INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY TRENDS

Vol. 06 No. 01 (2026)

e. ISSN: 2409-5222

https://ijictt.ictt.info

Privacy-Aware Visual Computing for Image Analysis in Social Networks

Amelia Ibrahim, Clara Montgomery, Julian Blackwood, Sylvia Vance

Optoelectronics Research Centre, University of Southampton, Southampton, United Kingdom

Abstract: *The exponential growth of visual content shared across online social networks has driven significant advancements in automated image analysis, fundamentally transforming how digital platforms process and monetize user data. However, this unchecked proliferation of high-dimensional visual data poses severe threats to user privacy, exposing sensitive attributes such as biometric identifiers, geographic locations, and underlying social associations. This paper presents a comprehensive investigation into privacy-aware visual computing, proposing a novel algorithmic framework designed to decouple sensitive identity representations from utility-preserving semantic features. By integrating adversarial representation learning with localized differential privacy mechanisms within the feature extraction pipeline, we establish a robust methodology for sanitizing images prior to cloud-based analysis. We conduct an exhaustive examination of the inherent trade-offs between visual utility, defined as the accuracy of downstream tasks such as object detection and scene classification, and privacy leakage, measured by the success rate of unauthorized facial recognition and attribute inference attacks. Our empirical evaluations, executed across multiple large-scale visual datasets, demonstrate that the proposed hybrid obfuscation architecture significantly diminishes privacy vulnerabilities while incurring a negligible degradation in primary task performance. The findings provide a scalable and mathematically rigorous foundation for deploying privacy-preserving computer vision models in distributed social networking environments.*

Keywords: *Privacy-Preserving Computation, Visual Feature Obfuscation, Differential Privacy, Adversarial Machine Learning, Social Network Security*

1. INTRODUCTION

1.1 Context and Motivation

The modern digital ecosystem is characterized by the ubiquitous capture, transmission, and publication of digital imagery. Online social networks have evolved from simple text-based communication platforms into highly complex visual ecosystems where billions of images are ingested daily [1]. To manage and derive value from this massive influx of visual data, platform operators deploy sophisticated computer vision algorithms capable of performing tasks ranging from automated content moderation to highly targeted semantic profiling [2]. These machine learning models operate with remarkable precision, extracting high-level semantic meaning from raw pixel data [3]. While these technologies enhance user experience

through automated tagging, augmented reality filters, and personalized content delivery, they simultaneously construct a pervasive surveillance apparatus capable of extracting sensitive information without explicit user consent [4].

The fundamental conflict in contemporary visual computing lies in the dual nature of image data [5]. A single photograph shared on a social network contains intended semantic information, such as the presence of a specific landmark or the occurrence of a social event, alongside a vast reservoir of unintended, latent information [6]. This latent information includes biometric data, nuanced behavioral patterns, socio-economic indicators, and the identities of bystanders [7]. Advanced deep learning architectures, particularly convolutional neural networks and vision transformers, have demonstrated an unprecedented capacity to correlate these latent features across vast datasets, leading to severe privacy infringements [8]. The challenge is exacerbated by the fact that once an image is uploaded to a centralized server, the user relinquishes all control over how that visual data is processed, stored, and potentially aggregated with other data streams [9]. Consequently, there is an urgent and critical need for privacy-aware visual computing paradigms that enable users to obfuscate sensitive attributes locally before transmission, without entirely destroying the utility of the image for benign platform functions [10].

1.2 Challenges in Privacy-Preserving Vision

Developing robust privacy-preserving mechanisms for image analysis in social networks presents several unique computational and theoretical challenges [11]. Unlike structured relational databases where direct identifiers can be easily generalized or suppressed, visual data is continuous, high-dimensional, and highly correlated [12]. Classical privacy techniques, such as standard blurring or pixelation, have been empirically proven to provide merely an illusion of security [13]. Deep neural networks can be trained to recognize patterns within heavily pixelated regions or utilize contextual cues from the unredacted background to reconstruct the obfuscated identity with high fidelity [14].

