information. This research proposes a novel method for text-aware sensitive information masking in visual data in order to overcome these limitations. Our objective is to design and implement a system capable of precisely identifying and categorising textual content present in videos and images. Subsequently, we intend to utilise context-aware masking methods to selectively obfuscate only the sensitive segments, while maintaining the visual signals and encompassing context. This methodology holds significant potential for achieving a vital equilibrium between safeguarding privacy and preserving the integrity of the content, guaranteeing the effective concealment of sensitive data while preserving the comprehensibility and utility of the multimedia material. In order to accomplish this, our research capitalises on developments in the fields of computer vision and natural language processing (NLP). We intend to develop a dependable and precise system for discerning sensitive information from textual data embedded in visual media by integrating text detection algorithms with NLP-based text classification techniques. This functionality allows for the targeted implementation of context-aware masking methods that are customised to suit the particular nature and degree of sensitivity of the identified text. An assortment of masking techniques are investigated, such as semantic-based keyword masking to selectively conceal particular categories of sensitive data, character-level redaction to obscure individual characters within sensitive words, and image inpainting methods that are specifically engineered to seamlessly fill in text regions while preserving the visual context.

The prospective impact of the research findings on the management of sensitive information in multimedia content is revolutionary. Our proposed methodology may find application in a wide range of contexts and settings, including video surveillance systems that selectively obscure licence plates or confidential documents captured by cameras, and social media platforms that automatically mask personally identifiable information (PII) in shared images. This is achieved by enabling the selective masking of sensitive text while maintaining visual integrity and context. The overarching objective of this research is to make a scholarly contribution that will enable us to experience the advantages of visual communication while safeguarding the privacy of

individuals and guaranteeing the ethical management of sensitive data in the era of digitalization.

II. LITERATURE REVIEW

A. Literature Review

Our literature review critically examines the nascent domain of sensitive text masking and detection in multimedia content. This paper presents an examination of the diverse methodologies developed to detect and obfuscate sensitive data across multiple media. It underscores the growing importance of these types of systems. Our evaluation comprises a discerning examination of the methodologies, their efficacy, and the intrinsic limitations they encounter. We engage in introspection regarding the development of this field, which served as the basis for our initial contributions and illuminated possible avenues for subsequent investigation.

The issue of text redaction in images while maintaining visual integrity and context is addressed in the paper [1]. It presents an innovative methodology that surpasses pixel-level masking through the application of semantic analysis. To begin, the system employs pre-existing text detection algorithms in order to discern textual areas present in the image. The detected text is then analysed by a natural language processing (NLP) module according to its semantic meaning and adjacent context. In the case of confidential documents or personally identifiable information, the system employs selective redaction on the text. Reducing the process of obscuring or blacking out the entire text region, this method selectively redacts pertinent keywords or entities while maintaining contextual cues in the surrounding area. By employing this technique, observers are able to preserve comprehension of the image's overarching message while safeguarding confidential data. The study assesses the efficacy of their methodology on a range of image datasets comprising various types of text. It reveals substantial enhancements in visual clarity and privacy protection when compared to conventional masking techniques.

The objective of this paper [2] is to redact and automatically identify sensitive regions in images according to their visual saliency. Prior to anything else, the system utilises saliency detection algorithms to determine which regions are most likely to capture human interest. These regions frequently encompass critical visual indicators or potentially confidential data. The detected saliency map is subsequently analysed by the system in order to identify regions that contain text. Following this, an inpainting technique based on deep learning is implemented to restore the obscured text regions effectively and smoothly, thereby guaranteeing a seamless merge between the redacted areas and the adjacent image content. By employing this methodology, the visual appeal of masked images is substantially enhanced in contrast to conventional pixel-level blurring, which frequently results in discernible artefacts. The efficacy of their approach is showcased by the authors across a range of image datasets, and its potential utility in safeguarding the privacy of online platforms and media sharing services is emphasised.

The paper [4] investigates the obstacle posed by safeguarding sensitive data that is exchanged on digital social networks. A deep learning-based framework is suggested by the authors, which enables the automatic detection and redaction of sensitive content in images that are submitted to social media platforms. The system utilises a convolutional neural network (CNN) structure that has been trained on an extensive collection of images that have been annotated with a diverse range of sensitive data. Sensitive text, including financial information, personal names, and addresses, can be accurately detected in uploaded images by CNN. After detecting the confidential text, the system employs a range of redaction techniques to remove it, such as context-aware blurring, pixel-level masking, and character-level redaction. Furthermore, the framework provides the capability to tailor the degree of privacy safeguarding in accordance with platform policies and user inclinations. The findings of this study have substantial ramifications for the improvement of privacy and security in virtual social media communities.

