The Visual Computer (2026) 42:388
https://doi.org/10.1007/s00371-026-04624-6

RESEARCH

# Rectangling stitched images via unsupervised warping

Yun Zhang[1] · Yao Lu[1] · Jialing Yang[1] · Zhe Zhu[2] · Yu-Kun Lai[3] · Fang-Lue Zhang[4] · Xinyuan Zheng[1]

Received: 5 May 2026 / Accepted: 21 June 2026


**Abstract**

Image stitching allows wide field-of-view images to be created. However, handheld shooting and alignment of overlapping regions in image stitching intrinsically result in irregular boundaries, compromising the wide-angle effect. To address this problem, we propose an unsupervised warping-based method for rectangling stitched images. We formulate irregular mesh prediction as a mesh motion regression task, constrained by three complementary objectives: shape-preserving, boundary-fitting, and content-preserving losses. This approach leverages geometric and semantic features of images to achieve rectangling without requiring labeled training data. Our primary contributions include (1) a label-free learning framework that improves rectification performance and generalization capability, and (2) a novel boundary-fitting scheme that reconstructs well-aligned meshes, producing visually natural rectangling results across diverse scenarios. Experiments demonstrate that our method achieves competitive or superior performance compared with state-of-the-art supervised methods.

**Keywords** Unsupervised · Rectangling · Warping-based · Mesh motion regression · Boundary fitting

## 1 Introduction

Image stitching seeks to create images with a wide field-of-view (FOV) for numerous applications, including virtual reality, video monitoring and medical imaging. Due to the limitations of cameras, wide-FOV images are usually created by stitching multiple images captured from different viewpoints with overlapping regions. However, since the images intended for stitching are typically captured by cameras with unrestricted movement, the resulting stitched images often exhibit irregular boundaries. Consequently, additional cropping is required, which significantly reduces the effective field-of-view of the final stitched images.

To address the issue of irregular boundaries while preserving the wide-angle effect, warping-based solutions have been proposed [6, 13, 20], which focus on constructing a well-fitted mesh on the stitched image. While effective, traditional methods such as [6, 20] rely on computationally expensive optimization procedures such as seam-carving, and they often fail to preserve the semantic structure of the scene. To improve both efficiency and visual quality, deep learning-based methods have recently been introduced. Notable works include the learning baseline for rectangling by Nie et al. [13], and methods using a diffusion model [25]. Nevertheless, most existing learning-based methods are supervised and require labor-intensive labeled datasets, for which a standard ground-truth definition is often unavailable. Furthermore, these supervised learning methods often exhibit limited generalization capabilities, restricting their applicability in real-world scenarios.

In this work, we address the rectangling problem by leveraging an unsupervised learning paradigm, offering a label-free alternative to existing supervised approaches. To avoid computationally intensive training and inference, we adopt a warping-based solution that yields highly efficient rectangling results via content-aware warping. The primary challenge, however, lies in effectively learning geometric transformations without access to labeled training data.

Specifically, we predefine a regular mesh on the rectangular output and predict the mesh for the irregular input by utilizing the geometric structure and semantic content of the input images. Specifically, we formulate the irregu-

✉ Yao Lu
luyao5916@gmail.com

1 Zhejiang Key Laboratory of Film and TV Media Technology, School of Media Engineering, Communication University of Zhejiang, Hangzhou 310018, China

2 Samsung Research America, Irvine, CA 92612, USA

3 School of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, Wales, UK

4 School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 2033, Australia

Published online: 04 July 2026

lar mesh prediction as a mesh motion regression problem, constrained by a set of complementary objective functions including shape-preserving, boundary-fitting, and content-preserving losses. To address the inherent limitations of mesh-based warping, we subsequently employ inpainting techniques to fill small gaps around the rectangular boundaries. Compared with supervised methods [13, 15, 25], our method exhibits enhanced generalization capabilities and can be applied to more practical real-world scenarios, including image stitching and image rectification. This improvement is mainly attributed to its effective geometric perception and robust content preservation. Experimental evaluations clearly demonstrate that our method effectively corrects irregular boundaries in a content-aware manner, achieving superior performance and rectification effects compared to previous state-of-the-art methods.

To the best of our knowledge, this work presents the first unsupervised mesh-warping framework for image rectangling. Our contributions are summarized as follows:

- We propose an unsupervised learning framework for image rectangling, eliminating the need for labor-intensive labeled datasets while improving generalization capability.
- We introduce an effective boundary perception strategy that reconstructs well-aligned meshes for images with irregular boundaries, enabling smooth and natural rectangling results across diverse scenarios.

## 2 Related work

In this section, we mainly review and summarize the works closely related to this paper.

### 2.1 Image stitching

Image stitching aims to expand the FOV of images by aligning multiple images with overlapping regions, and has long been an active topic in computer graphics and vision [22]. The primary objective is to achieve accurate alignment while minimizing geometric distortions and visual artifacts. Previous works can be divided into traditional and deep learning-based methods. Traditional methods focus on designing effective geometric transformation models to achieve accurate alignment between overlapping images. Representative methods include automatic stitching based on a single homography [1], smoothly varying affine stitching [11], as-projective-as-possible stitching [23], seam-guided stitching [10], geometric structure preserving optimization [3], manifold optimization [24], etc. Although these methods achieve promising results under ideal conditions, they often struggle in challenging scenarios involving large parallax, textureless regions, or inaccurate feature correspondences.

To overcome these limitations, deep learning techniques have recently been introduced to image stitching [21]. Deep learning-based methods typically extract high-level semantic features from input images and estimate warping parameters using neural networks. Due to the difficulty of generating accurate ground-truth labels for stitching tasks, many recent approaches adopt unsupervised learning frameworks. Nie et al. [14] proposed parallax-tolerant unsupervised method to handle large parallax by a robust and flexible mesh warping. Jia et al. [8] proposed pixel-wise alignment for stitching using the optical flow based warping. Very recently, Jin et al. [9] further improved pixel-wise alignment based stitching by introducing bidirectional warp for more balanced distortions over two views. Experimental results have shown that deep learning-based stitching methods often achieve superior performance compared with traditional approaches, particularly in challenging scenarios where reliable feature matching is difficult, such as low-light, motion blur, or weak-texture environments.

### 2.2 Image rectangling

Image rectangling is a fundamental image rectification task that aims to transform images with irregular boundaries into images with regular rectangular shapes. Such irregular boundaries are commonly introduced by warping-based image editing operations, particularly image stitching. The goal of rectangling is to generate a rectangular output while minimizing geometric distortions and preserving the visual content of the original image.

Early rectangling methods were primarily based on optimization-based warping techniques. He et al. [6] were among the first to study this problem and proposed a two-step content-aware warping scheme to rectify irregular panoramic images. Compared with cropping-based or inpainting-based approaches [4], their method produces visually plausible results while preserving more image content. Following this work, He et al. [5] further developed an optimization-based warping framework to correct rotated images while maintaining image content within an upright rectangular region. Wu et al. [19] extended the rectangling framework to video sequences by introducing a spatio-temporal warping formulation.

Although effective in many cases, optimization-based rectangling methods often suffer from high computational cost and limited ability to preserve semantic structures. To address these limitations, learning-based rectangling approaches have recently been proposed. Nie et al. [13] introduced the first deep learning-based rectangling framework for image stitching, which significantly improves efficiency and robustness compared with traditional optimization-

based solutions. More recently, Zhou et al. [25] proposed a diffusion-based rectangling framework that combines warping-based rectification with generative modeling to further improve rectangling performance. Despite their effectiveness, most existing learning-based rectangling methods rely on supervised learning and require large amounts of labeled training data, which is expensive and difficult to obtain. Moreover, diffusion-based approaches often introduce substantial computational overhead and may still suffer from distortions or blurry boundaries in certain cases. Motivated by these limitations, we propose an unsupervised rectangling framework that eliminates the need for labeled data while maintaining competitive performance and improved generalization capability.

## 3 Algorithm

As shown in Fig. 1, our proposed method adopts an unsupervised learning framework for the rectangling task. Compared with the supervised baseline proposed by Nie et al. [13], our method does not require ground truth mesh supervision. This design improves the generalization capability of the model while simplifying the learning pipeline.

