Structured Discrete Graph Generation Model for Fragmented Image Recovery

Yalong Zhu, **Liguo Zhang***, **Zhibo Wang**, **Ruyue Liu**,
Yue Cao*, **Yunfei Long** and **Jingang Wang**
College of Computer Science and Technology, Harbin Engineering University, China
{zhuyalong, zhangliguo, wangzb, cscaoyue, 2016064109}@hrbeu.edu.cn,
{liuruyue123, jingangw123}@gmail.com

Abstract

Fragmented image recovery is of significant importance in computer vision, such as cultural relic and artwork restoration, archival document recovery, and digital forensics. The goal is to recover the original image topology from an unordered set of fragments and spatially align and stitch them together. The adjacency relationships among fragments are discrete, sparse, and highly structured, making it difficult for traditional methods to effectively handle global topological consistency. To address this challenge, we propose a fragment adjacency recovery method based on a conditional graph diffusion model. First, we perform discrete denoising pretraining with structural masking to learn structure-aware node representations from perturbed adjacency matrices, using graph neural networks for message passing. Building on this, we design a masked discrete diffusion process tailored for fragment reconstruction, which progressively restores the connectivity between fragments. Furthermore, to enhance the controllability of the generation process, we introduce a topology-guided mechanism that steers the generation of adjacency structures via a topological scoring function, ensuring that the reconstructed fragment graph satisfies global topological constraints. Experimental results demonstrate that our method achieves state-of-the-art performance on hand-torn calligraphy, painting replica datasets and document datasets, outperforming existing approaches in both accuracy and robustness.

1 Introduction

Fragmented image recovery aims to reconstruct a complete image from a set of non-overlapping fragments, a problem that arises naturally in cultural relic restoration [Zhang *et al.*, 2023; Zhang *et al.*, 2024], archival document recovery [Salih *et al.*, 2019], and digital forensics [Paixão *et al.*, 2022]. Unlike conventional image stitching that relies on overlapping regions for alignment, fragment reassembly must infer spatial relationships purely from fragment boundaries and visual compatibility, making it a structured prediction problem over discrete adjacency relationships[Tang *et al.*, 2020].

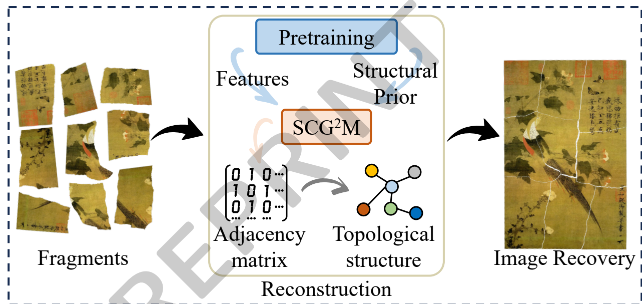


Figure 1: Overview of the fragment reconstruction pipeline.

The core challenge lies in balancing local compatibility with global topological consistency. While pairwise fragment matching can identify locally plausible connections, it lacks awareness of global structural constraints, such as sparsity, bounded degrees, and overall connectivity [Abitbol *et al.*, 2021; Pirrone *et al.*, 2021]. As a result, greedy or heuristic-based assembly strategies often propagate local errors, leading to topologically invalid configurations. This limitation becomes particularly severe in large-scale scenarios, where a single incorrect adjacency decision may cascade into widespread structural failure, while exhaustive combinatorial search remains computationally infeasible.

From a structural perspective, fragment adjacency relationships are inherently discrete, sparse, and highly constrained. Conventional reconstruction methods struggle to effectively model such structured dependencies, especially when global connectivity and topology consistency must be preserved. Although recent generative models, including GANs [Talon *et al.*, 2025], VAEs [Liu *et al.*, 2022; Liu *et al.*, 2024b], diffusion models [Xie *et al.*, 2023; Liu *et al.*, 2024a], and large vision-language models [Wu *et al.*, 2023; Yu *et al.*, 2023], can synthesize visually plausible content, the generated pixels lack objective authenticity. In sensitive applications such as art restoration and forensic analysis, relying on hallucinated content is unacceptable, as it may compromise historical credibility or legal validity. Consequently, most existing studies focus on recovering the geo-

*Corresponding author.

metric and topological structure of fragments rather than generating missing content.

To address this challenge, we propose a Structured conditional graph Generation framework (as show in Figure 1.) that formulates fragment adjacency recovery as reversing a controlled corruption process on graph structure. Our key insight is to leverage the iterative refinement nature of diffusion models to progressively denoise masked adjacency matrices while explicitly enforcing global topological constraints. Specifically, we introduce a structure-masked discrete denoising pretraining scheme that perturbs adjacency matrices and employs graph message passing to learn topology-aware node representations. Building upon this foundation, we design a masked discrete diffusion process that reconstructs fragment connectivity directly in adjacency space, enabling globally consistent topology recovery. Furthermore, we propose a topology-guided sampling mechanism that incorporates sparsity, degree constraints, and connectivity priors through explicit structural scoring, providing principled controllability during inference.

Extensive experiments on three public benchmarks and challenging real-world datasets, including manually torn artwork replicas and shredded documents, demonstrate that the proposed method consistently outperforms existing approaches in both reconstruction accuracy and robustness.

Our main contributions are summarized as follows:

- By formulating pretraining as structure-masked discrete denoising on adjacency matrices, our approach learns transferable topology-aware representations that capture both local pairwise compatibility and global graph structure, eliminating the need for hand-crafted features.
- By operating directly in discrete adjacency space with iterative diffusion-based refinement, our framework progressively reconstructs fragment connections through learned reverse processes, naturally preserving global topological consistency via graph Transformer message passing without requiring post-hoc correction.
- Through topology-guided sampling with explicit structural scoring, our method enforces hard constraints on sparsity, degree distribution, and connectivity during inference, providing principled controllability that bridges the gap between learned generation and domain requirements.

