Journal of Applied Engineering and Technological Science
Vol 7(2) 2026: 1609-1634

ENHANCING PRIVACY IN REAL-TIME VIDEO STREAMS: TECHNIQUES, CHALLENGES, AND BENCHMARK DATASETS POWERED BY DEEP LEARNING

Emad I. Nyaz[1*], Mohammed S.H. Al-Tamimi[2]
Computer Science Department, College of Science, University of Baghdad, Iraq[12]
emad.niaz2301@sc.uobaghdad.edu.iq[1*], m_altamimi75@yahoo.com[2]

Received: 31 May 2025, Revised: 13 February 2026, Accepted: 24 February 2026
**Corresponding Author*

ABSTRACT
The exponential growth of video surveillance, live streaming platforms, and AI-driven analytics has created unprecedented threats to visual privacy. Traditional de-identification methods (pixelation, blurring) fail to balance privacy protection with contextual utility in dynamic environments. This systematic review of 30+ peer-reviewed studies uses a taxonomical framework to classify machine learning-based privacy preservation techniques into three domains: intervention methods (sensor saturation, broadcasting commands), obfuscation strategies (encryption, morphing, adaptive blurring), and secure processing pipelines. We test convolutional neural networks (CNNs), YOLO-based object detection systems, and hybrid approaches including GAN-driven synthetic data substitution using public datasets (MARS, DukeMTMC, Market-1501). CNN-YOLO hybrid architectures provide 30+ FPS real-time performance with 92-98% detection accuracy, while GAN-based anonymization preserves visual usefulness better than traditional approaches. Dataset scalability, illumination variability handling (accuracy drops 15-23% in low-light settings), occlusion resilience, and adversarial attack vulnerability remain key shortcomings. Although promising, lightweight encryption approaches for edge devices cost 12-18% processing speed and lack defined privacy-utility trade-off measures. Implications: This work unifies computer vision, cryptography, and privacy engineering into a single taxonomy, showing that context-aware frameworks need multi-level security designs to manage varied threat scenarios. Our findings help practitioners choose strategies depending on deployment restrictions (computational resources, latency, privacy regulations), yet 67% of reviewed methods lack real-world validation outside controlled datasets.This review uniquely synthesizes intervention, obfuscation, and secure processing research to provide uniform standards, context-adaptive privacy frameworks, and adversarially-robust de-identification systems. Five key research directions—federated learning for distributed privacy, attention-mechanism-enhanced detection under occlusion, and explainable AI for privacy-utility optimization—will shape the next generation of ethical, scalable visual privacy solutions in pervasive video analytics.
Keywords : *Privacy, De-identification, CNN, YOLO, Datasets*

1. Introduction

With the advancement of identity recognition technologies through video live streaming in social media platforms, surveillance systems placed in public or private places or any other equipment's, which making collecting sensitive private data easier even without explicit user consent, these collected data facial identity, biometric recognition or any other sensitive information that may be misused for identity thefts, impersonation or fraudulent acts that raise concerns about individual privacy(Bochkovskiy et al., 2020). Machine learning (ML) has revolutionized privacy preservation by conducting context-aware systems that is able to identify and anonymize sensitive information in real-time videos. ML algorithms such as supervised and unsupervised learning has shown the ability to adapt to real-time environments (Mohammed et al., 2024), for instance (Y. Li & Lyu, 2019) demonstrated the efficacy of ML-based approaches in dynamically masking facial features in diverse conditions (lighting and occluded objects), the advancements of ML had some potential to mitigate privacy risks (Y. Li & Lyu, 2019). Deep learning has changed real-time video privacy. While computationally efficient, privacy-preserving methods like pixelation, blurring, and masking are limited in dynamic contexts and lack contextual relevance for downstream analytics (Hukkelas & Lindseth, 2023). Deep learning allows context-aware, adaptive algorithms to safeguard privacy while keeping crucial visual information for authorized surveillance and monitoring (Raj & Gopalan, 2025). CNNs are the

foundation for privacy-sensitive region recognition and segmentation. Modern CNN-based systems use Faster R-CNN, Mask R-CNN, and YOLO (You Only Look Once) versions to recognize faces, license plates, and human bodies in real time with exceptional accuracy and speed (Dhayea et al., 2024) (Hukkelas & Lindseth, 2023). In edge deployment settings, YOLOv8m-seg achieves adaptive anonymization depending on relative object size and depth while retaining real-time performance (Asres et al., 2026). BlurNet, an early pioneer, developed YOLOv3 for simultaneous face and license-plate identification in privacy-preserving data collection pipelines, proving detection-then-obfuscation procedures work (Dworak, 2020). GANs move from obfuscation to intelligent identity substitution. GAN-based methods create photorealistic surrogate faces that preserve non-identity attributes like pose, expression, age, and gender while replacing identity information (Hellmann et al., 2023); (Maximov et al., 2020). Traditional methods simply obscure sensitive information. CIAGAN (Conditional Identity Anonymization GAN) pioneered conditional generators that retain pose and expression consistency across video frames, achieving state-of-the-art de-identification performance while keeping downstream task utility (Maximov et al., 2020). FIVA (Facial Image and Video Anonymization) used temporal consistency techniques and explicit reconstruction defenses to demonstrate robust anonymization even under adversarial settings(Rosberg et al., 2023).

Recent developments in latent-space manipulation provide finer privacy-utility tradeoff control. StyleID and FALCO project faces into latent space, alter identification vectors while retaining other properties, and provide anonymized outputs with customizable privacy levels using pre-trained StyleGAN models. These methods are faster than custom GAN training while preserving visual quality and attributes. For edge devices and embedded systems, deep learning privacy systems must meet strict latency and computational restrictions. Recent research shows that resource-constrained platforms can support privacy-preserving deep learning without losing real-time performance. PhiNet-GAN achieves over 15 frames per second on a low-power embedded Kendryte K210 microprocessor while preserving photorealistic face swapping(Ancilotto et al., 2023). Generated models can run within IoT devices' power and memory budgets with careful architectural design, including depthwise separable convolutions and efficient residual blocks. (Lee et al., 2021) applied trained face-anonymization models directly to unmanned aerial aircraft to preserve SLAM performance and anonymize all captured faces before transmission or storage. The literature analysis cited DeepDish's edge-based YuNet face identification and tracking method using Raspberry Pi hardware. Tracking temporal coherence eliminates redundant detection calls, resulting in ~83% accuracy and 75% delay increase compared to non-private baselines, making it a reasonable compromise for various applications. Hybrid detection-tracking is an important resource-constrained deployment design concept. Split-computing architecture from PrivacyEye separates privacy-sensitive processing between edge devices and cloud servers for computationally effective video analytics while assuring raw identifiable data never reaches the edge. Architectural innovation can balance privacy and performance, since downstream workloads degrade less than 2% while giving strong privacy guarantees.

However, despite these advancements several gabs and limitations still persist in the field The research on deep learning privacy for real-time video streams has many important holes despite significant progress:

Unstandardized Video Benchmarks: Most datasets use static images or brief action clips instead of genuine streaming scenarios with temporal dynamics, occlusion patterns, and environmental heterogeneity. Unstandardized benchmarks make technique comparisons difficult. Poor Adversarial Evaluation Protocols: Few papers use common attacker models or evaluation methods, making cross-paper anonymity claims impossible to compare. Standardized adversarial evaluation procedures like adversarial robustness research are badly needed. Limited Real-World Deployment Studies: Most research uses controlled offline datasets. Large-scale field deployments with longitudinal privacy, utility, and system reliability evaluations are rare.

Cryptographic Practicality Gap: Homomorphic encryption and secure multi-party computation provide strong theoretical guarantees, but scaled demonstrations of real-time video anonymization are scarce. The computational overhead and systems engineering problems need

more study. Long-Tail Scenarios and Fairness: Few research analyze rare identities, occluded persons, or minority demographic groupings, and most benchmarks have identity imbalance. Privacy preservation that is fair is understudied.

The objective of this paper is to systematically review and synthesize state-of-the-art deep learning-based methods for real-time visual privacy preservation. Following the PRISMA methodology, we categorize approaches into a three-part taxonomy: intervention techniques (e.g., sensor saturation), obfuscation strategies (e.g., GAN-based identity substitution), and secure processing pipelines (e.g., federated learning, encryption). We also critically evaluate benchmark datasets (e.g., Market-1501, DukeMTMC, MARS) and analyze their fitness for evaluating privacy-preserving methods.

Our theoretical framework unifies cryptography, computer vision, and machine learning under a privacy engineering lens, enabling the comparative evaluation of strategies across accuracy, latency, robustness, and ethical viability. This review not only highlights state-of-the-art developments but also proposes a roadmap for future work in adaptive anonymization, adversarial robustness, explainable privacy, and fairness-aware systems.

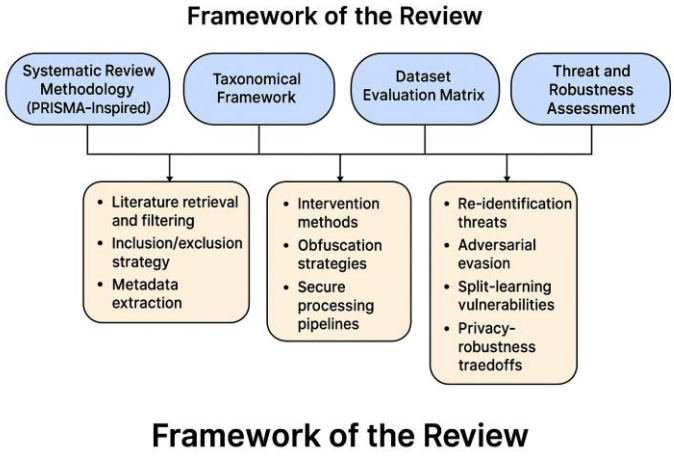

Fig. 1. Framework of the Review

This paper is structured as follows: Section 2 discusses recent related works; Section 3 outlines types of visual privacy preservation; Section 4 examines common visual obfuscation techniques; Section 5 describes detection methods; and Section 6 surveys major datasets used in the field.

2. Literature Review

Many researches achieved to manage and obscure important personal information's mainly hiding the appearance of individuals total body in the frames without compromising other data which it may be needed for analysis. Detecting Pedestrian could be achieved by many techniques Brkić, K., et al used in their study a Gaussian Mixture Models to detect moving pedestrians by modelling pixel distributions over time. Assumes static cameras and leverages motion cues to identify foreground (pedestrian) regions.an enhanced GrabCut algorithm is used in this study for pedestrian segmentation as the original GrabCut is semi-auto this improvement made it to work automatically by prioritizes pixels closer to the detected foreground boundary. Integrates background subtraction confidence to reduce misclassification. The de-identification via Neural Art algorithm, randomly selects a "style image" from a database, apply artistic/textured styles to alter clothing, hair, skin color, and facial features, transfers the style to the segmented pedestrian while retaining body pose and structure(Brkić et al., 2017).

