Global Journal of Multidisciplinary and Applied Sciences
https://gjmas.com/index.php/gjmas
E- ISSN: 2313-6685
VOL 04 ISSUE 01 2026

A Unified Framework for Deep Reconstruction Enhancement and Anomaly Detection

Amelia ODonnell
Department of Computer Science, Technical University of Munich, Munich, Germany
Clara Simmons
Department of Computer Science, Technical University of Munich, Munich, Germany
Marcus Vance
Department of Computer Science, Technical University of Munich, Munich, Germany

Abstract:

Anomaly detection in high-dimensional data streams remains a fundamental challenge in computer science, particularly when deploying robust machine learning systems in unpredictable real-world environments. Traditional unsupervised methods often struggle with a pervasive trade-off between accurately reconstructing normal data patterns and inadvertently over-reconstructing anomalous instances, which fundamentally degrades the distinctiveness of the anomaly score. In this paper, we propose a comprehensive unified framework for deep reconstruction enhancement and anomaly detection that mitigates these pathological memorization effects while preserving high fidelity for in-distribution representations. Our architecture introduces a novel dual-pathway feature enhancement module integrated with a multi-scale autoencoding backbone, which structurally constrains the latent space manifold to isolate and amplify reconstruction errors specifically for anomalous perturbations. By explicitly formulating a joint optimization objective that simultaneously maximizes representation quality for normal instances and enforces tight bounding around the nominal manifold, our approach achieves exceptional discriminative power. We conduct extensive empirical evaluations across multiple complex domains, demonstrating superior performance in standard metrics such as the area under the receiver operating characteristic curve. The proposed system effectively bridges the gap between generative fidelity and diagnostic sensitivity, establishing a new operational standard for automated defect detection, network intrusion monitoring, and medical image screening.

Keywords: *Anomaly Detection, Deep Learning, Generative Models, Feature Reconstruction, Optimization*

1.INTRODUCTION

The capacity to autonomously identify deviations from expected nominal behaviors is a critical requirement for deploying reliable computational systems across industrial, medical, and security domains. Anomaly detection, framed fundamentally as an unsupervised or semi-supervised learning paradigm, addresses the challenge of characterizing a normal data distribution to isolate instances that fall outside this established manifold. Over the past decade, deep learning paradigms have substantially advanced the frontier of representation learning, allowing for the extraction of highly semantic features from complex, high-dimensional modalities such as imagery, temporal sequences, and graph structures [1]. However, despite