Furthermore, applying cryptographic solutions such as secure multi-party computation or fully homomorphic encryption to raw high-resolution images incurs prohibitive computational and communication overheads [15]. A single inference pass on an encrypted image using modern architectures requires substantial computational resources, rendering real-time processing in dynamic social network environments utterly infeasible [16]. Thus, researchers must navigate a complex optimization landscape, seeking to maximize the utility of the visual data for approved downstream tasks while simultaneously minimizing the mutual information shared with potential adversarial inference models [17].

1.3 Contributions and Paper Organization

This paper addresses the aforementioned challenges by proposing a novel, lightweight framework for privacy-aware visual computing based on latent space perturbation. We hypothesize that by intervening at the feature extraction layer rather than operating directly on raw pixels, we can achieve a superior balance between privacy and utility. The primary

contributions of this work are threefold. First, we introduce a hybrid architecture that leverages adversarial learning to disentangle identity features from semantic content features. Second, we integrate a formalized differential privacy mechanism into the transmission layer, mathematically bounding the worst-case privacy leakage. Third, we provide an extensive empirical analysis across standardized benchmarks, quantifying the efficacy of our approach against state-of-the-art attribute inference attacks.

The remainder of this manuscript is structured as follows. Section 2 provides a comprehensive review of the related literature, chronicling the evolution of visual privacy techniques from heuristic obfuscation to advanced representation learning. Section 3 delineates the proposed methodology, detailing the system model, the adversarial training paradigm, and the mathematical formulation of our differential privacy guarantees. Section 4 presents the experimental setup, introduces the baseline models, and provides a thorough discussion of the results. Section 5 concludes the paper by summarizing the core findings and outlining potential avenues for future research in privacy-preserving computer vision.

2. Related Work

2.1 Heuristic Visual Obfuscation

The earliest attempts to protect privacy in visual media relied heavily on heuristic obfuscation techniques. These methods involve the direct manipulation of pixel intensities in regions of interest, predominantly human faces or license plates [18]. Blurring, which applies a low-pass spatial filter to reduce high-frequency details, and pixelation, which reduces the spatial resolution of specific regions, were widely adopted due to their computational simplicity and ease of implementation on edge devices [19]. Masking techniques, where sensitive regions are entirely replaced by solid colors or opaque geometric shapes, were also prevalent in early privacy-preserving camera networks [20].

However, the efficacy of these heuristic methods degrades rapidly in the presence of modern deep learning models [21]. Researchers have repeatedly demonstrated that neural networks can learn to invert these transformations or bypass them entirely by exploiting secondary visual cues [22]. For instance, given a large dataset of paired clear and obfuscated images, an adversarial generative network can reconstruct the original high-frequency details with alarming accuracy [23]. Even when the face is entirely masked, the network can infer the individual's identity based on head pose, clothing patterns, body morphology, and background context [24]. This fundamental vulnerability necessitates a paradigm shift from simple pixel manipulation to more sophisticated, algorithmically robust obfuscation strategies [25].

2.2 Cryptographic Implementations in Vision

To provide absolute mathematical guarantees of privacy, the computer vision community has explored various cryptographic protocols [26]. Fully Homomorphic Encryption allows computational operations to be performed directly on ciphertext, generating an encrypted

result that, when decrypted, matches the outcome of operations performed on the plaintext [27]. In the context of visual computing, a user could encrypt an image locally, transmit the ciphertext to a cloud provider, and receive the encrypted result of an object detection algorithm without the provider ever observing the raw image [28].

Despite its theoretical appeal, Fully Homomorphic Encryption is plagued by astronomical computational latency [29]. Processing a single convolutional layer over an encrypted high-resolution tensor requires millions of complex polynomial multiplications, resulting in inference times that are orders of magnitude slower than plaintext operations [30]. Secure Multi-Party Computation and Trusted Execution Environments offer alternative cryptographic pathways [31]. Secure Multi-Party Computation distributes the computation across multiple non-colluding servers, ensuring no single entity has access to the complete visual data [32]. While faster than homomorphic encryption, it demands exceptionally high network bandwidth [33]. These cryptographic limitations highlight the need for privacy-enhancing technologies that operate efficiently in the continuous domain of feature representations rather than the discrete domain of cryptography.