2 Related Work

2.1 Fragmented Image Recovery

Fragmented image recovery aims to reconstruct a complete image from a set of non-overlapping fragments and has been extensively studied in traditional jigsaw puzzles, shredded documents, and irregular fragment reconstruction. Early approaches primarily relied on geometric shape cues [Hoff and Olver, 2014] or hand-crafted visual features to establish fragment correspondences. With the development of deep learning, recent methods employ neural networks to predict fragment compatibility, relative positions, or matching scores [Sokic *et al.*, 2024].

Several learning-based approaches have been proposed, including coordinate regression models based on vision transformers [Kim *et al.*, 2025], evolutionary reinforcement learning frameworks [Song *et al.*, 2025], and compatibility-based matching networks [Rika *et al.*, 2025]. For mechanically shredded documents with regular boundaries, prior work explored spatial feature classification, gray-level co-occurrence analysis [Huang, 2025], and human-in-the-loop recommendation systems [Paixão *et al.*, 2022]. Self-supervised learning has also been introduced for fragment boundary matching, such as S3-Net based on shape similarity [Zhang *et al.*, 2022] and transformer-based models with rotation-aware positional encoding for adjacency prediction [Dongmo *et al.*, 2025]. To address fragment alignment and matching [Yang *et al.*, 2023], ReassembleNet [Islam *et al.*, 2025] reduces computational complexity using contour keypoints and graph neural networks, while PairingNet [Zhou *et al.*, 2024] integrates graph models with linear transformers for fragment pair search. Other representative methods include JigsawGAN [Li *et al.*, 2021], Jigsaw-ViT [Chen *et al.*, 2023], and JigsawNet [Le and Li, 2019]. However, except for JigsawNet, most of these methods are designed for regular square puzzles with limited boundary complexity, making them less effective for irregular fragments. CG^2AN [Zhang *et al.*, 2025b] formulates fragment reassembly as a graph reconstruction problem and generates adjacency matrices using conditional graph adversarial networks. While graph-based formulations provide a promising direction, existing methods still struggle to recover globally consistent topology under irregular boundaries, noise, and large-scale fragment settings, motivating the need for more structured and topology-aware graph generation approaches.

2.2 Discrete Graph Diffusion Models

Diffusion models have recently emerged as powerful generative frameworks, achieving remarkable success in image synthesis and structured data generation by learning a reverse denoising process from progressively corrupted data [Ho *et al.*, 2020]. This paradigm has been extended to graph-structured data, leading to graph diffusion models that explicitly model joint distributions over nodes and edges while preserving permutation invariance and enabling conditional generation [Niu *et al.*, 2020; Jo *et al.*, 2022; Vignac *et al.*, 2022; Huang *et al.*, 2022; Huang *et al.*, 2023]. Such models have demonstrated strong performance in applications including molecular graph generation and physical simulation [Hoogeboom *et al.*, 2022; Valencia *et al.*, 2025].

Graph diffusion methods can be broadly categorized into continuous embedding diffusion and discrete structure diffusion. The latter directly operates on discrete graph components, such as edge existence, making it particularly suitable for fragment adjacency recovery tasks with inherently discrete and sparse structures [Austin *et al.*, 2021].

However, existing graph diffusion models largely target general graph domains and insufficiently incorporate visual semantics or strong domain-specific topological priors, such as sparsity and global connectivity. DiffAssemble [Scarpellini *et al.*, 2024] applies diffusion to image assembly but is limited to idealized regular fragments and does not address real-world challenges such as boundary uncertainty, degradation, or missing fragments.

2.3 Topology-Guided Mechanism

The sampling stage of diffusion models plays a critical role in determining the quality and plausibility of generated structures. Unconstrained sampling may lead to topologically implausible graphs that violate domain-specific priors. To address this issue, recent studies have explored incorporating topological guidance during the sampling process. In topology optimization, TopoDiff [Mazé and Ahmed, 2023] introduces conditional diffusion with surrogate-guided sampling to generate structures that balance performance and manufacturability. In mesh analysis, MTGNet [Zhang *et al.*, 2025a] employs topology-aware graph neural networks to generate structured mesh representations. For trajectory generation, ControlTraj [Zhu *et al.*, 2024] integrates diffusion models with topological constraints imposed by road networks to enhance geographic consistency. Inspired by these works, we introduce topology-guided sampling tailored to fragment adjacency reconstruction. By explicitly incorporating sparsity, local degree constraints, and global connectivity into a unified scoring function, our method adjusts the reverse diffusion process without modifying model training. This design ensures that the generated adjacency structures satisfy task-specific global topological requirements while preserving stochasticity and flexibility during generation.

3 Methodology

In this section, we introduce the proposed SCG^{2}M framework in detail, including problem formulation, structured discrete graph diffusion, and topology-guided sampling.

3.1 Problem Formulation

Given a set of fragment images $\mathcal{X} = x_{i}{}_{i=1}^{N}$, our goal is to recover the true adjacency structure among the fragments, which serves as a topological constraint for subsequent geometric alignment and global reconstruction. We model the fragment set as an undirected graph $G = (V, E)$, where the node set $V = \{1, 2, \dots, N\}$ corresponds to individual fragments, and the edge set is represented by a binary symmetric adjacency matrix $E^{(0)} \in \{0, 1\}^{N\times N}$. An entry $E^{(0)}_{ij} = 1$ indicates that fragments i and j are adjacent in the target structure. Unlike continuous geometric parameters, fragment adjacency relations are discrete, sparse, and highly structured: each fragment is connected to only a small number of neighbors, and strong dependencies exist among different edges. Treating edge prediction as independent binary classification fails to capture global topological consistency. To address this issue, we formulate fragment structure recovery as a discrete graph generation problem, where the goal is to learn the underlying distribution of adjacency matrices and model it using a denoising paradigm. This enables the model to capture high-order structural dependencies among fragments beyond local edge-wise decisions.