This extensive survey paper [3] offers a significant synopsis of the current state of research concerning the safeguarding of privacy for multimedia content that contains text. This study examines and contrasts a range of methodologies implemented to safeguard privacy across different settings, encompassing scanned documents, videos, and images. The survey classifies these methodologies into distinct categories, including

redaction, text filtering, scrambling, and anonymization. Furthermore, it delves into sophisticated methodologies that exploit deep learning and natural language processing (NLP) to safeguard privacy in a context-aware manner. Through an in-depth analysis of the merits and drawbacks of each methodology, the article provides researchers and developers in this field with insightful advice and direction. This resource is of the utmost importance in comprehending the present condition of text-aware privacy protection and discerning possible directions for further investigation.

Although not specifically centred on text, this study [7] investigates a connected technology that is pertinent to safeguarding privacy in multimedia. This scholarly article explores the complex issue of reconciling the extraction of valuable information from facial expressions documented in images or videos with the protection of user privacy. The authors put forth an innovative methodology that acquires privacy-conscious representations of facial characteristics. This functionality enables systems to discern and classify emotions while safeguarding the identities of the subjects depicted in the images or videos. By utilising deep learning methodologies, this approach effectively extracts feature information from facial expressions while concurrently removing features that may contain personally identifiable data. These privacy-aware features are subsequently employed by the system to accurately identify emotions while safeguarding user confidentiality. This study presents promising prospects for the implementation of facial expression analysis across diverse domains, all the while safeguarding the confidentiality of individuals.

Our literature review provides a comprehensive examination of the progress and challenges associated with the identification and concealment of sensitive text in multimedia. The thorough examination we have conducted uncovers notable advancements in the domain, in addition to ongoing obstacles including the accuracy of detection algorithms, the efficacy of masking methods, and the management of heterogeneous data types. Despite these obstacles, the surveyed works provide a solid foundation for subsequent innovations. These findings emphasise the need for innovative methods that possess greater discernment, adaptability across various media formats, and are more suitable for the intricate nature of real-world scenarios. In summary, our review provides a comprehensive synthesis of existing knowledge and proposes a trajectory for future research that incorporates domain-general capabilities, context-awareness, abstractive summarization methods, and context-awareness. This will enable the development of more sophisticated privacy protection measures in the digital age.

B. Methodology

1. Dataset

The methodology section outlines the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, Data and Sources of Data, study's variables, and analytical framework. The details are as follows.

ICDAR 2013 Document Image De-identification (DID) Dataset: A collection of document anonymization technologies that have been rigorously curated for benchmarking purposes. The collection consists of a wide range of digitised documents, encompassing both typed and handwritten formats, which conceal a variety of confidential data, including personal addresses and social security numbers. The dataset, which contains more than two hundred images, simulates real-world privacy protection scenarios in paper-based documents in a challenging environment. The dataset's diversity in document layouts, typefaces, and text sizes provides an optimal training environment for algorithms designed to accurately identify and redact confidential information. This ensures that the remaining document content can be utilised for legitimate objectives. The dataset in question is an indispensable resource for researchers leading the charge in the development of sophisticated privacy-preserving methods for document processing. In document images, the ICDAR 2013 Document Image De-identification (DID) Dataset functions as an all-encompassing repository that facilitates the evaluation and advancement of text redaction systems. An assortment of document categories is incorporated in order to replicate the intricacy of document anonymization endeavours in practical situations. In addition to its extensive assortment of textual presentations, this dataset is additionally annotated with ground truth annotations that specify the locations of confidential data. Supervised learning approaches rely heavily on these annotations, as they enable the accurate

instruction of models regarding the specific locations and methods for redacting text. Furthermore, apart from serving as a fundamental training resource, the DID dataset facilitates a comparative evaluation of various anonymization algorithms by providing a benchmark. Its contributions have been crucial in advancing the study of automated techniques for identifying and redacting sensitive text in document images, thereby creating a setting conducive to the development, testing, and refinement of novel solutions. The ongoing engagement with the dataset guarantees that the algorithms it contributes to the development of are resilient and flexible in response to the diverse characteristics of authentic documents.