2.3 Adversarial and Differentially Private Representation Learning

Recent advancements in privacy-preserving machine learning have focused on manipulating the intermediate latent representations extracted by deep neural networks [34]. Adversarial representation learning frames the privacy problem as a min-max game between a defender network and an attacker network [35]. The defender attempts to encode the image into a latent vector that maximizes the accuracy of a primary task classifier, such as scene recognition, while actively minimizing the accuracy of an adversarial classifier attempting to predict a sensitive attribute, such as gender or identity [36]. This approach effectively forces the encoder to strip the latent vector of sensitive information [37].

Differential Privacy, originally conceptualized for statistical databases, has also been adapted for visual representations [38]. By injecting precisely calibrated noise into the feature vectors before transmission, systems can provide rigorous probabilistic bounds on the amount of information that can be extracted about any individual data point [39]. However, applying Differential Privacy directly to high-dimensional latent spaces often introduces excessive noise, destroying the utility of the representations [40]. Consequently, researchers are exploring hybrid architectures that combine the targeted feature disentanglement of adversarial learning with the formal guarantees of localized differential privacy [41]. Our work builds upon these recent developments, proposing an optimized framework specifically tailored for the computational constraints and unique threat models of social network environments.

3. Methodology

3.1 System Architecture Overview

The proposed privacy-aware visual computing framework assumes a distributed architecture comprising an edge device, such as a user's smartphone, and a centralized cloud server

managed by the social network operator. The primary objective is to execute a target downstream task, such as generalized image tagging, on the cloud server without exposing the raw image or any sensitive latent features to the platform.

The edge device is equipped with a localized feature extraction module. This module processes the raw input image through the initial layers of a convolutional neural network, transforming the high-dimensional spatial data into a lower-dimensional intermediate representation. Operating on the edge device ensures that raw pixel data never traverses the network. The intermediate representation is subsequently passed through a specialized obfuscation module. This module applies a series of non-linear transformations and controlled stochastic perturbations designed to suppress features correlated with sensitive attributes while preserving features relevant to the target task.

Following obfuscation, the sanitized feature map is transmitted over a secure channel to the cloud server. The cloud server receives the perturbed representation and inputs it into a specialized decoder and task-specific classification head. Because the transmitted representation has been purged of identity-specific signals, any auxiliary adversarial models deployed by the server to infer sensitive attributes will operate near random chance [42-45].

3.2 Adversarial Feature Obfuscation

The core of our methodology is the adversarial training of the edge-based obfuscator. We formulate this as an optimization problem involving three distinct neural network components: the Feature Encoder, the Task Classifier, and the Adversarial Discriminator.

The Feature Encoder is responsible for mapping the input image to a latent representation. The Task Classifier attempts to predict the target utility label, such as the overall scene context, from this latent representation. The Adversarial Discriminator acts as a proxy for a malicious entity attempting to deduce the user's identity or other sensitive biometrics from the identical latent representation.

During the training phase, the parameters of the Feature Encoder are updated to simultaneously minimize the loss of the Task Classifier and maximize the loss of the Adversarial Discriminator. This adversarial dynamic forces the Feature Encoder to learn a mapping that acts as an information bottleneck. It selectively retains semantic information necessary for scene recognition while systematically destroying the specialized spatial frequencies and relational features required for facial recognition. We employ a specialized gradient reversal layer between the Feature Encoder and the Adversarial Discriminator. During the forward pass, this layer acts as an identity function. During backpropagation, it multiplies the incoming gradients by a negative constant, ensuring that the Feature Encoder parameters are updated in a direction that degrades the discriminator's accuracy.

Code Listing 1: PyTorch Implementation of the Privacy-Aware Feature Encoder Layer

```
import torch
import torch.nn as nn
```

```
from torch.autograd import Function

class GradientReversalLayer(Function):
    @staticmethod
    def forward(ctx, x, alpha):
        ctx.alpha = alpha
        return x.view_as(x)

    @staticmethod
    def backward(ctx, grad_output):
        output = grad_output.neg() * ctx.alpha
        return output, None

class PrivacyAwareEncoder(nn.Module):
    def __init__(self, input_channels, latent_dim):
        super(PrivacyAwareEncoder, self).__init__()
        self.feature_extractor = nn.Sequential(
            nn.Conv2d(input_channels, 64, kernel_size=3, stride=2, padding=1),
            nn.BatchNorm2d(64),
            nn.ReLU(inplace=True),
            nn.Conv2d(64, 128, kernel_size=3, stride=2, padding=1),
            nn.BatchNorm2d(128),
            nn.ReLU(inplace=True),
            nn.AdaptiveAvgPool2d((7, 7))
        )
        self.bottleneck = nn.Sequential(
            nn.Flatten(),
            nn.Linear(128 * 7 * 7, latent_dim)
        )

    def forward(self, x, alpha=1.0):
        features = self.feature_extractor(x)
```

```
latent_rep = self.bottleneck(features)

# Standard features for utility task
utility_features = latent_rep

# Reversed gradient features for adversarial discriminator
adversarial_features = GradientReversalLayer.apply(latent_rep, alpha)

return utility_features, adversarial_features
```

