Lifting the Structural Morphing for Wide-Angle Images Rectification: Unified Content and Boundary Modeling

Wenting Luan[1] Siqi Lu[1] Yongbin Zheng[1*] Wanying Xu[1]
Lang Nie[2] Zongtan Zhou[1] Kang Liao[3]

[1]National University of Defense Technology
[2]Chongqing University of Posts and Telecommunications
[3]Nanyang Technological University

A. Supplemental Material

A.1. Overview

In this supplementary material, we provide the following content:

- Details of motivation for unifying the two cascade stage processes
- Details of dataset construction
- More qualitative comparison results
- More generalization evaluation Results
- More downstream tasks Performance

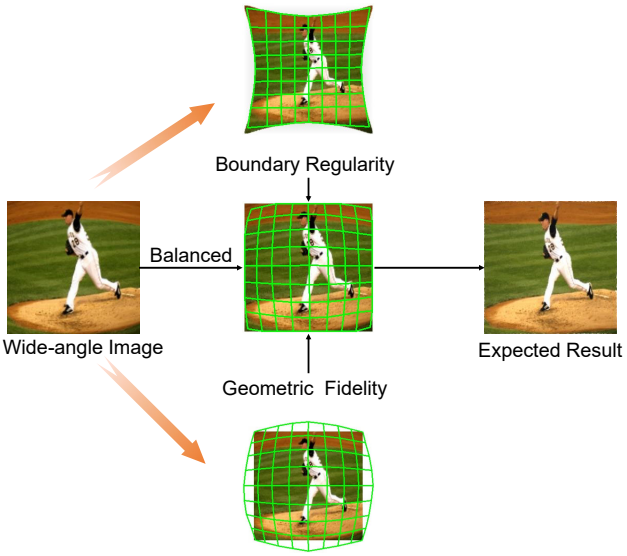

Figure 1. Rectification and rectangling represent two extreme TPS deformation fields. The optimal solution lies in a balanced intermediate state between these extremes.

*Corresponding author

A.2. Details of Motivation for Unifying the Two Cascade Stage Processes

Traditional two-stage pipelines for wide-angle distortion correction (*e.g.*, rectification followed by rectangling) mainly face three critical challenges:

1) Error Accumulation On one hand, independent processing propagates residual distortions from rectification to rectangling. For instance, inaccurate radial parameters in rectification lead to amplified warping artifacts during boundary alignment. On the other hand, cascaded methods exhibit a higher risk of pixel displacement error compared to unified approaches, which can cause blurred results.

2) Misaligned Optimization Objectives There is a conflict that rectification prioritizes geometric fidelity (minimizing central distortion) while rectangling emphasizes boundary regularity (maximizing peripheral deformation).

3) Ambiguous Physical Constraints Rectification methods (*e.g.*, physics-based regression [9, 11, 14]) struggle with nonlinear relationships between image content and lens parameters. Rectangling methods (*e.g.*, warping [6, 10, 12]) ignore lens physics, leading to content distortion.

To address the aforementioned challenges, we propose ConBo-Net to unify rectification and rectangling into a single optimization framework through Thin-Plate Spline (TPS) morphing. Our key insight stems from the observation that both tasks can be reparameterized as dual TPS transformations, whose interpolation inherently balances geometric fidelity and boundary regularity.

As illustrated in Fig. 1, rectification and rectangling correspond to two extreme TPS deformation fields: the rectification mesh preserves geometric fidelity by minimizing distortion near the image center. In contrast, the rectangling mesh enforces boundary regularity by deforming peripheral regions to align with a rectangular boundary. Intuitively, the optimal solution should reside in a continuous intermediate state between these extremes, balancing content preserva-

tion and boundary alignment. To achieve this, we propose a structural-aware morphing framework that dynamically balances geometric fidelity and boundary alignment through learnable spatial priors:

$$\mathcal{S}_{\text{final}} = \underbrace{\alpha(\mathbf{x})\mathcal{S}_{\text{rec}}}_{\text{Geometry}} + \underbrace{(1-\alpha(\mathbf{x}))\mathcal{S}_{\text{rect}}}_{\text{Boundary}}, \tag{1}$$

where $\alpha(\mathbf{x})$ is a spatially varying weight map learned from both image content and distortion priors, and $\mathcal{S}$ represents the specific task space. Theoretical analysis (see Sec 3.4 in the main manuscript) demonstrates that this unification is mathematically equivalent to jointly optimizing rectification and rectangling under shared physical constraints. Specifically, joint parameterization eliminates cascaded error propagation, while the inherent smoothness of the TPS energy function harmonizes conflicts between geometric fidelity and boundary regularity.