3.2 Structured-Masked Pretraining

To learn structure-aware node representations, we introduce a structure-masked pretraining strategy, as illustrated in Figure 2(a). Given the ground-truth adjacency matrix $E^{(0)}$, we randomly sample a set of node pairs $\mathcal{S} \subseteq V \times V$, and apply symmetric discrete perturbations to construct a corrupted adjacency matrix $\tilde{E}$:

$$\tilde{E}_{ij} = \begin{cases} E^{(0)}_{ij}, & \text{w.p. } 1-\epsilon, \\ 1 - E^{(0)}_{ij}, & \text{w.p. } \epsilon, \end{cases} (i, j) \in \mathcal{S}, \tag{1}$$

with $\tilde{E}_{ij} = \tilde{E}_{ji}$. This symmetric flipping noise perturbs existing edges and non-edges with equal probability, introducing local structural uncertainty without imposing systematic bias. As a result, the model is forced to infer masked relations from the contextual structure rather than memorizing individual edge labels.

Each fragment image is affinely normalized (AT) and processed by the deformable convolutional network (DCN) to extract visual features, which are mapped to initial node embeddings. The Message passing is then performed on the corrupted graph $\tilde{E}$ using a graph neural network. After L propagation layers, we obtain node embeddings$H^{(L)} = \{h^{(L)}_i \mid i \in V\}$. For each pair of masked nodes$(i, j) \in \mathcal{S}$, the model predicts the edge probability $\hat{p}_{ij} = \sigma\big(\phi([h^{(L)}_i ; h^{(L)}_j])\big)$, where $\phi(\cdot)$ denotes a learnable edge prediction function and $\sigma(\cdot)$ is the sigmoid function. The pretraining objective is a weighted Bernoulli negative log-likelihood:

$$\mathcal{L}_{\text{pre}} = -\sum_{(i,j)\in\mathcal{S}} w_{ij} \big[E^{(0)}_{ij} \log \hat{p}_{ij} + (1 - E^{(0)}_{ij}) \log(1 - \hat{p}_{ij})\big], \tag{2}$$

where the weights w_{ij} mitigate the imbalance between positive and negative edges.

After pretraining, the predicted probabilities $\hat{p}_{ij}$ form a soft structural prior $C = \hat{p}_{ij}$, which encodes latent dependencies among nodes. The adjacency matrix is only used as supervision during pretraining; in the subsequent generation stage, no adjacency information is provided. Instead, the learned structural prior C guides the discrete graph diffusion denoising process.

3.3 Structured Conditional Discrete Graph Generation Model

With structure-aware node representations and the soft structural prior C, we propose a structured discrete conditional graph generation model to generate the global fragment topology in the adjacency matrix space. Given the sparsity and limited connectivity of fragment adjacency graphs, we adopt a mask-based discrete diffusion framework, which progressively removes structural information and restores it under the guidance of the conditional prior.

Let $E^{(0)} \in 0, 1^{N\times N}$ denote the original adjacency matrix. In the forward diffusion process, edges are gradually masked to an unknown symbol $\varnothing$, simulating progressive structural degradation, as shown in Figure 2(b). To preserve undirected graph constraints, perturbations are applied only to edges with $i < j$, while enforcing $E^{(t)}_{ii} = 0$ and $E^{(t)}_{ji} = E^{(t)}_{ij}$.

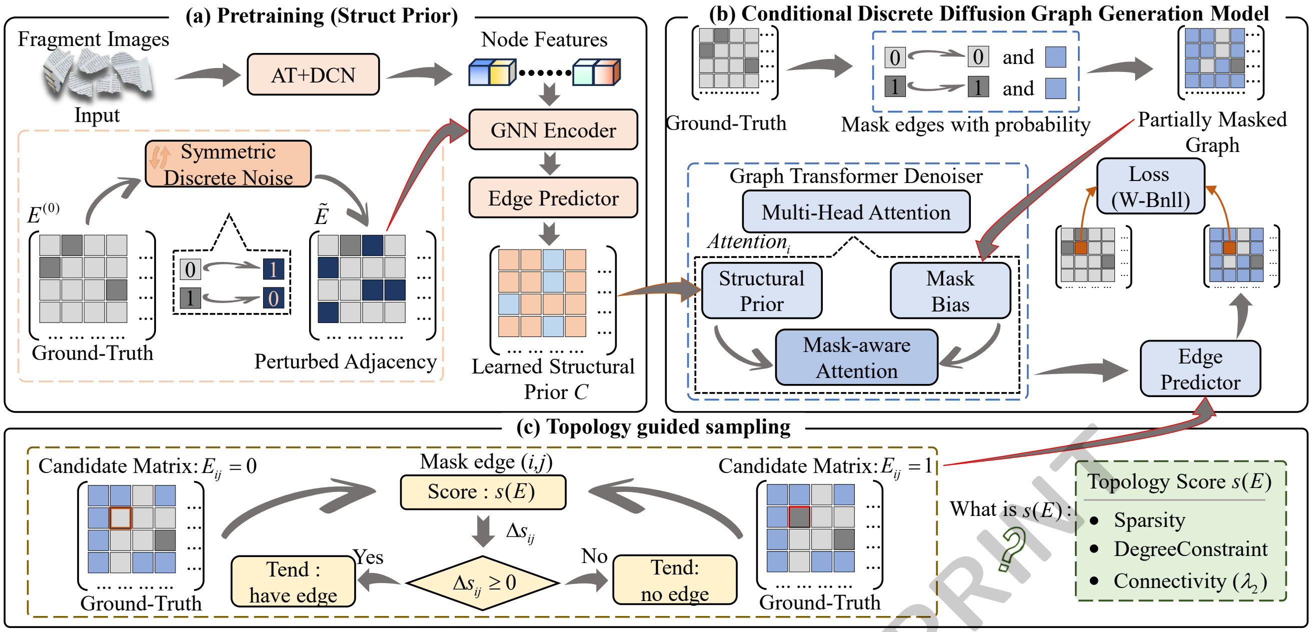


Figure 2: Overview of the proposed topology-guided conditional discrete diffusion framework for graph generation. **(a)**: Structural prior pretraining via symmetric discrete perturbation of adjacency matrices, where a GNN encoder and edge predictor capture global structural dependencies. **(b)**: Conditional discrete diffusion-based graph generation, in which a mask-aware graph transformer iteratively denoises partially observed graphs by integrating conditional cues and learned structural priors. **(c)**: Topology-guided sampling that explicitly incorporates sparsity, degree regularization, and connectivity constraints to guide edge selection during reverse diffusion.