Shifa, A. et al., presented in their study a hybrid method to obscure the skin of the individuals by encrypting the detected skin pixels. Each pixel goes through a color-space fusion, HSV (Hue, Saturation, Value), RGB (Red, Green, Blue), YCbCr (Luminance and Chrominance components). These color spaces are chosen to handle variations in skin tones, illumination, and

environmental conditions. Then a combined dynamic and explicit threshold is applied, dynamic thresholding adapts to skin tones using facial feature localization. explicit thresholding is used if dynamic detection fails (e.g., due to occluded faces or low-resolution videos). Pixels are classified as skin if at least two of the three-color spaces agree, reducing false positives. Advanced Encryption Standard in Cipher Feedback (CFB) encrypts detected skin pixels. This will allow only otherized users to decrypt and reveal the actual data using a proper key.(Shifa, Imtiaz, et al., 2020).

The study of Chen D, et al. presents a system providing privacy for a specific individuals in registered videos using an improved learning algorithm to identify people more accurately by face detection and identification with a small labeled data. Faces are detected, tracked forward and backward in video frames, and obscured using masks. The Edge Motion History Image (EMHI) removes body texture but preserves structural and motion cues for activity analysis. (Chen et al., 2007).

According to Climent-Pérez et al. (2020), body pose and motion analysis play a crucial role in active and assisted living systems, as they enable the detection of unexpected actions, particularly in hospitals, nursing homes, and other healthcare-related environments. This study uses RGB-only based visual privacy preservation filter. It uses based deep learning for segmentation DensePose and Mask R-CNN networks together combining the output for generating a robust human mask. For more insurance the merged mask will be dilated (15-pixls radius) further prevent information leakage from undetected edges or clothing. Several techniques are applied using the mask, Invisibility filter (completely removes user from scene), Pixelation, Blurring, Embossing, Replacement with an anonymizing avatar. The use of multiple filters offers different levels of privacy protection. The authors provide privacy for human beside of allowing for effective visual privacy protection in AAL applications while still enabling useful monitoring capabilities.(Climent-Pérez & Florez-Revuelta, 2021)

In the research of Shifa A, et al. For smart surveillance security systems, a five-level security (L1-L5) is proposed based on device capabilities and network requirements. Lower levels use lightweight encryption for constrained devices, while higher levels provide more comprehensive protection for capable devices. (L1/L2: Encrypt motion vectors (MVD) or texture coefficients for low-resource devices) to detect motion a Temporal difference between frames technique is used. (L3: Encrypt faces/human bodies for privacy protection) by using Histogram of Flows (HOF) with SVM classification for motion-based detection. (L4: Encrypt background to hide locations) Mixture of Gaussians (MOG2) algorithm to separate static and dynamic regions. (L5: Partial-full encryption for high-resource devices) at this level the entire frame is encrypted. AES (Advanced Encryption Standard) in Output Feedback (OFB) mode is applied to encrypt the detected FOIs (faces or full human bodies), ensuring privacy while preserving structural information for behavioral analysis (Shifa, Asghar, et al., 2020).

Asres MW, et al. study provides privacy in crowd video anomaly detection(VAD) proposing a lightweight adaptive anonymization for VAD that employs dynamic adjustment to enhance privacy protection. Different parameters used blurring , pixelization or these parameters adjusts anonymization based on the relative size/depth of detected human figures. YOLOv8m-seg for segmentation and to isolate regions requiring anonymization. The method evaluates privacy leakage through attribute detection (e.g., gender, face, skin color) and person re-identification (ReID) attacks. This study achieved to balancing privacy protection and utility preservation in real-time video surveillance systems(Asres et al., 2026).

3. Methods of Visual Privacy Preservation

Recent video analytics privacy research has explored cryptography and distributed learning methods beyond detection and anonymization. Video Processing Differential Privacy (DP) injects calibrated noise into model training or inference pipelines to guarantee privacy. On UCF-101 and HMDB-51 datasets, (Luo et al., 2023) built differentially private video activity identification systems that introduce noise to intermediate feature representations to safeguard action detection privacy while retaining accuracy. They showed that noise calibration and clipping can modify spatial-temporal transformers for DP training. In contrast, new investigations show complicated

connections between differential privacy and adversarial robustness. Boenisch et al. (2021) found that naive DP training can cause gradient masking, making models appear robust yet vulnerable to adaptive attacks. This shows the importance of evaluating DP systems against sophisticated opponents. Phan et al. (2019) developed computationally expensive yet provably robust DP training algorithms that optimize for privacy and adversarial robustness.

Distributed Privacy Federated Learning (FL) allows dispersed cameras and edge devices to train models collaboratively without centralizing raw video data, ensuring architectural privacy. Recent FL video surveillance deployments show viability, but many obstacles remain. Federated unsupervised video anomaly detection using Collaborative Learning of Anomalies with Privacy (CLAP) allows many surveillance systems to learn anomalous patterns without sharing video content (Al-lahham et al., 2024). Secure aggregation procedures give the system competitive anomaly detection performance over centralized baselines. FL systems struggle with non-IID (non-independent and identically distributed) data because multiple cameras capture varied scenes, lighting conditions, and activity patterns. Heterogeneity can hinder model convergence and affect client fairness. Federated video analytics still struggles with communication. Gradient compression, sparse updates, and model quantization can cut bandwidth by 35% at the cost of 3% accuracy loss (mentioned in more insights). Using selective encryption and compression, adaptive aggregation systems like PSSA reduce transmission costs by 40% and aggregation times by 25% while preserving accuracy.

Secure Computation and Homomorphic (HE) provides the best theoretical privacy guarantees. Recent research on privacy-preserving video analytics using CKKS (Cheon-Kim-Kim-Song) schemes has shown robustness to adversarial assaults even when 70% of servers are hostile. To achieve acceptable latency, practical deployment requires hardware acceleration, batching techniques, and parallel processing due to computational complexity.

Integration of HE with blockchain for federated video analytics has been proposed to guarantee privacy and integrity, however the cryptographic barrier is too high for real-time edge implementation. Current research employs selective encryption algorithms to apply HE solely to privacy-critical tasks and use lightweight encryption elsewhere. Other types of visual privacy prevention as shown in figure 2.

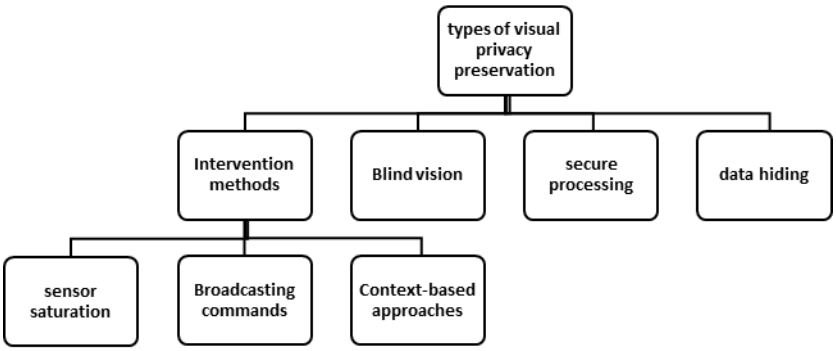

Fig. 2. Types of Privacy Preservation

3.1. Intervention Methods:

It is a technique that is added to interfere during data acquisition. That will prevent private visual data to be collected. Three different categories available, sensor saturation, broadcasting commands, and context-based approaches. Perez et al (Climent-Pérez et al., 2020)

3.1.1 Sensor Saturation:

This method is applied by feeding the sensor of the input device's with a signal of greater amplitude than the device can process and overwhelming or saturating sensors with excessive input signals to prevent them from capturing sensitive or identifiable information. We can see that in the work of (Zhu et al.) by proposing the LiShield method, which applies bright LEDs

producing a waveform not notable by the human eye, but the waveform will illuminate a scene and obscure any private information. (Zhu et al., 2017)

Harvey and Knight delineate anti-paparazzi apparatuses disguised as stylish clutch bags. The intervention device employs LEDs that provide a brightness sufficient to overexpose photographs captured by photographers. This will be engaged upon detection of the camera by an infrared light sensor, which identifies focusing lights. The camera activates this function by employing a light sensor in conjunction with an infrared sensor to detect autofocus illumination.

Table 1 - details of the reviewed papers

Author	Year	Method	aim	results	limitations	datasets
(Brkić et al., 2017).	2017	Improved GrabCut algorithm for pedestrian segmentation computer vision-based de-identification pipeline	de-identifies individuals in video sequences	The method reduces the face detection rate to 11% of the faces being detected in the de-identified images	The body shape is struggling issue, if two body shape are the same my cause re-identification. It need a huge database to avoid re-identification struggling with background subtraction false positive accrue when the clothing is same as background	Human3.6m : focuses on specific walking sequences from this dataset. CD-Net 2014 dataset for detection and segmentation. only the specific sequences that are relevant to their research
(Shifa, Imtiaz, et al., 2020)	2019	hybrid skin detection technique	selectively encrypting only sensitive skin regions	average encryption time is 8.268 s Encryption Space Ratio (ESR) is an average 7.25% for a high definition 91.5% for CTB-based segmentation 85.86%, HSV with 80.93% and YCbCr with an average 84.8%,	More encryption process time caused by false positive (non-skin pixels). Transforming into different color spaces produce complexity and inconsistent result with divers skin tons Consuming resources specially memory Detection accuracy could be impacted by environmental factors such as poor lighting the algorithm fails in detecting faces if been partially hided by mask or cellphones or half face is appeared in the frame which fail to capture the necessary coordinates for accurate skin detection	The study utilized part of the MOT17-11 dataset. Additional video sequences used in the experiments include "Miss-America," "Foreman," and "Paris."
(Chen et al., 2007)	2007	WPKLR A pseudo geometric model, edge motion history image (EMHI), is proposed for body obscuring.		88.89%.	The reliance on pairwise constraints that may be noisy or incorrect can introduce errors in the training phase. The user study indicated that while there was a high accuracy rate, there were still 20 errors out of 160 labeled constraints, which can negatively impact the learning process and the overall performance of the identification system	The study utilized a nursing home environment for video data. A pool of 102 silhouette images was used for user studies
(Climent-Pérez & Florez-Revuelta, 2021)	2021	mask R-CNN	Obscuring individuals in videos captured by RGB cameras for applications	87%	the dilation of the masks around detected individuals my obscure important contexed to be gathered. if the detection algorithms are not accurate, the dilation may inadvertently cover too much or too little, leading to either over-protection or	MuHAVi-MAS

					under-protection of sensitive information. This can result in revealing identifiable features or failing to obscure them adequately	
(Shifa, Asghar, et al., 2020)	2020	Histogram of Flows (HOF)Gaussian Mixture Model (GMM) for background modeling. MOG2 algorithm for robust background detection. Measurement Multiple Object Count (MOC) for human count accuracy.	five livel privacy prevention	detection rate of human faces 86.5%.Average human detection was 89.9%.Encryption encryption averaged 86.AES-OFB cipher with 128-bit key was implemented.	External factors like lighting conditions, camera angles, and the quality of the video feed may impact the effectiveness of feature of interest detection, accuracy may fail to detect sensitive area compromising the protection effort. In selective encryption important data may be unintentionally excluded from encryption leading to vulnerabilities in the system. Due to the complex architecture, limited processing power may struggle to process efficiently, leading to reduce performance in high demand situations.	Urban Tracker dataset PETS2009 dataset MOT17 dataset ABODA dataset
(Asres et al., 2026)	2024	lightweight adaptive anonymization for VAD (LA3D) yolo for segmentation	Video Anomaly Detection	Best Privacy Protection: Adaptive pixelization achieved ~40% cMAP and ~87% cMR reduction. Minimal Utility Loss: VAD models retained >95% AUC performance with adaptive methods. ReID Attack Mitigation: Adaptive anonymization reduced ReID accuracy by >90% in most scenarios.	with the heavier anonymization techniques may provide better privacy protection but can significantly degrade the utility of the images. the paper notes that skin color detection remains inadequately protected across various anonymization techniques, which could lead to potential privacy breaches	VISPR (10k training, 4k validation, 8k test0 . Market1501 for ReID evaluation UCF for VAD model training and testing (training: majority of videos, test: 290 videos).