This framework systematically addresses the three challenges of traditional cascaded pipelines. First, coupling rectification and rectangling into a single differentiable operation enables end-to-end error correction. Second, injecting lens-specific priors (*e.g.*, ordinal geometry constraints) into TPS constraints ensures physical plausibility. Our experiments outperform state-of-the-art two-stage solutions, enhancing PSNR by at least 0.6 dB.

A.3. Details of the Dataset Construction

We construct a rectangling-distortion& corresponding meshes dataset, and we would like to release it to promote the research development. To be specific, our dataset is built using the following four steps:

(i) Wide-angle image synthesis. To synthesize wide-angle images, we followed the existing distortion rectification hypothesis and used the polynomial distorted camera model[4, 7, 9, 13, 14] and synthesized the wide-angle images. To be specific, in an image coordinate system, the Euclidean distance between an arbitrary point $P_u(x, y)$ and the image center $P_0(x_0, y_0)$ on the perspective image is represented as r_u. In the fisheye image, the corresponding point $P_d(x_d, y_d)$ has an Euclidean distance r_d from the distortion center. We used a 4th-order polynomial model to describe the relationship between the perspective image and the fisheye image. This model is based on the equation:

$$\theta_u = \sum_{i=1}^{n} k_i \theta_d^{2i-1}, \quad n = 1, 2, 3, 4, \ldots \tag{2}$$

Where θ_u represents the angle of incident light and θ_d is the angle that light passes through the lens. The relationship between r_u and r_d can be expressed as:

$$r_u = f \sum_{i=1}^{n} k_i r_d^{2i-1}, \quad n = 1, 2, 3, 4, \ldots \tag{3}$$

Where f is the focal length of the fisheye camera. By merging the distortion parameters k_i and the focal length f, we obtained the final polynomial model:

$$r_u = \sum_{i=1}^{n} k_i r_d^{2i-1}, \quad n = 1, 2, 3, 4, \ldots \tag{4}$$

k_i denotes the distortion parameter, to be specific, we set the parameters range the same with PCN[13], which is randomly generated from the following ranges: $k_1 \in [1 \times 10^{-6}, 1 \times 10^{-4}], k_2 \in [1 \times 10^{-11}, 1 \times 10^{-9}], k_3 \in [1 \times 10^{-16}, 1 \times 10^{-14}], k_4 \in [1 \times 10^{-21}, 1 \times 10^{-19}]$.

(ii) Rectangling image find and selection. We noticed that there is classical panoramic image rectangling work[2], which makes the stitched image regular. It makes the stitched image regular by optimizing an energy function with line-preserving mesh deformation. Thus, we perform the same energy function on the rectified image to gain the rectangling images. We followed the same setting as RecRecNet[8], and there are 5160 pairs of rectified-rectangling images for the training process and 500 pairs for testing. The size of all images is 256×256.

(iii) Rectification and Rectangling Meshes Generation. RecRecNet[8] is a rectangling network for wide-angle image correction. It uses the TPS module to build nonlinear and non-rigid transformations, learning control points on the corrected image to warp the source structure to the target domain. We trained two RecRecNets, one for extracting the mesh of rectangling and the other for rectification.

(iv) Generalization Evaluation dataset. In addition to the synthesized dataset, we collect rectified wide-angle image results from the state-of-the-art rectification methods. Their results are derived from other types of datasets and real-world wide-angle lenses such as the Rokinon 8mm Cine Lens, Opteka 6.5mm Lens, and GoPro.

A.4. More Qualitative Comparison Results

As shown in Fig. 2, our ConBo-Net achieves simultaneous distortion rectification and boundary rectangling across diverse wide-angle scenes. The results preserve **structural integrity** (undistorted linear features) while maintaining **strict rectangular boundaries**, demonstrating a geometrically consistent representation. By contrast, previous two-stage methods struggle to reconcile these objectives: **(i) Content degradation** (*e.g.*, blurred textures in DR-GAN [7] and distorted shapes in RecRecNet [8]) arises from error accumulation during sequential rectification; **(ii) Boundary irregularity** (*e.g.*, black borders in SimFIR [1] and non-rectangular outputs in RecRecNet [8]) stems from misaligned optimization between stages.