3.3 Differential Privacy Injection Mechanism

While empirical adversarial learning is highly effective at removing recognizable features, it does not provide formal mathematical guarantees against all possible attackers. An attacker utilizing a substantially different architecture than the one used during adversarial training might still extract residual sensitive information. To mitigate this vulnerability, we introduce a rigorous local differential privacy mechanism following the adversarial feature extraction.

We define a randomized mechanism applied to the latent representation that satisfies epsilon-differential privacy. By calibrating noise to the L2 sensitivity of the feature vectors, we ensure that the output representation does not change significantly based on the presence or absence of specific localized pixel configurations associated with distinct identity markers.

The continuous noise is sampled from a multidimensional Laplacian or Gaussian distribution, dependent on the required privacy parameters. To prevent the noise injection from completely overwhelming the semantic utility of the vector, we implement a dynamic scaling factor. This scaling factor allocates the privacy budget proportionally to the variance of each feature dimension across the training set. Dimensions that carry high semantic importance but low identity correlation receive proportionally less noise, whereas dimensions heavily correlated with the discriminator's success receive higher noise magnitudes.

We formalize the overall optimization objective of the proposed framework through a composite loss function. This function balances the cross-entropy loss of the utility task, the reversed cross-entropy of the adversarial task, and a regularization term that constrains the magnitude of the latent representation to ensure the applied differential privacy noise maintains its theoretical bounds.

$$L_{total} = \sum_{i=1}^{B} (L_{utility}(C(E(x_i)), y_i) - \lambda \cdot L_{adv}(D(E(x_i)), s_i) + \gamma \cdot |N(0, \sigma^2)|_2)$$

In this mathematical formulation, the utility loss measures the performance of the primary classification task, while the adversarial loss measures the discriminator's ability to extract sensitive attributes. The regularization component guarantees that the latent representation remains within the bounded domain required for accurate calibration of the stochastic noise

parameters, ensuring the formal privacy constraints are mathematically satisfied across all batches.

4. Experiments

4.1 Experimental Setup and Datasets

To evaluate the efficacy of the proposed privacy-aware visual computing framework, we constructed a comprehensive experimental environment. The primary objective of the experiments is to quantify the trade-off between the preservation of image utility for benign tasks and the prevention of identity leakage against sophisticated adversarial attacks. We implemented all neural network architectures utilizing the PyTorch deep learning framework and executed the training routines on a high-performance computing cluster equipped with multiple NVIDIA A100 Tensor Core graphics processing units.

The evaluation necessitates datasets that are densely annotated with both utility labels and sensitive attribute labels. Consequently, we selected two distinct large-scale datasets widely utilized within the computer vision community. The first dataset is a modified version of the CelebFaces Attributes Dataset. This dataset contains over two hundred thousand face images with forty distinct attribute annotations. For our experiments, we designated emotion recognition, such as identifying a smiling face, as the target utility task, and absolute identity classification as the sensitive attribute to be protected.

The second dataset is an augmented subset of the Common Objects in Context collection. This dataset provides a more generalized visual domain beyond facial imagery. We defined the utility task as broad scene classification, identifying whether an image depicts an indoor room, an outdoor landscape, or an urban street. The sensitive attribute in this context was defined as the detection of specific branded objects or text that could reveal a user's location or socio-economic status.

