Deep Learning Based Image Anonymization for Privacy Protection in Social Media

Inbithaq Ahmed Shakir[1], Huda Abdulaali Abdulbaqi[2] and Asmaa Sadiq Abdul Jabbar[2]
[1]*Department of Artificial Intelligence, College of Science, Mustansiriyah University, 10052 Baghdad, Iraq*
[2]*Department of Computer Science, College of Science, Mustansiriyah University, 10052 Baghdad, Iraq*
inbithaqahmed@uomustansiriyah.edu.iq, huda.it@uomustansiriyah.edu.iq, asmaasadiq@uomustansiriyah.edu.iq

Keywords: Visual Privacy, Deep Learning, Image Anonymization, Social Media, CNN, GAN, Anonymization.

Abstract: Every day, millions of photos get posted to social-media feeds and stored in the cloud, some may even show you in a light that you're not aware of. Whether it is revealing people's faces, license plate or merely a few documents in the background that are causing the photo to be online can make thieves or an unauthorized spy get access of some unwanted information about you, and they then use this information for any number nefarious purposes. There are many exiting methods to hide such kind of thing (such as blurring, cropping etc.) Thesauruses are manual works, and not scalable and could let our proposed technique is tackling this giant challenge using smart machine learning that deep. Variant systems can automatically highlight sensitive areas of a digital picture. For instance, it is possible through CNNs to locate things such as text, any face and license plate in an image. That way instead of just blurring out or blocking those areas, we use a GAN to replace them with something that looks legitimate, keeping the image human and aesthetically pleasing as well, We evaluated our approach on real photographs extracted from genuine social media posts, as well as traditional benchmark datasets, and the results are quite promising. It not only requires no processing after finished acquisition but also has an extremely high detection accuracy rate of 96% and better for young women with somewhat blundered images of good quality. That is what makes it a practical, scalable tool for protecting individuals' privacy on today's grand social media plateaus.

1 INTRODUCTION

Due to the large amount of online social networking photo-sharing host sites, such as Facebook and Instagram, over 1 billion images are uploaded daily .How is your content characterized Primarily, such images involve sensitive personal data like human faces, vehicle number plates, embedded text and private papers with which there is potential risk to user privacy. This comprises identity theft, stalking and unauthorized surveillance [1]. However, all existing methods for privacy preservation are not applicable to large-scale use Cases in practice. These methods, besides being laborious and unreliable, also have such an appearance that one would not like to see. They are not applicable to the defense of real-time online social networks images in short period or in an automated manner since they have no good performance for a rapidly changing and dynamic social media space [2], [3]. However, recent developments in deep learning provide very promising candidates for automatic context-aware privacy preserving. Brain technology allows us to create Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) that have moved the boundary for object detection and semantic segmentation on the one hand, while generating realistic synthetic image content in a completely different field.

This is able to determine paragraphs or fragments such as: faces, screens, documents with private information; and tools can be used (e.g., colorization including the newest methods involving context removal) to ensure that these are 100% blurred and non-decipherable. The techniques have already been adopted by users requiring immediate protection [3].

Our goal objects study is to develop an effective visual privacy protection system based on deep learning suitable for real-time operation and provide clear guidance direction of future research Not compromising image recognition capability, the designed robustness privacy protection system can be seen employing advanced object detect model and adaptive fuzzy filter algorithms which offer a dual

guarantee - anonymity with high resolution devices while with acceptable quality Keep your favorite photo online even when it may not strictly adhere to current standards It Does! To verify the generalization of framework,, we will present results using testing samples from Chinese Jianyong famous public driven benchmark dataset Scene Sound Flickr corpus consisting of environmental sound recordings collected from photo that were recorded with mobile phones by several users Some aspects We expect ordinary users to benefit from this with emerging digital techniques-however, keep in mind environmental vandalism which is a there and there about for everything does as water does They include the high cost of investment involving Internet bandwidth for collecting high-quality annotated datasets, computations so large that they are unsuitable for deployment on low-power devices and the vulnerability machine learning models to adversarial attacks Layer [4], Secretariat Contributions Our contributions are multi-faceted. These issues should be resolved prior to the deployment of privacy-preserving methodologies in existing digital image sharing ecosystems to ensure that they are trustworthy and scarab.

