Transactions on Computational and Scientific Methods | Vol. 6, No. 01, 2026
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Cross-Domain Image Generation via Graph-Based Structural Modeling

Enis Sebe
Rutgers University, Piscataway, USA
es2290@rutgers.edu

Abstract: This paper addresses the problems of insufficient structural preservation and semantic drift in cross-domain image generation by proposing a cross-domain generation method based on graph-structured modeling. The method constructs node and edge representations of feature maps to explicitly model the semantic regions and structural relationships within images, enabling coordinated optimization of content consistency and style alignment during cross-domain feature transfer. Specifically, the model first uses a shared encoder to extract multi-scale semantic features from both the source and target domains, then builds topological relationships among nodes through a graph construction module. A graph convolution mechanism is applied to propagate and aggregate structured features, allowing the model to learn high-order semantic associations across domains. Finally, a decoder reconstructs the target-domain image. To enhance stability and realism, a joint optimization objective combining adversarial loss, content consistency constraint, and structural preservation term is designed to balance visual quality and structural coherence. Experimental results show that the proposed method outperforms mainstream generative models in MSE, MAE, PSNR, and SSIM, achieving high fidelity and semantic consistency under various cross-domain conditions. This study reveals the intrinsic patterns of cross-domain image generation from a structural modeling perspective and provides an effective new approach for graph-based generative visual modeling.

Keywords: Cross-domain image generation; graph structure modeling; feature alignment; structural consistency

1. Introduction

In today's field of artificial intelligence and computer vision, cross-domain image generation has become a major research focus. With the rapid development of deep learning models, image generation technologies have shown great potential in areas such as style transfer, data augmentation, medical image synthesis, and virtual reality[1]. However, in cross-domain scenarios, there are often significant distribution gaps between the source and target domains, including inconsistencies in content structure, lighting conditions, texture features, and semantic layouts. These discrepancies make it difficult for models to achieve high-quality transfer generation. Traditional generation frameworks based on convolution or adversarial learning often rely on large amounts of paired data or strict domain-matching assumptions, making it hard to capture global relationships and structural constraints between domains. As a result, the generated images may lack semantic consistency and visual realism. Therefore, how to balance content fidelity and style migration under cross-domain conditions has become a key scientific challenge that urgently needs to be addressed[2].

Against this background, graph-based modeling offers a new perspective for cross-domain image generation. Graph structures can represent high-level associations within and across domains through nodes and edges, thus overcoming the limitations of Euclidean space and enabling explicit modeling of complex relationships.

Compared with generation models that rely only on local convolutional features, graph-based modeling can integrate connections between semantic units at the global level, capturing topological relationships and contextual dependencies among objects. For example, by treating image patches, semantic regions, or feature points as nodes, and using edge weights to represent semantic or spatial similarity, a cross-domain feature alignment mechanism can be established. This mechanism maintains semantic consistency and structural integrity during generation. Such an approach strengthens the structural representation ability of the model and provides structural priors to address style distribution discrepancies between domains[3].

Moreover, cross-domain image generation has great practical significance. In scenarios where data annotation is costly or privacy-sensitive, it enables data transfer and expansion from the source domain to the target domain, effectively alleviating data scarcity. In applications such as medical image analysis, remote sensing reconstruction, and autonomous driving perception, variations in devices, weather, or lighting often cause significant domain shifts. Cross-domain generation can achieve domain adaptation through structural alignment, improving model generalization across diverse environments. In digital content creation and virtual simulation, it allows style fusion between real and virtual environments, supporting multi-modal visual generation. Thus, graph-based cross-domain image generation not only holds academic value but also promotes practical development in intelligent visual systems[4].

From a theoretical standpoint, introducing graph structures transforms cross-domain generation from traditional pixel-space mapping into relational modeling in graph space. Graph neural networks and other structured models can achieve cross-domain information flow through node representation propagation and aggregation, allowing the model to learn shared structural representations without relying on strictly aligned samples. This approach mitigates the impact of data distribution shifts, enabling the model to learn more robust cross-domain mapping patterns through topological relationships. Additionally, combining graph modeling with attention mechanisms allows adaptive weighting among nodes to strengthen key region generation, enhancing semantic consistency and structural interpretability. Through structured feature representations, the model can maintain content coherence and morphological stability when facing complex domain variations, thus providing a solid theoretical foundation for cross-domain generation[5].