COCO-Text: Compiling a vast assortment of images, the COCO-Text dataset aims to improve the performance of text detection and recognition systems when applied to natural settings. Unique to this dataset is its concentration on incidental text discovered in ordinary images, as opposed to text that constitutes the main subject matter of the image. Such incidental text can be found on billboards, signage, product labels, and street markings, among many other places. This variant presents a challenge for models attempting to identify and interpret text that is distorted, partially obscured, or incorporated into intricate backgrounds. Moreover, the dataset is extensive and diverse, consisting of more than 63,000 images accompanied by over 173,000 instances of text. Every individual text instance in COCO-Text is annotated with its location in the image and labelled with its transcription, thereby furnishing an extensive repository for supervised learning methodologies. This functionality empowers the development of models capable of comprehending the substance of text in diverse contextual environments, in addition to detecting its existence. Due to the wide variety of images contained within COCO-Text, the dataset plays a crucial role in the advancement of text detection systems that can withstand a multiplicity of real-life scenarios. The images comprise a wide range of situations, illumination conditions, and text orientations, which requires the implementation of advanced algorithms capable of delivering dependable results even in less-than-ideal conditions. In disciplines such as autonomous navigation, where the ability to interpret environmental text can be critical, this is of the utmost importance. Finally, COCO-Text finds application in diverse fields, including augmented reality (where the integration of live text and digital information in the physical environment is critical) and the advancement of assistive technologies for those with visual impairments. By virtue of its comprehensive representation of real-life scenarios, models trained on COCO-Text are applicable to a vast multitude of practical contexts. As a result, it has become an indispensable dataset for computer vision researchers and practitioners alike.

Street View Text (SVT): Specific imagery from the Street View Text (SVT) dataset has been compiled for the purpose of training and assessing text detection algorithms in urban environments. It contains a diverse collection of Google Street View street-level images that have been meticulously annotated with bounding boxes encircling visible text. The annotations play a pivotal role in the advancement of text recognition systems capable of functioning in the intricate visual environments commonly found in urban settings. The variety of images in SVT's collection exemplifies the multifaceted character of text in urban landscapes, encompassing street names, billboards, shop signs, and other forms of public signage. The presence of such diversity is crucial for the development of algorithms that possess the ability to adapt to the unpredictable nature of text appearance in real-world situations while maintaining accuracy in text detection. The SVT dataset, comprising thousands of images and more than 33 distinct sequences, presents models with an array of challenges, such as text that is distorted by perspective, varying degrees of occlusion, and fluctuating illumination conditions. This environment provides an optimal setting for refining the complexity of models to guarantee their dependability in various urban settings, ranging from the bustling visual activity of city centres to the more structured suburbs. Additionally, the development of context-aware concealing technologies is facilitated by the dataset. These technologies are indispensable for applications that necessitate the careful management of text. They guarantee that confidential details can be obscured while preserving the scene's contextual essence, which is frequently crucial for informational and navigational functions.

2. Our Approach

Our research endeavours to revolutionise the way privacy is protected in the digital domain by automating the identification and rectification of confidential data

contained in images. Our initiative is founded upon a complex, layered methodology that combines advanced image manipulation techniques, named entity recognition (NER), and cutting-edge image detection. At the outset, our system extracts text from images by employing cutting-edge detection algorithms. Following that, the text is analysed by the NER component to identify and extract sensitive information, including confidential data and personal identifiers. In conclusion, sophisticated image manipulation methods are implemented to obfuscate the detected sensitive information, thus safeguarding the confidentiality and integrity of the involved parties and organisations. Implementing this all-encompassing approach guarantees that our mechanised system can consistently obfuscate confidential data from being compromised, notwithstanding the intricacy of diverse visual material.

3. Data Pre-processing

During the data pre-processing stage, we diligently organise the dataset to establish ideal training conditions for our model. This is a fundamental aspect of our endeavour to protect privacy in digital imagery, which is the goal of our project. The pre-processing phase comprises several rigorous procedures:

Image Rescaling: A uniform scale is applied to the resolution of every image. By doing so, computational requirements are not only reduced but also consistency is maintained throughout the dataset, a critical factor for ensuring the accuracy of feature extraction and subsequent processing phases. To improve the legibility of text, noise reduction methods are implemented, including the use of Gaussian blur filters. The implementation of these methods is of the utmost importance in reducing the impact of visual distortions and artefacts that may impede the precision of text detection algorithms.