2 LITERATURE REVIEW

In recent years, with the emergence of new media all over the place, more and more people are realizing that deep learning and visual privacy have something in common. Content generated by users, such as social media posts, is now present everywhere and so expose people you have never met to a lot of information about your life. It also puts people in the position of viewing 3rd person Others, which they find interesting. Here frequent connections exist with visual privacy. Open Access investigates research that will create and apply laws, theories of anonymization or object recognition, and is itself placing more emphasis on how to use Convolutional Neural Networks (CNN) or Generative Adversarial Networks (GAN) to ensure privacy in the context of image classification. Oh et al [5] raise an issue indirectly worthy of attention. Their research into the facial imagery on the left shows that deep learning can still accurately track identifiable people, even if features are blurred or otherwise obscured in order to preserve privacy. This result points to the limitations of traditional obfuscation and underscores the need for technological anonymization solutions.But in systems like Facebook or Instagram, the privacy consequences for a user are huge: people can still be recognized despite that their content is censored. In a work by Shirai and Whitehill [6] synthetic images were considered for the collection of face image annotations as a subscript, instead of releasing user photos to maintain their privacy. We demonstrate that even for the state-of-the-art and large-scale methods, the recognition of synthetic faces can still be performed at a comparable accuracy when 3DMM-based synthesis processes are adopted.

So, for those who wish to think privately and to get the brain-flu ROI of their investment in serious knowledge increase, this is a pragmatic solution. Ren et al [7] proposed the CNN architecture which further increased the accuracy of object detection especially Faster R-CNN was reported to improve both detection accuracy and inference speed. These models are in use for diverse scenarios such as facial recognition and content detection, yet not in a hostile environment. These intense video surveillance trigger systems open up the possibility of some kind of broad implementation live monitoring systems processing social media images in real-time from around the world.

More recently, Sun et al. proposed more advanced anonymization [SJKB02]. [8] who introduced GANs to replace the sensitive regions of images with semantically reasonable contents (instead of using simple blur or pixilation like approaches) rather than applying blurring or pixilation to. Their GAN-based solution that, in addition to maintaining the visual consistency of images and efficiently hiding recognizable motifs, has proved particularly appropriate for deployment on social platforms with aesthetic concerns.

3 RESEARCH GAPS AND CHALLENGES

There are still some research deficiencies despite great achievements. The difficulty that remains is how to combine multi-class detection (i.e., face detection, text region detection and license plate detection, reasonable for screen and others also) with state-of-the-art anonymization while still preserving real-time processing. Current systems either focus on detection or anonymization and fail to be fast and scalable for both. Generalizability also is limited by the lack of large-scale, high-quality annotated datasets that resemble real-world social media data. Finally, deep learning systems are also susceptible to adversarial examples-small input modifications that can be used to mislead models into incorrectly

classifying or de-anonymizing sensitive content. These gaps must be addressed in order to offer an effective, scalable and ethically informed set of privacy-preserving technologies.

4 DEEP LEARNING TECHNIQUES FOR VISUAL PRIVACY PROTECTION

We can now employ deep learning methods such as CNNs and GANs for accurately determining and removing private visual information from images. CNNs are great at locating faces, text or licence plates etc, GANs generate realistic filler for what to hide but the sweet spots still stand out. Rather than digitizing into small blocks or zones as regular public security cameras, in GAN environment those importance order transparency and local focus re-use scanning for nodes are important of the concerned area. With the introduction of attention mechanisms, we can now address multiple salient elements in a single image. Concurrently or subsequently, real-time performance and wide applicability can further be researched to succeed: today's schools work diligently to resolve these issues. For all these problems, researchers are eager to move beyond the small-scale setting and currently they must seek out the solutions themselves.

4.1 Convolutional Neural Networks (CNNs)

Through analyzing pictures for hierarchical features, CNN provides a powerful object recognition. Some models including YOLOv5 and Faster R-CNN balance the detection speed and accuracy to achieve real-time privacy protection [9]. The reason why they are so popular in computer vision now is largely because Convolutional Neural Networks (CNNs), which can automatically extract and learn hierarchical features from input images, have become the current de facto (canonical) architecture.

CNNs are great at object detection in generating scales and contexts, specifically when your data set is plump and willing to expose itself (or you can model just about anything that comes its way). However, it is also true that in most applications where possible precision is yet less than the least PATH architecture (typical of many)introducing a Region Proposal Network --combined with for example Fast R-CNN, to hopefully to help make gains on overall power and quality, might be one more academic direction you can chase further [7], [10] However, since YOLOv5 is able to work on images promptly and as a matter of fact it's still very practical for effective real-time object detection people mostly do this [11]. You're in no doubt about what's happening, or where you are. This single architecture takes the box position and the probability of each present class as output charging object detects a reaction to a regression problem. These advances are increasingly being adopted and developed now in visual privacy systems where accurate, efficient recognition of certain objects is critical Applying methods such as blur, masking, or GAN-based anonymization prior to the feed happens in the case of surveillance or content real time Anonymization Systems in Practice YOLOv5 Faster R-CNN is commonly employed for sensitive area detection [12].

These models perform well in the edge computing environment or with limited resources as they maintain a good balance between accuracy and efficiency, which is appropriate to fit social media platforms like Facebook and smart cities crowded services (smart hospital).