3.1.2 Broadcasting commands

This is an alternative intervention strategy that safeguards sensitive information by transmitting commands across multiple communication protocols to disable input devices in the vicinity of the subject. Broadcasting orders are less efficacious than their physical counterparts, as user consent is requisite for these methods to function. Broadcasting directives are arguably less favored as intervention techniques compared to sensor saturation methods(Tran et al., 2023).

3.1.3 Context-based approaches

This method is considered with the context of the collected information. By triggering software action to decide whether the context to be protected or not. This method allows the users to control their digital information (Kapadia et al., 2007).

3.2 Blind vision

Allowing the processing of visual data without revealing sensitive information to any involved parties. This approach is particularly relevant in scenarios where privacy concerns are paramount, such as surveillance and data sharing. The method leverages secure multi-party computation to ensure that neither the data owner nor the algorithm provider gains unauthorized access to each other's sensitive information(J. Liu et al., 2021).

3.3 secure processing :

This technique focusses on protecting private information during process phase. It involves adding noise to sensitive areas to prevent leakage. This method is used to protect the algorithm used for privacy providing(Al-Obaidi et al., 2020).

3.4 Data Hiding

This approach conceals the original information within a changed object allowing retrieval solely by authorized users. Data hiding techniques encompass steganography, digital watermarking, and fingerprinting, which may be reversible or irreversible depending on the method employed.

Steganography employs a key to facilitate the retrieval of the concealed message. Digital watermarking encodes ownership information of an object through a visual pattern. Fingerprinting, on the other hand, conceals serial numbers that uniquely identify an object within an image, enabling the copyright owner to identify breaches of license agreements. (Petitcolas et al., 1999)

4. Visual Obfuscation Technique

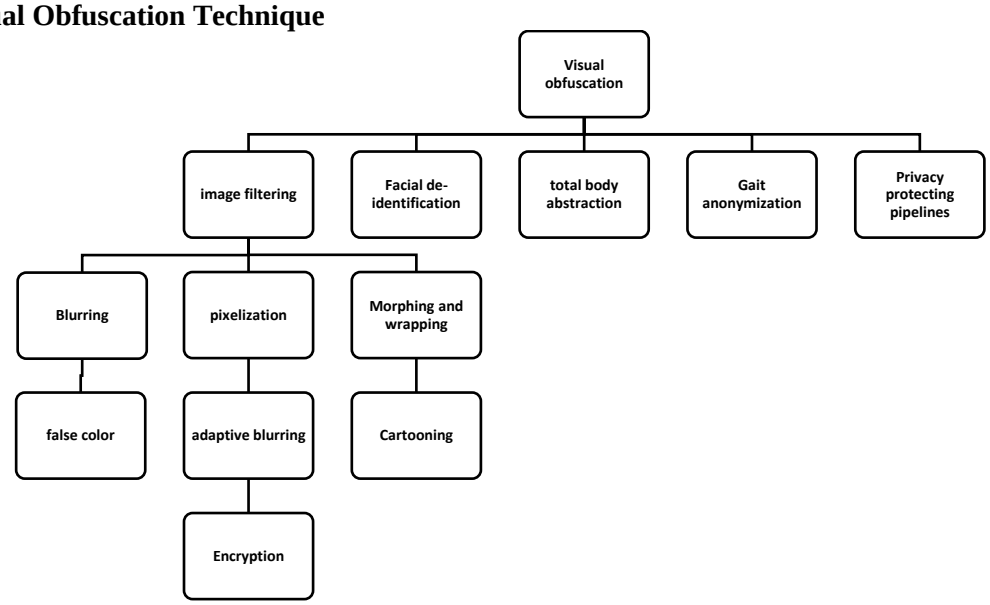

Fig. 3. Visual Obfuscation Technique

4.1 Image Filtering

Image filtering is a perceptual obfuscation technique that aims to protect privacy by altering visual objects. This technique can be applied to the entire image or video frame, or selectively to regions detected as containing sensitive or private information.

4.1.1 Blurring

It is a common method used in visual privacy prevention to obscure sensitive information. Making it imperceptible to human viewers while still allowing machines to extract necessary data. It also has limitations, such as the effectiveness of blurring, which can be influenced by various factors, such as the viewer's familiarity with the subject and the context in which the image is

viewed (Sawarkar & Sawarkar, 2024).

4.1.2 Pixelation

This method's advantage is its simplicity of application and its ability to provide a specific type of noise. It increases the dimensions of the chosen pixels in the selected area, which provides very low visibility on the sensitive private information, making it unrecognizable. So, it is suitable for moving objects and public locations. However, the body may appear bulky and sometimes unsuitable for small objects (Zhou & Pun, 2021).

4.1.3 Morphing and Wrapping

This technique is employed to provide privacy by altering identifiable information in the frame while retaining the natural appearance. In morphing, altering the body or face with other individuals features, generating a hybrid identity obscuring the sensitive details, hiding biometric identifiable information. By this a biometric recognition system will be disrupted. Wrapping works on region of interest altering biometric details (e.g., smile lines, stretching facial contours or bending libs). This method characrized by balancing privacy prevention with perceptual plausibility. On of the challenges of this method is to avoid over distortion that may lead to reveal identity of the individual (Deshmukh et al., 2018; Korshunov & Ebrahimi, 2013).

4.1.4 False Color

This is a retrieval method original frame to be restored if necessary. It works on the entire frame do not require region of interest, altering the color palette by applying color transformations rendering sensitive information unrecognizable while maintaining the overall structure. It widely used and particularly in video surveillance(Çiftçi et al., 2015).

4.1.5 Adaptive Blurring

The adaptive blurring approach employs Deeplab to produce segmentation masks and a scale-dependent Gaussian blur. The technique employs a bespoke symmetry-oriented approach to direct the application of Gaussian blur on object edges. These methods seek to conceal recognizable characteristics in photographs or videos while preserving a balance between privacy and functionality. The adaptive characteristics of these filters enable them to modify the degree of blurring according to particular factors, including the image resolution or the context of data utilization.

It possesses inherent limitations, as it fails to include camera distortion or depth ambiguity, which may result in insufficient or excessive blurring. Moreover, commercial tools has the capability to deblur obscured photos, so compromising the security of the pipeline (Ahn & Jang, 2021).

4.1.6 Cartooning

Cartooning enhances video privacy by substituting objects with clip art representations and abstracting background areas. This modification aids in concealing sensitive information while preserving the semantic integrity of the video. The cartooning process encompasses a synthesis of computer vision, deep learning, and image processing methodologies. These are employed to identify things, abstract details, and substitute them with cartoon clip art (Erdelyi et al., 2014).

4.1.7 Encryption

Advanced encrypting algorithm such as AES or RSA or any other type of encryption method is used to hide identity of individual. It provides to encrypt the entire frame or region of interest (e.g., face, body, license palate, etc.). these secure hidden and sensitive information only authorized users with the right key can reveal the actual content [22]. Encryption suffers from the complexity of applying it. Rather it compromises the privacy if the key is known by an unauthorized user. Many studies proposed methods of providing privacy with this technique. For instance (Mr. Nandish and Abdul Rahaman Pasha, (M, 2024)in there study work on providing selective encryption privacy prevention for a specific part of the video such as faces. (A. T.

Maolood, K. Gbashi, S. Mahmood, and C. Const, (Maolood et al., 2022) proposed a lightweight encryption method, like those using ChaCha20 stream cipher maps are designed to provide a high security by encrypting the entire frame which lowers the computational overhead.

4.2 Facial de-identification

Is the process of hiding person identity by concealing the face or even replacing it with another individual face of a cartoon figure, to protect the privacy of individuals so many techniques are proposed in the research. In Xie, et al.(Xie et al., 2017) survey gathered state of the art de-identification method, categorizing it into three levels : pixel-level, representation level, and semantic-level techniques. with recent advancements focusing on deep learning models such as Generative Adversarial Networks (GANs) and diffusion models for improved privacy and utility balance. They proposed a figure that gather the methods of facial de-identification as shown in the figure 4.

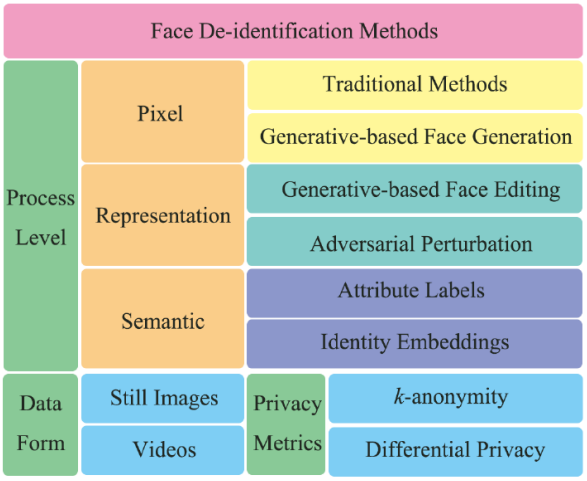

Fig. 4. Taxonomy of face de-identification methods.

4.3 Total Body Abstraction

Another privacy-preserving approach is to conceal the entire body of an individual while preserving the remaining scene information for subsequent processing and analysis. High-quality body abstraction typically relies on deep learning-based segmentation techniques. Various approaches, including semantic segmentation, have been proposed to accurately extract human subjects from video frames and replace the segmented regions with avatars, silhouettes, or other privacy-preserving representations. Such methods are widely employed in surveillance systems and active and assisted living applications. Several studies illustrating these approaches are reviewed in Section 2.

4.4 Privacy Protecting Pipelines

Another way of providing privacy by building end-to-end pipeline making it convenient with the method. Aim to provide privacy by combining various techniques in visual privacy prevention. These system typicaly implement models uses deep learning for detection and obscuring sensitive regions in the frame. One of the notable research Z. Qi, A. Maungmaung, and H. Kiya they propose a method of privacy-preserving image classification method using blockwise scrambled images and a modified ConvMixer. The encryption used is a block-wise scrambling technique. Which divide the image into blocks and scrumble it to obscure the original image. This study provide a classification of images without revealing the original content of the image.

4.5 Threats and Robustness Issues

Advanced adversaries can compromise anonymization in privacy-preserving systems. Critical research frontiers include understanding these risks and establishing strong defenses.

4.5.1 Reidentification Threats

Modern person re-identification models are powerful de-identification enemies. Recent research trains ReID models on anonymised data and measures residual re-identifiability to evaluate anonymization strategies. FIVA showed that adversaries can retrieve IDs from naive anonymization methods(Rosberg et al., 2023). FIVA uses noise injection and parameter perturbation to account for reconstruction opponents during training to protect against such attacks.

4.5.2 Split Learning Flaws

Gradient-based reconstruction attacks can target split learning architectures, which split model computation between edge devices and servers to avoid communicating raw data. Inversion can rebuild input images with surprising fidelity for adversaries with intermediate activations or gradients (Fan et al., 2023). This weakness compromises split architecture privacy claims and requires gradient noise injection, safe aggregation, or homomorphic encryption of intermediate representations.

