A. Problem formulation

Suppose some original images, after physical segmentation or natural damage, form a discrete fragment set, where each fragment can be attributed to a specific source image. Let the complete fragment set be denoted as $\{V_n^m | n \in \{1, 2, \cdots, N\}, m \in \{1, 2, \cdots, M_n\}\}$, where the subscript index $n \in [1, N]$ indicates the source image to which a fragment belongs, and the superscript $m \in [1, M_n]$ represents the number of fragments associated with that image.

The algorithm is required to output a set of weighted connected graphs $\{G_1(V_1, E_1, T_1), \ldots, G_N(V_N, E_N, T_N)\}$, where each connected graph G_i corresponds to the global assembly process for the i-th original image. Each connected graph must satisfy a tree structure constraint. The vertex set V_i of G_i forms a strict partition subset of the original fragment set, the edge set E_i contains edges (u_j, v_j) representing detected fragment pairs, and the edge weight $t_j \in T_i$ denotes the local affine transformation matrix required to align fragment u_j to v_j.

B. Metrics

This section describes the computation details of the evaluation metrics mentioned in Sec. 4.2.1. In our experiments, Precision and recall assess matching modules. Precision measures the proportion of true matching pairs among the screened candidate pairs. Recall reflects the proportion of true matching pairs successfully screened. Adjusted Rand Index (ARI) evaluates clustering in the ablation study, Area Under the Curve (AUC) measures score evaluation. Global assembly performance uses edge-based precision and recall.

For the global assembly stage, the evaluation focuses on the geometric rationality of the overall spliced structure, ensuring no cracks or overlapping regions. Building upon the metrics defined in JigsawNet, we propose edge-based precision and recall calculations. Global precision is defined as the proportion of true matching pairs among the pairs selected during maximum spanning tree construction, as given in Equation 2:

$$Prec_{global} = \frac{|E_{correct}|}{|E_{selected}|}. \tag{2}$$

where $E_{correct}$ is the set of correct matching pairs, and $E_{selected}$ is the set of pairs selected by the algorithm. Global recall is defined based on geometric error criteria, calculating the proportion of true matching pairs with a rotation error less than 5° and a translation error less than 100 pixels [18], as shown in Equation 3:

$$Rec_{global} = \frac{|E_{valid}|}{|E_{truth}|}. \tag{3}$$

where E_{valid} is the set of pairs passing the error criteria, and E_{truth} is the set of all true matching pairs. This edge-based metric, unlike JigsawNet's node-based recall, focuses on relative fragment positions, unaffected by global translation or rotation.

C. Experiment details

C.1. Dataset generation

The art_2192 dataset [38] consists of 2,192 high-quality traditional Chinese landscape paintings. All paintings are sized 512×512, from the four sources: Princeton University Art Museum, Harvard University Art Museum, Metropolitan Museum of Art and Smithsonian's Freer Gallery of Art. The pex_2000 dataset [23] contains resized and cropped free-use stock photos from Pexels. All the images have minimum dimensions of 768p and maximum dimensions that are multiples of 32.

For fragment details, each original image is segmented into 20–40 fragments to simulate damage. The hard dataset incorporates simulated stains, mold spots, and contour defects to assess robustness. Stains are modeled as dark circular spots, with radii $\sim \mathcal{N}(0, 5)$ pixels, and 10–15 stains per fragment. Mold spots use irregular green-to-brown textures, generated with Perlin noise. Contour defects involve retracting contour segments (2–30 pixels) along the normal direction by 1–5 pixels with a 30% probability per fragment, simulating contour corrosion.

C.2. Architecture details

All hyperparameters of the multilayer ResGCN network in the coarse matching stage are identical to those in PairingNet. The feature vector dimension dim is 64. In the fine-grained matching stage, we stacked 2 layers of Cross Attention Decoder ($n = 2$ in Figure 4), with each decoder containing 8 attention heads.

For all experiments conducted on the art_2192 dataset, we set the maximum number of contour points L to 1280. For all experiments conducted on the pex_2000 dataset, we set L to 2400. These settings ensure that no fragment is truncated due to an excessively long contour.