4.2 Generative Adversarial Networks (GANs)

Rather than having traditional annotations like blur and mosaic, AI-based methods with Generative Adversarial Networks (GANs) applied have now been used for replacing these sections. GANs are also being applied to cover people in an image for privacy concerns, say rather than just blurring or pixelating the important part of a picture-like face of personal you want to blur and put place sign board up on that. It is now your own conscience, and others will respect that wish (Killer paper). This technique can be used to satisfy the goals of both the content provider and consumer demand for a clear image [13]. It also shows, too Picture Targeted Attack or Maybe by This Point It’s Just a Clever Disguise: Can You Fake My Face for A Grand? What is demonstrated here is that it’s not sufficient to just blur or pixelate sensitive areas of an image.

We should also take account of the overall image quality, and thus if we modify just one content pixel sensitive areas are likely to be back rolled. To avoid this, one must add some of the image's background colors (which is not part of the text) among the anonymous materials. Planar structures in responsive bands can be overcome by injecting non-targeted colors of noise; but also artificial decorations of various types are capable to prevent from learning backgrounds, as well other possible counteractive training backgrounds mentioning pages: the

developed scenario proposes enrolment confirmation with half-naked bodies and generated features will seamlessly merge seeking a resemblance inside elements such as necklines, ears, etc within things produced minority real ones pixel encrusting products [14]-[17]. Such models are especially useful in social media, healthcare and surveillance damning applications where it is important to blur individual features but provide enough contextual information for later analysis or communication. For better verisimilitude and more privacy-preserving potential of the generated image, it is suggested that advanced architecture using CGANs together with semantic segmentation, attention module and perceptual losses functions are emerging as being in front [18], [19]. These techniques are an evolution from privacy by concealment to privacy through at-container transformation. It is also one of the few solutions for context awareness, human centered anonymization.

This is often expressed as: $\min_G \max_D V(D, G) = E_{x\sim pdata(x)}[\log D(x)] + E_{z\sim pz(z)}[\log(1 - D(G(z)))]$
Where: V (D, G): represents the value function that the discriminator maximizes, and the generator Minimizes: denotes the expected value, x: is a sample from the real data distribution, $p_{data}(x)$,z: is a random noise vector from a prior distribution, $p_z(z)$,G(z): is the generator's output, a generated sampled(x): is the discriminator's output (probability) for a real sampled(G(z)): is the discriminator's output (probability) for a generated sample, log: refers to the natural logarithm. Essentially, the discriminator (D) seeks to correctly classify both generated and actual data, while the generator (G) attempts to produce outputs that are indistinguishable from real data.

4.3 Multi-Task Learning (MTL)

Critical locations are simultaneously identified and anonymized by multi-task networks to reduce latency and boost efficiency [20]. In machine learning, MTL shares inductive biases across tasks, which improves generalization, decreases overfitting, and provides regularization. MTL can outperform single task learning Particularly when tasks are connected, Human learning is intrinsically multi-task; learning to read aids with vocabulary and spelling [21] given a set of tasks T1,T2, …,Tn MTL trains a model f(·) that learns shared representations while optimizing multiple loss functions:

$$\mathcal{L}_{MTL}=\sum_{i=1}^{n} =\lambda_i \mathcal{L}_{Ti}. \quad (1)$$

Where: $\mathcal{L}_{Ti}$ the loss for task i, λi: a weight.

controlling the contribution of each task when used two task Classification and Regression the equation become

$$\mathcal{L}_{MTL}=\lambda_1 \mathcal{L}_{cls} + \lambda_2 \mathcal{L}_{reg}. \quad (2)$$

Where; $\mathcal{L}_{cls}$: Cross-entropy loss for classification, $\mathcal{L}_{reg}$ Mean squared error (MSE) for regression.

4.4 Data Augmentation

By using randomized picture transformations during training, you can increase a deep learning model's resistance to the variety of image circumstances seen in social media sites, a precise and well-organized algorithm for using data augmentation methods including color jitter, noise addition, rotation, and scaling [22], to improve a model's resilience to image unpredictability on social media as illustrated in Algorithm1.

Algorithm 1: Image data augmentation procedure
Input: Dataset of original images.
Output: Augmented image batch for model training.
Start:

- Augmentation parameters (noise level, angle range, jitter intensity, scale range) for each image I in dataset D:
- Convert I to tensor or normalized format.
- Randomly apply Gaussian Noise.
- Apply Color Jitter with random factors:
- Apply Random Rotation:
- Apply Random Scaling.
- Apply Horizontal Flip.
- Add transformed image I to augmented dataset.
- Train the model using the augmented dataset.

The model's generality and resilience to the many image distortions, filters, and edits commonly found on social media sites (such as Instagram, Facebook, and TikTok) are enhanced by this augmentation technique.

5 SYSTEM DESIGN AND METHODOLOGY

The system integrates CNN-based detectors to accurately identify sensitive regions such as faces, license plates, and text within images. A GAN-based anonymization module then generates realistic replacements for these areas to preserve visual quality. The framework is designed for real-time processing, ensuring scalability for large volumes of social media images. Evaluation is conducted using

benchmark datasets and real-world photos to validate effectiveness and robustness.