4.5.3 Adversarial Changes and Evasion

Active adversaries might create perturbations to avoid detection or anonymization in addition to passive attacks on private data. Recent research on robust deep learning models shows that semantic-preserving adversarial attacks can mislead privacy-preserving systems while maintaining visual naturalness (Zhao et al., 2023). Adversarial training with various attack models, ensemble detection systems, and verified robustness procedures are defenses, but they add computational overhead and reduce value.

4.5.4 Privacy-Robustness Tradeoff

Recent academics' information-theoretic approach formalizes an adversarial robustness, task utility, and attribute privacy tradeoff. This theoretical result suggests that representation capacity and leakage assurances are limited by robustness and privacy. Practical systems must carefully balance robustness and privacy, and selective de-memorization, which targets high-risk samples for privacy protection rather than uniform noise, has shown promise in balancing robustness and privacy better than blanket differential privacy approaches.

5. Methods of Object Detection

Localizing objects is widely used in many areas such as face detection in video surveillance, autonomous driving etc. many techniques are proposed for object detection, basically in traditional detection methods goes in three steps. First targeting objects by scanning the input image or frame with sliding windows. Secondly using extraction method to extract some features. Then conducting a classifier using the selected features producing classifications. Traditional methods suffers from specifying the number of sliding windows.

The CNN-based object detection can be mainly divided into two major parts: two-stage object detection and one-stage object detection.The two-stage models mainly use region of interest(RoI) to generate candidate bounding boxes and then extract features from bounding boxes to find objects. In comparison, the one-stage methods do not use RoI mechanism to locate objects, both steps are combined in just one stage(M. and T. H. Y. 2024 Al-Tamimi, 2024).

5.1 Two-stage object detection

5.1.1 R-CNN (Region-based Convolutional Neural Network)

R-CNN, a two-stage object identification technique, employs the AlexNet model for automatic feature extraction, enabling the abstraction of information from pictures and frames, hence enhancing detection accuracy. Extracting and classifying the item solely enables the system to learn more effectively. Despite its greater complexity compared to other contemporary algorithms, it demonstrates a lower error rate than standard CNNs.[30] It comprises four primary

parts. The initial module suggests a candidate region utilizing the selective search method to pinpoint detection zones that encompass the target within each region. The second module has five convolutional layers and two fully connected layers, utilized to extract a 4096-dimensional feature vector from each region [31]. The third module classifies the extracted features using Support Vector Machine (SVM) to ascertain whether the target is within the selected region and to which target it belongs. Ensuring the quality of the chosen region. The less significant features are taken out by non-maximum suppression (NMS).(Shepley et al., 2023)

5.1.2 SPP-net (spatial pyramid pooling layer)

Convolutional Neural Networks (CNNs) comprise convolutional layers and fully linked layers that necessitate fixed-size inputs(Jassim et al., 2025). R-CNN frequently crops the input to accommodate the dimensions of the fully linked layer. SPP has eliminated the requirement for fixed input dimensions in the fully connected layer by integrating spatial pyramid matching (SPM) with CNN. The SPP divides the input into many scales, ranging from finer to coarser levels, and subsequently consolidates local information into higher-level representations. This technique enhances the adaptability and efficiency in object detection. SPP possesses the ability to generate output of a predetermined length, regardless of the input size. The architecture comprises a sequence of convolutional layers, spatial pyramid pooling (SPP) layers, and fully linked layers. (He et al., 2014)

5.1.3 Fast R-CNN

While SPP improves accuracy and efficiency compared to R-CNN, certain constraints still require refinement. The SPP-Net operates within the conventional R-CNN framework, encompassing classifier training, feature extraction, network fine-tuning, and bounding box regression. Nonetheless, the inability to adjust the weights of convolutional layers imposes considerable constraints on deep networks, leading to diminished accuracy. Fast R-CNN initially recovers pictures or frames via convolutional layers, after which the feature vectors, along with the bounding boxes, are transmitted to the RoI pooling layer to acquire fixed-size features. Each feature vector is entered into the bounding box regressor and classifier. (Girshick, 2015)

5.1.4 Faster R-CNN

The Region Proposal Network (RPN), proposed by Ren et al. in 2015, is a fully convolutional neural network designed to produce candidate target regions. This evolved into Faster R-CNN. By disseminating the convolutional features to the detection network. The RPN reduces the cost of region proposals, addressing the challenges faced by previous methods that created candidate boxes and region proposals through sliding windows, which resulted in issues related to cost and accuracy. Faster R-CNN addresses these difficulties by sharing convolutional features with the detection network, hence minimizing costs to nearly negligible levels. RPN operates as a Fully Convolutional Network (FCN), allowing it to concurrently anticipate objects and their scores. RPN accepts inputs of arbitrary sizes, generating various proposals for rectangular objects. This design generates region proposals efficiently. Despite these enhancements, it faces hurdles, including the necessity of time to refine the anticipated region proposals prior to estimating the bounding box offsets(Zhong et al., 2019).

5.1.5 Feature pyramid network (FPN)

As the previous method has improved detection accuracy and speed leveraging efficiency and more robust in object detection. When the dataset contains small scale or multiscale features the accuracy of detection my minimized. To solve this issue FPN is proposed with top-down and bottom up pathways, the top-down pathway up samples deeper features recovering high resolution, while Bottom up pathways processes the input object with a successive layers extracting spatial coarse semantically high level semantic features generating feature maps in multiple scales(Z. Li et al., 2023).

5.2 One-Stage Object Detection

5.2.1 YOLO (You Only Look Once)

As shown in the previous methods how detection accuracy and computational performance has been enhanced, real time application needed speed and timeline as priority request. YOLO is proposed to solve this issue; it is widely used for real-time detection scenarios. YOLO structure is based on GoogleNet (X. Huang et al., 2021) and enhanced it by modifying the general CNN network of GoogleNet, YOLO contains 24 convolutional layers followed by 2 fully connected layers, and 1 × 1reduction layer for inception, a 3 × 3 convolutional layers. Which allow to predict location of bounding boxes directly categories it, This is proposed with the first version of YOLO in 2015. many versions are proposed in the following years in each version enhances the previous one. The initial versions introduce single neural network that predicts bounding boxes and class predicting directly from full image or video fram. YOLOv3 has an improvement by using complex architecture, enhancing feature extraction. The following versions(Mulajkar & Yede, 2024) YOLOv4 to YOLOv6 has focused on better performance on constrained devices by optimizing the network, which improved the balance between speed and accuracy (Hussain, 2024)(M. S. H. Al-Tamimi et al., 2023). YOLOv7 and YOLOv8 these versions are designed to meet the demands of modern application, while continuing enhancing detection capabilities, incorporating transformers and Neural Architecture Search(NAS).researches continue to evolve for more robustness and efficiency, while more advancement made on the architecture more complex become consuming resources and high computation, also improving adaptability and generalization across diverse environments(Hussain, 2023) (Casas et al., 2023). in the survey of (Raees & Al-Tamimi, 2024) they conduct a table of pros and cons of each version (from v1 to v8). In there recommend to use YOLO for object detection and tracking along with blurring technique to provide privacy

4.2.2 Single Short Detector (SSD)

In parallel with YOLO series SSD is utilized for predicting multiple categories in real-time object detection scenarios. The SSD defines a set of anchor boxes for varying ratios and scales of objects in the frame, discretizing the bounding box output spaces fusing multiple features with different resolutions so it is more adaptable for detecting objects in different scales or ratios. The SSD architecture include VGG16 network as a backbone utilizing feature extraction with fully convolutional layers improving feature mapping process to capture features on input layers enabling detection of different scales object detection, which by this point SSD have the advancement over YOLOv1 of detecting different scale objects. This method primary limitation is detecting small objects in the complex background that obscure them. It also struggles with limited resource and computational power devices which is needed for achieving balance between speed and accuracy.

Table 2 - YOLO versions

Version	Backbone	Accuracy	features	Limitations and challenges
YOLOv1 2015 (Redmon et al., 2016)	Darknet-23	45 FPS 63.4 AP_{50} On COCO dataset, input resolution 448x448	single-pass detection 24 convolutional Layers stacked sequentially Unified architecture	This version suffers from the detection of small objects specially when the object near to each other or overlapped, this is caused due to assigning multiple object to the same grid cell. For training dataset need to be well annotated and balanced which these datasets are difficult to compile this lead to my cause generalization issue. These challenges lead to high computational needs (Ragab et al., 2024)
YOLOv2 2016 (R. Li & Yang, 2018)	Darknet-19	67 FPS Accuracy 76.8 AP_{50} on COCO dataset input resolution 416x416	Improved accuracy Anchor boxes And multi-scale training for stability. 19 convolutional layers without using batch normalization (BN) Multi-scale training	Using Non-Maximum Suppression (NMS) process, which eliminate redundant bounding boxes this reflects negatively on speed and accuracy. It also struggle from detecting small objects or overlapped and complex background. the complexity of architecture consumes memory which make it less sutable for limited computational power and memory platforms.(Baoyuan et al., 2021; Xu & Wu, 2020)

YOLOv3 2018 (Kim et al., 2018)	ResNet-50 & ResNet-101	30–45 FPS Accuracy 57.9 AP_{50} on COCO dataset input resolution 416x416 /608x608	Better small-object detection Multi-scale predictions for small objects 19 convolutional layers without using batch normalization (BN) predictions with 3 heads	Differences in accuracy when significant size differences between objects in the bounding box, also difficulties to detect object in complex environments. The enhancement of accuracy effects the computational demands leads to difficulties to balance between speed and accuracy.(Baoyuan et al., 2021)
YOLOv4 2020 (Baoyuan et al., 2021)	CSPDarknet53	62 FPS Accuracy 43.5 mAP on COCO dataset in resolution 608x608	Mosaic augmentation + CIoU loss for robust tracking 53 convolution layers. With batch normalization (BN)	Same as te previous versions it also struggles from detecting small or overlapped objects or in complex environments. Although it has a complex architecture this requires high computational cost and consume large memory and need longer training time. Due to these challenges, it requires powerful hardware, so it is not suitable for small or limited platforms (Erlina & Fikri, 2023) (Yu & Zhang, 2021)
YOLOv5 2020 (Sukkar et al., 2021)	EfficientNet	140 FPS Accuracy 50.7 mAP on COCO dataset in resolution 640x640	Improving speed PyTorch optimizations. Varies (53–155) layer AutoAnchor and mosaic augmentation. Ease of deployment.	Small and occluded objects detection still remains a challenge due to spatial resolution constraints. Although this version is characterized by faster training time with the previous versions, it requires high memory and processing equipment's. the performance vary significantly when it is applied to different environments dataset from those trained to. (T.-H. Wu et al., 2021)
YOLOv6 2022 (C. Li et al., 2022)	EfficientNet-L2	520 FPS Accuracy 52.8 mAP on COCO dataset in resolution 640x640	Proximally 498 convolutional layers due to the environment Anchor-free and RepVGG backbone	Architecture complexity of this version needs more power and more resources to balance speed and accuracy. With edge and small platforms can decrease detection speed. it also struggle in detection of small and variable scale objects that appear in diverse distances affecting efficacy.
YOLOv7 2022 (C.-Y. Wang et al., 2023)	ResNet-ACmix	161 FPS Accuracy 56.8 mAP on COCO dataset in resolution 640x640	-With variety of proximate 150 layer. -Proposed state of the art speed accuracy -Dynamic label assignment	While introducing improvement in speed and detection accuracy, architecture are more complexed so it consume more recourses, it also struggle generalizing in highly dynamic scenarios as scene with blur and motion objects and detection of occluded objects. With the high resource requirement make it less suitable for edge devices.
YOLOv8 2023 (Huangfu & Li, 2023)	CSPDarknet53	123 FPS Accuracy 53.9 mAP on COCO dataset in resolution 640x640	-Anchor free with improved data augmentation	Significant improvement has shoen in real-time applications detection objects the challenge of consuming recurses are still in the field particularly in lower-end hardware , which in these plat forms my decrease accuracy efficacy. Varying resolutions and different environments still a challenge in this version
YOLOv9 2024 (C.-Y. Wang et al., 2024)	GELAN **(**Generalized ELAN)	76 FPS Accuracy 56.8 mAP on COCO dataset in resolution of 640x640	Programmable Gradient Information PGI GELAN for feature retention in deep networks	Although the improvement of accuracy and speed with its predecessors, detecting small or overlapped and occluded object are still a struggling issue specially in adverse weather conditions, the overall accuracy my drop in weather like snowing. This model also rely on NMS for post-processing which my cause increasing inference latency and computational redundancy. For improving precision the use of data augmentation my enhance detection accuracy and robustness, with a carful selecting and tuning of these techniques
YOLOv10 2024	Enhanced GELAN	257 FPS Accuracy 56.8 mAP on COCO	-Rank-based label assignment	Although this model has not been adopted by wide researches, some of the limitation are noted. Relying on NMS (a novel model) using dual label