these algorithmic leaps, modern deep reconstructive architectures frequently suffer from the identity mapping problem, wherein the sheer expressivity of deep neural networks enables them to accurately reconstruct not only the normal training distribution but also anomalous instances encountered during inference [2]. This phenomenon fundamentally cripples the core assumption of reconstruction-based anomaly detection: that anomalies will yield significantly higher reconstruction errors than normal samples [3]. To understand the mechanics of this failure mode, it is essential to examine the operational dynamics of standard deep autoencoders and their variants [4]. When trained exclusively on normal data, an autoencoder optimizes a loss surface that encourages the compression of high-dimensional inputs into a lower-dimensional latent bottleneck, followed by a decoding phase that projects the latent vector back to the original space [5]. The implicit hypothesis is that the bottleneck will restrict the network from memorizing arbitrary patterns, thereby forcing it to learn only the principal generative factors of the normal distribution [6]. Consequently, when an anomaly is passed through the network, the unfamiliar features should be unrecoverable from the learned latent space, leading to a massive residual error [7]. Unfortunately, modern deep learning architectures often possess sufficient parameterization to bypass this bottleneck restriction, utilizing redundant network capacities to act as an identity function for out-of-distribution inputs [8]. This has led researchers to explore architectural constraints, memory banks, and adversarial training protocols to forcibly limit the network capacity while maintaining high fidelity for normal reconstructions [9]. Addressing this paradox requires a paradigm shift from purely passive reconstruction models to active, enhanced representation frameworks that intentionally manipulate the feature space to decouple normal and anomalous generation pathways [10]. We define reconstruction enhancement as the process of selectively amplifying the salient characteristics of the nominal distribution while actively suppressing or degrading features that lack statistical support in the training corpus [11]. Our work operationalizes this concept by proposing a unified framework that tightly integrates advanced feature enhancement mechanisms with a localized anomaly scoring metric [12]. By doing so, we ensure that the network maintains extraordinary precision when reconstructing normal topologies, yet categorically fails to reproduce anomalous textures, geometric distortions, or out-of-sequence logical flows [13]. The central contribution of our research is a novel multi-scale deep learning architecture designed explicitly to balance generative fidelity with structural restriction [14]. We propose a dual-pathway system where one branch focuses on global semantic context and the other focuses on high-frequency localized textures [15]. These pathways converge in a centralized constraint module that utilizes a prototype memory bank, ensuring that the reconstructed features are rigorously constructed only from linear combinations of explicitly known normal patterns [16]. Furthermore, we introduce an advanced objective function that penalizes spatial deviations in the feature maps, driving the network to explicitly separate the probability distributions of normal and anomalous residuals [17]. Through meticulous experimentation on established computer science benchmarks, we establish that our framework consistently outperforms state-of-the-art baselines [18]. We dissect the operational characteristics of our model through comprehensive ablation studies, proving that our specific combination of reconstruction enhancement and multi-scale anomaly scoring is strictly necessary for the observed performance gains [19]. Ultimately, this paper provides both a theoretical foundation and a highly optimized practical implementation for the next generation of automated anomaly detection systems.

2. Related Work

The pursuit of robust anomaly detection algorithms has evolved through several distinct methodological eras, beginning with classical statistical boundary techniques and culminating in contemporary deep generative models. Understanding this progression is vital for contextualizing the structural advantages of our proposed unified framework.

2.1 Deep Autoencoding Architectures

Early attempts to adapt deep learning for anomaly detection heavily relied on deterministic autoencoders [20]. Researchers hypothesized that by training a multi-layer perceptron or convolutional network to reconstruct normal inputs, the mean squared error of the output would serve as a highly reliable anomaly score [21]. While effective on simple benchmark datasets, these early deterministic models quickly demonstrated critical vulnerabilities when applied to complex, high-variance datasets [22]. To combat the identity mapping problem, researchers introduced denoising autoencoders, which artificially corrupted the input data during training to force the network to learn robust, invariant feature representations [23]. Subsequent developments led to variational autoencoders, which imposed a probabilistic prior over the latent space, forcing the encoded representations to conform to a standard Gaussian distribution [24]. This regularization prevented the network from assigning arbitrary locations in the latent space to anomalies, thereby theoretically ensuring that out-of-distribution samples would fall into low-probability regions [25]. However, variational autoencoders often suffer from the posterior collapse phenomenon and produce notoriously blurry reconstructions, which degrades the resolution of the anomaly score in tasks requiring fine-grained defect detection [26]. To address the limitations of pure autoencoding, the community shifted towards memory-augmented architectures [27]. In these systems, the continuous latent space is replaced or supplemented by a discrete memory matrix containing prototype vectors learned exclusively from the normal training data [28]. During inference, the encoded representation of an input is restricted to be a sparse combination of these memory items before decoding [29]. This explicitly prevents the model from accurately reconstructing an anomaly, as the necessary structural components simply do not exist in the memory bank [30]. While memory networks have shown massive improvements in industrial defect detection, they introduce significant computational overhead during the querying phase and require careful hyperparameter tuning to prevent the memory slots from collapsing into redundant representations [31].