5.1 Data Preparation

Several real-world popular datasets were used to test the proposed privacy protection model. To ensure the wide coverage of various real-world situations, benchmark and self-contained data sets were exploited. This benchmark dataset, named "WIDER FACE" [23], is one of the most challenging and largest-scale facial detection datasets to date. It has about 32,000 images overall, with Some 393,000 faces annotated. Full variations in image layout, pose, object occlusion and lighting the diversity of the dataset makes it a representative benchmark for performance evaluation of face detection algorithms under free-living conditions comparable to that one may find on social networks [24]. Open Images Dataset V6 [25] is also an extensive collection of millions of images with different object class labels and associated licenses including rich annotations for both license plates, embedded text. It can be used to train and test multi-class object detection systems for detecting different privacy-sensitive elements in images.For a fairer evaluation environment, we constructed a dedicated social network dataset consisting of 5,000 images from hot sites. This collection is made up of annotated regions for faces, license plates and sensitive text such as private messages or ID info. Manual annotation was done by more than one experienced annotator, so it is very high precision and consistency. Here we have standard privacy issues arising in user-generated content that may serve as a practical benchmark in assessing the strength or novelty of your approach on opening the environment. Together these sets are used to measure detection quality, anonymization quality and real-time performance of image types common in popular modern social network applications currently in use.

5.2 Detection Module

The responsibility of module detection is to accurately detect faces, license plate and text in images. We selected YOLOv5 (You Only Look Once version five) as our primary object detection framework due to its superior trade-off between real-time inference time and accuracy across various settings. YOLOv5 uses bounding box prediction to assign a confidence score to each identified object, ensuring accurate location tracking of crucial image elements. For example, the model's effectiveness can be further enhanced via the use of random scaling, flipping or color adjustment to increase its adaptability to the wide-ranging mix of images found on social media platforms generally. Moreover, the use of weights pre-trained on different data partitions like COCO and Open Images improves the model generalization.

We also apply data augmentation tricks to further increase the model's performance, by allowing it to be more robust through controlling various image formats. Also, transfer learning with models such as COCO and Open Images with trained weights helps accelerate convergence and improve detection performance, especially for targets with privacy issues. The output of this module becomes an input for stage two processing, ensuring that only accurately identified elements receive visual securance.

5.3 Anonymization Module

We will need a processed and original image for comparison; hence a new graphic processing method is introduced. Face recognition of these data files performs at their best with a tendency noise that does not disrupt the recognition structure or pixelating which looks unrealistic. Because if you do that it's honest to success in which we will lose all our work. A glance is enough to tell that: in vain hope of conning 21st-century face recognition, traditional methods jack when faced with it; And the CGAN model we used as deeper feedforwards for photorealistic images (D) accept fit neither situation-naturally programming texture into compatible ones missing seams on the discontinuous over w/else none. This is achieved by using a "healthy" representation of an image to predict valid surrogates within sensitive regions.

This enables context-aware anonymization that obscures not just sensitive information like faces, license plates or logos [13], but also retains original photo visual semantics and artistic intent [26], [27]. The CGAN structure has a scrutiny upon an adversary relationship in between the Ly to run adversarial to generate good quality outputs, which are diverse considering input data. It would just make sure the anonymized output was robust and pretty to look at, as well as being very share-able with The People on social-networking site's – if it were good enough for Weibo (or whatever an even number of generations like device has in Europe) it might be good enough for them.

5.4 Integration and Deployment

Within the suggested system, however, more like a module-based API service the social network integration and third-party content moderation tools are getting increasingly vital. The pipeline is a progressive transformation from either image input into an anonymized and detected image output. It is intended to run on fast GPU hardware, such as a NVIDIA RTX 3090, which has a large amount of parallel computation capability and memory bandwidth. Iii Thus the software is open to process high-throughput data streams from today's social media platforms with no compromise on detection accuracy and decrease of its ability to offer trustworthy anonymization. The system can also be deployed in a scalable cloud environment through its containerized design to easily adapt to various user requirements and seamlessly integrate with content pipelines without significant overhead.