C.3. Training details

The coarse matching network is trained on true matching fragment pairs using InfoNCE [5] loss (temperature coefficient 0.12), with an initial learning rate of 10^{-4}, batch size of 75, and 128 epochs. The fine-grained matching network uses the same data but employs FocalLoss [20] ($\alpha = 0.55$, $\gamma = 8$), with an initial learning rate of 10^{-3}, batch size of 54, and 128 epochs. Both processes take approximately 5 days. The score evaluation network is trained on balanced positive and negative sample pairs (negative samples from random non-matching pairs of the same source image) us-

ing binary cross-entropy loss, with an initial learning rate of 10^{-4}, batch size of 20, and 10 epochs, lasting about 1 day.

Training and testing are conducted on a single NVIDIA TITAN RTX with 24 GB memory.

C.4. Testing details

We select the checkpoint with the lowest validation loss for testing. The test set consists of fragment data independent of the training and validation sets, accounting for 20% of the total data. The generation process for the test set follows the same synthetic strategy as the training set to ensure distributional consistency. All tests use a fixed random seed of 1024 to guarantee reproducible results.

We employed different testing hyperparameters for the test sets of different datasets. For the art_2192 dataset, in the coarse matching stage, K was set to 20, meaning 20 fragments were selected for each fragment to form candidate matching pairs. In the fine-grained matching stage, the score evaluation threshold was set to 0.5. For the pex_2000 dataset, in the coarse matching stage, K was set to 30, meaning 30 fragments were selected for each fragment to form candidate matching pairs. In the fine-grained matching stage, the score evaluation threshold was also set to 0.5.

In comparative experiments, we made minor adjustments to the baseline methods to enable them to perform multi-source manuscript restoration tasks. For JigsawNet, since its proposed HLM global assembly algorithm tends to splice all fragments into a single image, which is unsuitable for multi-source manuscript restoration, we replaced it with the global assembly module used in this work. For PairingNet, we queried the Top-K adjacent fragments for each fragment, using the same K values as in our proposed algorithm. As matching scores for fragment pairs could not be obtained, in the global assembly stage, we arbitrarily selected a spanning tree for splicing.

D. Generated dataset examples

Here we provide some examples of synthetic datasets(shown in Figure 12), where white curves represent the generated contours, and green edges indicate the adjacency relationships between fragments.

E. Robustness and real-world case study

For stains and molds, we gradually increase the total number of stains and molds from none to severe (35 spots per fragment). Throughout this process, the number of stains and molds is kept equal (even sum) or differed by 1 (odd sum). For contour defects, we progressively increase the contour defect ratio from none to severe (35%). Other settings remain consistent with the corresponding experiments in Table 1.

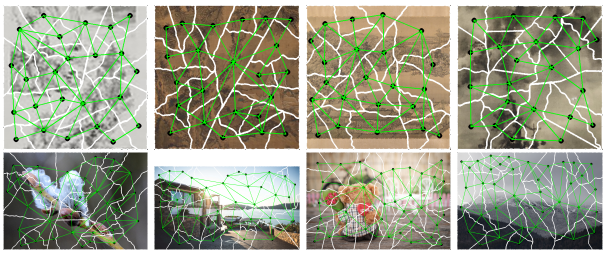


Figure 12. The fragments generated by the dataset generation algorithm are shown. The upper part of the figure displays the generation results for the art_2192 dataset, while the lower part shows the results for the pex_2000 dataset.

As shown in Figure 13, the decline in Global Assembly F1 Score remains approximately linear and relatively gradual for both stains&molds and contour defect, demonstrating the strong robustness of our method. Figure 14 intuitively illustrates the model's restoration results under varying degrees of degradation. As shown in the figure, the model maintains robust restoration performance across different levels of degradation. Furthermore, the table reveals that stains&molds have a greater adverse impact on assembly performance compared to contour defect. This is likely because stains&molds simultaneously degrade both the contour and textural features of fragments, whereas edge breakage primarily affects only the contour features.