In summary, graph-based cross-domain image generation represents an important extension of deep generative models and a key direction for achieving visual intelligence across domains. It integrates concepts from structural modeling, representation learning, and generative mechanisms, aiming to reveal the essential patterns of cross-domain feature transformation from a graph-structural perspective. This research promotes a paradigm shift from data-matching to structure-understanding in cross-domain generation, enhancing model interpretability, stability, and generalization in complex visual scenarios. With the continued development of graph neural networks and generative adversarial frameworks, this direction is expected to play a central role in multi-modal generation, domain adaptation, and intelligent image reconstruction, paving the way for more cognitively capable generative visual systems.

2. Related work

In recent years, cross-domain image generation has gradually shifted from pixel-level mapping toward structure-aware modeling. To alleviate issues such as semantic drift and structural distortion, enforcing structural consistency has become an important direction for improving generation quality. For example, [6] introduces a structure-consistency mechanism that effectively reduces semantic distortion in unsupervised image-to-image translation. Similarly, [7] proposes a graph-structured generation framework, Graph2Pix, which represents image semantics using nodes and edges and achieves cross-domain generation under explicit structural alignment.

Meanwhile, graph neural networks (GNNs) have demonstrated strong capabilities in modeling structural information. In related studies, GNNs have been applied to structural generalization modeling for microservice routing, validating their scalability and generalization ability in complex systems [8]. In

distributed systems, GNN-based approaches have also been used to predict spatiotemporal traffic patterns [9], further demonstrating the effectiveness of graph structures in modeling dynamic changes and enabling consistent feature transfer across domains.

In sequence modeling and structure learning, Transformer architectures have also been widely adopted for structure-aware tasks. For instance, [10] proposes a Transformer-based change-point detection framework to monitor structural variations in cloud-native systems, while [11] leverages Transformers to model user interaction sequences, improving the accuracy of dwell-time prediction in user interfaces. These works indicate that combining sequential modeling with structural modeling in dynamic environments enhances representation stability and generalization, which aligns well with the structural preservation requirements in cross-domain image generation.

To improve robustness against data distribution shifts, various self-supervised and attention-driven deep learning frameworks have been proposed. In [12], a self-supervised mechanism is introduced for anomaly detection in heterogeneous time series, while [13] employs attention mechanisms to enhance anomaly identification accuracy in ETL pipelines. In addition, [14] presents a dictionary-based few-shot anomaly segmentation framework that emphasizes generalization ability and sample efficiency under cross-domain conditions, which resonates with the goal of using graph structures to enhance cross-domain semantic consistency.

Furthermore, to achieve unified representations across tasks and domains, [15] introduces a dynamic prompt fusion mechanism to adapt to varying tasks and domains, and [16] proposes a trust evaluation mechanism in multi-agent systems, highlighting the importance of structural relationships in maintaining stable collaborative generation. Although these methods are applied in different contexts, they offer valuable insights into structure-aware modeling and robust generation.

In information retrieval and recommendation systems, methodological advances also reflect the importance of structural and consistency modeling. Specifically, [17] proposes a multi-objective contextual ranking mechanism to improve generation faithfulness, while [18] employs causal modeling to mitigate relevance bias in advertising recommendation. Similarly, [19] introduces uncertainty estimation to enhance the reliability of large language model outputs. These approaches provide theoretical support for structural generalization and risk-aware modeling, which are beneficial for strengthening consistency control in cross-domain generation.

Although some studies originate from control systems or industrial automation, the robust control strategy proposed in [20] demonstrates effective approaches for maintaining performance stability in dynamic systems, offering useful references for handling structural variations in cross-domain scenarios.