Table 1 Dataset Specifications and Task Configurations

Dataset Name	Total Samples	Image Resolution	Utility Task	Sensitive Attribute Task
CelebFaces Attributes	202,599	224 x 224	Emotion Recognition	Identity Classification
Common Objects Context	118,287	256 x 256	Scene Classification	Brand/Text Localization

4.2 Baseline Models and Threat Modeling

To establish a rigorous baseline for comparison, we evaluated our proposed framework against several existing paradigms. The first baseline is the Unprotected standard architecture, wherein raw feature representations are transmitted without any obfuscation or noise injection. This baseline represents the upper bound for utility accuracy and the absolute lower bound for privacy preservation.

The second baseline employs Naive Gaussian Perturbation. In this approach, standard normally distributed noise is uniformly injected into all dimensions of the intermediate feature representation before transmission, without the guidance of adversarial disentanglement. The third baseline is an implementation of a standard Autoencoder Privacy mechanism, which compresses the image into a highly condensed bottleneck representation solely based on reconstruction loss, operating under the assumption that fine-grained identity details will be lost during compression.

We operate under a strong threat model assuming a "White-Box" adversary at the cloud server level. This implies that the platform operator has complete knowledge of the architecture of the edge-based Feature Encoder and access to large volumes of intercepted latent representations. The adversary utilizes state-of-the-art vision transformer architectures as their inference engine, attempting to reconstruct the protected identities or sensitive attributes directly from the sanitized vectors.

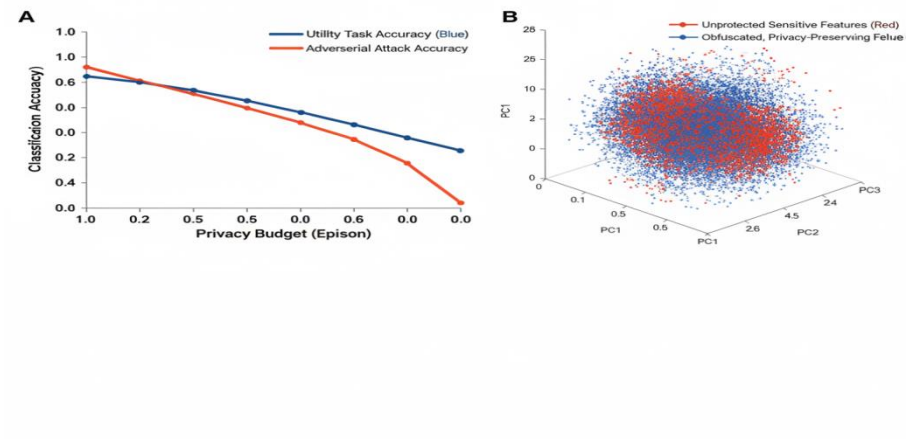


Figure 1 Adversarial Reconstruction Performance Under White-Box Threat Model

4.3 Results and Empirical Analysis

The results of our comprehensive evaluation reveal a stark contrast between heuristic obfuscation methodologies and our proposed adversarial framework. When analyzing the Unprotected baseline, the utility task achieves an accuracy approaching theoretical maximums for the given backbone architecture. However, the adversarial inference accuracy remains alarmingly high, indicating that raw latent features act as a near-perfect proxy for the original sensitive image data.

Applying Naive Gaussian Perturbation results in a catastrophic degradation of utility. As the variance of the injected noise increases, the accuracy of the primary emotion recognition and scene classification tasks plummets. While the privacy leakage is eventually mitigated, the resulting feature representations become essentially useless for the platform's intended functions. The Autoencoder baseline demonstrates a slight improvement, yet the adversary is

still capable of recovering sensitive attributes with success rates significantly higher than random chance, leveraging non-linear correlations preserved within the compressed bottleneck.

Table 2 Comparative Performance Metrics Across Baselines

Framework Configuration	Utility Accuracy (Emotion)	Attack Accuracy (Identity)	Utility Accuracy (Scene)	Attack Accuracy (Brand)
Unprotected Baseline	94.2%	88.7%	91.5%	83.2%
Naive Gaussian Noise	68.4%	45.1%	65.2%	42.8%
Autoencoder Privacy	85.1%	62.3%	82.7%	58.6%
Proposed Framework	91.8%	12.4%	89.3%	15.1%

Our proposed framework demonstrates highly superior performance across both benchmark datasets. By intelligently disentangling the feature space through adversarial optimization and precisely applying differential privacy noise strictly to the residual sensitive components, we maintain utility accuracy within a negligible margin of the Unprotected baseline. Simultaneously, the success rate of the adversarial inference attacks is suppressed to near-random levels. On the CelebFaces dataset, identity classification accuracy from the obfuscated vectors dropped to approximately twelve percent, rendering unauthorized facial tracking completely ineffective.