At diffusion step t, the transition probability is defined as

$$q\left(E_{ij}^{(t)} \mid E_{ij}^{(t-1)}\right) = \begin{cases} 1-\beta_t, & E_{ij}^{(t)} = E_{ij}^{(t-1)}, \\ \beta_t, & E_{ij}^{(t)} = \varnothing, \\ 1, & E_{ij}^{(t-1)} = \varnothing, \end{cases} \tag{3}$$

where $\beta_t \in (0,1)$ is the masking probability at step t. As t increases, observable structure gradually vanishes, and the adjacency matrix degenerates into a representation dominated by unknown states. The forward process defines a valid Markov chain, while the reverse process is parameterized as a conditional denoising model that directly predicts the original adjacency relations rather than explicitly reversing the chain.

During reverse denoising, the model progressively restores masked relations under the guidance of the structural prior C. At each diffusion step, it predicts $\hat{E}_{ij}^{(0)}(t) = p_\theta(E_{ij}^{(0)} = 1 \mid E^{(t)}, C)$, enabling recovery of the original topology from partial structures. Unlike conventional denoising approaches, the model is trained to recover the same original graph $E^{(0)}$ from all noise levels, ensuring global topological consistency.

To enhance denoising performance, we adopt a graph Transformer architecture. Fragment images are processed by a DCN to extract visual features and mapped to initial node embeddings $h_i^{(0)}$. Node representations are then updated through MLPs and multi-layer graph Transformer attention.To explicitly encode masking information, we introduce a binary indicator matrix $m^{(t)}$, where $m_{ij}^{(t)} = 1$ indicates a masked edge and $m_{ij}^{(t)} = 0$ denotes a known state.

The masking indicator and structural prior jointly influence attention computation, preventing masked edges from being misinterpreted as negative connections.

At each attention layer, queries, keys, and values are defined as $Q_i = W_Q h_i^{(l)}$, $K_j = W_K h_j^{(l)}$, $V_j = W_V h_j^{(l)}$. Using graph multi-head attention (GMHA), the attention score integrates structural and masking biases:

$$\mathrm{Attn}_{ij} = \frac{Q_i K_j^\top}{\sqrt{d}} + \gamma C_{ij} - \delta m_{ij}^{(t)}, \tag{4}$$

where γ controls the influence of the soft structural prior and δ adjusts the reliability of masked positions. After Softmax normalization, node updates (with time embedding e_t) are computed as

$$\tilde{H}^{(l+1)} = \mathrm{GMHA}(H^{(l)}, C, m^{(t)}, t) + H^{(l)} + W_t e_t, \tag{5}$$

followed by a feed-forward network:

$$H^{(l+1)} = \mathrm{FFN}(\tilde{H}^{(l+1)}) + \tilde{H}^{(l+1)}. \tag{6}$$

After stacking L layers, the final node representations $H^{(L)}$ are used to predict edge probabilities:

$$\hat{E}_{ij}^{(0)}(t) = \sigma\left(\phi\left([h_i^{(L)}; h_j^{(L)}]\right)\right), \tag{7}$$

where $\phi(\cdot)$ is a two-layer MLP. During training, supervision is applied only to edges masked at the current diffusion step $M_t = \{(i,j) \mid E_{ij}^{(t)} = \varnothing\}$, by minimizing a weighted

Bernoulli negative log-likelihood:

$$\mathcal{L}_{\text{diff}} = -\mathbb{E}_t \sum_{(i,j)\in M_t} w_{ij} \Big[E_{ij}^{(0)} \log \hat{E}_{ij}^{(0)}(t) + (1 - E_{ij}^{(0)}) \log(1 - \hat{E}_{ij}^{(0)} \quad (8)$$

Through discrete denoising, the model integrates local visual cues with global structural priors to recover topologically consistent fragment adjacency graphs, which subsequently guide geometric alignment and image stitching.

3.4 Topology-Guided Sampling

Although the discrete conditional graph generation model in Section 3.3 can predict fragment adjacency under joint visual and structural conditions, independently sampling edges based solely on local probabilities may lead to globally implausible structures during generation, such as overly dense connections, excessive node degrees, or disconnected graphs. These issues arise not from poor local edge prediction, but from the lack of explicit global coordination among local decisions.

To address this, we introduce a topology-guided sampling mechanism during reverse denoising, as shown in Figure 2(c). Without modifying model parameters or training, this mechanism explicitly adjusts the sampling process to enforce global topological priors.

At reverse step t, given the current structure $E^{(t)}$, the model predicts $p_\theta(E^{(0)} \mid E^{(t)}, C)$. Topology-guided sampling rewrites the generation objective as maximizing

$$\log p_\theta(E^{(t-1)} \mid E^{(t)}, C) + \lambda_{\text{guide}} s(E^{(t-1)}), \quad (9)$$

where the first term represents model confidence and the second term $s(E^{(t-1)})$ measures global topological plausibility. The coefficient λ_{guide} balances prediction fidelity and topological constraints. This score is used only during sampling and does not affect training or backpropagation.

The topology score is defined as a linear combination of computable structural metrics:

$$s(E) = \alpha_{\text{sparse}}\, s_{\text{sparse}}(E) + \alpha_{\text{deg}}\, s_{\text{deg}}(E) + \alpha_{\text{conn}}\, s_{\text{conn}}(E), \quad (10)$$

corresponding to sparsity, degree constraints, and global connectivity. To discourage redundant connections, we define a sparsity score:

$$s_{\text{sparse}}(E) = -\sum_{i<j} E_{ij}. \quad (11)$$

To limit excessive local connectivity, the degree of node i is

$$d_i(E) = \sum_{j\neq i} E_{ij}, \quad (12)$$

and a soft degree constraint is imposed:

$$s_{\text{deg}}(E) = -\sum_i \big[\max(0, d_i(E) - d_{\max})\big]^2. \quad (13)$$

Global connectivity is encouraged using a Laplacian spectral measure. Given the symmetric normalized Laplacian

$$L(E) = I - D^{-1/2} E D^{-1/2}, \quad (14)$$

the connectivity score is

$$s_{\text{conn}}(E) = \lambda_2(L(E)), \quad (15)$$

where $\lambda_2(\cdot)$ is the second smallest eigenvalue, which is zero for disconnected graphs and increases with connectivity. During reverse sampling, for each masked edge (i, j), we evaluate the score difference

$$\Delta s_{ij} = s(E \text{ with } E_{ij} = 1) - s(E \text{ with } E_{ij} = 0), \quad (16)$$

which adjusts the predicted edge probability to favor decisions that improve global topology while preserving stochasticity.