3. Method

This paper proposes a cross-domain image generation method based on graph structure modeling. The core idea is to achieve high-quality mapping between the source and target domains by constructing a cross-domain feature graph and performing structural constraints and semantic propagation in the graph space. Specifically, the proposed framework first extracts semantic representations from both domains through a shared encoder to ensure that features are comparable at multiple levels. These features are then transformed into graph nodes, while the relationships among them are represented as weighted edges that capture spatial and semantic dependencies. Within this graph space, the model employs a graph convolution mechanism to propagate information across nodes, enabling the transfer of structural knowledge and domain-invariant features. Finally, a decoder reconstructs the target domain image under the guidance of the learned graph representation, ensuring that the generated content preserves semantic consistency while adapting to the target domain's structural and visual characteristics. Figure 1 illustrates the overall architecture and data flow of the proposed method.

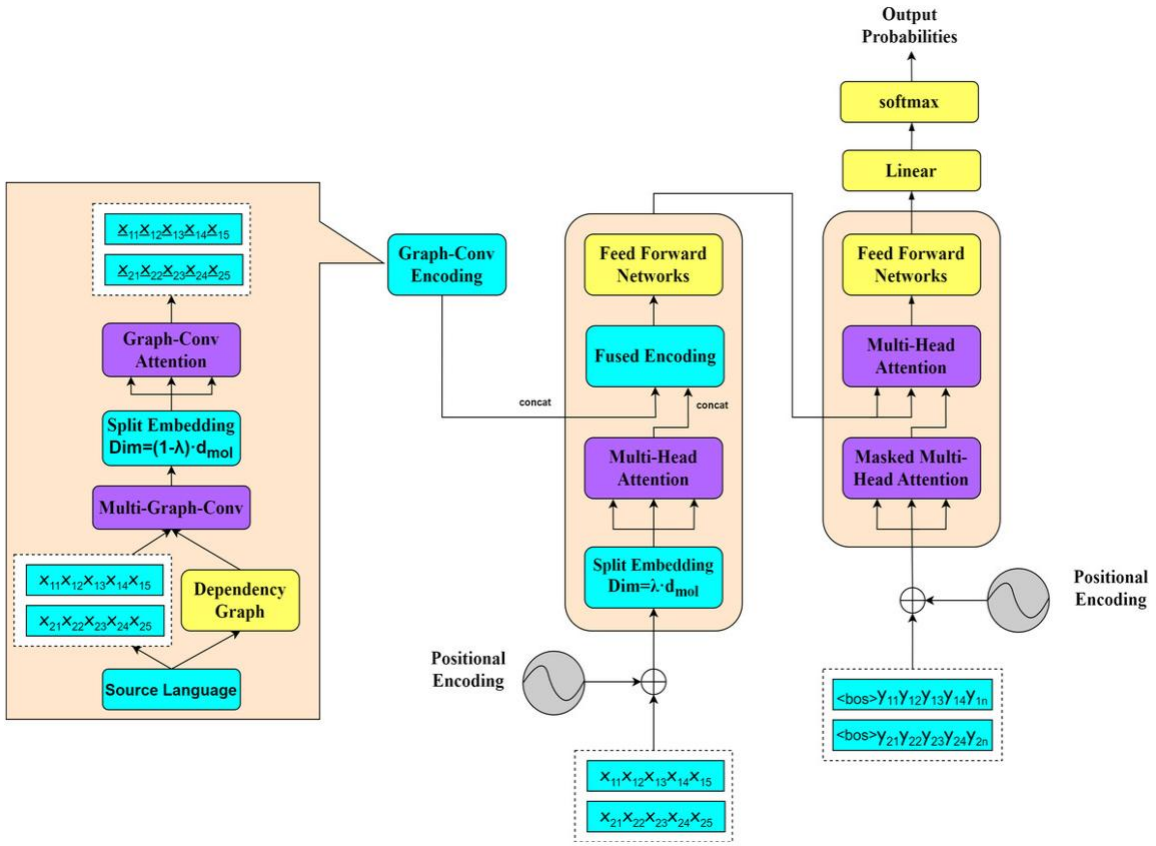


Figure 1. Overall model architecture

First, given a source domain image set X_s and a target domain image set X_t, the model extracts semantic feature representations $F_s = E(X_s)$ through a shared encoder $E(\cdot)$. In the feature space, these representations are further transformed into a node set $V = \{v_i\}_{i=1}^{N}$, where each node represents a local semantic region or key structural unit. The adjacency matrix A is constructed using the similarity metric function:

$$A_{ij} = \exp(-\frac{\| v_i - v_j \|_2^2}{\sigma^2})$$

It is used to represent the strength of structural associations between nodes. This graph structure not only preserves the spatial dependencies of images but also provides a basis for subsequent cross-domain feature propagation.