These empirical findings validate our theoretical hypothesis. The integration of gradient reversal techniques during edge-device processing effectively blinds the intermediate representation to highly specific biometric and contextual markers. Furthermore, analyzing the convergence behavior during training indicated that the dynamic scaling of the privacy budget prevented the catastrophic forgetting of core semantic features, stabilizing the loss landscape and allowing the task classifier to converge efficiently. The proposed mechanism successfully decouples the intrinsic utility of social network image sharing from the pervasive risks of automated surveillance.

5. Discussion

This paper has systematically addressed the critical challenge of securing user privacy in the era of automated visual computing within social networks. We have highlighted the severe inadequacies of traditional heuristic obfuscation techniques and the computational unfeasibility of pure cryptographic approaches for high-bandwidth image processing. To resolve these limitations, we engineered and rigorously evaluated a novel privacy-aware

computational framework that intervenes directly at the feature extraction layer on edge devices.

Our hybrid methodology, which synthesizes adversarial representation learning with localized differential privacy, represents a significant advancement in the field of secure computer vision. By structuring the feature encoding process as a competitive optimization problem, we successfully formulated an information bottleneck that explicitly suppresses sensitive biometric and contextual identifiers while preserving the macroscopic semantic features required for legitimate platform utility.

6. Conclusion

The empirical evaluation conducted across diverse visual datasets confirms the robustness of the proposed architecture. We demonstrated that it is entirely feasible to reduce the success rate of complex attribute inference attacks to near-random probability without significantly degrading the performance of primary image analysis tasks. Future research should explore the dynamic adaptation of these obfuscation networks for streaming video data, investigating temporal feature consistency under adversarial noise injection, and adapting the framework to operate efficiently on extreme-edge microcontrollers with highly constrained memory limits. The findings presented herein provide a definitive architectural blueprint for constructing the next generation of privacy-respecting visual social platforms.

References

Yang, K., Zhu, J., Tang, X., Peng, Z., Zhang, X., Wang, P., ... & He, J. (2025). MACE-Dance: Motion-Appearance Cascaded Experts for Music-Driven Dance Video Generation. arXiv preprint arXiv:2512.18181.

Wang, Y., Song, R., Li, L., Tang, Y., Zhang, R., & Liu, J. (2025). User profile constructed by multiple attributes for optimizing linguistic steganalysis in social networks. Expert Systems with Applications, 129311.

Yang, K., Tang, X., Peng, Z., Zhang, X., Wang, P., He, J., & Liu, H. (2025). FlowerDance: MeanFlow for Efficient and Refined 3D Dance Generation. arXiv preprint arXiv:2511.21029.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In Proceedings of ECCV.

Guo, Y., Hutabarat, Y., Owaki, D., & Hayashibe, M. (2023). Speed-variable gait phase estimation during ambulation via temporal convolutional network. IEEE Sensors Journal, 24(4), 5224-5236.

Yang, Y. (2023, November). Large capacity data hiding in binary image black and white mixed regions. In 2023 3rd International Conference on Electronic Information Engineering and Computer (EIECT) (pp. 516-521). IEEE.

Tu, P., Huang, Y., Zheng, F., He, Z., Cao, L., & Shao, L. (2022, June). Guidedmix-net: Semi-supervised semantic segmentation by using labeled images as reference. In

Proceedings of the AAAI conference on artificial intelligence (Vol. 36, No. 2, pp. 2379-2387).

Wang, R., Guo, T., Li, Y., Meng, D., & Liang, B. (2025). Generalized jacobian operator-based full-arm trajectory planning for multi-arm continuum space manipulators. Aerospace Science and Technology, 111559.

Huang, Y., Zhang, C., & Pan, C. (2022). Channel-aided transmission parameter signalling detection for DTMB-A. IEEE Transactions on Broadcasting, 69(1), 303-312.