Furthermore, our method adopts a single-stage regression framework, resulting in both faster training and more efficient inference.

### 3.1 Network structure

As illustrated in Fig. 2, we propose an unsupervised regression network that takes an irregular stitched image and its corresponding binary mask as input. The network predicts the mesh motion field, which is subsequently utilized to warp the irregular image into a rectangular output while preserving image content. Different from previous methods [13, 25], our network operates in an unsupervised manner, requiring only irregular stitched images without any ground-truth rectangled images for training. We give details of our algorithm as follows.

*Feature extractor:* To extract high-level semantic features from the concatenated image&mask input, we utilize a modified ResNet-18 [7] backbone pretrained on ImageNet [2]. Specifically, we modify the first convolutional layer to accommodate the 4-channel input (3-channel RGB image and 1-channel binary mask), keeping all subsequent layers such as BatchNorm, ReLU, max-pooling, and the first three residual blocks (layer1 to layer3) unchanged. This deep encoder incrementally reduces the spatial resolution by a factor of 16, resulting in a feature map with 256 channels at the bottleneck. In practice, employing a pretrained backbone with residual connections provides robust feature representations and ensures stable gradient flow during training, which facilitates learning complex geometric transformations necessary for rectangling.

*Mesh motion regression:* After feature extraction, we employ a regression network to predict vertex displacement vectors for the mesh. This network is composed of three convolutional blocks followed by three fully connected layers, and each block contains two $3 \times 3$ convolutional layers with ReLU activation, followed by a $2 \times 2$ max-pooling layer, which iteratively reduces the spatial dimensions of the feature maps. The compressed features are subsequently flattened and passed through fully connected layers with dimensions 2048, 1024, and $(U+1) \times (V+1) \times 2$, respectively. Here, $U \times V$ denotes the resolution of the mesh grid, which is set to $8 \times 6$ in our implementation. This network outputs displacement vectors for all $(U+1) \times (V+1)$ mesh vertices relative to the predefined regular mesh. These displacement vectors represent the horizontal and vertical movements required to rectify the irregular image into a rectangular output.

### 3.2 Irregular boundary perception

In unsupervised rectangling, accurately aligning the predicted mesh with irregular image boundaries is particularly challenging due to the absence of ground-truth supervision. To address this issue, we introduce an irregular boundary perception mechanism that transforms unstructured boundary information into structured geometric guidance for mesh alignment.

The key idea is to convert the binary mask of the irregular stitched image into a set of geometric priors that guide the mesh deformation process. Our perceptual process comprises two key components: scanline-based boundary localization and topological corner extraction, the details of which are elaborated as follows.

#### 3.2.1 Scanline-based boundary localization

To estimate the irregular boundaries of stitched images accurately, we propose a scanline-based boundary localization strategy. Since irregular boundaries cannot be easily represented by simple analytical functions, we discretize the boundary detection process by scanning the binary mask $\mathcal{M}$ along orthogonal scanlines.

Taking the upper boundary in Fig. 3a as an example, we describe the algorithm in Algorithm 1. For each vertical scanline $x = x_k$ ($k \in \{0, \ldots, W\}$), we conduct a vertical search to identify the ordinate $y_{top}$ of the first content pixel where the binary mask value is equal to 1. This ordinate represents the vertical starting boundary of the content at the current horizontal position $x_k$. Meanwhile, to mitigate the boundary "holes" caused by discrete sampling and estimation errors, we apply an inward geometric margin $\tau$ to the captured coordinate to derive the target anchor, denoted as $\hat{y}^k_{top} = y_{top} + \tau$.

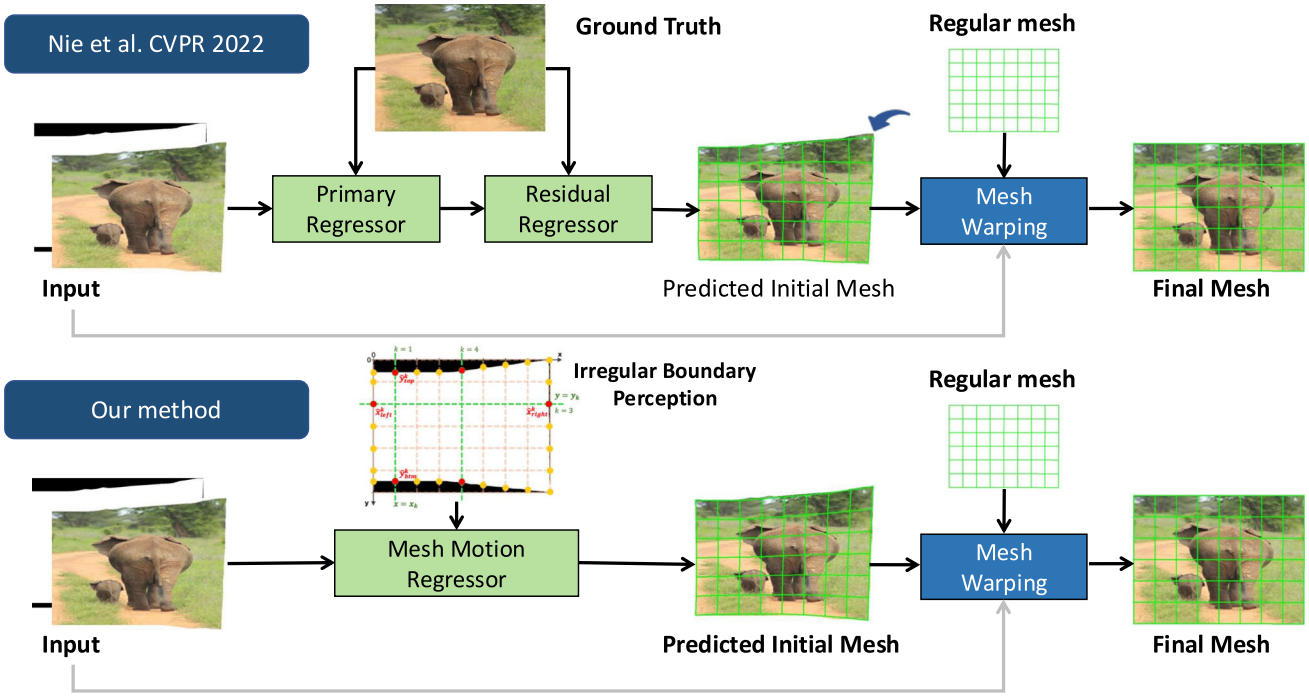



**Fig. 1** Comparison of supervised learning baseline [13] and our unsupervised learning baseline. The supervised baseline uses a two-stage regressor with ground truth supervision, while our method directly predicts mesh motions guided by the boundary priors through the unsupervised irregular boundary perception. The blue arrow highlights the artifacts of the initial mesh predicted by [13]