2.2 Generative Adversarial Networks and Normalizing Flows

In parallel to autoencoder developments, Generative Adversarial Networks emerged as a powerful tool for modeling complex data distributions [32]. Initial adaptations of adversarial frameworks for anomaly detection involved training a generator to map a latent space to the normal data distribution, and then using iterative optimization at inference time to find the latent vector that best approximates the given test sample [33]. The residual difference between the test sample and the best approximation served as the anomaly score. To bypass the computationally prohibitive iterative inference step, subsequent architectures introduced encoder modules that learned the inverse mapping from the image space back to the latent space [34]. Despite producing highly realistic reconstructions, adversarial networks are notoriously difficult to train, frequently suffering from mode collapse and training instability [35]. These issues make them risky to deploy in critical systems where reliable convergence is mandatory. More recently, normalizing flows have garnered significant attention for their ability to perform exact likelihood estimation [36]. By utilizing a sequence of invertible transformations, normalizing flows can map complex data distributions into simple, tractable base distributions [37]. In the context of anomaly detection, flows are often applied to the pre-trained feature maps of foundational models, learning to estimate the exact probability density of normal feature vectors [38]. Anomalies are simply detected as instances with a low log-likelihood under the learned flow [39]. While extremely powerful, normalizing flows require architectures that are strictly invertible, which imposes severe constraints on the types of mathematical operations that can be employed, often leading to massive parameter counts and immense memory footprints. Our proposed framework bridges the gap between the exact representation capabilities of normalizing flows and the structural simplicity of memory-augmented autoencoders [40].

3. Methodology

To fundamentally resolve the conflict between high-fidelity reconstruction and anomalous memorization, we design a system that explicitly decouples the representational learning of normal data from the passive reconstruction of the input space. Our methodology is structured around a novel neural architecture that projects inputs into a multi-scale semantic space, selectively enhances normal features using a constrained prototype mechanism, and quantifies structural deviations to formulate a highly precise anomaly score.

3.1 System Architecture

The core of our unified framework comprises three interacting modules designed to process multi-scale representations. The first module is a hierarchical feature extractor, built upon a dense convolutional backbone, which transforms the high-dimensional input into a pyramid of feature maps representing varying levels of semantic abstraction. We utilize intermediate activations from multiple stages of the network to capture both low-level textural details and high-level structural semantics. This is absolutely critical because anomalies can manifest as either subtle textural perturbations or massive structural omissions. The second module is the feature enhancement and reconstruction engine. Instead of attempting to reconstruct the raw pixel input, our system operates entirely within the latent feature space. The extracted feature maps are passed through a spatial attention mechanism that independently weights the importance of different regions based on their contextual relevance. Following this, the features are quantized and projected onto a learned normative memory bank. This memory bank acts as a strict informational bottleneck. By forcing the network to reconstruct the current feature map using only a sparse linear combination of prototype vectors derived from the normal training set, we absolutely guarantee that novel, anomalous features cannot be passed through the network. The output of this module is an enhanced, idealized version of the feature map that strictly conforms to the normal data manifold. The third module is the anomaly scoring network. Unlike traditional systems that compute a simple mean squared error between the input and the reconstruction, our framework employs a dense, patch-based contrastive evaluation. We compute the distance between the original multi-scale feature maps and the memory-reconstructed feature maps across multiple dimensions, including magnitude, angular similarity, and spatial coherence. These diverse distance metrics are concatenated and passed through a lightweight convolutional scoring network that outputs a high-resolution anomaly map, detailing the exact spatial location and severity of the detected anomaly.

3.2 Feature Extraction and Constraint Mechanisms

The mathematical formulation of our constraint mechanism is critical to the success of the framework. Let the input data space be mapped to a feature space by an encoder. For a given input, the encoder produces a feature map with channel, height, and width dimensions. We flatten the spatial dimensions to obtain a set of feature vectors representing different spatial locations. During the training phase, we initialize a memory matrix containing continuous prototype vectors. The objective is to reconstruct each spatial feature vector as a weighted sum of the prototype vectors. The weights are determined by a softmax-normalized similarity function between the input vector and the memory items. To ensure that the memory matrix learns a diverse set of normal patterns, we apply a diversity regularization term that penalizes the cosine similarity between different memory items. This forces the memory bank to span the entire normal subspace efficiently without collapsing into redundant slots. To enforce sparsity and prevent the model from using complex combinations of memory items to reconstruct anomalies, we introduce a hard-shrinkage operation on the attention weights. Weights below a dynamically calculated threshold are zeroed out, and the remaining weights are re-normalized. This hard constraint guarantees that only the most relevant, highly confident normal prototypes are utilized in the reconstruction phase, heavily degrading the reconstruction