By incorporating topology-guided sampling, the model explicitly enforces global structural preferences during generation, effectively mitigating topological degeneration caused by accumulated local predictions and producing adjacency structures better suited for fragment reassembly.

4 Experiments

4.1 Datasets and Experimental Settings

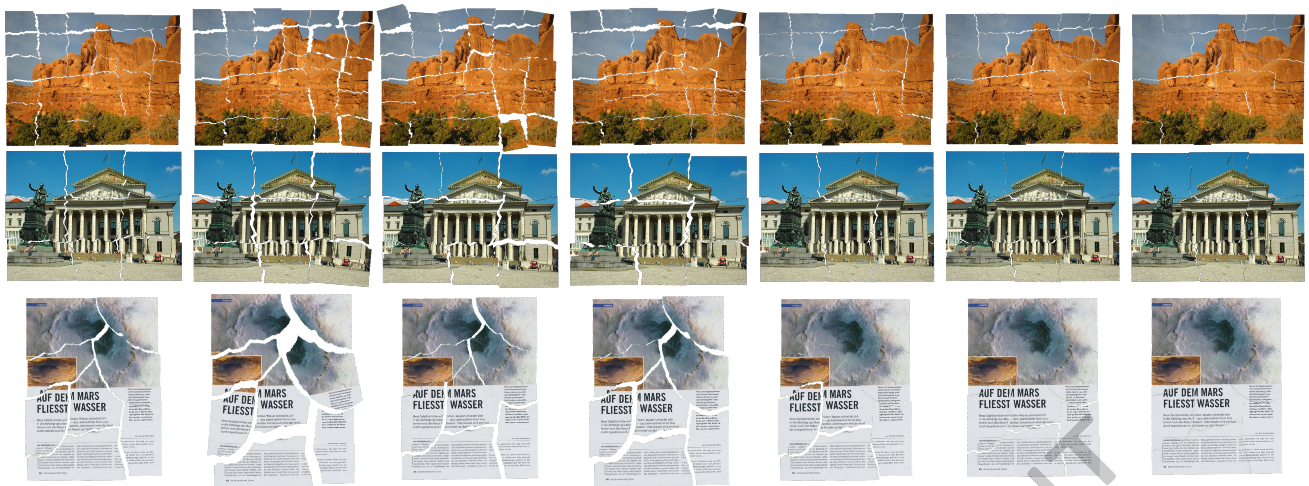
To evaluate the effectiveness of the proposed Conditional Graph Diffusion Model, we conduct experiments on both public benchmarks and real-world datasets: (i) BGU and MIT, containing digitally cropped regular fragments with well-defined boundaries as controlled testbeds; (ii) BDW, consisting of scanned manually torn scientific journal pages with highly irregular boundaries representing realistic degradation; and (iii) self-collected datasets including manually torn calligraphy, painting reproductions, and fragmented documents to evaluate robustness under diverse fragment types. All images are resized to 256×256. The DCN feature extractor is trained for 1000 epochs with batch size 32. Each dataset is split with 80% for training and 20% for testing, with data augmentation with data augmentation (random erosion, rotation, brightness perturbation) applied during training.

4.2 Evaluation Metrics

For fair comparison with existing fragment reassembly methods, we adopt standard evaluation metrics widely used in prior work, assessing performance from two complementary perspectives: topological consistency of predicted adjacency structures and visual quality of final assembly results.

Topological Consistency. This metric measures the similarity between the predicted adjacency matrix and the ground truth. We adopt Cosine Similarity(CS) and Jaccard Index(JI) to quantify adjacency matrix similarity.

Assembly Accuracy. This metric evaluates both local matching quality and global reconstruction validity through three metrics: (i) Pose Correct Rate (PCR), measuring the proportion of fragments with correct spatial alignment; (ii) Alignment Correct Rate (ACR), defined as the ratio of correctly predicted pairwise adjacencies in the final assembly; and (iii) Error Rate (ER), computed as $ER = (R - S)/S$, where R denotes the area of the assembled result and S is the sum of all fragment areas. Higher PCR and ACR indicate superior local matching, while lower ER reflects effective suppression of accumulated global errors, ensuring topologically valid reconstruction.



JigsawNet | JigsawGAN | Pirrone's | Jigsaw-ViT | ParingNet | CG²AN | Ours

Figure 3: Qualitative comparison of fragment reassembly results produced by seven methods on the BGU, MIT, and BDW datasets. While most methods can recover fragment adjacency to some extent, inaccurate alignment and stitching errors are still observed. In contrast, the proposed method reconstructs more coherent structures with fewer alignment errors, especially on hand-torn fragments.

	BGU		MIT		BDW	
Method	CS	JI	CS	JI	CS	JI
JigsawNet	86.3	74.1	88.7	76.5	71.8	59.4
PairingNet	89.5	77.9	91.2	80.6	78.6	66.8
CG2AN	91.8	81.4	93.5	83.7	82.3	70.9
Ours	**94.2**	**85.6**	**96.1**	**87.9**	**89.7**	**79.3**

Table 1: Topological consistency comparison on different datasets (%). Best results are in **bold**.

4.3 Comparison Experiments

Table 1 presents the topological consistency results across all datasets in terms of CS and JI. Our method consistently outperforms all baselines, achieving the highest structural accuracy on both metrics. Notably, on the challenging BDW dataset with irregular fragment boundaries, our approach surpasses the best competing method CG^{2}AN by 7.4% in CS and 8.4% in JI, demonstrating superior robustness to boundary irregularity and shape variability. Similar improvements are observed on BGU and MIT datasets. These results validate that the proposed graph diffusion framework, combined with topology-guided sampling, effectively recovers globally consistent adjacency structures that prior deterministic or adversarial approaches fail to capture.