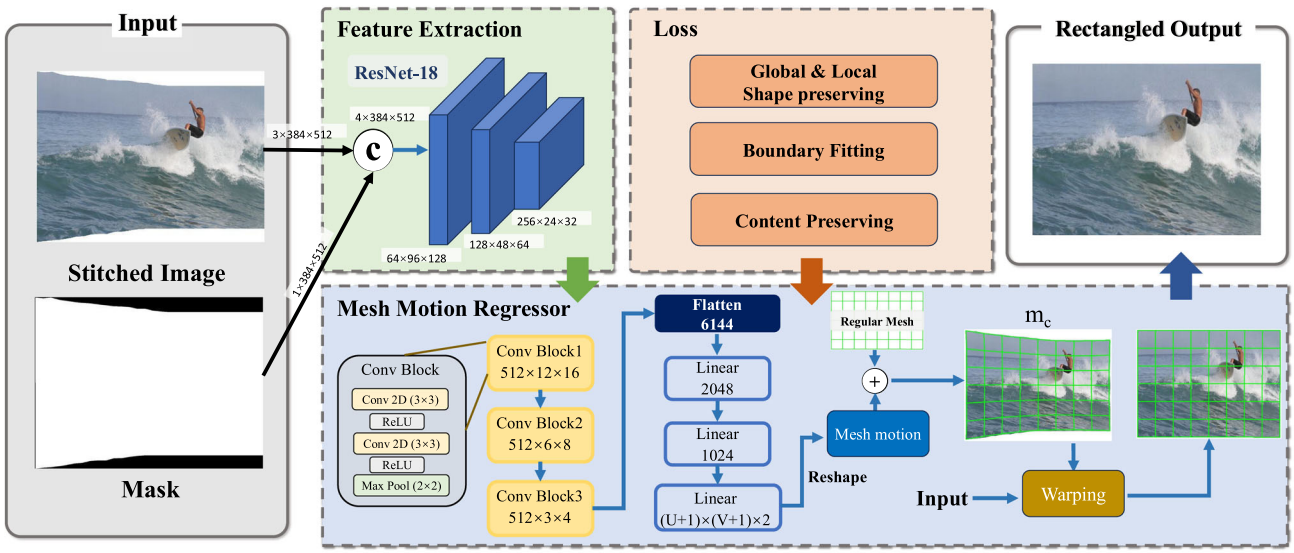



**Fig. 2** The overall structure of our unsupervised learning architecture. Given a stitched image and its one-zero mask as input, the network first extracts features based on ResNet-18, then predicts mesh displacement vectors through the mesh motion regressor, constrained by shape-preserving, boundary-fitting and content-preserving terms. Finally, we warp the predicted mesh to the target regular mesh, and obtain the rectangled output

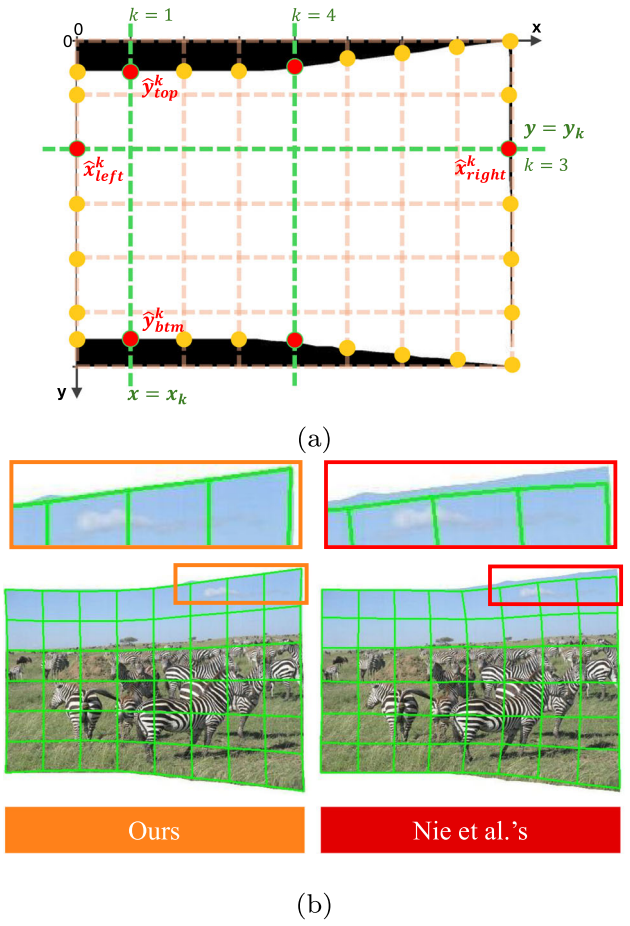



**Fig. 3** Boundary anchors localization. **a** illustrates how target boundary anchors are obtained via orthogonal scanlines; **b** demonstrates that our method achieves superior boundary fitting through irregular boundary perception compared with that of Nie et al. [13]

**Algorithm 1** Boundary anchor localization algorithm

**Input:** Mask $\mathcal{M}$, Margin $\tau$, Mesh Size $(H, W)$
**Output:** Boundary Anchor Set $\mathcal{T}$
**for** each $k \in \{0, \ldots, W\}$ **do** ▷ $k$ is the scanline index
  $y_{\text{top}} \leftarrow \min\{y \mid \mathcal{M}(x_k, y) = 1\}$
  $\mathcal{T}_{\text{top}}(k) \leftarrow y_{\text{top}} + \tau$
  $y_{\text{btm}} \leftarrow \max\{y \mid \mathcal{M}(x_k, y) = 1\}$
  $\mathcal{T}_{\text{btm}}(k) \leftarrow y_{\text{btm}} - \tau$
**end for**
**for** each $k \in \{0, \ldots, H\}$ **do**
  $x_{\text{left}} \leftarrow \min\{x \mid \mathcal{M}(x, y_k) = 1\}$
  $\mathcal{T}_{\text{left}}(k) \leftarrow x_{\text{left}} + \tau$
  $x_{\text{right}} \leftarrow \max\{x \mid \mathcal{M}(x, y_k) = 1\}$
  $\mathcal{T}_{\text{right}}(k) \leftarrow x_{\text{right}} - \tau$
**end for**
**return** $\mathcal{T} = \{\mathcal{T}_{top}, \mathcal{T}_{btm}, \mathcal{T}_{left}, \mathcal{T}_{right}\}$

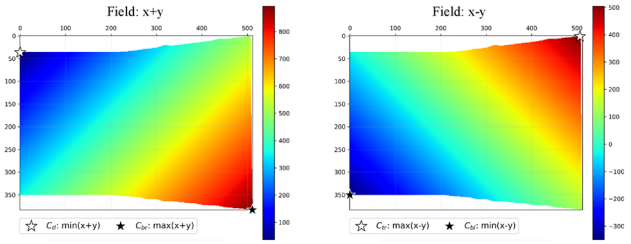



**Fig. 4** Coordinate extremum field on the effective mask region Ω. The left panel illustrates the extremum field of the pixel coordinate sum $(x + y)$, while the right panel illustrates that of the difference $(x - y)$

Subsequently, all target anchors identified on the top boundary are organized in the array $\mathcal{T}_{top}$ based on the index $k$. Analogously, the coordinates of boundary anchors may be localized by conducting symmetric searches on the *bottom*, *left*, and *right* boundaries, resulting in the boundary anchor set $\mathcal{T} = \{\mathcal{T}_{top}, \mathcal{T}_{btm}, \mathcal{T}_{left}, \mathcal{T}_{right}\}$. This method discretizes the complex boundaries into a sequence of differentiable target anchors, effectively directing the mesh edges to perform subsequent mesh fitting, as demonstrated in Fig. 3b

### 3.2.2 Topological corner extraction

To ensure that the rectified image maintains a proper topological structure and avoids rotation drift, we employ the geometric extremum principle to identify the extreme vertices of the content, subsequently constructing a coordinate extremum field within the effective mask region $\Omega = \{(x, y) \mid \mathcal{M}(x, y) = 1\}$. In particular, we calculate $(x + y)$ and $(x - y)$ for all coordinates within the region Ω. The spatial distribution of these values is visually represented in the heatmap of Fig. 4, where intensity indicates the field value. Consequently, we define the four topological corners $C = \{C_{tl}, C_{tr}, C_{bl}, C_{br}\}$ as the pixels that reach the extreme $(x+y)$ and $(x-y)$ values within this geometric field. Specifically, the top-left corner $C_{tl}$ and bottom-right corner $C_{br}$ correspond to the minimum and maximum values of $(x+y)$, while the top-right corner $C_{tr}$ and bottom-left corner $C_{bl}$ correspond to the maximum and minimum values of $(x - y)$. This mechanism effectively suppresses local boundary fluctuations and enables robust estimation of the global image geometry.

Through the irregular boundary perception, we determined the corresponding optimal alignment coordinates $\mathcal{T}$ for each mesh boundary point, as well as the topological corners $C$ for the four vertices. These identified geometric priors provide effective guidance for the subsequent boundary-fitting optimization.

## 3.3 Objective functions

We optimize our network parameters by means of a comprehensive objective function with three terms. The associated

optimization goal is formulated as follows:

$$\mathcal{L}_{total} = \lambda_s \mathcal{L}_s + \lambda_b \mathcal{L}_b + \mathcal{L}_c, \tag{1}$$

where $\mathcal{L}_s$, $\mathcal{L}_b$ and $\mathcal{L}_c$ represent the shape preserving term, boundary fitting term and content preserving term, respectively, and $\lambda_s$, $\lambda_b$ and $\lambda_c$ are their corresponding weights.