Data Augmentation: To enhance the resilience of our model, we incorporate fluctuations into the training data that replicate irregularities encountered in the real world. By employing techniques such as random cropping, rotation, and skewing, the training set is enhanced in scope and the model is better equipped to deal with the diverse transformations that text might experience in practical situations.

Normalisation and Contrast Adjustment: To enhance the visibility of text against backgrounds and optimise the performance of our text detection algorithms, we perform pixel value normalisation and contrast adjustment on the images. Under annotation and labelling, bounding boxes are utilised to annotate and label each text segment within the images in preparation for the NER module. This process establishes the foundation for accurate entity recognition and subsequent data masking.

By meticulously carrying out these pre-processing procedures, we strengthen the groundwork of our undertaking. The establishment of this fundamental basis is crucial for the subsequent phases of sensitive data detection and redaction, which utilise an intricate combination of image manipulation, NER, and image detection to ensure dependable privacy safeguarding in multimedia streams.

4. Text Detection

In addition to identifying visible text, our system's text detection capability must also comprehend the complexities of text presentation in a variety of environments. Trained on an extensive dataset, our sophisticated deep learning detector is capable of processing text that is obscured by shadows, distorted by perspective, or superimposed on dynamic backgrounds. The ability to adapt is of the utmost importance in practical scenarios where text is presented in less-than-ideal conditions. Furthermore, our system incorporates contextual analysis, which allows it to differentiate between essential text within the image and incidental text, thereby augmenting the pertinence of its detections. By means of iterative testing and feedback, we enhance the accuracy of our bounding boxes, guaranteeing that every identified text area is faithfully represented in subsequent phases of processing. The systematic methodology employed in text detection serves as the fundamental basis for our system's resilient functionality in safeguarding confidential data across a multitude of multimedia file types.

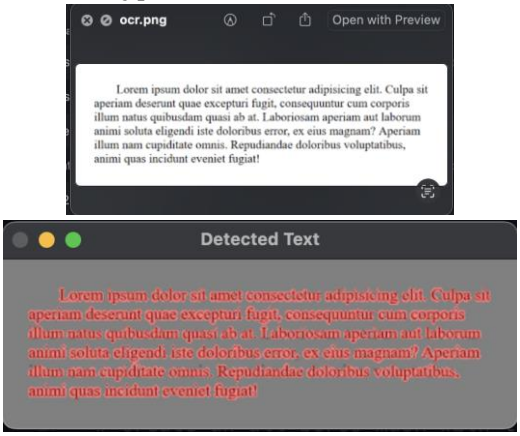

Fig 1: Text Detection in Multimedia (photos, videos)

5. Name Entity Recognition

After detecting the text within the image, our system proceeds to a critical stage in which it discerns any confidential data that may be present. A sophisticated Named Entity Recognition (NER) model is employed to accomplish this. The model undergoes rigorous training using a wide range of annotated data, which includes a variety of sensitive entities. Personal names, residential and business addresses, email addresses, telephone numbers, credit card information, and unique personal identifiers are examples of such entities. The exhaustive training of the model guarantees its ability to accurately identify and categorise a wide range of sensitive data types, including those that are concealed within intricate textual formats. After identifying the text contained within the bounding box of each image, the NER model conducts a thorough analysis of said text. Sophisticated algorithms are utilised to effectively detect and classify every instance of sensitive data, assigning them to predetermined sensitive categories.

In addition to detection, this procedure entails a meticulous categorization of the characteristics of the sensitive information. As an illustration, the system discerns between an email address and a phone number by comprehending the distinct context and format associated with each category of sensitive entity. It is critical for instituting effective data protection and privacy measures, particularly in situations where handling sensitive information requires strict regulatory conformance, that the system can differentiate between these distinctions.