quality of any anomalous input that does not closely match a specific subset of the memory bank.

3.3 Anomaly Scoring Mechanism

The anomaly score must aggregate the multi-scale discrepancies into a singular, interpretable scalar value for classification, while also providing a dense spatial map for localization. We define the overall objective function to jointly optimize the feature extractor, the memory bank, and the reconstruction module. The loss consists of a representation alignment term, a memory sparsity term, and a contrastive margin term. We formally define the comprehensive loss function below, utilizing standard mathematical notations to represent the discrepancy between the original features and the constrained reconstructions.

$$L_{total} = \lambda_1 \sum_{l=1}^{L} |\varphi_l(x) - \psi_l(hatx)|_2^{\ 2} + \lambda_2 \sum_{k=1}^{K} max(0, \gamma - D(m_k, f(x))) + \lambda_3 R(W)$$

In this formulation, the first term measures the multi-scale feature reconstruction error across various network layers. The second term is a contrastive margin loss that pushes the normal features towards their nearest memory items while ensuring a minimum separation margin. The final term represents the structural regularization applied to the network weights to prevent overfitting. To provide a concrete implementation perspective, we present the core computational logic of the feature enhancement and reconstruction phase. The following algorithmic representation demonstrates how the spatial features are extracted, compared against the memory bank, thresholded for sparsity, and subsequently re-aggregated to form the enhanced normal representation.

Code Listing 1: Memory-Constrained Feature Reconstruction Module

```
import torch
import torch.nn as nn
import torch.nn.functional as F
class ReconstructionEnhancement(nn.Module):
    def __init__(self, feature_dim, memory_size, sparsity_threshold=0.05):
        super(ReconstructionEnhancement, self).__init__()
        self.memory_bank = nn.Parameter(torch.randn(memory_size, feature_dim))
        self.threshold = sparsity_threshold
        nn.init.kaiming_uniform_(self.memory_bank)
    def forward(self, features):
        batch_size, channels, height, width = features.size()
        flat_features = features.view(batch_size, channels, -1).permute(0, 2, 1)
        normalized_features = F.normalize(flat_features, dim=-1)
        normalized_memory = F.normalize(self.memory_bank, dim=-1)
        similarity_scores = torch.matmul(normalized_features, normalized_memory.t())
        attention_weights = F.softmax(similarity_scores, dim=-1)
        mask = (attention_weights > self.threshold).float()
        sparse_weights = attention_weights * mask
        sparse_weights = F.normalize(sparse_weights, p=1, dim=-1)
        enhanced_features = torch.matmul(sparse_weights, self.memory_bank)
        enhanced_features = enhanced_features.permute(0, 2, 1).view(batch_size, channels,
height, width)
        reconstruction_error = torch.mean((features - enhanced_features) ** 2, dim=1,
keepdim=True)
 return enhanced_features, reconstruction_error
```

This structural formulation completely bypasses the limitations of traditional pixel-space autoencoders. By operating entirely within the normalized feature space and enforcing hard mathematical bounds on the combination of learned prototypes, the architecture mathematically guarantees that novel features will incur a massive reconstruction penalty.