### 3.3.1 Shape-preserving term

Warping-based rectangling can easily introduce geometric distortions if the mesh deformation is not properly constrained. To ensure smooth and consistent mesh deformation, we introduce a shape-preserving loss that regulates both local and global mesh structures.

*Local shape preserving:* To preserve the local shape, we aim to prevent the folding and flipping of individual mesh grids and ensure the consistent size of each element. As shown in Fig. 5, we design a unified penalty term $P_e$ that detects mesh flipping by evaluating the cross product between adjacent edges. We formulate the penalty term as follows:

$$P_e = \begin{cases} \alpha - \vec{e}_u \times \vec{e}_v, & \vec{e}_u \times \vec{e}_v \leq \alpha \\ 0, & \vec{e}_u \times \vec{e}_v > \alpha, \end{cases} \tag{2}$$

where $\vec{e_u}$, $\vec{e_v}$ refer to the mesh edges oriented in the horizontal rightward and vertical downward directions, respectively. The cross product $\vec{e_u} \times \vec{e_v}$ represents the signed area of adjacent edges, with non-positive values indicating a flipping. We penalize negative cross products using $(\alpha - \vec{e}_u \times \vec{e}_v)$, where $\alpha = 0.0001$ is introduced to prevent flipping and to avoid excessive compression, and to penalize edges that are close to flipping. To enhance the consistency and smoothness of warping for high-quality rectangling, we promote uniformity in the size of each mesh grid. This is achieved by minimizing the $L_1$ norm of the difference between an element's dimension and the global average dimension. We define the local shape preserving term as follows:

$$\begin{aligned} \mathcal{L}_{local} = \beta \cdot \frac{1}{(H-1)(W-1)} \sum_{i=1}^{H-1} \sum_{j=1}^{W-1} P_e + \\ \sum_{\vec{e_u} \in \Gamma_u} \|\vec{e_u} \times \vec{i} - \overline{w}\|_1 + \sum_{\vec{e_v} \in \Gamma_v} \|\vec{e_v} \times \vec{j} - \overline{h}\|_1, \end{aligned} \tag{3}$$

where $H$ and $W$ denote the dimension of the mesh, and $\vec{i}$ and $\vec{j}$ are the unit vectors along the horizontal and vertical directions, respectively. We use $\beta = 40$ to balance the importance of these terms.

*Global shape preserving:* While local constraints prevent mesh flipping, they do not guarantee global geometric smoothness. Therefore, we further impose second-order smoothness constraints on the mesh vertices to encourage globally consistent deformation.

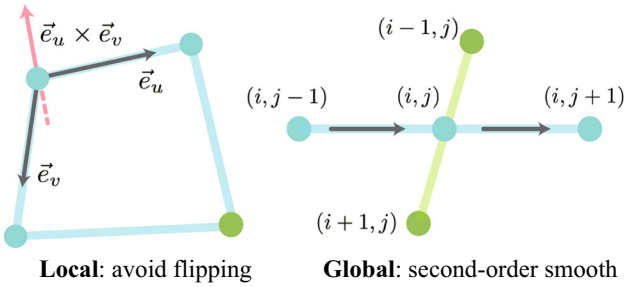



**Fig. 5** Local and global shape preserving. Local shape preservation aims to prevent unwanted mesh flipping by the cross product operation, while global shape preservation ensures adjacent mesh edges are constrained to exhibit second-order smoothness

Specifically, we minimize the second-order differences between neighboring vertices along both horizontal and vertical directions. These constraints suppress abrupt vertex displacement and promote smooth mesh transitions.

Additionally, we apply stronger smoothness constraints to mesh boundary vertices to improve boundary alignment stability. This enables the predicted mesh edges to follow irregular image boundaries while maintaining globally smooth deformation. We define the global shape preserving term as follows:

$$\begin{aligned} \mathcal{L}_{global} = & \sum_{i,j} \|m_c(i-1,j) - 2m_c(i,j) + m_c(i+1,j)\|_1 \\ & + \sum_{i,j} \|m_c(i,j-1) - 2m_c(i,j) + m_c(i,j+1)\|_1 \\ & + \eta \sum_{i,j} \|m_c^{\mathcal{B}}(i-1,j) - 2m_c^{\mathcal{B}}(i,j) + m_c(i+1,j)^{\mathcal{B}}\|_2^2 \\ & + \eta \sum_{i,j} \|m_c^{\mathcal{B}}(i,j-1) - 2m_c^{\mathcal{B}}(i,j) + m_c(i,j+1)^{\mathcal{B}}\|_2^2, \end{aligned} \tag{4}$$

where the first two lines define the second-order smoothness of the mesh edges, and the last two lines evaluate the second-order smoothness of the mesh boundaries in both horizontal and vertical directions. Here, $m_c(i,j)$ denotes the coordinate of the vertex $(i,j)$, and $m_c^{\mathcal{B}}(i,j)$ indicates the coordinate of a vertex located on the mesh boundary. We set $\eta = 5$ to balance the contributions of these four terms.

In sum, the total shape preserving term is concluded as follows:

$$\mathcal{L}_s = \mathcal{L}_{local} + \lambda_g \mathcal{L}_{global}, \tag{5}$$

where $\lambda_g = 2$ is set to balance the two loss terms. By assigning a higher weight to the global shape-preserving loss, we

encourage the network to prioritize the rectification of the overall geometric structure effectively.

#### 3.3.2 Boundary-fitting term

The boundary fitting term is designed to encourage precise alignment between the predicted mesh boundary and the irregular boundary of the stitched image. Since no ground-truth rectangled images are available in the unsupervised setting, the key challenge lies in estimating reliable boundary priors from the input mask. To address these challenges, we propose a scanline-based boundary localization and topological corner extraction scheme to effectively determine the reliable priors for the irregular boundary vertices.

To ensure more accurate boundary fitting, we further propose a refined sampling strategy to align precisely with the boundary of the zero–one mask. Specifically, we uniformly sample an additional $20-25$ points along each edge of the mesh boundary, and encourage each sample $q \in \Psi$ to locate on the outer boundary of the mask.

We define the boundary fitting loss as follows:

$$\mathcal{L}_b = \|m_c^{\mathcal{B}} - \mathcal{T}\|_1 + \gamma \|m_c^{\mathcal{C}} - \mathcal{C}\|_1 + \sum_{q \in \Psi} \mathrm{ReLU}(\theta - \mathcal{M}(q)), \tag{6}$$

where $\mathcal{T}$ and $\mathcal{C}$ denote the mesh boundary and corner priors, respectively, while $m_c^{\mathcal{B}}$ and $m_c^{\mathcal{C}}$ represent the predicted mesh boundary and corner vertices. We set $\theta = 0.98$, and utilize the ReLU function to impose penalties on points situated within the black regions of the mask. A weight of $\gamma = 2$ is used to balance these terms.

#### 3.3.3 Content-preserving term

Although geometric constraints ensure stable mesh deformation, they may still produce artifacts near image boundaries, such as small holes or missing pixels. To address this issue, we introduce a content-preserving loss that encourages the rectified output to maintain visual consistency with the input image.

First, we enforce canvas coverage by comparing the warped mask with an all-one matrix, ensuring that the rectified image fully occupies the output canvas. This constraint prevents undesired gaps along the rectified boundary.

Second, we incorporate a perceptual similarity constraint using features extracted from the VGG-19 network. By minimizing the distance between high-level feature representations of the input and rectified images, the network is encouraged to preserve semantic structures during rectification. We define the content preserving loss as follows:

$$\mathcal{L}_c = \sigma \|E - \mathcal{W}(\mathcal{M}, m_c)\|_1 + \varphi \|\mathcal{V}(\mathcal{W}(I, m_c)) - \mathcal{V}(I)\|_2, \tag{7}$$

where $E$ denotes the all-one matrix, and we warp the mask $\mathcal{M}$ using the predicted mesh $m_c$. $\mathcal{V}$ refers to the $13^{\text{th}}$ layer of VGG-19 features for semantic extraction. The weighting factors $\sigma = 500$ and $\varphi = 0.001$ are used to balance the two terms.