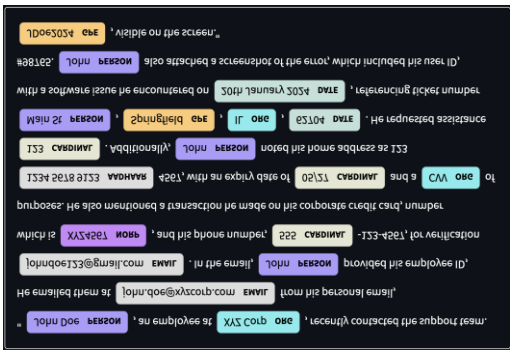

Fig 2- NER to detect sensitive labels in text

6. Masking and Blurring

With the aim of maintaining privacy within our system, we have made enhancements to the method by which sensitive data identified by our NER model is obscured. As a result, we have developed a resilient privacy protection solution that is not only efficient but also visually appealing.

Intelligent Masking: The masking capabilities of our system have undergone an advancement to integrate intelligent algorithms that dynamically select masking patterns and colours in accordance with the surrounding image region. By employing intelligent masking, not only is confidential information concealed, but the modifications are also rendered visually seamless, thereby reducing visual disturbances, and preserving the image's organic appearance and texture.

Context-Sensitive Blurring: An intelligent blur level adjustment mechanism is implemented by the system, which is responsive to the viewer's focus areas within the image and the significance of the text. This method is intended to preserve the integrity of the image, guaranteeing that the principal visual message is communicated while preventing the exposure of the confidential information.

Privacy-enhancing technologies: To further bolster privacy, our methodology incorporates cutting-edge technologies that enhance privacy. This process entails further safeguarding privacy by employing encryption methods to render the text unintelligible and reversible solely by authorised parties, as opposed to mere obscuring or concealing.

Dynamic Obfuscation: A novel technique has been implemented: dynamic obfuscation, which alters the degree of obscuring and concealing in accordance with the sensitivity level of the content. In contrast to lower-risk data, high-risk information is subjected to a more rigorous obfuscation process, which guarantees that the degree of safeguarding is proportional to the potential consequences of data exposure.

User-centric Design: User-centric design continues to be the cornerstone of our approach to safeguarding privacy. We consistently solicit feedback to enhance the equilibrium between visual coherence and data protection. This ensures that our system upholds both the security of sensitive information and the provision of a pleasurable and functional user interface.

Protecting the Future via Innovation: We are investigating state-of-the-art methodologies in the fields of machine learning and image processing in order to maintain a competitive edge. Our system is designed to be adaptable to new categories of sensitive data and emergent threats to privacy by incorporating technological advancements.

The comprehensive nature of this methodology emphasises our dedication to protecting privacy without compromising the functionality or quality of visual content. The adaptability and sophistication of our system establish a novel benchmark within the domain of automated privacy protection.

7. Fine Tuning and Evaluation

The continuous improvement of our model is an essential component of its functionality. This process entails an ongoing cycle of evaluating performance by utilising a specialised validation dataset. By doing so, we can optimise the training process by adjusting the model's hyperparameters, with an emphasis on improving accuracy and decreasing the occurrence of false positives and negatives. In addition, practical feedback is integrated into our processes via user perception surveys. The utilisation of these surveys is crucial for assessing the efficacy of the system as perceived by its users, thereby yielding insights that may not be discernible solely through technical analysis. The significance of this feedback cycle lies in the fact that it guarantees not only the technical integrity of our system, but also its alignment with the requirements and anticipations of users. Maintaining agility and receptiveness to the continuously evolving challenges in automated privacy protection is essential. In summary, our methodology is comprehensive, encompassing rigorous data pre-processing, strong text detection, sophisticated NER functionalities, and efficient privacy safeguarding strategies. Through ongoing system improvement and the incorporation of user feedback, our objective is to establish a leading position in the domain of automated privacy protection. This will empower individuals to exert greater authority over their sensitive data in a society increasingly influenced by digital technologies.

III. RESULTS AND DISCUSSION

A. Technology Used

We have integrated an array of cutting-edge models and technologies into the development of our comprehensive system for the detection and concealing of sensitive text in multimedia to guarantee its high performance and dependability.

1. **Web Application Technology:** An advanced web application has been developed to function as the user interface for our system. For users who require the processing of image and video streams, the cross-platform compatibility and usability of this application are guaranteed by its foundation on robust web technologies.