(A. Wang et al., 2024)		dataset in resolution of 640x640	-natively anchor-free design	assignments which can impact inference latency which is a crucial for real-time applications. This model suffers from computational redundancy.
YOLOv11 2024 (Sapkota et al., n.d.)	C3k2 along with Spatial Pyramid Pooling Fast (SPPF) and Cross Stage Partial with Spatial Attention (C2PSA)	290FPS Accuracy : 76.8% mAP@0.5. 68.1% mAP@0.75. 48.5% mAP@[0.5:0.95] . In resolution randomly scaled between 320 × 320 and 640 × 640	Enhancing feature extraction capabilities with the use of C3k2 block better handling of spatial information for object detection and segmentation.	Small, occluded and rotated object still a struggle to detect in this version as in its predecessors. Overfitting is a potential to happen particularly if trained on limited datasets.

5.2.3 RetinaNet

Is one stage detector method which has advantages such as speed and accuracy. Although it considered slower from some two stage methods. To overcome the multi scale object detection problem FPN is included(W. Huang et al., 2021), with the full steps, top-down and bottom-up pathway, of FPN. Limitations and challenges face RetinaNet such as fusion of features from different stages especially from complex scenes, leading to insufficient extraction reducing detection accuracy. small object detection still a challenge in particular in aerial and remote sensing applications due to low resolution and weal representative features. Poor localization of objects is noted due to the inaccurate bounding box regression.(Ahmed et al., 2022; Q. Wei et al., 2022)

Famous Datasets

Evaluation Challenges and Benchmark Datasets

Datasets with ground-truth identity labels, diverse environmental conditions, and standardized evaluation protocols are needed to evaluate privacy-preserving video systems. Present benchmarks have significant limitations that prevent reproducible and realistic assessment.

6.1 Reidentification People Datasets Market-1501, DukeMTMC-ReID, and MARS, the most popular benchmarks, were designed for person re-identification rather than privacy evaluation, leaving gaps in their privacy-relevant scenarios

Market-1501 contains 32,668 images of 1,501 identities from six cameras. It provides a large-scale benchmark with standardized evaluation protocols, but its image-based nature limits temporal consistency in video anonymization, and its relatively controlled capture conditions do not reflect real-world deployments' lighting variability, occlusion patterns, and environmental diversity (Karanam et al., 2019) .DukeMTMC-ReID uses eight cameras to re-identify 1,404 people with 36,411 images. Recent research has found that model performance varies greatly across cameras in this dataset due to differences in viewing angle, illumination, and camera quality. This imbalance invalidates aggregate metrics and necessitates camera-aware evaluation protocols. Instead of static images, MARS (Motion Analysis and Re-identification Set) provides video tracklets for temporal modeling and frame consistency evaluation. The dataset has 1,191 identities and 20,715 tracklets from six cameras. It lacks long-term appearance variation (clothing changes, seasonal effects) and diverse occlusion scenarios found in real surveillance deployments. The recently proposed CHIRLA dataset addresses these limitations by increasing temporal coverage and environmental diversity (Dominguez-Dager et al., 2026).

6.2 Action Recognition and Privacy-Annotated Datasets:

Post-hoc privacy annotation has been added to action recognition datasets like UCF-101 and HMDB-51 for privacy research. STPrivacy added five privacy attribute labels (face visibility, gender, nudity, relationship, skin color) to these datasets to evaluate privacy-utility tradeoffs in

action recognition tasks. PA-HMDB51 annotated HMDB-51 with privacy attributes and adversarial evaluation protocols to test anonymization against re-identification. These datasets were not designed for privacy evaluation, resulting in several limitations: Privacy attribute annotations are added post-hoc rather than part of the original collection protocol, identity diversity is limited compared to dedicated re-identification datasets, temporal dynamics and occlusion patterns do not reflect surveillance scenarios, and anonymization strength is not measured using standardized adversarial evaluation protocols.

6.3 Simulate environments and synthetic datasets

Researchers propose large-scale simulation engine-generated synthetic datasets to address privacy concerns in collecting and distributing real surveillance data. These datasets are statistically reliable and eliminate real-world privacy risks, but rendering fidelity is needed to capture authentic lighting, occlusion, and appearance variations. This tradeoff between synthetic realism and real-world validity needs further study.

This section presents a list of datasets utilized for assessing the effectiveness of privacy preservation methods, accompanied by details of their composition in Table 3.

Table 3 - datasets details

Author	Dataset name	Method	indoor or outdoor	description	limitations
(Gevers & Smeulders, 2016)	MARS	The paper employs Convolutional Neural Networks (CNN) for re-identification. It utilizes ID-discriminative Embedding (IDE) for model training. Motion features like HOG3D and gait are evaluated. Max pooling is used to combine tracklet features. The dataset includes distractor tracklets for realistic evaluation.	outdoor	MARS, an expansion of Market-1501, is a large-scale video-based person reidentification dataset. It came from six near-synchronized cameras. It has 1,261 pedestrians captured by at least 2 cameras. Due to pedestrian stances, colors, and illuminations, as well as poor image quality, matching accuracy is low. To add realism, the dataset has 3,248 distractors. Deformable Part Model with GMMCP tracker generated 25-50-frame tracklets automatically.	Motion features are less effective in complex backgrounds. Limited effectiveness of motion features on large datasets. Variations in viewpoint and pose reduce accuracy. False detection and tracking results create distractor tracklets
(T. Wang et al., 2016)	iLIDS-VID	Per-Camera Tracklet Discrimination Learning and Cross-Camera Tracklet Association Learning are two significant improvements in the Tracklet Association Unsupervised Deep Learning (TAUDL) system proposed in the study. The Per-Camera Tracklet Discrimination Learning maximizes local within-camera tracklet label discrimination, whereas the Cross-Camera Association Learning maximizes global association. To eliminate tracklet labeling ID duplication, Sparse Space-Time Tracklet (SSTT) sampling and label assignment is presented. Also, the Cross-Camera Tracklet Association (CCTA) loss function aligns cross-view tracklet feature	outdoor	The person re-identification dataset iLIDS-VID includes 300 pedestrians from two disjunct camera perspectives in public area. It has 600 image sequences of 300 people, each with two camera viewpoints. Image sequences average 73 frames and range from 23 to 192. Personal clothing similarities, illumination and viewpoint differences across camera views, crowded backgrounds, and unpredictable occlusions make the iLIDS-VID dataset difficult.	The paper highlights that existing person re-identification (re-id) methods often rely on supervised learning with manually labeled pairwise training data, which limits scalability in practical deployments due to the exhaustive identity labeling required for every camera pair. It notes that classical unsupervised learning models perform significantly weaker than supervised models because they lack cross-view pairwise ID labeled data, which is crucial for learning context-aware ID discriminative information. Additionally, the paper

		distributions to improve model learning.			acknowledges challenges in optimizing cross-camera tracklet association without camera pairwise ID labels, making the learning process more complex
(Y. Wu et al., 2018)	DukeMTMC-VideoReID	The paper proposes a one-shot learning approach for video-based person re-identification (re-ID) that utilizes unlabeled tracklets through a stepwise learning method. It employs a progressive sampling strategy to select pseudo-labeled candidates from unlabeled data, enhancing the model's robustness. The methodology includes two iterative steps: sampling reliable pseudo labels and updating the Convolutional Neural Network (CNN) model accordingly. The approach contrasts with existing static sampling strategies, aiming to improve label estimation and overall performance in one-shot settings.	outdoor	DukeMTMC-VideoReID is a subset of DukeMTMC for video-based person re-ID. High-resolution movies from 8 cameras comprise the collection. Hand-drawn bounding boxes crop photos in one of the largest pedestrian video collections. The dataset comprises 4832 tracklets with 1812 IDs and average 168 frames per tracklet.	The paper highlights that the initial labeled data in the one-shot setting are too few to accurately depict the underlying distribution, which limits the model's performance. It notes that incorporating excessive not-yet-reliable pseudo-labeled data during the initial iterations can hinder subsequent improvements in the model. The reliability of pseudo labels is challenged by the accumulation of estimation errors during iterations, which can prevent robust model training. Additionally, the static sampling strategy used in previous works is deemed inappropriate, as it may lead to the selection of unreliable predictions.
(Cho & Yoon, 2017)	PRID 2011	The paper proposes a novel framework called Pose-aware Multi-shot Matching (PaMM) for person re-identification, which utilizes pose information for efficient multi-shot matching. It incorporates various feature extraction methods, including Histogram of Oriented Gradient (HoG), dcolorSIFT, and LOMO, to enhance re-identification performance. The methodology also employs metric learning techniques such as KISSME, ITML, and LMNN to improve matching accuracy. Additionally, the paper evaluates the performance of PaMM against other single-shot and multi-shot matching methods, demonstrating its superior effectiveness.	outdoor	A person reidentification dataset, PRID 2011, contains numerous human trajectories from two static surveillance cameras monitoring crosswalks and sidewalks. The dataset has a clean background and few obscured persons. The dataset shows 200 persons in both views. 178 of 200 have appeared more than 20 times.	The paper notes that reidentifying people is problematic due to large appearance changes between camera views and positions. Existing approaches struggle to address ambiguities induced by viewpoint and posture variations in real-world surveillance scenarios. Although multi-shot matching algorithms have been presented, they still face challenges such as different looks and the need for accurate posture estimation. The article highlights that stationary persons may complicate angle estimation and reidentification.