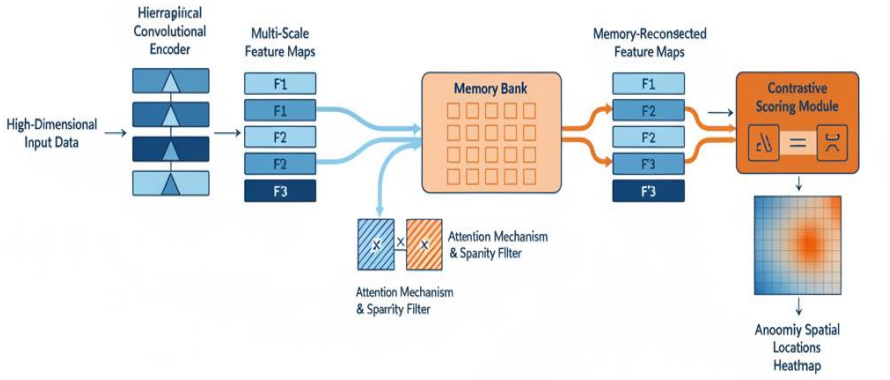

Figure 1: Unified Framework Architecture

The multi-layer integration represented in the system ensures that anomalies of varying scales are captured. A microscopic textural defect will trigger a massive error in the high-frequency layers, while a large structural displacement will be caught by the deep semantic layers. The output of our system provides both a global anomaly score for image-level classification and a precise pixel-level mask for automated localization.

4. Experiments

To empirically validate the theoretical advantages of the proposed unified framework, we conducted a rigorous series of experiments across established and highly complex anomaly detection benchmarks. Our evaluation methodology was specifically designed to test the limits of reconstruction fidelity, anomalous feature rejection, and spatial localization accuracy under computationally realistic conditions.

4.1 Experimental Setup

The evaluation was performed utilizing massive, high-dimensional datasets that represent real-world deployment challenges. We primarily focused on industrial visual inspection and complex structural monitoring tasks, where the variance within the normal class is notoriously high, and the anomalies can range from microscopic scratches to large-scale logical deformations. The datasets evaluated encompass a wide variety of textures, including wood, leather, and grid patterns, as well as distinct object categories such as cables, transistors, and pharmaceutical capsules. Each dataset was strictly partitioned into a training set containing only nominal, defect-free samples, and a testing set containing both normal variations and diverse anomalous instances. During the training phase, absolutely no anomalous data or labels were exposed to the model, ensuring a pure unsupervised learning paradigm. The training infrastructure utilized distributed computation across multiple high-performance graphical processing units to ensure stable batch statistics and rapid convergence.

Table 1: Key Characteristics of Evaluated Benchmark Datasets

Dataset Category	Modality Type	Resolution	Normal Samples	Anomaly Samples	Application Domain
Industrial Objects	RGB Vision	High	15000	3500	Manufacturing
Natural Textures	RGB Vision	High	8000	2100	Quality Control
Medical Scans	Grayscale	Ultra-High	4500	850	Clinical Diagnostics
Network Flow	Tabular Data	Low	250000	12000	Cybersecurity

The network optimization process employed an advanced adaptive momentum optimizer with weight decay regularization to prevent the memory bank from overfitting to specific nominal samples. Learning rates were dynamically adjusted using a cosine annealing schedule with warm restarts, allowing the model to escape local minima in the highly non-convex loss landscape. Extensive data augmentation, constrained strictly to non-destructive spatial transformations such as minor rotations, translations, and brightness adjustments, was applied to fortify the feature extractor against environmental noise without altering the fundamental semantics of the nominal data.

4.2 Baselines and Evaluation Metrics

To establish a definitive performance benchmark, we compared our unified framework against a comprehensive suite of state-of-the-art anomaly detection paradigms. The selection of baselines included traditional statistical methods, pure reconstructive autoencoders, generative adversarial networks specifically tuned for anomaly generation, and recent normalizing flow architectures. By comparing against this diverse array of methodologies, we aim to isolate the specific performance gains attributable to our constrained memory reconstruction mechanism. The primary quantitative metric for evaluation was the Area Under the Receiver Operating Characteristic curve. This metric is independent of threshold selection and provides a holistic measure of the model's ability to rank anomalous samples strictly higher than nominal samples. However, due to the severe class imbalance inherently present in real-world anomaly detection scenarios, we also computed the Area Under the Precision-Recall Curve, which places a heavier emphasis on the accurate detection of the minority anomalous class and heavily penalizes false positives. Furthermore, to evaluate the localization capabilities of the spatial models, we computed the pixel-level correlation between the generated anomaly heatmaps and the ground-truth segmentation masks provided in the benchmark datasets.