We further treat real scenarios as OOD (out-of-distribution) settings. Strong OOD performance indicates practical efficacy. We test a pex_2000-trained model on art_2192, comparing to baselines (rule-based result unchanged from main text). Results are in Table 5 and Figure 15. ShreddingNet outperforms baselines in OOD generalization, with $<3\%$ drop vs. in-distribution performance.

Table 5. OOD results on art_2192 using pex_2000-trained model.

Difficulty	IC	Metrics/Baselines	JigsawNet	PairingNet	Ours
normal	3	Prec(%)	82.99	92.71	**96.29**
		Rec(%)	63.84	84.39	**92.72**
	6	Prec(%)	80.73	91.10	**93.88**
		Rec(%)	61.80	83.34	**91.44**
hard	3	Prec(%)	77.18	77.73	**84.36**
		Rec(%)	54.18	53.57	**66.01**
	6	Prec(%)	73.90	76.00	**81.61**
		Rec(%)	51.33	51.93	**65.28**

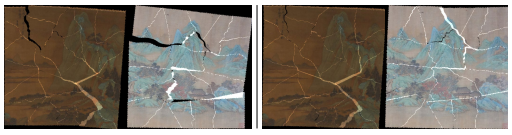


Figure 15. Visual result of OOD (left) and original (right) settings.

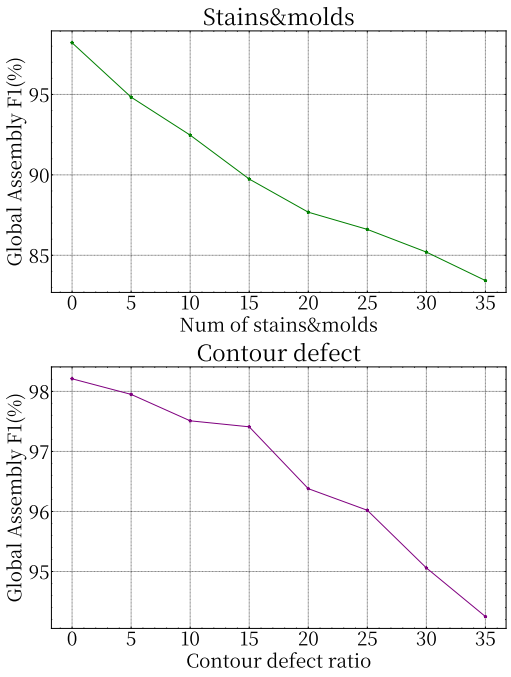


Figure 13. The left figure illustrates the impact of an increasing number of stains and molds on model performance, while the right figure shows the impact of progressively severe contour defects on model performance. In both figures, the horizontal axis represents the severity of degradation, and the vertical axis represents the Global Assembly F1 Score.

F. Ablation study

Figure 16 illustrates the probability density distributions (obtained via Gaussian kernel density estimation) of matching scores assigned by the CNN-based score evaluation module and the rule-based score evaluation module for correct and incorrect matching pairs. The rule-based score is not normalized, resulting in scores that may exceed 1.

G. Additional experiments

We adopt the fragment feature extraction method from PairingNet, but to demonstrate that this method is more effective than directly applying pre-trained models, we compare our feature extraction with pre-trained models (ResNet50, ViT-B, ViT-L) on art_2192(3 images). The results are shown in Table 6. Our method outperforms these alternatives in fine-grained matching F1-scores.

Table 6. Fine-grained matching F1-scores of different feature extraction methods

Model	ResNet50	ViT-B	ViT-L	ResGAN(Ours)
F1-score(%)	34.5	38.8	44.5	**77.8**

Because manuscript restoration tasks prioritize detailed features near fragment contours, while other pre-trained models focus more on extracting semantic information from images, pre-trained feature extractors perform poorly.

We test our algorithm on subsets of art_2192 and pex_2000 with similar content. Specifically, we collect ShuiMo paintings from the art_2192 test set and portraits from the pex_2000 test set to form two new test sets, and evaluate the model on each. The global metrics obtained are shown in Table 7.