(Song et al., 2017)	Wild (LPW)	The paper introduces the Region-based Quality Estimation Network (RQEN), which utilizes a fully convolutional network to generate middle representations of input images. It employs a landmark detector to identify key points on the human body, allowing for a division of images into upper, middle, and lower regions. The network includes a region-based quality predictor that estimates the quality of regional features, facilitating the aggregation of complementary information across frames. The methodology also incorporates joint end-to-end training to evaluate the validity of information from different image regions.	outdoor	Two–four cameras per annotated identification detect 2,731 individuals in three settings with Labelled Pedestrian in the Wild (LPW). Clean 7,694 tracklets with 590,000 photographs make up the LPW. Large scale, cleanliness, automatically detected boundary boxes, more congested scenes, and a wider age spectrum distinguish it from previous datasets. This dataset makes testing more complex algorithms more realistic and difficult. .	The paper highlights that existing person re-identification datasets often suffer from issues related to scale and cleanliness, with sizes ranging from 200 to 1500 identities, which is insufficient for extensive experiments. It notes that while larger datasets exist, they frequently contain poor quality data due to detection or tracking failures, undermining algorithm performance. Additionally, the paper points out that most datasets rely on manually aligned bounding boxes, which may not reflect real-world scenarios where misalignment or missing parts can occur.
(L. Wei et al., 2017)	MSMT17	The research proposes a Person Transfer Generative Adversarial Network (PTGAN) to bridge the domain gap between ReID datasets. To preserve individual identities while matching target dataset backdrops and lighting, PTGAN transforms styles. Using a huge dataset termed MSMT17, which provides lighting, scene, and pose problems, PTGAN is tested to reduce the domain gap. The research compares DukeMTMC-reID, Market-1501, and CUHK03 datasets.	both	Multi-scene, multi-time person re-identification dataset MSMT17. It has 180 hours of video from 12 outdoor and 3 indoor cameras over 12 time intervals. The movies have 4,101 tagged IDs and 126,441 bounding boxes throughout time and complicated lighting. .	The largest person ReID collection has 8 cameras and less than 2,000 identities, which is insufficient for real surveillance applications. Most datasets represent solitary interior or outdoor scenarios, missing real-world complexity. Current datasets include short-time surveillance footage without major lighting fluctuations, simplifying ReID. These datasets have inaccurate bounding boxes due to expensive hand-drawing or outdated detection algorithms. .
(Arkushin et al., 2024)	42Street	new method GEFF for face feature extraction the proposed model combines face and ReID modules for improved performance.	indoor 42 steer theatre	For instance, 42Street is based on a theatre play. The 42Street stage production [42street] is recorded publicly for the dataset. The play is 1.5 hours long and divided into 5 20-minute segments with costume changes.	GEFF may introduce only slight improvements in certain scenarios with limited query samples. The 42Street dataset lacks a training set and does not conform to standard settings
(Y. Liu et al., 2019)	The iQIYI-VID	Head detection using YOLO V2 for clip validation Video-level feature transformation using NetVLAD or Average Pooling	both types of environments are represented in the	The iQIYI-VID collection contains clips from variety shows, films, and dramas. The dataset includes 500,000 celebrity videos from 5,000.	Traditional methods focus on single modal information only. Face features are more reliable than head features.

			iQIYI-VID datase	Each video lasts 1-30 seconds.	Actors frequently change hairstyles, complicating identification
(Xiao et al., 2017)	CUKL-SYSY	A new deep learning framework is proposed. Joint handling of detection and identification in a single CNN. Online Instance Matching (OIM) loss function for training. Use of a stem CNN for feature extraction. Pedestrian proposal network predicts bounding boxes. RoI-Pooling layer for feature extraction from proposals. Multi-task learning for joint training of networks.	both	CUKL-SYSY is a large person search benchmark with 18,184 photos and 8,432 identities. Instead of matching query persons with manually cropped pedestrians, this dataset searches people from entire gallery photographs, closer to real application scenarios.	Existing methods focus on cropped pedestrian images only. Performance limited by handcrafted features in earlier approaches. Sliding window framework is not scalable. Learning large classifier matrix is difficult. False alarms and misalignments in pedestrian proposals.
(Mirzazadeh et al., 2023)	MEVID	The research compares state-of-the-art video person ReID algorithms such CAL, AGRL, BiCnet-TKS, TCLNet, PSTA, PiT, STMN, Attn-CL, and AP3D utilizing mean Average Precision (mAP) and Cumulative Matching Characteristic (CMC). A semi-automatic annotation system and GUI were created to simplify annotation for the MEVID dataset, which uses real-time models for object identification, posture estimation, person ReID, and multi-object tracking. The paper addresses video person ReID's wardrobe, scale, and location issues. .	both	MEVID is a huge wild-video person re-identification (ReID) collection. It spans a big indoor and outdoor area nine times in 73 days, with many camera angles and costume changes. It distinguishes 158 unique people wearing 598 apparel from 8, 092 tracklets, average length 590 frames, visible in 33 camera views from the massive MEVA human activities dataset. .	The research highlights the change-of-clothing problem as a major obstacle in real-world applications, where previous approaches achieve only 5 accuracy in complicated environments. Video person ReID methods function better when the same outfit is worn, but location and scale changes make the results unsatisfactory, with the best scores being 39.0 mAP and 56.6 top-1. The ReID problem is complicated by more clothes per person, which increase intra-class variations and lower top method performance. .
(Zheng et al., 2016)	PRW	The research uses cascaded fine-tuning to improve re-identification (re-ID) accuracy by training a detection model before the classification model. It improves re-ID performance with a Confidence Weighted Similarity (CWS) metric that incorporates detection scores. Modern image descriptors and metric learning approaches including Bag-of-Words (BoW), LOMO, gBiCov, and IDE are tested. The paper compares DPM, ACF, and LDCF pedestrian detection techniques with their RCNN equivalents.	outdoor	PRW is a large dataset for raw video frame pedestrian detection and human recognition. PRW tests Wild Person Re-identification using six synchronized cameras. At least 932 IDs and 11,816 frames have pedestrian bounding box identities. .	it implies that existing datasets lack thorough annotations for combined evaluation of person detection and re-identification (re-ID) systems, a major gap in the area. Additionally, typical datasets often lack ID annotations from several cameras, limiting their real-world usefulness. While pedestrian detection models perform well on benchmark datasets, the paper shows that their effectiveness for person

		.			re-ID tasks is understudied. .
(G. Wang et al., 2018)	Market-1501	The research offers a two-stream spatial-temporal person Re-identification (st-ReID) paradigm that blends visual semantic and spatial-temporal data. It unifies heterogeneous data using a shared similarity metric and Logistic Smoothing (LS). Rapid Histogram-Parzen (HP) approximation of the spatial-temporal probability distribution is devised. It uses a visual feature stream, spatial-temporal stream, and joint metric sub-module in a two-stream architecture. The st-ReID model reduces gallery database size by removing irrelevant photos. .	outdoor	Public person re-identification benchmark Market-1501 is big. Six cameras recorded 1501 IDs, and the Deformable Part Models pedestrian detector created 32,668 pedestrian picture bounding boxes. Persons average 3.6 photographs per viewpoint. The dataset comprises 751 testing identities and 750 training IDs. 3,368 query photos are compared to 19,732 reference gallery images using the approved testing method.	The paper highlights that conventional person re-identification (ReID) methods struggle with appearance ambiguity across different camera views, particularly when the gallery database is large. This leads to poor performance in identifying individuals. It notes that the spatial-temporal probability is unreliable due to uncertain walking trajectories and velocities, which can result in low recall rates. Additionally, the paper mentions that existing models often make strong assumptions for simplification, which may not effectively address the complexities of real-world scenarios.

Development of privacy prevention technology:

Maintaining privacy for individuals in real-time environments is essential for protecting sensitive personal information, necessitating ongoing research. The progress of neural networks designed to enhance individual privacy include concealing the human body, masking face features, or substituting significant objects using privacy techniques. Training neural networks with diverse datasets in complicated contexts involves discovering and tracking weaknesses in the networks. This represents a critical field of research aimed at enhancing the accuracy and efficiency of recognizing and concealing significant regions. Despite the substantial importance of neural networks and machine learning in identifying and detecting items inside frames, this capability streamlines the process, resulting in time savings compared to manual handcrafted procedures.

Application of individual's privacy

In an era characterized by ubiquitous digital content sharing, surveillance, and artificial intelligence, the preservation of individual visual privacy in videos is paramount. Visual privacy is employed in the blurring or masking of faces in security camera footage and in the real-time anonymization of live video broadcasts. Privacy-preserving techniques facilitate security monitoring while safeguarding identities under public surveillance. Automated facial or skin recognition aids in preventing unintentional exposure on social media and internet platforms, hence safeguarding users' image management. Video anonymization assists law enforcement and investigative journalists in protecting vital witnesses and whistleblowers. Privacy-enhancing methodologies, including YOLO-based skin detection and encryption, safeguard individuals' identities in artificial intelligence and deep learning while maintaining video functionality. Privacy filters assist telemedicine applications in maintaining patient anonymity. Ensuring visual privacy fosters a balance among security, innovation, and human rights as video-based artificial intelligence advances in capability.

Suggested future work directions :

Several promising research directions are emerging to address the identified gaps: On-Device Generative Anonymization: Compact GAN architectures designed for microcontrollers and edge devices enable privacy-by-design systems that eliminate identify at the sensor before transmission or storage (Ancilotto et al., 2023).

Latent-Space Identity Disentanglement: Pre-trained generative models' latent spaces can be used to provide customizable privacy at lower computational cost than custom anonymization models (StyleID, FALCO). STPrivacy uses transformer-based architectures to jointly process video tubelets for context-aware anonymization that respects action semantics and removes identification information across temporal sequences. Federated and Split Learning with Cryptographic Guarantees: Secure aggregation, differential privacy, and selective homomorphic encryption enable scalable, privacy-preserving video analytics for distributed surveillance networks.

Adversarial Training Ensembles: Training anonymization models against varied attacker models, including unknown adaptive adversaries, can increase real-world attack resilience (FIVA, DeMem) .Uu-net and Invertible Mask Network allow authorized re-identification for law enforcement or forensic purposes while maintaining strong privacy for general access, but secure key management and governance frameworks need further development.

6. Result and Discussion

The systematic evaluation divided state-of-the-art privacy preservation approaches into a novel, three-part taxonomy—Intervention, Obfuscation, and Secure Processing—that encompassed the entire data pipeline rather than just de-identification. This analytical paradigm showed various crucial performance trade-offs and deep learning's relative originality in over 30 peer-reviewed research. The analysis reveals that modern one-stage detectors, such as CNN-YOLO hybrid architectures, outperform traditional methods and achieve real-time rates of $30+$ FPS with high accuracy (92-98%). Adaptive anonymization methods using YOLOv8m-seg (Asres et al., 2026) outperform non-deep-learning methods like GrabCut (Brkić et al., 2017) and early EMHI (Chen et al., 2007), which face re-identification risks and body segmentation inaccuracies.