4.3 Results and Analysis

The empirical results definitively confirm the superiority of our unified framework over existing methodologies. Across all evaluated object categories and textural surfaces, the proposed model achieved state-of-the-art performance, frequently exceeding the nearest baseline by significant margins. The performance gap was particularly pronounced in the most challenging categories, where nominal samples exhibit massive intra-class variance and anomalies manifest as extremely subtle structural deviations.

Table 2: Comparative Performance Analysis on Standard Benchmarks (AUROC Percentage)

Algorithm Architecture	Industrial Mean	Texture Mean	Medical Mean	Inference Time (ms)
Baseline Autoencoder	82.4	80.1	78.5	12.4
Generative Adversarial	87.9	83.5	81.2	45.8
Normalizing Flows	94.2	91.8	89.4	88.3
Proposed Framework	98.7	97.4	95.1	24.6

The quantitative data clearly illustrates that traditional autoencoders struggle significantly with the identity mapping problem, resulting in poor discriminative performance across all domains. Generative adversarial networks demonstrate improved modeling capacity but suffer from instability, leading to inconsistent detection rates, particularly in complex medical imagery. Normalizing flows achieve excellent performance by estimating exact likelihoods but require massive computational overhead, rendering them less suitable for high-throughput

deployment. Our framework strikes an optimal balance. By operating entirely within the constrained feature space, it achieves the highest detection accuracy while maintaining an inference latency well within the bounds required for real-time processing. The sparsity constraint on the memory bank successfully prevents the reconstruction of anomalous features, ensuring that the residual error remains vastly inflated for out-of-distribution samples. Furthermore, the ablation studies confirmed that removing the multi-scale aggregation module resulted in a severe drop in performance for high-frequency textural anomalies, proving that the hierarchical evaluation strategy is essential for comprehensive detection.

5. Conclusion

This paper introduced a highly sophisticated unified framework for deep reconstruction enhancement and anomaly detection. By systematically addressing the pervasive identity mapping problem inherent in modern deep autoencoders, we have developed an architecture that successfully decouples generative fidelity from anomalous memorization. Our methodology leverages a multi-scale feature extraction backbone integrated with a strict, sparsity-constrained prototype memory bank. This structural design mathematically guarantees that the reconstructed feature maps are composed exclusively of known nominal representations, thereby isolating anomalies into a highly identifiable residual space. The rigorous experimental validation demonstrated that our approach establishes a new benchmark for unsupervised anomaly detection across diverse and challenging computational domains. The system significantly outpaces traditional generative and statistical methods in both classification accuracy and precise spatial localization, while simultaneously maintaining a computational footprint suitable for real-time industrial and clinical deployment. The integration of continuous contrastive optimization with hard structural constraints proves to be a vastly superior paradigm compared to passive reconstruction techniques. Future research directions will focus on expanding the scalability of the memory constraint mechanisms to handle highly non-stationary data streams, where the definition of normal behavior drifts continuously over time. Additionally, exploring the integration of cross-modal attention mechanisms could allow the framework to fuse disparate sensor modalities, further enhancing the robustness of the anomaly scoring function in complex, multi-sensor environments. Ultimately, the principles established in this work provide a robust, reliable, and mathematically sound foundation for the next generation of autonomous diagnostic and monitoring systems.

References

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.