Table 7. Global metrics on the test set subsets of content-similar source images and the full test set. A refers to art_2192, P refers to pex_2000.

Dataset	A(ShuiMo)	A(all)	P(portrait)	P(all)
Precision(%)	99.23	99.15	98.37	98.2
Recall(%)	97.02	97.59	98.18	96.78

Both subsets contain images with high content similarity. The Global metrics demonstrate consistent performance regardless of content overlap, since the model leverages both contour and texture features for assembly. Even when two fragments exhibit nearly identical textures, the distinct contour features enable correct matching.

We further create art_2192 (same) by segmenting one-third of art_2192 with seeds 256/512/1024, inputting three fragment sets from *the same image* at once. Results are in Table 8 and Figure 17. When fragments share similar content, the number of non-homologous pairs may increase, which can weaken fragment clustering and allow more non-homologous mismatches to enter fine-grained matching. These pairs are still rejected during fine-grained matching and thus have limited impact on the final results. Consistently, Table 4 (main text) shows that disabling fragment clustering only mildly affects overall performance.

Table 8. Results on art_2192 (same) vs. original art_2192. DSEP refers to the proportion of cross-source pairs, Prec/Rec refers to the Global Assembly metrics.

Metrics	DSEP(%)	Prec(%)	Rec(%)
art_2192 (same)	10.86	94.85	94.40
art_2192 (origin)	0.69	99.15	97.59

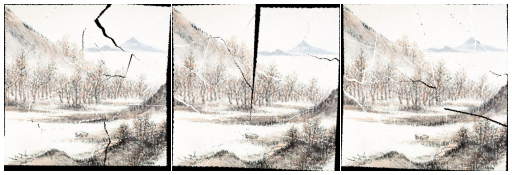
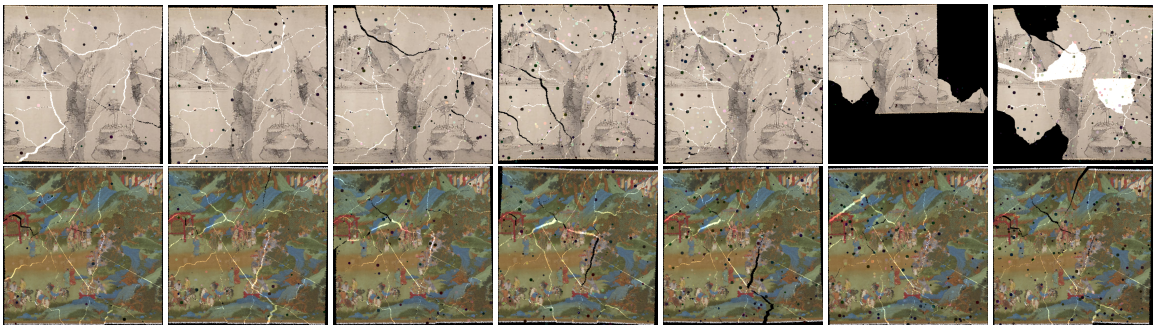
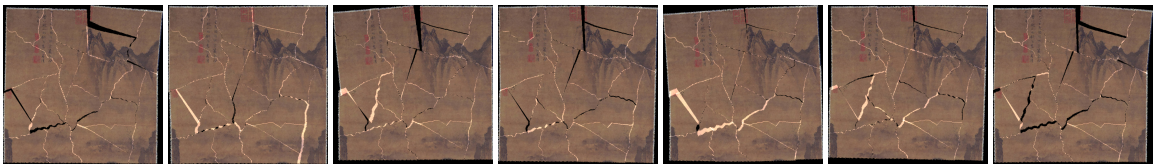


Figure 17. Visual result of art_2192 (same).