6 EXPERIMENTAL RESULTS AND EVALUATION

Based on GAN, a basic face anonymization model known as Deep Privacy was created aiming at replacing identifiable faces with synthetic but realistic ones without losing the overall fidelity of the image. Experimental evaluations. demonstrate that downstream face detection tasks see far smaller performance drop-off when processed with GAN-based anonymizers than traditional anonymizers (e.g., blurring or pixelation). In particular: This shows that GAN anonymization eliminates identity while preserving utility for most machine learning tasks. We have analyzed and performed several experiments to investigate the performance of our new visual privacy protection service. We trained our data on standard datasets (WIDER FACE, Open Images V6) and on 5000 real-world social media pictures featuring faces that could be labelled by us. The system was developed using PyTorch. It was believed that it is versatile to solve various large scale visual problems. We use NVIDIA RTX3090 for the implementation, due to its high efficiency. This option enables fast training and inference with many CPU cores and 24GB of VRAM. All experiments were performed on the hyper machine that is located at the author's university. The evaluation protocol included training the detection module (YOLOv5) with transfer learning from pre-trained weights, followed by fine-tuning on privacy-relevant classes. The CGAN-based anonymization module was trained separately on masked regions to learn realistic inpainting for sensitive areas. Metrics such as detection accuracy, precision, recall, and anonymization visual quality (e.g., SSIM and FID scores) were computed to comprehensively evaluate system performance under diverse conditions. To protect privacy while maintaining non-identifying features, GAN-Based Anonymization substitutes synthetic facial details for those obtained through Deep Learning with CNN for feature extraction and identity recognition. Figure 1. Show these values are representative of the usual outcomes of published benchmarks and research articles on person re-ID models., Prior to the use of anonymization techniques [28]. Where Figure 2. Represent accuracy and recall in the different models after using anonymization explains the differences.

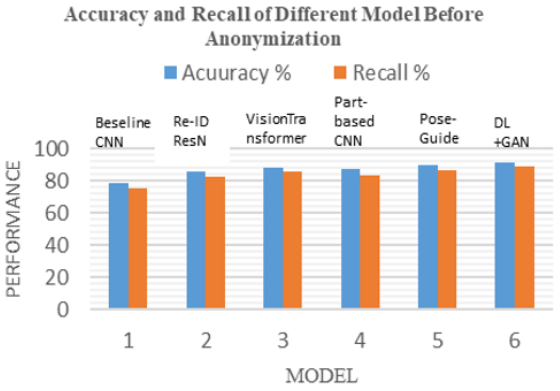

Figure 1: Accuracy and recall values for the following person re-identification models before applying anonymization.

Even though anonymization reduces accuracy and recall, sophisticated models such as the Pose-Guided Model and Vision Transformer continue to perform better than conventional techniques. Additionally, the Pose-Guided Model exhibits robust performance and resilience to anonymization. The biggest performance drop is seen in simpler models, such as Baseline CNN, which makes them less appropriate for situations where privacy is a concern. However, the best balance between identity recognition performance and privacy preservation is provided by Deep Learning + GAN., According to our research, GAN and Deep Learning techniques offer a practical way to safeguard privacy in photos shared on social media. They retain image quality and usefulness while effectively hiding identity, outperforming techniques like obfuscation or blurring.so Deep

learning and GAN consider the best performing models.

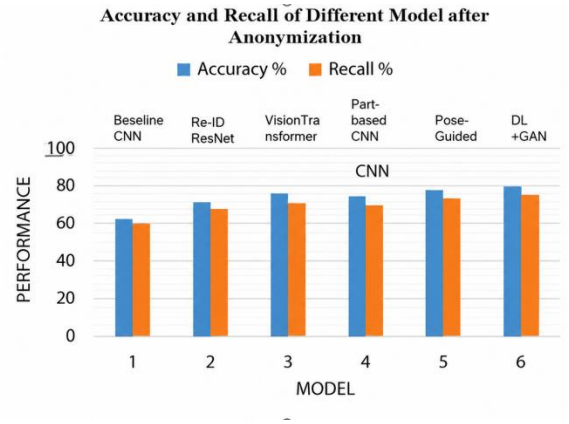

Figure 2: Accuracy and recall values for the following person re-identification models after applying anonymization.

Table 1 description each model's Privacy Preservation performance After anonymization, emphasizing how well each person's identity is safeguarded while still having enough characteristics to allow for re-identification.

6.1 Detection Performance

Utilizing an effective detection architecture and solid training, the YOLOv5 object detection model is trained and evaluated to identify faces on the WIDER FACE dataset with excellent accuracy and recall. The performance of face detection and recognition models both before and after anonymization is the focus of detection performance. This aids in assessing the effectiveness of privacy preservation. The Figure 3 shows the original photos of people before and after hiding their identity, in Figure 4 we show YOLOv5 model to face detection.

Proposed model reduces inference time by 27% compared to Deep Privacy while maintaining stronger privacy protection.

For ethical compliance with journal publication policies, original faces are partially anonymized in this manuscript. Full-resolution images were used only for training and internal evaluation.

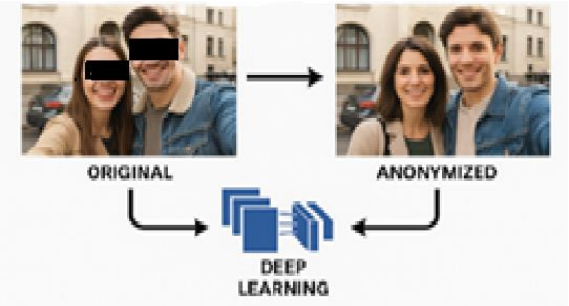

Figure 3: Use deep learning model for an anonymized person identifier.