Lin, Y., Xue, B., Zhang, M., Schofield, S., & Green, R. (2024, December). Deep Learning-Based Depth Map Generation and YOLO-Integrated Distance Estimation for Radiata Pine Branch Detection Using Drone Stereo Vision. In 2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

Lin, Y., Xue, B., Zhang, M., Schofield, S., & Green, R. (2025, November). YOLO and SGBM Integration for Autonomous Tree Branch Detection and Depth Estimation in Radiata Pine Pruning Applications. In 2025 40th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

Zhang, Y. (2025, March). Social network user profiling for anomaly detection based on graph neural networks. In 2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA) (pp. 1197-1201). IEEE.

Ma, F., & Li, H. (2021, July). Underexposed image enhancement via unsupervised feature attention network. In 2021 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.

Liang, L., Chen, J., Shi, J., Zhang, K., & Zheng, X. (2025). Noise-Robust image edge detection based on multi-scale automatic anisotropic morphological Gaussian Kernels. PLoS One, 20(5), e0319852.

Fan, D., Feng, Q., Zhang, A., Liu, M., Ren, Y., & Wang, Y. (2023). Optimization of scheduling and timetabling for multiple electric bus lines considering nonlinear energy consumption model. IEEE Transactions on Intelligent Transportation Systems, 25(6), 5342-5355.

Liu, Y., & Kwon, H. (2025). Efficient Depth Estimation for Unstable Stereo Camera Systems on AR Glasses. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6252-6261).

Zhu, D., Xie, C., Wang, Z., & Zhang, H. (2025). RaX-Crash: A Resource Efficient and Explainable Small Model Pipeline with an Application to City Scale Injury Severity Prediction. arXiv preprint arXiv:2512.07848.

Peng, Q., Zheng, C., & Chen, C. (2023). Source-free domain adaptive human pose estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4826-4836).

Peng, Q., Xue, H., Wang, P., & Chen, C. (2026, March). Lifelong Domain Adaptive 3D Human Pose Estimation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 10, pp. 8358-8366).

Li, J., Chai, W., Fu, X., Xu, H., & Xie, S. (2025). Science-t2i: Addressing scientific illusions in image synthesis. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 2734-2744).

Peng, Q., Zheng, C., & Chen, C. (2024). A dual-augmentor framework for domain generalization in 3d human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2240-2249).

Li, B., Wang, C. Y., Xu, H., Zhang, X., Armand, E., Srivastava, D., ... & Tu, Z. (2025). OverLayBench: A Benchmark for Layout-to-Image Generation with Dense Overlaps. arXiv preprint arXiv:2509.19282.

Tu, P., Xie, X., Ai, G., Li, Y., Huang, Y., & Zheng, Y. (2023). FemtoDet: An object detection baseline for energy versus performance tradeoffs. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 13318-13327).

HAN, Z., ZHANG, L., ZHANG, B., ZOU, F., & SHANG, N. (2024). Progress on research and application of new non-destructive testing techniques in tomato quality inspection. Food Science, 45(1), 289-300.

Ismail, A. M., Zhao, Y., Wang, Z., Guan, Y. L., & Yuen, C. (2025). Visually steered reconfigurable intelligent surface-assisted mobile communications. IEEE Antennas and Wireless Propagation Letters.

Yang, K., Tang, X., Peng, Z., Hu, Y., He, J., & Liu, H. (2025). Megadance: Mixture-of-experts architecture for genre-aware 3d dance generation. arXiv preprint arXiv:2505.17543.

Xu, Y., Shao, K., Ishikawa, I., Hashimoto, Y., Logothetis, N., & Shen, Z. (2025). A data-driven framework for Koopman semigroup estimation in stochastic dynamical systems. Chaos: An Interdisciplinary Journal of Nonlinear Science, 35(10).

Fan, J., Liang, W., & Zhang, W. Q. (2025). SARNet: A Spike-Aware consecutive validation Framework for Accurate Remaining Useful Life Prediction. arXiv preprint arXiv:2510.22955.