## 4 Experiments

In this section, we evaluate the performance of the proposed method through comprehensive experiments. We first describe the implementation details in Sect. 4.1, followed by quantitative and qualitative comparisons with state-of-the-art (SOTA) methods in Sect. 4.2. Ablation studies are presented in Sect. 4.3 to analyze the contribution of individual components. Finally, we analyze the computational efficiency and provide in-depth discussions of the proposed method in Sects. 4.4 and 4.5.

### 4.1 Implementation details

Our framework is implemented in PyTorch on a single NVIDIA A40 GPU. The network is trained with the Adam optimizer, using a learning rate initially set to $10^{-4}$, which decreases exponentially over time. The training proceeds for 200,000 iterations, with a batch size of 16. The learning rate reduction is controlled by an ExponentialLR scheduler, decaying at a rate of 0.97 per epoch to facilitate stable convergence. Furthermore, gradient clipping with a maximum norm of 5 is employed to avoid gradient explosion during backpropagation.

We employ the 'DIR-D' dataset, introduced in the rectangling baseline [13], for both model training and inference. Unlike [13], which relies on ground-truth labels for constraints, our approach operates without such supervision, leading to superior generalization across diverse scenarios.

In this paper, we employ the $L_1$ norm for most loss functions to enforce strict structural and geometric constraints, enabling robust mesh warping. Inspired by the constraint norm selection strategy in [13], we employ the $L_2$ norm to define effective perceptual similarity constraint in Eq. 7 for smoother optimization and improved visual consistency. In terms of loss weights, the shape-preserving loss $\mathcal{L}_s$, boundary-fitting loss $\mathcal{L}_b$, and content-preserving loss $\mathcal{L}_c$ are assigned weights of 50, 1000, and 1, respec-


**Fig. 6** Results and comparisons on the test dataset of DIR-D. The yellow arrows mark irregular boundary regions, the green rectangles indicate distorted regions, such as roof and windows, and the red rectangles highlight regions with missing input content

tively. This is because the primary challenge of the task is to stretch irregular edges into a standard rectangle. We apply a sufficiently strong pulling force to overcome the mesh's resistance to deformation, ensuring that the edges fit tightly and accurately to the outermost boundary. Additionally, the shape-preserving loss serves as an elastic regularization constraint, preventing the mesh from folding or undergoing extreme distortion under strong stretching forces, thereby ensuring a smooth and natural deformation process. Finally, the content-preserving loss serves as a fine-grained refinement upon the two major constraints above, thus requiring only a small weight.

## 4.2 Comparative results

To comprehensively demonstrate the superiority of our method, we conduct comparative experiments from three aspects: quantitative comparison, qualitative comparison and cross-dataset evaluation Table 1.

**Table 1** Image quality comparisons on the DIR-D [13] dataset

| Method | PIQE↓ | Rect.↑ | EMD↓ |
|---|---|---|---|
| Nie et al.'s [13] | 33.5184 | 0.9978 | 14.3996 |
| Zhou et al.'s [25] | 42.5062 | **0.9999** | 14.2694 |
| Ours | **31.8827** | 0.9994 | **13.7995** |

The metrics for quantitative comparison include PIQE (lower is better), Rect. (higher is better), and EMD (lower is better). The bold numbers indicate the best results

### 4.2.1 Quantitative comparison

We perform a comparative evaluation of our approach against SOTA methods [13, 25] utilizing the testing dataset of DIR-D [13]. The assessment includes metrics such as PIQE [12](perceptual image quality), Rect. (Rectangularity) [17](structural regularity), and EMD (Earth Mover's


**Fig. 7** Qualitative comparison on cross-dataset images. The left and right panels show results for bidirectional warp and unidirectional warp, respectively. The yellow arrows highlight the uneven boundary regions in the results of Nie et al.'s [13]. In contrast, our method effectively rectifies these stitched images into rectangled images, exhibits stronger generalization ability

Distance) [18] (content and detail preservation), all of which are crucial for evaluating image rectangling results.

Although our unsupervised method does not present significant advantages, it performs comparably to, or slightly better than, fully supervised approaches. With respect to structure regularity [17], while [25] achieves marginally better results due to its generative framework, our outcome (0.9994) remains highly competitive and is very close to the ideal value of 1. In conclusion, the evaluations herein showcase the effectiveness and competitiveness of our method regarding perceptual quality, structural regularity, and content preservation.

#### 4.2.2 Qualitative comparison

Figure 6 presents qualitative comparison with SOTA rectangling methods. Taking the first row of the results as an example, it contains large missing content regions that tend to induce unnatural distortion in boundary warping. As observed, our result exhibits slight distortion in the elephant area, whereas the other methods also show obvious distortion and content loss in the roof and windows region, as well as irregular boundaries, as highlighted in the green boxes and yellow arrows. Compared with [13], our proposed approach not only produces more regular and visually consistent boundaries, but also more effectively preserves the detailed structural information and rich semantic content from the original input image. Compared with Zhou et al. [25], our method achieves more thorough content preservation while minimizing distortions and content loss.

#### 4.2.3 User study

As image rectangling aims to improve the visual quality of images for users, we conduct a user study to quantitatively evaluate the effectiveness of our proposed approach based on visual preference. The participants include 15 college students and 15 non-expert volunteers.

To ensure effective and fair user ratings, each group of results is displayed on a separate page, with the order randomized for each presentation. Additionally, users are allowed to zoom in or out on the stitching results to assist their evaluations. Each example was rated by participants on a scale of 1–5 ("1 = poor, 5 = excellent") based on two core

**Table 2** User study on the DIR-D dataset [13]

| | Content | Boundary |
|---|---|---|
| Nie et al.'s [13] | $\mu(3.63)\ \sigma(0.61)$ | $\mu(3.60)\ \sigma(0.50)$ |
| Ours | $\mu(4.03)\ \sigma(0.56)$ | $\mu(3.90)\ \sigma(0.48)$ |
| P-value | 0.0052 | 0.0174 |
| Ours | $\mu(4.03)\ \sigma(0.56)$ | $\mu(3.90)\ \sigma(0.48)$ |
| Ours+Inpainting | $\mu(4.33)\ \sigma(0.48)$ | $\mu(4.47)\ \sigma(0.63)$ |
| P-value | 0.0046 | 0.0002 |

We present the scores of 30 participants on the image rectangling results of different methods, which are evaluated from two aspects: content preservation and boundary regularity

criteria: Content Preservation and Boundary Regularity. To verify the statistical reliability of user preferences, we conducted a two-tailed paired t-test on the collected rating data. As shown in Table 2, our method outperforms [13] in both *content preservation* (4.03 vs. 3.63) and *boundary regularity* (3.90 vs. 3.60), with statistically significant differences ($P < 0.05$). Moreover, our inpainting-augmented results achieve a notably higher boundary score of 4.47, with strong statistical significance ($P = 0.0002$). The results of the user study demonstrate that our rectanlging results are preferred by users, and the inpainting operation further improves user satisfaction.

#### 4.2.4 Cross-dataset evaluation

In this cross-dataset evaluation, our model is trained on the DIR-D dataset [13], while its performance is evaluated on public image stitching datasets [14]. We utilize distinct image stitching strategies, such as bidirectional and unidirectional warping, to stitch images from conventional image stitching datasets. Subsequently, these stitched images undergo rectangling through various algorithms. The outcomes depicted in Fig. 7 clearly demonstrate that our approach produces images with substantially more regular boundaries than other supervised methods [13, 25], validating the strong and robust generalization ability of our model. Furthermore, as illustrated in Fig. 8, we conduct cross-dataset comparisons with the method in [25]. The results demonstrate that our method exhibits better generalization in cross-dataset scenarios than the diffusion-based method [25]. Specifically, our approach better preserves regular image boundaries and geometric structures, such as straight lines Fig. 9.

### 4.3 Ablation studies

Despite its simple architecture, our method achieves competitive performance. As shown in Table 3, we conduct ablation studies on the DIR-D dataset [13] to demonstrate the contribution of each constraint.