A key finding is the privacy-utility balance. Generative Adversarial Networks (GANs) (Rosberg et al., 2023) preserve contextual usefulness better than fixed pixelation (Zhou & Pun, 2021). Simple encryption solutions like multi-level security (Shifa, Asghar, et al., 2020) provide lightweight protection for restricted edge devices, whereas GAN-based identity substitution is more sophisticated and attribute-preserving. According to the taxonomy, this review emphasizes the growing importance of Secure Processing, which includes Federated Learning(Al-lahham et al., 2024) and Differential Privacy (Luo et al., 2023), which address architectural and inference threats, a domain previously neglected by reviews focused solely on post-capture techniques. Analysis of benchmark datasets (Table 3), such as Market-1501 and MARS (Zheng et al., 2016)(Mirzazadeh et al., 2023), highlights a persistent challenge: unstandardized metrics and critical accuracy degradation ($15-23%$) in real-world variables like low-light and occlusion, indicating current solutions are optimized for controlled environments.

This systematic synthesis suggests that visual privacy's future requires a multi-level security design that mixes robust, real-time deep learning detection (YOLOv9/v10) with secure processing protocols (FL/HE) to handle various threat situations. Complex, context-aware frameworks have replaced simple data masking, demonstrating a potential shift from obscurity to proactive data governance. This requires a change to adversarially-aware evaluation frameworks since non-standardized datasets limit the verifiability and ethical compliance of present solutions.

7. Conclusion

Growing deep learning-powered video analytics in public and private settings brings privacy opportunities and challenges. Intervention approaches, obfuscation strategies, and safe processing pipelines were evaluated in this article to protect real-time video privacy. We found

key strengths and weaknesses in the field by reviewing over 30 peer-reviewed articles and benchmark datasets. GAN-based anonymization models like CIAGAN and FIVA provide better identity obfuscation and task utility but poor deployment scalability and adversarial robustness. Under different settings, real-time detection methods like YOLOv8m-seg offer adaptive anonymization but struggle with fairness and generalizability. This paper presents a taxonomy and evaluation matrix that highlights privacy, usefulness, and real-time performance trade-offs. Our findings link privacy preservation to model architecture, dataset design, and deployment context, advancing theory. The findings can help developers create privacy-aware surveillance systems and legislators, AI ethics committees, and regulators shape data governance frameworks. We outline a research roadmap for next-generation visual privacy systems by highlighting underexplored directions like fairness-aware obfuscation, adversarial robustness, and federated architectures. Next steps should include standardizing privacy-specific datasets, establishing explainable anonymization methods, and balancing edge-based deployment delay with formal privacy assurances. In the age of intelligent video systems, privacy requires a holistic approach that combines technical rigor with ethical accountability.

References

Ahmed, M., Wang, Y., Maher, A., & Bai, X. (2022). Fused RetinaNet for small target detection in aerial images. *International Journal of Remote Sensing*, *43*(8), 2813–2836. https://doi.org/10.1080/01431161.2022.2071115

Ahn, B., & Jang, S.-W. (2021). Context-adaptive blocking for protecting personal information exposed to social multimedia content. *Multimedia Tools and Applications*, *80*(26–27), 34249–34267. https://doi.org/10.1007/s11042-020-10042-0

Al-lahham, A., Zaheer, M. Z., Tastan, N., & Nandakumar, K. (2024). Collaborative Learning of Anomalies with Privacy (CLAP) for Unsupervised Video Anomaly Detection: A New Baseline. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12416–12425. https://doi.org/10.1109/CVPR52733.2024.01180

Al-Obaidi, S., Al-Khafaji, H., & Abhayaratne, C. (2020). Modeling Temporal Visual Salience for Human Action Recognition Enabled Visual Anonymity Preservation. *IEEE Access*, *8*, 213806–213824. https://doi.org/10.1109/ACCESS.2020.3039740

Al-Tamimi, M. and T. H. Y. 2024. (2024). An Exhaustive Survey of Deep Learning Techniques in ECG Signals. *Ibn AL-Haitham Journal For Pure and Applied Sciences.*, *37*(3), 428–441. https://doi.org/https://doi.org/10.30526/37.3.3901

Al-Tamimi, M. S. H., Amer, F., & Ali, M. (2023). Face mask detection based on algorithm YOLOv5s. *Int. J. Nonlinear Anal. Appl*, *14*, 2008–6822. https://doi.org/10.22075/ijnaa.2022.28178.3824

Ancilotto, A., Paissan, F., & Farella, E. (2023). PhiNet-GAN: Bringing real-time face swapping to embedded devices. *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops)*, 677–682. https://doi.org/10.1109/PerComWorkshops56833.2023.10150292

Arkushin, D., Cohen, B., Peleg, S., & Fried, O. (2024). GEFF: Improving Any Clothes-Changing Person ReID Model Using Gallery Enrichment with Face Features. *2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 143–153. https://doi.org/10.1109/WACVW60836.2024.00021

Asres, M. W., Jiao, L., & Walter Omlin, C. (2026). Low-Latency Video Anonymization for Crowd Anomaly Detection: Privacy Versus Performance. *IEEE Transactions on Information Forensics and Security*, *21*, 1–16. https://doi.org/10.1109/TIFS.2025.3630347

Baoyuan, C., Yitong, L., & Kun, S. (2021). Research on Object Detection Method Based on FF-YOLO for Complex Scenes. *IEEE Access*, *9*, 127950–127960. https://doi.org/10.1109/ACCESS.2021.3108398

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. https://doi.org/10.48550/arXiv.2004.10934

Brkić, K., Hrkać, T., & Kalafatić, Z. (2017). Protecting the privacy of humans in video sequences using a computer vision-based de-identification pipeline. *Expert Systems with Applications*, *87*, 41–55. https://doi.org/10.1016/j.eswa.2017.05.067

Casas, E., Ramos, L., Bendek, E., & Rivas-Echeverría, F. (2023). Assessing the Effectiveness of YOLO Architectures for Smoke and Wildfire Detection. *IEEE Access*, *11*, 96554–96583. https://doi.org/10.1109/ACCESS.2023.3312217

Chen, D., Chang, Y., Yan, R., & Yang, J. (2007). Tools for protecting the privacy of specific individuals in video. *Eurasip Journal on Advances in Signal Processing*, *2007*. https://doi.org/10.1155/2007/75427

Cho, Y.-J., & Yoon, K.-J. (2017). *PaMM: Pose-aware Multi-shot Matching for Improving Person Re-identification*. https://doi.org/10.1109/TIP.2018.2815840

Çiftçi, S., Korshunov, P., Akyüz, A. O., & Ebrahimi, T. (2015). *Using false colors to protect visual privacy of sensitive content* (B. E. Rogowitz, T. N. Pappas, & H. de Ridder, Eds.; p. 93941L). https://doi.org/10.1117/12.2083189

Climent-Pérez, P., & Florez-Revuelta, F. (2021). Protection of visual privacy in videos acquired with RGB cameras for active and assisted living applications. *Multimedia Tools and Applications*, *80*(15), 23649–23664. https://doi.org/10.1007/s11042-020-10249-1

Climent-Pérez, P., Spinsante, S., Mihailidis, A., & Florez-Revuelta, F. (2020). A review on video-based active and assisted living technologies for automated lifelogging. *Expert Systems with Applications*, *139*, 112847. https://doi.org/10.1016/j.eswa.2019.112847

Deshmukh, S., Doshi, K., & Borse, Y. (2018). Securing Images Using Layered Morphing. *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697888

Dhayea, A. M., El Abbadi, N. K., & Hasan, Z. G. A. (2024). Human Skin Detection and Segmentation Based on Convolutional Neural Networks. *Iraqi Journal of Science*, *65*(2), 1102–1116. https://doi.org/10.24996/ijs.2024.65.2.40

Dominguez-Dager, B., Escalona, F., Gomez-Donoso, F., & Cazorla, M. (2026). CHIRLA: Comprehensive High-resolution Identification and Re-identification for Large-scale Analysis. *Scientific Data*, *13*(1), 109. https://doi.org/10.1038/s41597-025-06425-3

Dworak, D. (2020). *BlurNet: Keeping Collected Data Private with a Neural Network Based Pipeline* (pp. 1237–1248). https://doi.org/10.1007/978-3-030-50936-1_103

Erdelyi, A., Barat, T., Valet, P., Winkler, T., & Rinner, B. (2014). Adaptive cartooning for privacy protection in camera networks. *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 44–49. https://doi.org/10.1109/AVSS.2014.6918642

Erlina, T., & Fikri, M. (2023). A YOLO Algorithm-based Visitor Detection System for Small Retail Stores using Single Board Computer. *Journal of Applied Engineering and Technological Science (JAETS)*, *4*(2), 908–920. https://doi.org/10.37385/jaets.v4i2.1872

Fan, M., Chen, C., Wang, C., Zhou, W., & Huang, J. (2023). *On the Robustness of Split Learning Against Adversarial Attacks*. https://doi.org/10.3233/FAIA230330

Gevers, T., & Smeulders, A. (2016). Foreword. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): 9914 LNCS* (p. V). Springer Verlag. https://doi.org/10.1007/978-3-319-46466-4

Girshick, R. (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440–1448. https://doi.org/10.1109/ICCV.2015.169

He, K., Zhang, X., Ren, S., & Sun, J. (2014). *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition* (pp. 346–361). https://doi.org/10.1007/978-3-319-10578-9_23

Hellmann, F., Mertes, S., Benouis, M., Hustinx, A., André, E., Hsieh, C., De, T.-B., Hsieh, T.-C., Conati, C., & Krawitz, P. (2023). *GANonymization: A GAN-based Face Anonymization Framework for Preserving Emotional Expressions ACM Reference Format* (Vol. 1, Number 1). https://doi.org/doi:10.1145/3641107.

Huang, W., Li, G., Chen, Q., Ju, M., & Qu, J. (2021). CF2PN: A Cross-Scale Feature Fusion Pyramid Network Based Remote Sensing Target Detection. *Remote Sensing*, *13*(5), 847. https://doi.org/10.3390/rs13050847

Huang, X., Chen, W., & Yang, W. (2021). Improved Algorithm Based on The Deep Integration of Googlenet and Residual Neural Network. *Journal of Physics: Conference Series*, *1757*(1), 012069. https://doi.org/10.1088/1742-6596/1757/1/012069

Huangfu, Z., & Li, S. (2023). Lightweight You Only Look Once v8: An Upgraded You Only Look Once v8 Algorithm for Small Object Identification in Unmanned Aerial Vehicle Images. *Applied Sciences*, *13*(22), 12369. https://doi.org/10.3390/app132212369

Hukkelas, H., & Lindseth, F. (2023). DeepPrivacy2: Towards Realistic Full-Body Anonymization. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1329–1338. https://doi.org/10.1109/WACV56688.2023.00138

Hussain, M. (2023). YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines*, *11*(7), 677. https://doi.org/10.3390/machines11070677

Hussain, M. (2024). YOLOv1 to v8: Unveiling Each Variant–A Comprehensive Review of YOLO. *IEEE Access*, *12*, 42816–42833. https://doi.org/10.1109/ACCESS.2024.3378568

Jassim, A. H., Ali, N. H., Al-Taie, A., & Majed, D. M. (2025). Accelerating Face Mask Detection Training Model Based on Multi-GPUs and Multi-core CPU. *Baghdad Science Journal*, *22*(6), 2099–2118. https://doi.org/10.21123/2411-7986.4979

Kapadia, A., Henderson, T., Fielding, J. J., & Kotz, D. (2007). Virtual walls: Protecting digital privacy in pervasive environments. In *H.-W. Gellersen, R. Want, & A. Schmidt (Eds.), Pervasive Computing* (pp. 162–179). Springer. https://doi.org/10.1007/978-3-540-72037-9_10