To evaluate assembly quality, we randomly sample 50 images per dataset and partition them into 36, 9, and 8 fragments for BGU, MIT, and BDW, respectively, following the assembly protocol of CG^{2}AN. Table 2 reports the quantitative results in terms of PCR, ACR and ER, while Figure 3 provides qualitative comparisons of reconstruction results. Several baseline methods produce invalid assemblies on certain samples; these failure cases are excluded from metric computation but are visible in the qualitative visualizations. Our method consistently achieves the best performance across all datasets and metrics. We observe several key findings: (i) Methods relying on regular structure assumptions (Jigsaw-ViT, JigsawGAN) perform reasonably on BGU and MIT but suffer significant degradation on BDW, with PCR dropping by 10-17%, as irregular boundaries violate their grid-based assumptions. (ii) Domain-specific methods such as Pirrone achieve competitive performance on BDW (ACR: 80.1%) but generalize poorly to digitally cut datasets (ACR: 73.3-76.9%), indicating over-specialization. (iii) Local matching methods (JigsawNet, PairingNet) demonstrate stable pairwise accuracy but struggle with global consistency, reflected in higher ER (6.1-7.8% vs. our 4.8-5.1%). (iv) Graph generation methods (CG^{2}AN) achieve strong baseline performance but still produce suboptimal structures due to lack of explicit topological constraints.

In contrast, our method progressively recovers fragment adjacency through a graph diffusion process combined with topology-guided sampling, ensuring global topological consistency. On the BDW dataset, our approach improves assembly accuracy by at least 10% over CG^{2}AN, demonstrating strong robustness in irregular fragment scenarios.

4.4 Robustness and Ablation Studies

To validate the effectiveness of each component, we systematically ablate graph diffusion (GD), topology-guided sampling (TGS), and structure-masked pretraining (SMP). Table 3 reports the results. We observe three key findings:

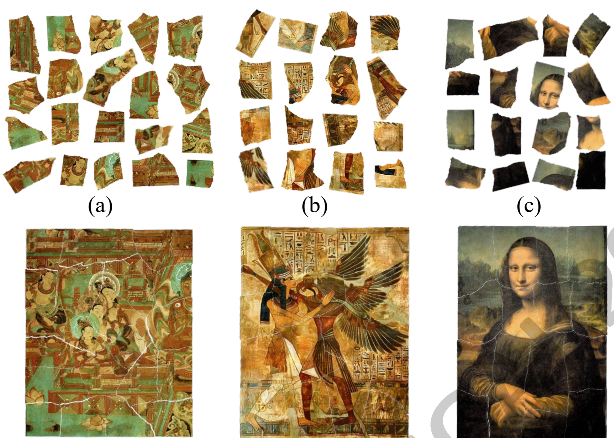
(i) Removing GD (i.e., w/o GD) causes the most significant performance drop, with CS and JI decreasing by 15% and 18%, respectively. Without diffusion, the model degenerates to deterministic prediction, failing to capture the inherent ambiguity in fragment relationships. (ii) Disabling TGS (i.e., w/o TGS) leads to moderate degradation, with CS and JI dropping by 5% and 6%, respectively. While the diffusion process remains intact, the lack of explicit structural con-

Method	PCR (%) ↑			ACR (%) ↑			ER (%) ↓		
	BGU	MIT	BDW	BGU	MIT	BDW	BGU	MIT	BDW
JigsawNet	85.2	90.1	76.3	89.4	85.7	78.5	9.8	8.5	12.3
JigsawGAN	79.4	72.8	62.5	82.1	79.3	68.2	11.1	10.4	15.6
Pirrone	70.5	81.9	74.8	73.3	76.9	80.1	15.4	14.3	11.5
Jigsaw-ViT	73.1	74.9	65.3	75.5	79.2	71.1	13.6	12.8	15.4
PairingNet	91.3	94.5	88.2	94.2	95.3	89.7	6.1	5.4	7.8
CG^{2}AN	93.2	97.4	92.5	96.1	97.3	92.1	5.3	4.3	6.2
Ours	**94.7**	**98.2**	**93.8**	**97.5**	**98.4**	**94.7**	**4.8**	**3.7**	**5.1**

Table 2: Comparison of assembly accuracy on different datasets. ↑ indicates higher is better, ↓ indicates lower is better.

Method	CS(%)	JI(%)	RT (min)
GD+TGS+SMP	**88**	**79**	18.3
w/o TGS	84	74	16.9
w/o SMP	81	72	17.2
w/o GD	75	65	15.4

Table 3: Ablation results on topological consistency (CS, JI) and runtime (RT).



(a) (b) (c)
(d) (e) (f)

Figure 4: Reconstruction results on real-world artwork datasets. (a, b, c) Input fragments from Dunhuang murals, Egyptian wall paintings, and Portrait oil paintings. (d, e, f) Corresponding reconstructions by our method.

straints allows generation of invalid graphs that violate sparsity or connectivity requirements, demonstrating that hard topological priors are essential. (iii) Removing SMP (i.e., w/o SMP) causes 8% and 9% performance loss in CS and JI. This indicates that large-scale self-supervised learning of structural patterns significantly improves generalization across diverse fragment shapes, compared to training from scratch. Regarding computational efficiency, ablated variants reduce average runtime (RT) by 1.1-2.9 minutes due to simplified architectures, but at the cost of substantial accuracy degradation, validating the necessity of each component.

In applications such as art restoration, we further evaluate our method on self-collected datasets containing fragments from Portrait oil paintings, Dunhuang murals, and ancient

(a) (b) (c)

Figure 5: Robustness evaluation under image degradation and missing fragments.

Egyptian wall paintings, as illustrated in Figure 4. These datasets simulate challenging real-world restoration scenarios. Our method achieves high reconstruction accuracy and topological consistency, effectively restoring damaged artworks while preserving authenticity.