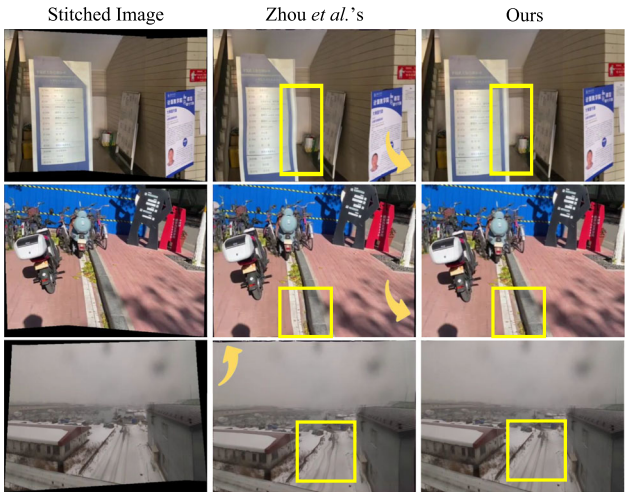



**Fig. 8** Qualitative comparison between the method of Zhou et al. [25] and our method on cross-dataset images. From left to right: original stitched images, rectangling results obtained by Zhou et al.'s method and our method. Yellow arrows and boxes highlight that our method achieves better preservation of boundary regularity and straight-line structure

**Without the shape-preserving term,** the model yields the poorest performance across all metrics. This demonstrate that the network fails to capture the fundamental geometric structure required f or rectification, leading to severe distortions. To further verify the individual effects of the local and global shape constraints, we conduct additional ablation studies. As shown in Fig. 10, removing both global and local shape constraints leads to severe distortion, while local constraints alone are effective in preventing mesh flipping but inadequate for suppressing global structural deformation, as highlighted by the red rectangles. In contrast, our method maintains geometric stability during rectangling by jointly enforcing local and global shape constraints. This indicates that the global and local constraints play complementary roles in our framework.

**Without the boundary-fitting term,** the model achieves the highest rectangularity. However, this comes at the cost of the integrity of the image content, as indicated by a higher EMD score.

**Without the content-preserving term,** the model struggles to maintain the structural and spatial completeness of the input, which tends to cause undesired content loss and irregular boundaries. As illustrated in Fig. 11, our content-preserving loss is designed to promote boundary regularity. In contrast to the shape-preserving loss, which mitigates stretching artifacts, the content constraint focuses on improving boundary regularity, thereby ensuring high-quality warping-based rectangling.

**Table 3** Ablation studies on DIR-D dataset [13]

| Loss Function | | | Mesh resolution | | | Metric | | |
|---|---|---|---|---|---|---|---|---|
| $\mathcal{L}_s$ | $\mathcal{L}_b$ | $\mathcal{L}_c$ | 4×3 | 8×6 | 16×12 | $PIQE \downarrow$ | $Rect. \uparrow$ | $EMD \downarrow$ |
| | ✓ | ✓ | | ✓ | | 39.7912 | 0.9961 | 14.6525 |
| ✓ | | ✓ | | ✓ | | 35.6900 | 0.9996 | 14.3574 |
| ✓ | ✓ | | | ✓ | | 31.9569 | 0.9993 | 13.8759 |
| ✓ | ✓ | ✓ | | ✓ | | 31.8827 | 0.9994 | 13.7995 |
| ✓ | ✓ | ✓ | ✓ | | | 32.4376 | 0.9966 | 14.1323 |
| ✓ | ✓ | ✓ | | | ✓ | **31.5695** | **0.9997** | **13.7595** |

Rows 1–3 refer to the models trained without the shape-preserving loss, boundary-fitting loss, and content-preserving loss, respectively. With all loss constraints jointly optimized, our method in row 4 achieves the optimal perceptual quality with the lowest PIQE score, superior content and detail preservation reflected by the minimum EMD value, and meanwhile maintains competitive rectangling performance, as indicated by the Rect. value. The last three rows present comparative results for different mesh sizes while all constraints are kept enabled

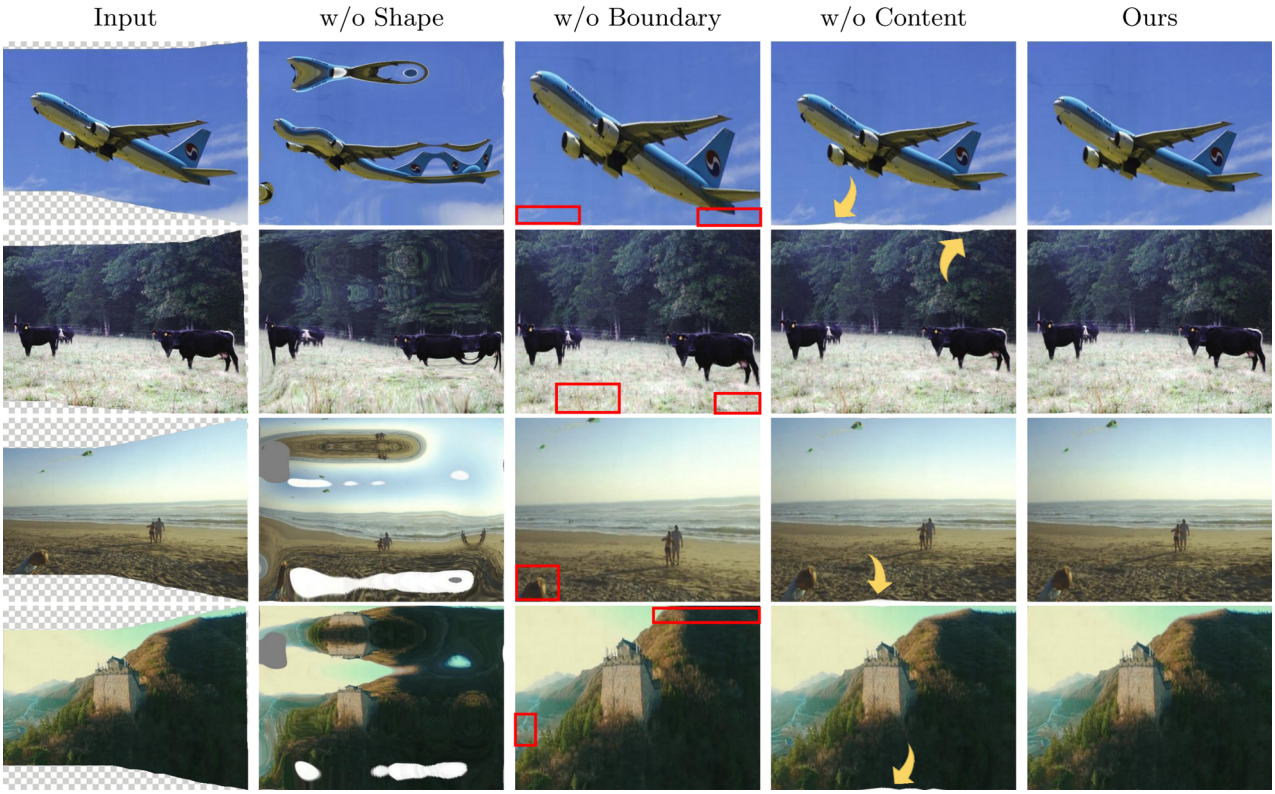



**Fig. 9** Ablation study of loss function. Removing the shape-preserving loss results in severe geometric collapse and artifacts. Without the boundary-fitting loss, the boundaries are not constrained to the canvas edges, leading to missing content (highlighted by red rectangles). Removing the content-preserving loss introduces unnatural distortions along the edges (pointed out by yellow arrows). Our full method effectively corrects the geometry while maintaining semantic fidelity

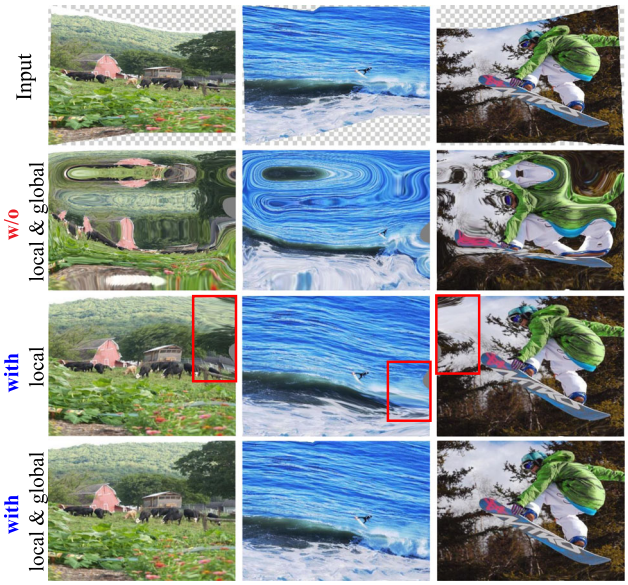



**Fig. 10** Ablation study of local and global shape-preserving losses. As shown in rows 2–3, removing both local and global shape constraints leads to severe distortion, and local constraints alone cannot fully prevent global structural deformation, as highlighted by the red rectangles. Our result in row 4 preserves geometric stability during rectangling by jointly enforcing local and global shape constraints