In the graph representation learning phase, the model uses a graph convolution mechanism to achieve feature aggregation and semantic propagation between nodes. Specifically, given the node feature matrix $H^{(l)}$ at layer l, it is updated to:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)})$$

Where $\tilde{A} = A + I$ represents the adjacency matrix with self-connection, $\tilde{D}$ is its degree matrix, $W^{(l)}$ is the learnable weight parameter, and $\sigma(\cdot)$ is the nonlinear activation function. The core function of this layer

is to achieve feature sharing between nodes across domains through topological constraints, enabling the model to learn domain-invariant semantic relationships at the structural level.

In the generation phase, the model integrates the structural features into a unified graph embedding representation $Z = Aggregate(H^{(L)})$ and inputs it to the decoder $G(\cdot)$ for reconstruction to generate the target domain image $\hat{X}_t = G(Z)$. To ensure that the generated result is semantically consistent with the source domain content and has the appearance characteristics of the target domain, this paper designs a joint optimization objective function that includes adversarial constraints and structural consistency terms. The overall loss function is defined as:

$$L_{total} = L_{adv} + \lambda_1 L_{content} + \lambda_2 L_{structure}$$

Among them, L_{adv} is used to improve the domain adaptability of generated samples, $L_{content}$ ensures the semantic consistency of cross-domain mapping, $L_{structure}$ strengthens the fidelity of geometric topology through graph constraints, and D is the weight coefficient.

Finally, the entire model is jointly optimized within an end-to-end framework. The encoder extracts structural features shared across domains, the graph convolution module models high-order dependencies between nodes, and the decoder, guided by this structure, performs target domain generation. By modeling the graph structure, the model simultaneously preserves content semantics and global geometric structure across domains, achieving high-fidelity mapping from structural space to visual space. This approach provides a unified structured representation framework for cross-domain image generation and lays a theoretical foundation for adaptive learning of multi-domain vision tasks.

4. Experimental Results

4.1 Dataset

This study employs the Cityscapes dataset as the primary data source for the cross-domain image generation task. The dataset consists of high-resolution urban street-view images that cover diverse scenes across different cities, seasons, and weather conditions. Each image has a resolution of 2048 × 1024 and includes multiple semantic categories such as buildings, roads, pedestrians, vehicles, and vegetation. These characteristics comprehensively reflect the spatial structures and semantic distributions in complex urban environments. The dataset also provides fine-grained pixel-level annotations, offering a reliable foundation for learning semantic consistency and structural preservation in the model.

In the experimental design, the Cityscapes dataset is divided into a source domain and a target domain to evaluate the model's cross-domain transfer capability. The source domain typically contains images captured under normal lighting or sunny conditions, while the target domain consists of images from rainy, foggy, or nighttime scenarios. This setup effectively tests the model's robustness against variations in lighting, color shifts, and structural noise. In the cross-domain generation process, the model is required to generate images that exhibit the characteristics of the target domain while maintaining the original structural content, demonstrating its adaptability and mapping ability across different domains.

The Cityscapes dataset is chosen for its high-quality annotations, rich scene diversity, and strong research relevance. As an important benchmark for cross-domain image generation, it ensures reproducibility and fairness in model training and provides a unified standard for performance comparison among different methods. Furthermore, its clear semantic boundaries and multi-scale object distributions make it an ideal platform for validating graph-based feature modeling, enabling the model to learn more robust semantic and structural relationships in complex urban environments.

4.2 Experimental Results

This paper first gives the results of the comparative experiment, as shown in Table 1.