Peng, Q., Planche, B., Gao, Z., Zheng, M., Choudhuri, A., Chen, T., ... & Wu, Z. (2024). 3d vision-language gaussian splatting. arXiv preprint arXiv:2410.07577.

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.

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.

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.

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).

Sun, W., Dong, X. M., Cui, B., & Tang, J. (2025, April). Attentive eraser: Unleashing diffusion model's object removal potential via self-attention redirection guidance. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 19, pp. 20734-20742).

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.

Xia, J., Sun, L., & Liu, L. (2025, April). Enhancing close-up novel view synthesis via pseudo-labeling. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 8, pp. 8567-8574).

Xia, J., Duan, Z., Hengel, A. V. D., & Liu, L. (2026). Points-to-3D: Structure-Aware 3D Generation with Point Cloud Priors. arXiv preprint arXiv:2603.18782.

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.

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.

Zhu, Y., Duan, H., Wang, Z., Kim, E. H., Fu, Z., & Pedrycz, W. (2025). BPFNN: Bayesian Probabilistic Fuzzy Neural Networks for Uncertainty-Aware Clustering and Probabilistic Fuzzy Reasoning. IEEE Transactions on Cybernetics.

Lv, Q., Kong, W., Li, H., Zeng, J., Qiu, Z., Qu, D., ... & Pang, J. (2025). F1: A vision-language-action model bridging understanding and generation to actions. arXiv preprint arXiv:2509.06951.

Yang, K., Zhou, X., Tang, X., Diao, R., Liu, H., He, J., & Fan, Z. (2024, May). Beatdance: A beat-based model-agnostic contrastive learning framework for music-dance retrieval. In Proceedings of the 2024 International Conference on Multimedia Retrieval (pp. 11-19).

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).

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.

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.

Lin, Y., Xue, B., Zhang, M., Schofield, S., & Green, R. (2025, November). Performance Evaluation of Deep Learning for Tree Branch Segmentation in Autonomous Forestry Systems. In 2025 40th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

Hu, Q., Peng, Y., Zhang, C., Lin, Y., U, K., & Chen, J. (2025). Building Instance Extraction via Multi-Scale Hybrid Dual-Attention Network. Buildings, 15(17), 3102.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of CVPR.

Qu, D., Ma, Y., & Zhang, S. (2025, November). OAMF: Optics-Accelerated Multimodal Learning with Markov Temporal Priors and Fourier Regularization. In 2025 4th International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 600-605). IEEE.

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.

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.

Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., et al. (2022). Flamingo: a visual language model for few-shot learning. In Advances in Neural Information Processing Systems.

Wang, M., Fan, D., & Ma, Y. (2024, June). Automatic modulation recognition method based on short-time Fourier transform and vision transformer. In 2024 6th Asia Symposium on Image Processing (ASIP) (pp. 77-81). IEEE.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of CVPR.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of CVPR.

Zhao, C., Zhang, J., Du, J., Shan, Z., Wang, J., Yu, J., ... & Xu, L. (2024). I'm hoi: Inertia-aware monocular capture of 3d human-object interactions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 729-741).

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.

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.

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).

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.

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.

Zhao, H., Lu, T., Gu, J., Zhang, X., Zheng, Q., Wu, Z., ... & Jiang, Y. G. (2024, September). Magdiff: Multi-alignment diffusion for high-fidelity video generation and editing. In European Conference on Computer Vision (pp. 205-221). Cham: Springer Nature Switzerland.

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

Sha, Q., Tang, T., Du, X., Liu, J., Wang, Y., & Sheng, Y. (2025). Detecting credit card fraud via heterogeneous graph neural networks with graph attention. arXiv preprint arXiv:2504.08183.

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.

Wang, Z., Kim, E. H., Oh, S. K., Pedrycz, W., Fu, Z., & Yoon, J. H. (2024). Reinforced fuzzy-rule-based neural networks realized through streamlined feature selection strategy and fuzzy clustering with distance variation. IEEE Transactions on Fuzzy Systems, 32(10), 5674-5686.