To further assess robustness, we conduct three perturbation experiments: overexposure degradation, missing fragments, and color contamination of fragments. As shown in Figure 5(a), even when fragment images suffer from severe overexposure, our model successfully recovers the underlying adjacency topology, demonstrating strong robustness to boundary degradation. In Figure 5(b), we simulate the absence of two fragments; despite incomplete data, the reconstruction quality remains stable with no significant performance degradation. Figure 5(c) further illustrates a scenario where fragment colors are heavily contaminated, yet the proposed method remains capable of inferring correct fragment adjacency, indicating robustness to appearance corruption. These results confirm that our method maintains reliable reconstruction performance under noise and incomplete data conditions.

5 Conclusion

We propose SCG^{2}M, a structured discrete graph generation framework for fragmented image recovery. It formulates fragment reassembly as a conditional graph diffusion process to model adjacency relationships with visual and topological guidance. A topology-guided sampling strategy further constrains sparsity, node degrees, and global connectivity, reducing structural errors. Experiments on benchmarks and real-world cases show that SCG^{2}M achieves state-of-the-art performance and strong robustness under degradation and incomplete data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62502112), the Hainan Provincial Natural Science Foundation of China (No. 626QN0683), the Outstanding Master's and Doctoral Dissertations Program of Longjiang in the New Era (No. LJYXLZR2025-045), the Fundamental Research funds for the Central Universities (No. 3072025YC0602).

References

[Abitbol *et al.*, 2021] Roy Abitbol, Ilan Shimshoni, and Jonathan Ben-Dov. Machine learning based assembly of fragments of ancient papyrus. *Journal on Computing and Cultural Heritage (JOCCH)*, 14(3):1–21, 2021.

[Austin *et al.*, 2021] Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems*, 34:17981–17993, 2021.

[Chen *et al.*, 2023] Yingyi Chen, Xi Shen, Yahui Liu, Qinghua Tao, and Johan AK Suykens. Jigsaw-vit: Learning jigsaw puzzles in vision transformer. *Pattern Recognition Letters*, 166:53–60, 2023.

[Dongmo *et al.*, 2025] Kevin Dongmo, Fadel Toure, Alain Goupil, and Etienne Beaulac. Fragments adjacency prediction: A contour based approach using transformer models with rotary positional encoding. In *Proceedings of the 2025 3rd International Conference on Machine Learning and Pattern Recognition*, pages 50–58, 2025.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[Hoff and Olver, 2014] Daniel J Hoff and Peter J Olver. Automatic solution of jigsaw puzzles. *Journal of mathematical imaging and vision*, 49(1):234–250, 2014.

[Hoogeboom *et al.*, 2022] Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pages 8867–8887. PMLR, 2022.

[Huang *et al.*, 2022] Han Huang, Leilei Sun, Bowen Du, Yanjie Fu, and Weifeng Lv. Graphgdp: Generative diffusion processes for permutation invariant graph generation. In *2022 IEEE International Conference on Data Mining (ICDM)*, pages 201–210. IEEE, 2022.

[Huang *et al.*, 2023] Han Huang, Leilei Sun, Bowen Du, and Weifeng Lv. Conditional diffusion based on discrete graph structures for molecular graph generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4302–4311, 2023.

[Huang, 2025] Mingyang Huang. Reconstruction of shredded figures based on grayscale matrix. *International Journal of Pattern Recognition and Artificial Intelligence*, 2025.

[Islam *et al.*, 2025] Adeela Islam, Stefano Fiorini, Stuart James, Pietro Morerio, and Alessio Del Bue. Reassemblenet: Learnable keypoints and diffusion for 2d fresco reconstruction. *arXiv preprint arXiv:2505.21117*, 2025.

[Jo *et al.*, 2022] Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International conference on machine learning*, pages 10362–10383. PMLR, 2022.

[Kim *et al.*, 2025] Garam Kim, Hyeonseong Cho, and Hyoungsik Nam. Solving jigsaw puzzles by predicting fragment's coordinate based on vision transformer. *Expert Systems with Applications*, 272:126776, 2025.

[Le and Li, 2019] Canyu Le and Xin Li. Jigsawnet: Shredded image reassembly using convolutional neural network and loop-based composition. *IEEE Transactions on Image Processing*, 28(8):4000–4015, 2019.

[Li *et al.*, 2021] Ru Li, Shuaicheng Liu, Guangfu Wang, Guanghui Liu, and Bing Zeng. Jigsawgan: Auxiliary learning for solving jigsaw puzzles with generative adversarial networks. *IEEE Transactions on Image Processing*, 31:513–524, 2021.

[Liu *et al.*, 2022] Qiankun Liu, Zhentao Tan, Dongdong Chen, Qi Chu, Xiyang Dai, Yinpeng Chen, Mengchen Liu, Lu Yuan, and Nenghai Yu. Reduce information loss in transformers for pluralistic image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11347–11357, 2022.

[Liu *et al.*, 2024a] Jinyang Liu, Wondmgezahu Teshome, Sandesh Ghimire, Mario Sznaier, and Octavia Camps. Solving masked jigsaw puzzles with diffusion vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23009–23018, 2024.

[Liu *et al.*, 2024b] Qiankun Liu, Yuqi Jiang, Zhentao Tan, Dongdong Chen, Ying Fu, Qi Chu, Gang Hua, and Nenghai Yu. Transformer based pluralistic image completion with reduced information loss. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(10):6652–6668, 2024.

[Mazé and Ahmed, 2023] François Mazé and Faez Ahmed. Diffusion models beat gans on topology optimization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 9108–9116, 2023.

[Niu *et al.*, 2020] Chenhao Niu, Yang Song, Jiaming Song, Shengjia Zhao, Aditya Grover, and Stefano Ermon. Permutation invariant graph generation via score-based generative modeling. In *International conference on artificial intelligence and statistics*, pages 4474–4484. PMLR, 2020.

[Paixão *et al.*, 2022] Thiago M Paixão, Rodrigo F Berriel, Maria CS Boeres, Alessandro L Koerich, Claudine Badue, Alberto F De Souza, and Thiago Oliveira-Santos. A human-in-the-loop recommendation-based framework for reconstruction of mechanically shredded documents. *Pattern Recognition Letters*, 164:1–8, 2022.