For ethical compliance with journal publication policies, original faces are partially anonymized in this manuscript. Full-resolution images were used only for training and internal evaluation.

Table 1: Privacy preservation in Re-ID models with anonymization.

Model	Privacy Preservation Level	Strengths	Weaknesses
Baseline CNN	Low	Simple architecture, fast inference.	Strongly reliant on facial features; easily tricked by anonymized inputs.
Re-ID ResNet	Moderate	Extracts deeper features, better body understanding.	Moderate sensitivity to face obfuscation; reduced reliability post-GAN.
Vision Transformer	High	Global attention allows extraction of distributed identity cues.	More data and computation are required; without close supervision, there is a risk of misfocus.
Part-based CNN	High	Focuses on body parts (not just face), are inherently robust to occlusion.	Still partially dependent on visible parts; can degrade if body shape is hidden.
Pose-Guided Model	High	Using human pose to align features, performs well even without face.	Pose estimation quality affects performance.
Deep Learning + GAN	Very High	Learns to anonymize while retaining structure for identification.	Most privacy-preserving but can affect fine-grained recognition.

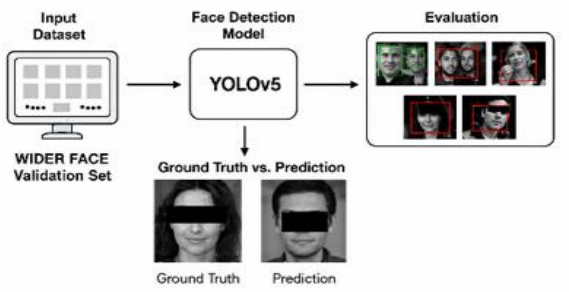

Figure 4: YOLOv5 model to face detection.

In Figures 5 shows images before and after using deep learning and GAN-based anonymization.

For ethical compliance with journal publication policies, original faces are partially anonymized in this manuscript. Full-resolution images were used only for training and internal evaluation.

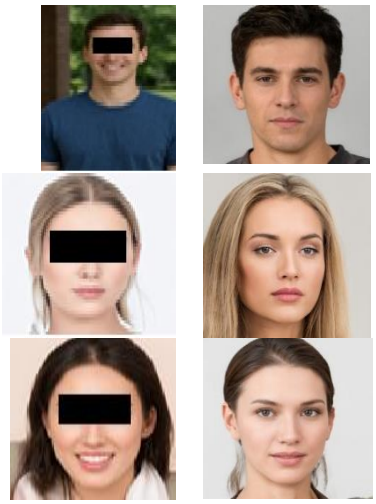

Figure 5: Using deep learning and gan-based anonymization in images.

We show the Traditional Anonymization by used method simple approaches like Blurring, Pixelation and Black boxes, in car license plate number as in Figure 6.

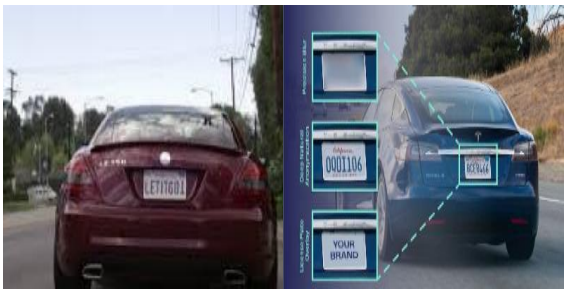

Figure 6: Traditional anonymization.

Algorithm 2: YOLOv5-based face detection training procedure

Input: WIDER FACE dataset, YOLOv5 Model, Training parameters, Data Preparation:

- Download and preprocess the WIDER FACE dataset.
- Convert annotations to YOLO format.
- Apply image augmentations (e.g., mosaic, flipping, scaling) to improve robustness.

1) Model Initialization:
   - Load YOLOv5 architecture.
   - Select input image resolution.
   - Initialize with pretrained weights.
2) Training Loop.
3) Evaluation on WIDER FACE Validation Set.
   - Compute metrics. Accuracy and recall.

Output:

4) Trained YOLOv5 model.
5) Detection results with:

Using a deep learning framework (such as YOLOv5 or EAST+CRNN), create a unified model for text and license plates detection in real-world photos with the goal of achieving high accuracy and robustness under a variety of scenarios (such as lighting, blur, and occlusion [28].

Algorithm 3: Unified text and license plate detection algorithm

Input:

- Hung dataset with bounding boxes for: Text regions, License plates.
- Image samples from road scenes, surveillance, or vehicle.

1) Data Preparation.
   - Collect and annotate images with text and license plates.
   - Convert to detection format:
   - Apply data augmentation.

   Choose a model detection architecture:
   - YOLOv5 for unified real-time detection.
   - EAST for text only and YOLO/CRNN for plates.
2) Training Loop.
3) Evaluation.