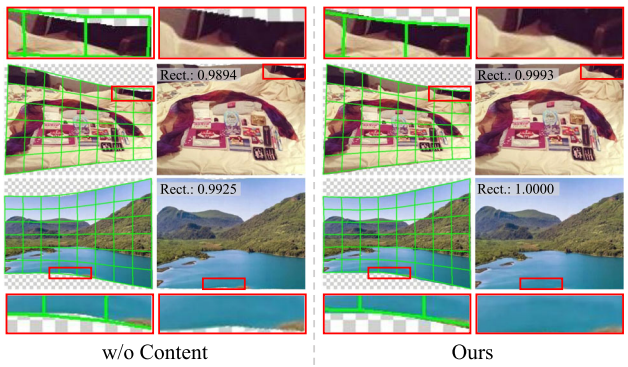



**Fig. 11** Analysis of content-preserving loss for rectangling. As illustrated in the zoom-in views, different from the shape-preserving loss that suppresses stretching artifacts, the content-preserving loss concentrates on improving boundary regularity to guarantee high-quality warping-based rectangling

By incorporating all three constraints together with the optimized 8 × 6 mesh resolution, our method achieves competitive performance across all evaluation metrics and delivers superior overall performance, demonstrating the effectiveness of our framework.

**Ablation Study on Mesh Size.** In addition to the aforementioned constraint ablation, we further investigate the impact of mesh size on rectangling performance, with three typical configurations: 4 × 3, 8 × 6, and 16 × 12. As reported in the last three rows of Table 3, the 4 × 3 mesh leads to degraded rectangling performance. By contrast, the 8 × 6 and 16 × 12 meshes achieve comparable results with marginal performance differences. Nevertheless, the 16 × 12 mesh introduces higher computational overhead. Considering both effectiveness and efficiency, we therefore adopt the 8 × 6 mesh in our implementation.

**Table 4** Average runtime comparison (in seconds) among different methods, including Nie et al. [13], Zhou et al. [25], and our proposed method

| | Nie et al. [13] | Zhou et al. [25] | Ours |
|---|---|---|---|
| AvgTime | 0.1025 | 447.4730 | 0.0102 |

## 4.4 Performance

To evaluate the performance of our method, we compare our processing efficiency with [13] and [25] on a single NVIDIA A40 GPU, utilizing the DIR-D test dataset [13]. As detailed in Table 4, the average runtime of our method is 0.0102*s* per image, whereas the method of Nie et al. [13] has an average runtime of 0.1025*s* per image. In contrast, processing an image by Zhou et al.'s method needs 447.4730*s* (1.0750*s* for MDM and 446.3980*s* for CDM), which imposes an extremely high computational cost and limits its practical feasibility. In conclusion, it is evident that our method not only surpasses the speed of that presented in [13], but also avoids the substantial computational demands typically associated with diffusion models [25].

## 4.5 Discussions

Different from previous supervised rectangling methods [13, 25], we propose an effective solution to rectify the irregular boundaries of stitched images using a label-free learning strategy. As the key of the regression network is to reconstruct meshes for images with irregular shapes, we propose a novel strategy to localize the boundary points from the input masks, and take them as a prior for effective boundary fitting. To obtain natural rectangling results, we design smooth and uniform warping schemes via local and global shape-preserving constraints. Benefiting from such constraints, our method can effectively preserve image content even under large mesh deformation, as highlighted by the yellow boxes in Fig. 12. We further discuss the impact of stitching seam artifacts on the rectangling results. As shown in Fig. 13, stitching inevitably causes inherent seam artifacts. Nevertheless, these artifacts remain stable after the stitching process, and will not be aggravated during the subsequent unsupervised warping, as observed in the zoom-in view. This enables our method to produce high-quality rectangling results with reliable feature alignment.

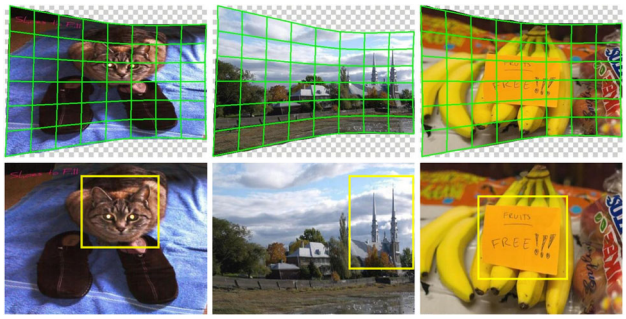


**Fig. 12** Performance under large mesh deformation. The first row presents the predicted meshes for irregular input images, which require large deformation for effective rectangling, and the second row shows our rectangling results, demonstrating effective content preservation under large mesh deformation, as highlighted in the yellow boxes


**Fig. 13** Influence of unsupervised warping on stitching artifacts. Columns 1–2 present the original input image pairs and stitched results with ghosting errors at the seams (highlighted in yellow boxes); Column 3 refers to our rectangularized results

Although effective and robust across various scenarios, our method also has some inherent limitations. As shown in Fig. 14, when the input image has very sharp boundaries due to the content loss and large parallax in image stitching, our method may fail to reconstruct a satisfactory mesh to fit the sharp boundaries, and this is also a major challenge for previous methods, such as [13]. To deal with this challenging problem, we leverage the Stable Diffusion Inpainting model [16] to reconstruct the missing regions and refine the boundary transitions. To ensure seamless fusion, we first use morphological dilation to expand valid regions, enabling the diffusion model to reconstruct original edges from local context. We then propose a multi-stage blending strategy with solid coverage and Gaussian feathering. As shown in the last column of the results in Fig. 14, by filling the boundary "hole" with generated pixels and smoothing via a gradient alpha-mask, our method removes visual discontinuities and

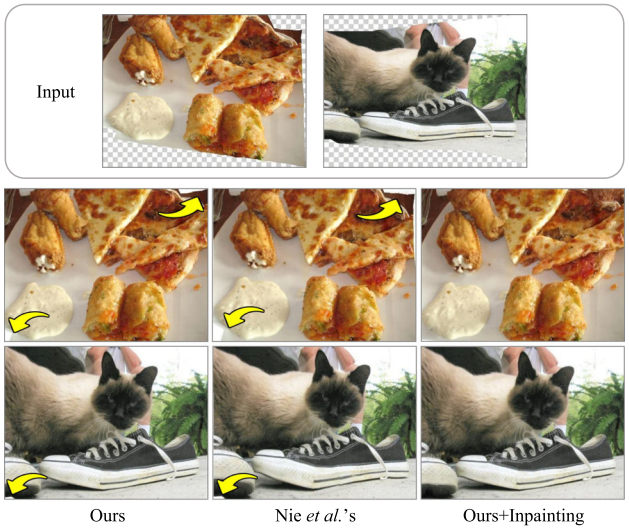



**Fig. 14** Challenging cases with severe content loss. As indicated by yellow arrows, both Nie et al.'s [13] method and ours suffer from boundary imperfections, yet our results are comparatively better. Moreover, the subtle artifacts produced by our method can be effectively repaired via inpainting

produces rectangling results with high-fidelity details. In the future we plan to explore enhanced feature modeling and improved network architecture to fundamentally address the challenges of rectangling. We also plan to further extend the unsupervised framework to video rectangling.

## 5 Conclusions

In this paper, we present an unsupervised learning framework for rectangling stitched images with irregular boundaries. Unlike existing supervised rectangling approaches, the proposed method eliminates the need for labeled training data by introducing an irregular boundary perception mechanism that extracts geometric priors directly from the input masks. Based on these priors, our framework predicts mesh motion fields through a regression network and performs content-aware warping to obtain rectified images with regular rectangular boundaries. The combination of shape, boundary, and content constraints enables the model to achieve smooth mesh deformation while preserving important visual structures.