2. **YOLO v8 Object Detection Software:** Central to our text detection module is the state-of-the-art YOLO v8 (You Only Look Once) algorithm that we have implemented. The specialised objective of identifying text within intricate multimedia content has been tailored to this cutting-edge model, which is renowned for its speed and precision in object detection.

3. To optimise the learning process, we have conducted experiments with transfer learning methodologies with ResNet models. By utilising pre-learned visual features extracted from large datasets, our system has been able to enhance the precision of text detection without requiring substantial computational resources.

4. **Tesseract OCR for Image Detection:** Tesseract OCR has been integrated to convert identified text regions into machine-encoded text. The optical character recognition engine plays a crucial role in the processing of text images derived from our multimedia streams, thereby establishing a dependable basis for subsequent analysis.

5. **Named Entity Recognition and Natural Language Processing (NLP):** Our system employs sophisticated NLP methodologies to discern and categorise confidential text. By analysing the context and semantics of the text extracted by Tesseract OCR with tools such as NLTK and SpaCy, named entities that necessitate masking are identified.

6. **Ongoing Model Training and Optimisation:** Our models undergo ongoing training and optimisation processes to effectively respond to the dynamic nature of text presentation in multimedia. This process entails the consistent updating of datasets, the iterative improvement of model parameters, and the integration of user feedback to maintain the cutting edge of privacy protection technology within our system.

7. **Design with the End-User in Mind and Feedback Loop:** With the end-user in consideration, we have developed a robust and user-friendly system. User input is systematically gathered and integrated into the system

via an iterative feedback cycle to consistently improve its functionality and user experience.

By incorporating these technologies, our objective is to provide a resilient and intuitive system that can effectively safeguard privacy in the growing domain of digital media.

B. Algorithm Flow for Model Training

A multi-step procedure is required to develop an automated text detection and privacy protection system for multimedia, which incorporates technologies such as natural language processing and deep learning. In the beginning, the system is configured by importing a deep learning model that has been specifically trained to detect text (e.g., YOLO v8) in conjunction with an Optical Character Recognition (OCR) engine (Tesseract) and a Named Entity Recognition (NER) model (typically constructed utilising frameworks such as SpaCy or NLTK) that can identify sensitive entities.

The pre-processing of the input image or video frame initiates the procedure. This process entails achieving image resolution standardisation to ensure consistent processing, implementing Gaussian blur to diminish noise, and normalising the image to amplify the contrast between the text and the background. By following these procedures, the image is prepared in an ideal manner for text detection.

After pre-processing, text within the image is detected by the system. The process entails applying the YOLO v8 model to the image to detect prospective text regions. The likelihood that a detected region contains text is subsequently predicted for each region, and areas with low confidence scores are eliminated. A list of bounding frames is generated to encircle the identified text regions.

OCR commences following the identification of text regions. The picture portion associated with each bounding box is extracted and subsequently transformed into digital text via Tesseract OCR processing. The extraction of legible text from the image regions is dependent on this phase.

Subsequently, sensitive information contained within the extracted text is identified. The NER model is applied to the OCR text to identify and classify sensitive entities, such as financial, personal, or confidential. This stage ascertains the precise text segments that necessitate masking to safeguard privacy.

To obscure the sensitive data, the system delineates perimeters enclosing these areas and populates them with a pattern or colour that is neutral in nature. Subsequently, these areas are subjected to a Gaussian haze effect, the intensity of which is modified in accordance with the desired degree of privacy. By taking this action, it is guaranteed that the sensitive data is securely obscured.

In the concluding stage, the image is post-processed using obfuscated text. This may involve implementing supplementary colour correction or image enhancement techniques to preserve the image's visual coherence without compromising its overall interpretability during the obfuscation process.

To ascertain the efficacy and dependability of the system, a mechanism for quality control and feedback is incorporated. A visual inspection is performed on every privacy-protected image to verify that all sensitive information is sufficiently safeguarded. Parameter adjustments are implemented as required, and user feedback is gathered to facilitate ongoing system enhancement. Furthermore, to uphold the effectiveness of the system, regular retraining of the text detection and NER models with fresh data improves precision and enables them to accommodate novel forms of sensitive text. Continuous improvement and updating is essential for the system to maintain its efficacy amidst the constantly changing realm of multimedia privacy protection and text detection.