[Pirrone *et al.*, 2021] Antoine Pirrone, Marie Beurton-Aimar, and Nicholas Journet. Self-supervised deep metric learning for ancient papyrus fragments retrieval. *International Journal on Document Analysis and Recognition (IJDAR)*, 24(3):219–234, 2021.

[Rika *et al.*, 2025] Daniel Rika, Dror Sholomon, Eli David, Alexandre Pais, and Nathan S Netanyahu. A generic hybrid framework for 2d visual reconstruction. *arXiv preprint arXiv:2501.19325*, 2025.

[Salih *et al.*, 2019] Nbhan D Salih, Wan Noorshahida Mohd Isa, and Abdelouahab Abid. An efficient method for hand-torn document reconstruction. *International Journal of Engineering Research and Technology*, 12(7):986–992, 2019.

[Scarpellini *et al.*, 2024] Gianluca Scarpellini, Stefano Fiorini, Francesco Giuliari, Pietro Moreiro, and Alessio Del Bue. Diffassemble: A unified graph-diffusion model for 2d and 3d reassembly. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28098–28108, 2024.

[Sokic *et al.*, 2024] Emir Sokic, Isam Vrce, Armin Zunic, Nedim Osmic, and Almir Salihbegovic. Automated image processing approach for solving semi-apictorial 2d puzzles. In *2024 International Symposium ELMAR*, pages 351–354. IEEE, 2024.

[Song *et al.*, 2025] Xingke Song, Xiaoying Yang, Chenglin Yao, Jianfeng Ren, Ruibin Bai, Xin Chen, and Xudong Jiang. Erl-mpp: Evolutionary reinforcement learning with multi-head puzzle perception for solving large-scale jigsaw puzzles of eroded gaps. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 6968–6977, 2025.

[Talon *et al.*, 2025] Davide Talon, Alessio Del Bue, and Stuart James. Ganzzle++: Generative approaches for jigsaw puzzle solving as local to global assignment in latent spatial representations. *Pattern Recognition Letters*, 187:35–41, 2025.

[Tang *et al.*, 2020] Hui Tang, Guohua Geng, and Mingquan zhou. Application of digital processing in relic image restoration design. *Sensing and Imaging*, 21(1):6, 2020.

[Valencia *et al.*, 2025] Mario Lino Valencia, Tobias Pfaff, and Nils Thuerey. Learning distributions of complex fluid simulations with diffusion graph networks. In *The Thirteenth International Conference on Learning Representations*, 2025.

[Vignac *et al.*, 2022] Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022.

[Wu *et al.*, 2023] Chenfei Wu, Shengming Yin, Weizhen Qi, Xiaodong Wang, Zecheng Tang, and Nan Duan. Visual chatgpt: Talking, drawing and editing with visual foundation models. *arXiv preprint arXiv:2303.04671*, 2023.

[Xie *et al.*, 2023] Shaoan Xie, Zhifei Zhang, Zhe Lin, Tobias Hinz, and Kun Zhang. Smartbrush: Text and shape guided object inpainting with diffusion model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22428–22437, 2023.

[Yang *et al.*, 2023] Mingxin Yang, Yonatan Svirsky, Zhanglin Cheng, and Andrei Sharf. Self-supervised fragment alignment with gaps. *IEEE Transactions on Visualization and Computer Graphics*, 30(9):6235–6246, 2023.

[Yu *et al.*, 2023] Tao Yu, Runseng Feng, Ruoyu Feng, Jinming Liu, Xin Jin, Wenjun Zeng, and Zhibo Chen. Inpaint anything: Segment anything meets image inpainting. *arXiv preprint arXiv:2304.06790*, 2023.

[Zhang *et al.*, 2022] Chongsheng Zhang, Bin Wang, Ke Chen, Ruixing Zong, Bo-feng Mo, Yi Men, George Almpanidis, Shanxiong Chen, and Xiangliang Zhang. Data-driven oracle bone rejoining: A dataset and practical self-supervised learning scheme. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 4482–4492, 2022.

[Zhang *et al.*, 2023] Yuqing Zhang, Zhou Fang, Xinyu Yang, Shengyu Zhang, Baoyi He, Huaiyong Dou, Junchi Yan, Yongquan Zhang, and Fei Wu. Reconnecting the broken civilization: Patchwork integration of fragments from ancient manuscripts. In *Proceedings of the 31st ACM international conference on multimedia*, pages 1157–1166, 2023.

[Zhang *et al.*, 2024] Yuqing Zhang, Hangqi Li, Shengyu Zhang, Runzhong Wang, Baoyi He, Huaiyong Dou, Junchi Yan, Yongquan Zhang, and Fei Wu. Llmco4mr: Llms-aided neural combinatorial optimization for ancient manuscript restoration from fragments with case studies on dunhuang. In *European Conference on Computer Vision*, pages 253–269. Springer, 2024.

[Zhang *et al.*, 2025a] Haoxuan Zhang, Haisheng Li, Xiaoqun Wu, and Nan Li. Mtgnet: multi-label mesh quality evaluation using topology-guided graph neural network. *Engineering with Computers*, 41(1):321–333, 2025.

[Zhang *et al.*, 2025b] Liguo Zhang, Zhibo Wang, Yuxin Dong, and Yalong Zhu. *CG2AN: Conditional Graph Generative Adversarial Networks for Shredded Image Restoration*. 10 2025.

[Zhou *et al.*, 2024] Rixin Zhou, Ding Xia, Yi Zhang, Honglin Pang, Xi Yang, and Chuntao Li. Pairingnet: A learning-based pair-searching and-matching network for image fragments. In *European Conference on Computer Vision*, pages 234–251. Springer, 2024.

[Zhu *et al.*, 2024] Yuanshao Zhu, James Jianqiao Yu, Xiangyu Zhao, Qidong Liu, Yongchao Ye, Wei Chen, Zijian Zhang, Xuetao Wei, and Yuxuan Liang. Controltraj: Controllable trajectory generation with topology-constrained diffusion model. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4676–4687, 2024.