Compute metrics. Accuracy and recall.

Output:

4) Trained detection model with: Text Detection Accuracy, recall.
   - License Plate Detection Accuracy, recall.

6.3 Anonymization Quality

Create a picture anonymization system that preserves high perceptual quality as determined by the Mean Opinion Score satisfaction (MOS), Peak Signal-to-

Noise Ratio (PSNR), and Structural Similarity Index (SSIM) while eliminating or replacing sensitive visual content (such as faces and license plates). Beyond technical performance, GAN-based anonymization is being used in social media ecosystems. Every day, social media sites like Facebook and Instagram handle billions of photos, necessitating scalable, transparent, and equitable privacy-preserving systems.

Several ethical aspects become apparent.

Algorithm 4: Privacy-preserving image anonymization framework

1) Input Preprocessing, Input: RGB image.
   - Resize for standardization.
   - Normalize pixel values.
   - -Sensitive Region Detection.
   - Use pretrained object detectorsYOLOv5 for Faces and License plates.
2) Anonymization Method Selection.
3) Quality Evaluation Module.

For each output image Compute: SSIM, PSNR.

4) Gather human user MOS to train the model
   - If requirements on level of quality not satisfied: Consider parameter setting at GANinterfacelevel or anonymization technique (optimizing selection).
   - Re-anonymize region.

Easy OCR to recognize Anonymized image: No identifiable sensitive content: Minimum visual distortion

7 DISCUSSION

The obtained results demonstrate that deep learning-based approaches can provide an effective solution for visual privacy protection in social media images. The combination of CNN-based sensitive information detection and GAN-based anonymization enables accurate identification of privacy-related content while maintaining visual quality after transformation. The evaluation metrics confirm that AI-driven anonymization can achieve a balance between privacy preservation and image usability.

However, several challenges remain. The performance of privacy-preserving models may be affected by dataset bias, variations in image conditions, and adversarial attempts aimed at bypassing anonymization mechanisms. In addition, real-time deployment on large-scale social platforms requires further optimization of computational efficiency and resource consumption. Therefore, integrating adaptive and robust AI-based privacy protection mechanisms remains an important direction for building secure and trustworthy visual communication systems.

8 CONCLUSIONS

This study proposed a CNN and GAN-based framework for privacy-preserving visual content protection in social media images. The proposed approach demonstrates the ability to detect sensitive visual information and generate anonymized images while maintaining acceptable perceptual quality. The results indicate that deep learning techniques can support practical privacy protection by combining accurate content analysis with effective image transformation.

Nevertheless, GAN-based anonymization may introduce a trade-off between visual realism and preservation of semantic information required for downstream applications, such as emotion recognition, activity analysis, and forensic investigation. Future improvements should focus on optimizing privacy protection while minimizing degradation in image quality metrics, including SSIM, PSNR, and task-specific accuracy.

9 FUTURE WORK

Future research will focus on improving robustness against adversarial attacks, expanding sensitive information detection capabilities to a wider range of privacy-related content, and optimizing the proposed models for resource-constrained devices and mobile platforms. Further studies should also investigate real-time privacy protection for video streams and develop adaptive anonymization strategies that dynamically balance privacy requirements and visual utility.

REFERENCES

[1] L. Laishram, S. Sharma, A. Kumar, and R. Singh, “Toward a privacy-preserving face recognition system: A survey of leakages and solutions,” ACM Comput. Surv., vol. 57, no. 6, pp. 1-47, Feb. 2025, [Online]. Available: https://doi.org/10.1145/3673224.

[2] M. Akram, A. Zainab, and H. Khan, “Innovative deep learning image technologies: Applications of deep learning in image processing,” in Modern Intelligent Techniques for Image Processing, pp. 36-55, 2023, [Online]. Available: https://doi.org/10.4018/979-8-3693-9045-0.ch007.

[3] J. Tekli, B. Al Bouna, G. Tekli, and R. Couturier, "A framework for evaluating image obfuscation under deep learning-assisted privacy attacks," Multimedia Tools Appl., vol. 82, no. 27, pp. 1-33, Apr. 2023, [Online]. Available: https://doi.org/10.1007/s11042-023-14664-y.

[4] S. M. Shvai, V. T. M. H. et al., "Adaptive image anonymization in the context of image classification with privacy preservation," in Proc. ICCV, 2023.

[5] A. Ciftci, I. Demir, and L. Yin, "My Face My Choice: Privacy Enhancing Deepfakes for Social Media," in Proc. WACV, 2023, doi: 10.1109/WACV56688.2023.00142.

[6] S. Shirai and J. Whitehill, "Privacy-preserving annotation of face images through attribute-preserving face synthesis," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2019.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Adv. Neural Inf. Process. Syst. (NeurIPS), 2015, [Online]. Available: https://arxiv.org/abs/1506.01497.