IV. CONCLUSION

In summary, our study represents a substantial advancement in the domain of safeguarding privacy in multimedia content. A highly developed system has been introduced, which effectively detects and obscures confidential textual data within video and image transmissions. By incorporating advanced deep learning models (e.g., YOLO v8) and transfer learning architectures (ResNet), in conjunction with the accuracy of Tesseract OCR and the sophistication of natural language processing (NLP) tools (NLTK and SpaCy), our system sets a novel standard in the automated identification and redaction of sensitive content. Our methodology not only places a high emphasis on

safeguarding privacy but also acknowledges the pragmatic requirement of preserving the visual and operational authenticity of the initial multimedia material. By utilising a dual-layered obfuscation strategy that incorporates both masking and obscuring techniques, confidential data is effectively protected while the contextual information remains understandable. This equilibrium demonstrates our dedication to honing privacy safeguarding techniques that are simultaneously efficient and user focused. As we contemplate the future, it is critical that our system continues to evolve. By adopting developments in artificial intelligence and machine learning, we expect our system's precision and adaptability to be further enhanced. Our research establishes the foundation for continuous advancements in privacy technology, paving the way for the development of more resilient, flexible, and user-centric systems that effectively address the expanding complexities of safeguarding digital information.

V. FUTURE SCOPE

Our research anticipates that in the future, our privacy protection system will be enhanced to include a wider range of multimedia elements, including 3D content and immersive environments, which are more intricate and varied. Integration of sophisticated AI models, pursuance of real-time processing capabilities, cross-platform compatibility, user-centric customisation options, adherence to evolving privacy standards, and a steadfast dedication to feedback-driven development will comprise this evolution. By undertaking these efforts, we can guarantee that our system remains at the forefront of protecting sensitive data across various multimedia platforms and technologies, thereby making a valuable contribution to the establishment of a more secure digital environment.

REFERENCES

[1]. Amini, A. A., Roussos, A., & Katsaggelos, A. K. (2019). Context-aware privacy preserving for images via semantic region redaction. In Proceedings of the IEEE International Conference on Image Processing (ICIP) (pp. 3671-3675). IEEE.

[2]. Wang, F., Chen, F., Zhang, F., & Liu, X. (2020). Privacy-preserving image redaction using salient object detection and attention-based inpainting. IEEE Transactions on Circuits and Systems for Video Technology, 31(7), 1505-1518.

[3]. Zhou, X., Yao, C., Wen, H., Liu, Y., & Tian, S. (2014). Text detection and recognition in imagery: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7), 1489-1509.

[4]. Hossain, M. S., Hasan, M. A., & Pickering, M. D. (2019). Deep learning-based privacy protection for online social networks. arXiv preprint arXiv:1909.02007.

[5]. El Bouchti, M. A., Foukar, F., & Khelladi, M. T. (2018). De-identification for privacy protection in multimedia content: A survey. Computers & Security, 74, 308-332.

[6]. Jia, X., Gong, N., Liu, X., Xiang, Y., & Zhou, Y. (2019). Privacy-preserving image sharing via homomorphic encryption with secure computation. arXiv preprint arXiv:1905.04465.

[7]. Li, X., Wang, A., Shi, Z., & Chai, T. (2019). Towards privacy-preserving facial expression recognition. In Proceedings of the 2019 ACM Multimedia Conference (pp. 2567-2575).

[8]. K. Menaka and B. Yogameena (2021). Face Detection in Blurred Surveillance Videos for Crime Investigation. In Journal of Physics: Conference Series. DOI 10.1088/1742-6596/1917/1/012024.

[9]. Li, M., Zhang, S., Wang, W., & Feng, W. (2019). A survey of privacy protection techniques for text in images and videos. arXiv preprint arXiv:1903.04517.

[10]. Zhao, J., Xu, Y., Yang, Y., & Wang, Y. (2019, June). Towards context-aware masking of sensitive information in images. In Proceedings of the International Conference on Multimedia (pp. 3098-3102).

Cite this article as :

Ardon Kotey, Tejan Gupta, Shivendra Bharuka, Abhishek Singh, Nikhil Ghugare, Lalith Samanthapuri, "Preserving Privacy in Multimedia : Text-Aware Sensitive Information Masking for Visual Data", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 10, Issue 1, pp.166-174, January-February-2024.
Available at doi : https://doi.org/10.32628/CSEIT2410117
Journal URL : https://ijsrcseit.com/CSEIT2410117