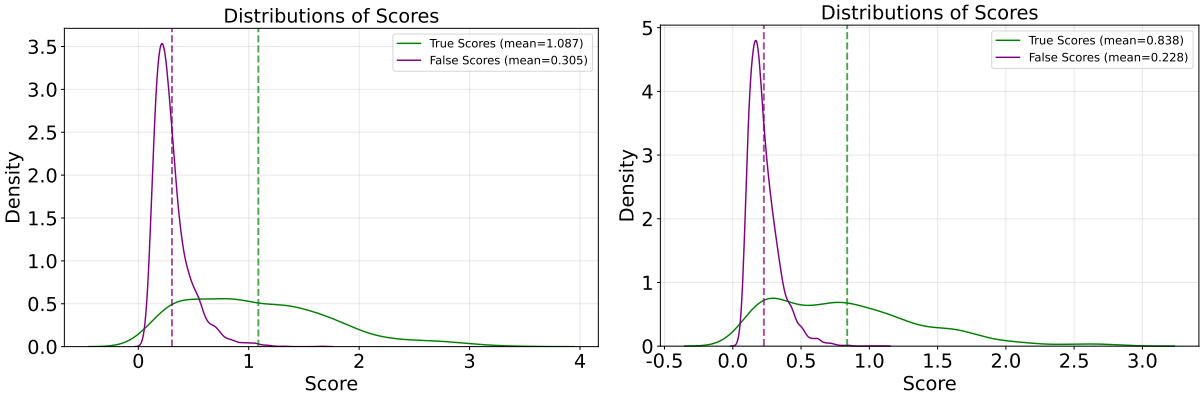
(a) stains&molds



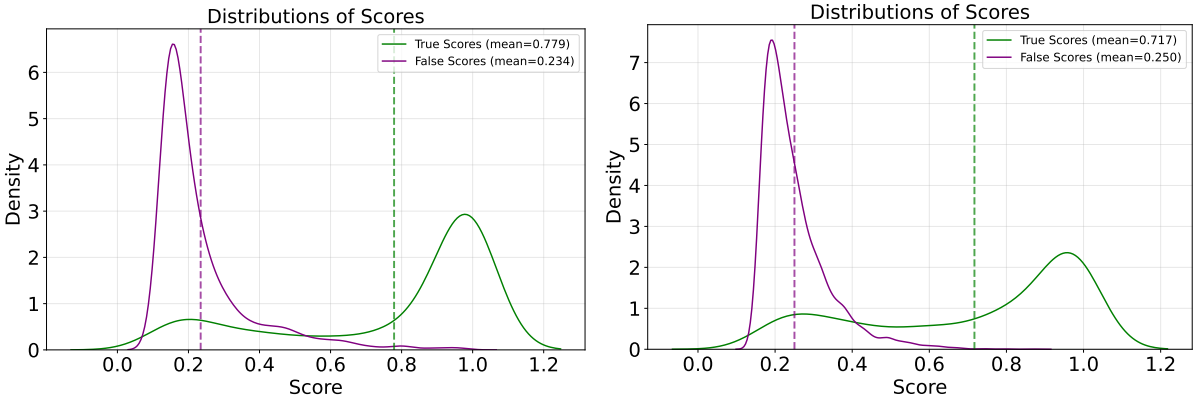
(b) contour defect

Figure 14. This figure intuitively illustrates the model's restoration results under varying degrees of degradation. The top two rows display the model's output images when the number of stains and mold varies from 5 to 35. The bottom row shows the model's output images when the degree of contour defects varies from 5 to 35.

Our model is designed to be insensitive to fragment orientation and patch size. During training and testing, all fragments undergo random rotation from 0° to 360°. For patch size, since we used the feature extraction method of PairingNet which had conducted relevant studies and proved that 7x7 was the best patch size setting and insensitive to other patch sizes, we did not conduct repeated experiments and directly used the settings of PairingNet.



(a) Rule-based score distribution



(b) CNN-based score distribution

Figure 16. The left side of the figure shows the score distribution for the art_2192 dataset, while the right side shows the score distribution for the pex_2000 dataset. In each subplot, the horizontal axis represents the scores of fragment pairs, and the vertical axis represents the probability density.