Karanam, S., Gou, M., Wu, Z., Rates-Borras, A., Camps, O., & Radke, R. J. (2019). A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(3), 523–536. https://doi.org/10.1109/TPAMI.2018.2807450

Kim, K.-J., Kim, P.-K., Chung, Y.-S., & Choi, D.-H. (2018). Performance Enhancement of YOLOv3 by Adding Prediction Layers with Spatial Pyramid Pooling for Vehicle Detection. *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–6. https://doi.org/10.1109/AVSS.2018.8639438

Korshunov, P., & Ebrahimi, T. (2013). Using face morphing to protect privacy. *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 208–213. https://doi.org/10.1109/AVSS.2013.6636641

Lee, H., Kim, M. U., Kim, Y., Lyu, H., & Yang, H. J. (2021). Development of a Privacy-Preserving UAV System With Deep Learning-Based Face Anonymization. *IEEE Access*, *9*, 132652–132662. https://doi.org/10.1109/ACCESS.2021.3113186

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). *YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications*. https://doi.org/10.48550/arXiv.2209.02976

Li, R., & Yang, J. (2018). Improved YOLOv2 Object Detection Model. *2018 6th International Conference on Multimedia Computing and Systems (ICMCS)*, 1–6. https://doi.org/10.1109/ICMCS.2018.8525895

Li, Y., & Lyu, S. (2019). De-identification Without Losing Faces. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 83–88. https://doi.org/10.1145/3335203.3335719

Li, Z., Lu, S., Dong, Y., & Guo, J. (2023). MSFFA: a multi-scale feature fusion and attention mechanism network for crowd counting. *The Visual Computer*, *39*(3), 1045–1056. https://doi.org/10.1007/s00371-021-02383-0

Liu, J., Xia, Y., & Tang, Z. (2021). Privacy-preserving video fall detection using visual shielding information. *The Visual Computer*, *37*(2), 359–370. https://doi.org/10.1007/s00371-020-01804-w

Liu, Y., Peng, B., Shi, P., Yan, H., Zhou, Y., Han, B., Zheng, Y., Lin, C., Jiang, J., Fan, Y., Gao, T., Wang, G., Liu, J., Lu, X., & Xie, D. (2019). iQIYI-VID: A Large Dataset for Multi-modal Person Identification. *Proceedings of the 26th ACM International Conference on Multimedia*, 1360–1363. https://doi.org/10.48550/arXiv.1811.07548

Luo, Z., Zou, Y., Yang, Y., Durante, Z., Huang, D.-A., Yu, Z., Xiao, C., Fei-Fei, L., & Anandkumar, A. (2023). Differentially Private Video Activity Recognition. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 20029–20038. https://doi.org/https://doi.org/10.48550/arXiv.2306.15742

M, Mr. N. (2024). Protecting Privacy in Surveillance Systems via Selective Video Encryption. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, *08*(05), 1–5. https://doi.org/10.55041/IJSREM34554

Maolood, A. T., Gbashi, E. K., & Mahmood, E. S. (2022). Novel lightweight video encryption method based on ChaCha20 stream cipher and hybrid chaotic map. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(5), 4988. https://doi.org/10.11591/ijece.v12i5.pp4988-5000

Maximov, M., Elezi, I., & Leal-Taixé, L. (2020). *CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks*. https://doi.org/10.1109/CVPR42600.2020.00549

Mirzazadeh, A., Dubost, F., Pike, M., Maniar, K., Zuo, M., Lee-Messer, C., & Rubin, D. (2023). ATCON: Attention Consistency for Vision Models. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1880–1889. https://doi.org/10.1109/WACV56688.2023.00192

Mohammed, N. A., Abdulateef, O. F., Hamad, A. H., & Abdullah, O. I. (2024). Performance Analysis of Different Machine Learning Algorithms for Predictive Maintenance. *Al-Khwarizmi Engineering Journal*, *20*(2), 26–38. https://doi.org/10.22153/kej.2024.11.003

Mulajkar, R., & Yede, S. (2024). YOLO Version v1 to v8 Comprehensive Review. *2024 International Conference on Inventive Computation Technologies (ICICT)*, 472–478. https://doi.org/10.1109/ICICT60155.2024.10544452

Petitcolas, F. A. P., Anderson, R. J., & Kuhn, M. G. (1999). Information hiding-a survey. *Proceedings of the IEEE*, *87*(7), 1062–1078. https://doi.org/10.1109/5.771065

Raees, S., & Al-Tamimi, M. (2024). The Role of Artificial Intelligence in Providing People With Privacy: Survey. *Journal of Applied Engineering and Technological Science (JAETS)*, *5*(2), 813–829. https://doi.org/10.37385/jaets.v5i2.4013

Ragab, M. G., Abdulkadir, S. J., Muneer, A., Alqushaibi, A., Sumiea, E. H., Qureshi, R., Al-Selwi, S. M., & Alhussian, H. (2024). A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access*, *12*, 57815–57836. https://doi.org/10.1109/ACCESS.2024.3386826

Raj J, J. S., & Gopalan, A. (2025). Compact Bi-slot Patch Antenna with Tapered Edges for Ka-Band Applications Featuring Machine Learning-Assisted Performance Prediction. *International Journal of Environment, Engineering and Education*, *7*(3), 196–207. https://doi.org/10.55151/ijeedu.v7i3.326

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. https://doi.org/10.1109/CVPR.2016.91

Rosberg, F., Aksoy, E. E., Englund, C., & Alonso-Fernandez, F. (2023). FIVA: Facial Image and Video Anonymization and Anonymization Defense. *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 362–371. https://doi.org/10.1109/ICCVW60793.2023.00043

Sapkota, R., Qureshi, R., Flores-Calero, M., Badgujar, C., Nepal, U., Poulose, A., Zenorobotics, P. Z., Billings, L., Bhanu, U., Vaddevolu, P., Khan, S., Shoman, M., Yan, H., & Karkee,

M. (n.d.). *YOLOv12 to Its Genesis: A Decadal and Comprehensive Review of The You Only Look Once (YOLO) Series*. Retrieved doi: 10.48550/arXiv.2406.19407

Sawarkar, C. D., & Sawarkar, G. B. (2024). Face identification and blurring the face using deep learning based approaches in videos. *International Journal for Multidisciplinary Research (IJFMR)*. https://www.ijfmr.com/research-paper.php?id=14479

Shepley, A. J., Falzon, G., Kwan, P., & Brankovic, L. (2023). Confluence: A Robust Non-IoU Alternative to Non-Maxima Suppression in Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(10), 11561–11574. https://doi.org/10.1109/TPAMI.2023.3273210

Shifa, A., Asghar, M. N., Fleury, M., Kanwal, N., Ansari, M. S., Lee, B., Herbst, M., & Qiao, Y. (2020). MuLViS: Multi-level encryption based security system for surveillance videos. *IEEE Access*, *8*, 177131–177155. https://doi.org/10.1109/ACCESS.2020.3024926

Shifa, A., Imtiaz, M. B., Asghar, M. N., & Fleury, M. (2020). Skin detection and lightweight encryption for privacy protection in real-time surveillance applications. *Image and Vision Computing*, *94*. https://doi.org/10.1016/j.imavis.2019.103859

Song, G., Leng, B., Liu, Y., Hetang, C., & Cai, S. (2017). *Region-based Quality Estimation Network for Large-scale Person Re-identification*. doi:10.1609/aaai.v32i1.12305

Sukkar, M., Kumar, D., & Sindha, J. (2021). Real-Time Pedestrians Detection by YOLOv5. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 01–06. https://doi.org/10.1109/ICCCNT51525.2021.9579808

Tran, B., Reddy Kona, S. H., Liang, X., Ghinita, G., Summerour, C., & Batsis, J. A. (2023). VPASS: Voice Privacy Assistant System for Monitoring In-home Voice Commands. *2023 20th Annual International Conference on Privacy, Security and Trust (PST)*, 1–10. https://doi.org/10.1109/PST58708.2023.10320179

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-Time End-to-End Object Detection*. doi.org/10.48550/arXiv.2405.14458%0A

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*. https://doi.org/10.48550/arXiv.2402.13616

Wang, G., Lai, J., Huang, P., & Xie, X. (2018). *Spatial-Temporal Person Re-identification*. https://doi.org/10.48550/arXiv.1812.03282

Wang, T., Gong, S., Zhu, X., & Wang, S. (2016). Person Re-Identification by Discriminative Selection in Video Ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(12), 2501–2514. https://doi.org/10.1109/TPAMI.2016.2522418

Wei, L., Zhang, S., Gao, W., & Tian, Q. (2017). *Person Transfer GAN to Bridge Domain Gap for Person Re-Identification*. https://doi.org/10.48550/arXiv.1711.08565

Wei, Q., Hu, X., Wang, X., & Wang, H. (2022). Improved RetinaNet Target Detection Model. *2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI)*, 470–476. https://doi.org/10.1109/AHPCAI57455.2022.10087635

Wu, T.-H., Wang, T.-W., & Liu, Y.-Q. (2021). Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network. *2021 3rd World Symposium on Artificial Intelligence (WSAI)*, 24–28. https://doi.org/10.1109/WSAI51899.2021.9486316

Wu, Y., Lin, Y., Dong, X., Yan, Y., Ouyang, W., & Yang, Y. (2018). Exploit the Unknown Gradually: One-Shot Video-Based Person Re-identification by Stepwise Learning. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5177–5186. https://doi.org/10.1109/CVPR.2018.00543

Xiao, T., Li, S., Wang, B., Lin, L., & Wang, X. (2017). Joint Detection and Identification Feature Learning for Person Search. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3376–3385. https://doi.org/10.1109/CVPR.2017.360

Xie, S., Girshick, R., Dollar, P., Tu, Z., & He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5987–5995. https://doi.org/10.1109/CVPR.2017.634

Xu, D., & Wu, Y. (2020). Improved YOLO-V3 with DenseNet for Multi-Scale Remote Sensing Target Detection. *Sensors*, *20*(15), 4276. https://doi.org/10.3390/s20154276

Yu, J., & Zhang, W. (2021). Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4. *Sensors*, *21*(9), 3263. https://doi.org/10.3390/s21093263

Zhao, Y., Gao, D., Yao, Y., Zhang, Z., Mao, B., & Yao, X. (2023). Robust Deep Learning Models against Semantic-Preserving Adversarial Attack. *2023 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN54540.2023.10191198

Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y., & Tian, Q. (2016). *Person Re-identification in the Wild*. https://doi.org/10.48550/arXiv.1604.02531

Zhong, Z., Sun, L., & Huo, Q. (2019). An anchor-free region proposal network for Faster R-CNN-based text detection approaches. *International Journal on Document Analysis and Recognition (IJDAR)*, *22*(3), 315–327. https://doi.org/10.1007/s10032-019-00335-y

Zhou, J., & Pun, C.-M. (2021). Personal Privacy Protection via Irrelevant Faces Tracking and Pixelation in Video Live Streaming. *IEEE Transactions on Information Forensics and Security*, *16*, 1088–1103. https://doi.org/10.1109/TIFS.2020.3029913

Zhu, S., Zhang, C., & Zhang, X. (2017). Automating Visual Privacy Protection Using a Smart LED. *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 329–342. https://doi.org/10.1145/3117811.3117820