Table1: Comparative experimental results

Model	MSE	MAE	PSNR	SSIM
VAE[21]	0.021	0.089	25.83	0.821
GAN[22]	0.017	0.076	27.46	0.854
WGAN[23]	0.015	0.072	28.12	0.871
Stable-Diffusion[24]	0.013	0.068	29.05	0.887
Ours	0.010	0.061	30.42	0.914

Overall, Table 1 presents the performance comparison of different generative models on the cross-domain image generation task. It can be observed that the proposed graph-based modeling method achieves the best performance across all metrics. In particular, it attains an MSE of 0.010 and an MAE of 0.061, which are significantly lower than those of traditional generative models. This demonstrates that the proposed method can effectively reduce reconstruction errors during cross-domain mapping and ensure pixel-level accuracy in the generated images. In contrast, traditional VAE and GAN models, which lack structural constraints, often suffer from texture blurring or semantic misalignment in complex scenes. By introducing graph-based modeling, the proposed approach achieves more stable cross-domain feature alignment and reconstruction.

Further analysis of the PSNR metric shows that the proposed method performs exceptionally well in image sharpness and detail preservation. It achieves a PSNR of 30.42, improving by 1.37 units compared with Stable Diffusion, indicating that the model can maintain high-quality visual reconstruction when handling variations in lighting, color, and texture between domains. This advantage stems from the graph convolution mechanism, which enables the model to capture high-level semantic structures. As a result, the model not only focuses on local textures but also learns spatial dependencies between objects, ensuring globally consistent semantic layouts during generation.

In terms of SSIM, the proposed method also achieves a remarkable improvement, reaching 0.914. This indicates that the generated images show better structural similarity and visual coherence than those produced by other methods. Traditional GAN-based models often produce distortions or artifacts when faced with large domain gaps. In contrast, the introduction of graph structures allows explicit modeling of relationships between nodes, effectively preserving topological continuity and geometric consistency. This property enables the model to maintain the spatial integrity of objects and generate semantically coherent structures within the target domain style, improving the overall realism of generated images.

In summary, the proposed model achieves superior results across multiple metrics, including reconstruction error, structural similarity, and image quality. These results strongly validate the effectiveness of the graph-based cross-domain generation mechanism. By incorporating graph topological constraints in the feature space, the model facilitates efficient information transfer and structural alignment across domains, overcoming the limitations of pixel-level mapping in traditional generative methods. This approach not only provides a new technical pathway for cross-domain image translation but also offers practical support for advancing the theoretical development of structured generation.

This paper also presents a sensitivity experiment on the number of graph convolution layers to the single-metric MAE, and the experimental results are shown in Figure 2.

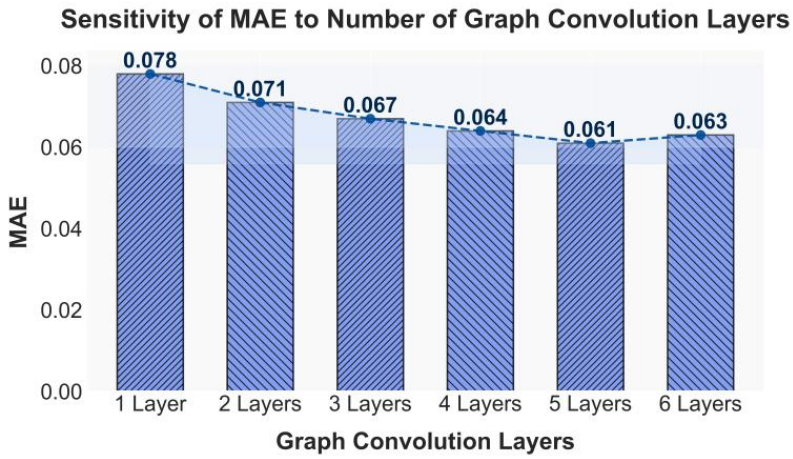


Figure 2. Sensitivity experiment of the number of graph convolution layers to single-index MAE

As shown in the figure, with the increase in the number of graph convolution layers, the model's MAE first decreases and then slightly increases. Specifically, when the number of layers increases from 1 to 5, the MAE decreases from 0.078 to its lowest value of 0.061. This indicates that adding more graph convolution layers effectively enhances the structural representation of cross-domain features, enabling the model to achieve more accurate semantic alignment across domains. This trend suggests that shallow networks cannot fully capture higher-order dependencies among graph nodes, while deeper graph convolutions strengthen semantic propagation and structural constraints, thereby reducing generation errors.