Extensive experiments demonstrate that the proposed method achieves competitive or superior performance compared with existing SOTA rectangling approaches, while offering improved generalization capability and significantly lower computational cost.

**Acknowledgements** The authors would like to thank all anonymous reviewers for their valuable comments. This work was supported

by "Pioneer" and "Leading Goose" R&D Program of Zhejiang (No.2025C02014).

**Author Contributions** Y. Z. was responsible for the conceptualization of the research idea, algorithm research, and manuscript writing & polishing. Y. L. and J. Y. undertook algorithm implementation, experimental evaluation, model training and inference, as well as manuscript writing and figure preparation. Z. Z. participated in algorithm research and manuscript polishing. Y. L. and F. Z. were in charge of manuscript polishing and experimental design. X. Z. was responsible for funding support and dataset preparation.

**Data Availability** No datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

1. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vision **74**(1), 59–73 (2007)
2. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 248–255. IEEE (2009)
3. Du, P., Ning, J., Cui, J., Huang, S., Wang, X., Wang, J.: Geometric structure preserving warp for natural image stitching. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3678–3686. IEEE (2022)
4. Elharrouss, O., Almaadeed, N., Al-Maadeed, S., Akbari, Y.: Image inpainting: A review. Neural Process. Lett. **51**(2), 2007–2028 (2020)
5. He, K., Chang, H., Sun, J.: Content-aware rotation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 553–560. IEEE (2013)
6. He, K., Chang, H., Sun, J.: Rectangling panoramic images via warping. ACM Transactions on Graphics (TOG) **32**(4), 1–10 (2013)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. IEEE , (2016)
8. Jia, Q., Feng, X., Liu, Y., Fan, X., Latecki, L.J.: Learning pixel-wise alignment for unsupervised image stitching. In: A. El-Saddik, T. Mei, R. Cucchiara, M. Bertini, D.P.T. Vallejo, P.K. Atrey, M.S. Hossain (eds.) Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023, pp. 1392–1400. ACM (2023)
9. Jin, H., Nie, L., Lin, C., Feng, X., Zhao, Y.: Pixelstitch: Structure-preserving pixel-wise bidirectional warps for unsupervised image stitching. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 28125–28134. IEEE (2025)
10. Lin, K., Jiang, N., Cheong, L., Do, M.N., Lu, J.: SEAGULL: seam-guided local alignment for parallax-tolerant image stitching. In: European conference on computer vision (ECCV), Part III, pp. 370–385. Springer (2016)
11. Lin, W., Liu, S., Matsushita, Y., Ng, T., Cheong, L.F.: Smoothly varying affine stitching. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR), pp. 345–352. IEEE (2011)
12. N., V., D., P., Bh., M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: Twenty First National Conference on Communications, NCC, pp. 1–6. IEEE (2015)
13. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Deep rectangling for image stitching: A learning baseline. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5730–5738. IEEE (2022)
14. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Parallax-tolerant unsupervised deep image stitching. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 7365–7374. IEEE (2023)
15. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Semi-supervised coupled thin-plate spline model for rotation correction and beyond. IEEE Trans. Pattern Anal. Mach. Intell. (2024)
16. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 10684–10695. IEEE (2022)
17. Rosin, P.L.: Measuring rectangularity. Mach. Vis. Appl. **11**(4), 191–196 (1999)
18. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision **40**(2), 99–121 (2000)
19. Wu, J., Shi, J., Zhang, L.: Rectangling irregular videos by optimal spatio-temporal warping. Comput. Vis. Media **8**(1), 93–103 (2022)
20. Wu, J.L., Shi, J.J., Zhang, L.: Rectangling irregular videos by optimal spatio-temporal warping. Computational Visual Media **8**(1), 93–103 (2022)
21. Yan, N., Mei, Y., Xu, L., Yu, H., Sun, B., Wang, Z., Chen, Y.: Deep learning on image stitching with multi-viewpoint images: A survey. Neural Process. Lett. **55**(4), 3863–3898 (2023)
22. Yang, Z., Yin, Y., Xu, H., Jing, Q., Jiang, Z., Liao, T., Guedes Soares, C.: Advancements of image and video stitching techniques: A review. IEEE Sens. J. **25**(13), 23526–23551 (2025)
23. Zaragoza, J., Chin, T., Tran, Q., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving DLT. IEEE Trans. Pattern Anal. Mach. Intell. **36**(7), 1285–1298 (2014)
24. Zhang, L., Huang, H.: Image stitching with manifold optimization. IEEE Trans. Multim. **25**, 3469–3482 (2023)
25. Zhou, T., Li, H., Wang, Z., Luo, A., Zhang, C., Li, J., Zeng, B., Liu, S.: Recdiffusion: Rectangling for image stitching with diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 2692–2701. IEEE (2024)

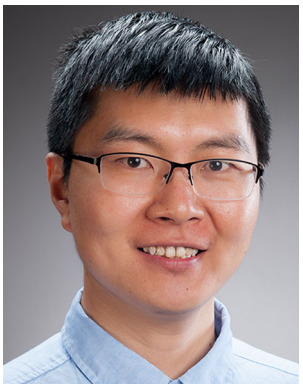
**Yun Zhang** received his bachelor's and master's degrees in Computer Science from Hangzhou Dianzi University, in 2006 and 2009, respectively, and his doctoral degree in Computer Science from Zhejiang University, in 2013. He is currently a Professor and Vice Dean with the School of Media Engineering, Communication University of Zhejiang, China. He visited the Visual Computing Group of Cardiff University, in 2018 and 2023. His research interests include computer graphics, image and video editing, and virtual reality. He is a senior member of CCF.

**Yao Lu** received her bachelor's degree from Communication University of Zhejiang in 2025. She is currently a master's candidate with the School of Media Engineering, Communication University of Zhejiang, China. Her research interests include computer graphics and Artificial Intelligence.

**Jialing Yang** received her bachelor's degree from Sichuan Agricultural University in 2025. She is currently a master's candidate with the School of Media Engineering, Communication University of Zhejiang, China. Her research interests include computer graphics, image, and video editing.

**Zhe Zhu** is currently a Staff Engineer at Samsung Research America. Before that he was a Senior Research Associate at Duke University. He got his Ph.D. in the Department of Computer Science and Technology, Tsinghua University, in 2017. He received his bachelor's degree in Wuhan University in 2011. His research interests are in computer vision and computer graphics.

**Yu-Kun Lai** received his bachelor's and Ph.D. degrees in Computer Science from Tsinghua University in 2003 and 2008, respectively. He is currently a Professor in the School of Computer Science & Informatics, Cardiff University. His research interests include computer graphics, geometry processing, image processing, and computer vision. He is on the editorial boards of IEEE Transactions on Visualization and Computer Graphics, The Visual Computer and Computers & Graphics.



**Fang-Lue Zhang** is currently a Senior Lecturer in the School of Computer Science and Engineering at the University of New South Wales (UNSW), Australia. He received a bachelor's degree from Zhejiang University, Hangzhou, China, in 2009, and a doctoral degree from Tsinghua University, Beijing, China, in 2015. His research interests include learning-based visual content generation, structure-aware image understanding, and panoramic scene reconstruction for VR/MR applications. He received the Victoria Early-Career Research Excellence Award in 2020. He serves on the editorial boards of Computers & Graphics and Visual Intelligence. He has also served as program chair for Pacific Graphics (2020 and 2021) and Computational Visual Media (2024). He is a senior member of IEEE.

**Xinyuan Zheng** received his bachelor's degree in Mechanical Design and Manufacturing (Mechatronics) from Zhejiang University of Technology in 1997. From 1997 to 2016, he worked at Zhejiang Radio & TV Research Institute, successively serving as Technician, Person-in-Charge of HFC Projects, and Director of the Research Department. He obtained the qualification of Professor-level Senior Engineer in 2013. Since 2016, he has been a Professor and Master's Supervisor at Communication University of Zhejiang. His main research focuses on media convergence technologies.