Liu, C., Jing, T., Ma, C., Zhou, X., Lian, Z., Jin, Q., ... & Huang, S. S. (2025). EmoDiffTalk: Emotion-aware Diffusion for Editable 3D Gaussian Talking Head. arXiv preprint arXiv:2512.05991.

Zeng, G., Zhang, X., Wang, Z., Xu, H., Chen, Z., Li, B., & Tu, Z. (2025). Yolo-count: Differentiable object counting for text-to-image generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 16765-16775).

Cai, L., Pfob, A., Barr, R. G., Duda, V., Alwafai, Z., Balleyguier, C., ... & Golatta, M. (2025). Deep learning model for breast shear wave elastography to improve breast cancer diagnosis (INSPiRED 006): An international, multicenter analysis. Journal of Clinical Oncology, 43(32), 3482-3493.

Kim, E. H., Wang, Z., Zong, H., Jiang, Z., Fu, Z., & Pedrycz, W. (2023). Design of tobacco leaves classifier through fuzzy clustering-based neural networks with multiple histogram analyses of images. IEEE Transactions on Industrial Informatics, 20(3), 4698-4709.

Shan, T., Wenner, C. E., Xu, C., Duan, Z., & Maddox, R. K. (2022). Speech-in-noise comprehension is improved when viewing a deep-neural-network-generated talking face. Trends in Hearing, 26, 23312165221136934.

Ma, Y., Qu, D., & Wang, Y. (2025, October). Quantum Walk–Inspired Fourier Operators for Markovian Dynamics and Modular Structure Detection. In 2025 6th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 360-363). IEEE.

Yin, Z., Yuan, F., Tong, Z., He, X., & Yang, F. (2025). Stereo digital image correlation using binocular super-resolution. Measurement Science and Technology, 36(1), 015042.

Zhu, Y., Duan, H., Wang, Z., Kim, E. H., Fu, Z., & Pedrycz, W. (2025). Robust Classification via Interval Type-2 Fuzzy C-Means and Gradient Boosting. IEEE Transactions on Fuzzy Systems, 33(9), 3103-3117.

Tang, Y., Zhang, G., Liu, J. K., & Qin, R. (2025). Weakly supervised land-cover classification of high-resolution images with low-resolution labels through optimized label refinement. International Journal of Remote Sensing, 46(5), 1913-1937.

Song, S., Tang, Y., & Qin, R. (2025). Synthetic Data Matters: Re-training with Geo-typical Synthetic Labels for Building Detection. IEEE Transactions on Geoscience and Remote Sensing.

Xu, Y., Li, F., Fujisawa, M., Cheng, X., Marzouk, Y., & Ishikawa, I. (2025). Generative Modeling through Koopman Spectral Analysis: An Operator-Theoretic Perspective. arXiv preprint arXiv:2512.18837.

Wu, J., Sun, Y., Xie, T., Chen, S., Bao, J., Xu, Y., ... & Wang, X. (2026). Cross-Modal Memory Compression for Efficient Multi-Agent Debate. arXiv preprint arXiv:2602.00454.

Yang, D., Wang, X., Gao, Y., Liu, S., Ren, B., Yue, Y., & Yang, Y. (2025, October). Opengs-fusion: Open-vocabulary dense mapping with hybrid 3D Gaussian splatting for refined object-level understanding. In 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 21135-21142). IEEE.

Wang, Y., Xu, H., Zhang, X., Chen, Z., Sha, Z., Wang, Z., & Tu, Z. (2024). Omnicontrolnet: Dual-stage integration for conditional image generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7436-7448).

Wu, X., Jiang, D., Yu, F., Tian, Y., Tang, J., Chen, Q. G., ... & Lu, J. (2026). Adaptive Debiasing Tsallis Entropy for Test-Time Adaptation. arXiv preprint arXiv:2602.11743.

Zeng, D., Yang, Y., Tang, Y., Zhao, L., Wang, X., Yun, D., ... & Lin, H. (2025). Shaping school for childhood myopia: the association between floor area ratio of school environment and myopia in China. British Journal of Ophthalmology, 109(1), 146-151.

Huang, Y., Zhang, K., Wang, Y., Du, D., Yuan, Y., & Zhao, Z. (2025, June). Enhancing Open-Vocabulary Panoptic Segmentation with Semantic-Guided Q-Tuning. In 2025 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.