When the number of layers exceeds 5, the MAE slightly increases to 0.063, indicating that overly deep graph convolution layers may cause an over-smoothing effect, which weakens the distinction between node representations. This feature degradation makes it difficult for the model to preserve local structural differences, leading to minor information loss during cross-domain mapping. Therefore, the improvement in model performance tends to saturate or even decline beyond a certain depth, showing that proper control of the number of graph convolution layers is crucial for maintaining generation quality.

The overall trend demonstrates that the depth of the graph structure is closely related to the model's generalization ability in cross-domain generation tasks. A moderate number of layers achieves a balance between global and local information, ensuring both semantic consistency and geometric integrity. By aggregating features across graph convolution layers, the model can better capture complex structural relationships between domains, maintaining content coherence and style consistency during visual transformation. This structure-based feature propagation mechanism is one of the key reasons why the proposed method outperforms traditional pixel-level mapping approaches.

In summary, the sensitivity experiment on the number of graph convolution layers reveals the trade-off between structural modeling capacity and feature smoothing. Deeper graph structures can enhance the model's ability to capture global dependencies, while moderate depth helps avoid structural information loss caused by excessive smoothing. These findings further validate the advantages of the proposed graph-based cross-domain image generation framework in structural representation and semantic consistency, providing valuable guidance for future model design and structural optimization.

5. Conclusion

This study focuses on the task of cross-domain image generation. To address the limitations of traditional generative models in structural preservation and semantic consistency, it proposes a cross-domain image generation method based on graph-structured modeling. The proposed method introduces nodes and edges

into the feature space to explicitly model the relationships among different semantic regions, enabling high-quality feature transfer between the source and target domains. Experimental results show that the proposed method outperforms mainstream generative models across multiple evaluation metrics and effectively alleviates common problems such as semantic drift and geometric distortion in cross-domain mapping. With global constraints imposed by the graph structure, the model maintains both content consistency and style authenticity, providing a structure-driven solution for cross-domain generation tasks.

From a theoretical perspective, this research extends the modeling paradigm of cross-domain image generation by transforming traditional pixel-space mapping into relational learning in graph space. The introduction of graph convolution allows the model to capture high-order dependencies and topological features among nodes, achieving structurally consistent semantic reconstruction at the global level. This structured generation framework improves both interpretability and generalization, and it provides new directions for related tasks such as style transfer, image restoration, and structural reconstruction. The stable performance of the model across different domains demonstrates that graph-structured modeling can effectively enhance the adaptability of deep generative networks to complex distribution shifts, offering theoretical support for cross-modal visual tasks.

From an application perspective, this research has significant practical value. Cross-domain image generation plays an important role in fields such as medical image synthesis, autonomous driving adaptation, remote sensing enhancement, and virtual reality content creation. By embedding geometric and semantic constraints within the graph structure, the model can generate high-quality images under varying data sources and acquisition conditions, improving system robustness and applicability across multiple environments and tasks. In scenarios with high annotation costs or large domain discrepancies, the proposed method can effectively achieve data augmentation and domain adaptation, promoting the deployment and scalability of intelligent visual systems in real-world engineering applications.

Future research can be further extended in several directions. One direction is to explore the integration of graph-structured modeling with diffusion-based generative models and variational autoencoders to enhance fine-grained details and output diversity. Another direction is to introduce dynamic graph structures or adaptive adjacency modeling, allowing the model to automatically learn structural topologies based on input features and improve its adaptability in complex scenes. In addition, future work may incorporate multi-modal information such as text, depth maps, or semantic labels to enable controllable generation and multi-dimensional collaborative learning under cross-domain conditions. By further advancing the theory of structure-aware generation, this research is expected to promote the development of cross-domain visual generation toward intelligence, interpretability, and practical application, providing a solid technical foundation for multi-scenario artificial intelligence systems.

References

[1] Peng D, Hu P, Ke Q, et al. Diffusion-based image translation with label guidance for domain adaptive semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2023: 808-820.

[2] Li B, Xue K, Liu B, et al. Bbdm: Image-to-image translation with brownian bridge diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition. 2023: 1952-1961.

[3] Sun S, Wei L, Xing J, et al. SDDM: score-decomposed diffusion models on manifolds for unpaired image-to-image translation[C]//International Conference on Machine Learning. PMLR, 2023: 33115-33134.