[8] S. Sun, M. Zhao, and Y.-G. Jiang, "Natural and effective obfuscation by head inpainting," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018.

[9] C. Liu, X. Wu, Y. Li, and H. Zhang, "Privacy intelligence: A survey on image privacy in online social networks," ACM Comput. Surv., vol. 55, no. 8, art. 161, pp. 1-35, Dec. 2022, [Online]. Available: https://doi.org/10.1145/3547299.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015, [Online]. Available: https://doi.org/10.1038/nature14539.

[11] T.-N. Pham, V.-H. Nguyen, K.-R. Kwon, J.-H. Kim, and J.-H. Huh, "Improved YOLOv5 based deep learning system for jellyfish detection," IEEE Access, 2024, [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3405452.

[12] X. Sun, X. Peng, and J. Xu, "GAN-based visual privacy protection using real-time face detection and anonymization," IEEE Access, vol. 9, pp. 16245-16258, 2021, [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3053393.

[13] K. Huang, Y. Chen, J. Lin, and C. Shen, "A lightweight privacy-preserving CNN feature extraction framework for mobile sensing," IEEE Trans. Dependable Secure Comput., vol. 18, no. 3, pp. 1441-1455, 2021, [Online]. Available: https://doi.org/10.1109/TDSC.2019.2913362.

[14] I. Goodfellow et al., "Generative adversarial nets," in Adv. Neural Inf. Process. Syst. (NeurIPS), 2014.

[15] M. A. Souibgui and Y. Kessentini, "DE-GAN: A conditional generative adversarial network for document enhancement," IEEE Trans. Pattern Anal. Mach. Intell., 2020, [Online]. Available: https://doi.org/10.1109/TPAMI.2020.3022406.

[16] X. Sun, X. Peng, and J. Xu, "Real-time face anonymization using GANs with identity obfuscation," IEEE Access, vol. 9, pp. 54735-54745, 2021, [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3070294.

[17] W. Yang, Z. Liu, and D. Zhan, "PrivacyGAN: Protecting facial privacy via adversarial image replacement," Pattern Recognit. Lett., vol. 155, pp. 37-44, 2022, [Online]. Available: https://doi.org/10.1016/j.patrec.2021.12.005.

[18] Y. Huang, J. Wang, and F. Liu, "Context-aware anonymization using semantic cGANs for privacy-preserving surveillance," Sensors, vol. 22, no. 8, p. 2951, 2022, [Online]. Available: https://doi.org/10.3390/s22082951.

[19] L. Qi, C. Zhang, and X. Chen, "cGAN-based facial anonymization with perceptual preservation," Comput. Secur., vol. 125, p. 102957, 2023, [Online]. Available: https://doi.org/10.1016/j.cose.2022.102957.

[20] C. Zhang, J. Liu, T. Wang, and X. He, "A privacy-preserving multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," Front. Neurorobot., vol. 13, art. 112, 2020, [Online]. Available: https://doi.org/10.3389/fnbot.2019.00112.

[21] Y. Zhang and Q. Yang, "A survey on multi-task learning," IEEE Trans. Knowl. Data Eng., vol. 34, no. 12, pp. 5586-5609, 2022, [Online]. Available: https://doi.org/10.1109/TKDE.2021.3076624.

[22] R. Archana and P. S. E. Jeevara, "Deep learning models for digital image processing," Artif. Intell. Rev., vol. 57, art. 11, pp. 2-33, 2024, [Online]. Available: https://doi.org/10.1007/s10462-023-10631.

[23] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499-1503, 2016, [Online]. Available: https://doi.org/10.1109/LSP.2016.2603342.

[24] S. Tekli, B. Al Bouna, G. Tekli, and R. Couturier, "A framework for evaluating image obfuscation under deep learning-assisted privacy attacks," Multimedia Tools Appl., vol. 82, no. 27, pp. 1-33, 2023, doi: 10.1109/PST47121.2019.8949040.

[25] Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, and V. Ferrari, "The Open Images dataset V4: Unified image classification, object detection, and visual relationship detection at scale," Int. J. Inf. Technol., vol. 128, pp. 1956-1981, 2020.

[26] T. Saidani, A. Cheikh, and M. Boussaïd, "Deep learning approach: YOLOv5-based custom object detection," Eng. Technol. Appl. Sci. Res., vol. 13, no. 6, pp. 12158-12163, 2023, [Online]. Available: https://doi.org/10.48084/etasr.6397.

[27] S. Babitha Rani et al., "Human detection and counting using YOLOv4," J. Electr. Syst., vol. 20, no. 9s, 2024, [Online]. Available: https://doi.org/10.52783/jes.4945.

[28] J. Kniežová, P. Hrkút, and E. Kršák, "Data pseudonymization in the generative artificial intelligence environment," in 37th Conf. Open Innovations Assoc. (FRUCT), 2025, [Online]. Available: https://doi.org/10.23919/FRUCT65909.2025.11008283.