[4] Kim B, Kwon G, Kim K, et al. Unpaired image-to-image translation via neural schr\" odinger bridge[J]. arXiv preprint arXiv:2305.15086, 2023.

[5] Ko M, Cha E, Suh S, et al. Self-supervised dense consistency regularization for image-to-image translation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18301-18310.

[6] J. Guo, J. Li, H. Fu, X. Jiang, S. Yang and C. Wang, "Alleviating semantics distortion in unsupervised low-level image-to-image translation via structure consistency constraint", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18249–18259, 2022.

[7] D. Gokay, E. Simsar, E. Atici, A. Ozturk and M. U. Gulec, "Graph2Pix: A graph-based image to image translation framework", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2001–2010, 2021.

[8] C. Hu, Z. Cheng, D. Wu, Y. Wang, F. Liu and Z. Qiu, "Structural generalization for microservice routing using graph neural networks", Proceedings of the 2025 3rd International Conference on Artificial Intelligence and Automation Control (AIAC), pp. 278–282, 2025.

[9] Z. Qiu, F. Liu, Y. Wang, C. Hu, Z. Cheng and D. Wu, "Spatiotemporal Traffic Prediction in Distributed Backend Systems via Graph Neural Networks", arXiv preprint arXiv:2510.15215, 2025.

[10] C. Hua, N. Lyu, C. Wang and T. Yuan, "Deep Learning Framework for Change-Point Detection in Cloud-Native Kubernetes Node Metrics Using Transformer Architecture", unpublished, 2025.

[11] R. Liu, R. Zhang and S. Wang, "Transformer-Based Modeling of User Interaction Sequences for Dwell Time Prediction in Human-Computer Interfaces", arXiv preprint arXiv:2512.17149, 2025.

[12] Y. Shu, K. Zhou, Y. Ou, R. Yan and S. Huang, "A Self-Supervised Learning Framework for Robust Anomaly Detection in Imbalanced and Heterogeneous Time-Series Data", unpublished, 2025.

[13] H. Wang, C. Nie and C. Chiang, "Attention-Driven Deep Learning Framework for Intelligent Anomaly Detection in ETL Processes", unpublished, 2025.

[14] Z. Qu, X. Tao, X. Gong, S. Qu, X. Zhang, X. Wang and G. Ding, "Dictas: A framework for class-generalizable few-shot anomaly segmentation via dictionary lookup", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 20519–20528, 2025.

[15] X. Hu, Y. Kang, G. Yao, T. Kang, M. Wang and H. Liu, "Dynamic prompt fusion for multi-task and crossdomain adaptation in LLMs", Proceedings of the 2025 10th International Conference on Computer and Information Processing Technology (ISCIPT), pp. 483–487, 2025.

[16] K. Gao, H. Zhu, R. Liu, J. Li, X. Yan and Y. Hu, "Contextual Trust Evaluation for Robust Coordination in Large Language Model Multi-Agent Systems", unpublished, 2025.

[17] T. Guan, S. Sun and B. Chen, "Faithfulness-Aware Multi-Objective Context Ranking for Retrieval-Augmented Generation", unpublished, 2025.

[18] S. Li, Y. Wang, Y. Xing and M. Wang, "Mitigating Correlation Bias in Advertising Recommendation via Causal Modeling and Consistency-Aware Learning", unpublished, 2025.

[19] S. Pan and D. Wu, "Trustworthy summarization via uncertainty quantification and risk awareness in large language models", Proceedings of the 2025 6th International Conference on Computer Vision and Data Mining (ICCVDM), pp. 523–527, 2025.

[20] X. T. Li, X. P. Zhang, D. P. Mao and J. H. Sun, "Adaptive robust control over high-performance VCM-FSM", unpublished, 2017.

[21] Gatopoulos I, Tomczak J M. Self-supervised variational auto-encoders[J]. Entropy, 2021, 23(6): 747.

[22] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[23] Rana S, Gatti M. Comparative Evaluation of Modified Wasserstein GAN-GP and State-of-the-Art GAN Models for Synthesizing Agricultural Weed Images in RGB and Infrared Domain[J]. MethodsX, 2025, 14: 103309.

[24] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10684-10695.