These artifacts originate from cascaded pipelines where independent rectification and rectangling lack error feedback, propagating geometric inconsistencies irreversibly. For instance, SimFIR's black pixel edges directly result from

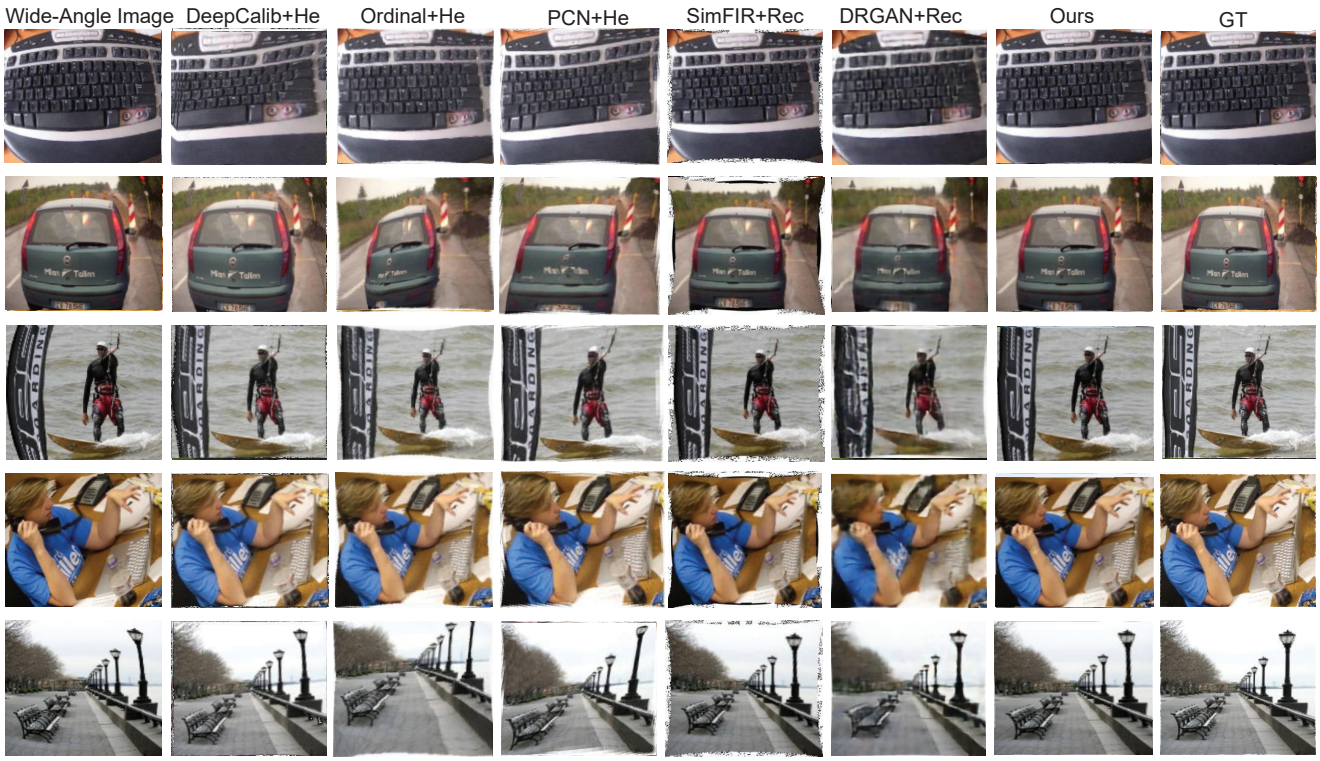

Figure 2. More qualitative results compared to two-stage approaches. Results demonstrate the superior effectiveness of our approach in correcting wide-angle distortions across diverse scenes. Examples include surfing scenarios, urban streets, vehicle photography, and close-up keyboard imagery. Compared to two-stage approaches, our method consistently produces results closest to the ground truth, particularly in preserving linear structures (*e.g.*, buildings and keyboards) and mitigating curvature artifacts in dynamic scenes (*e.g.*, moving vehicles and ocean waves).

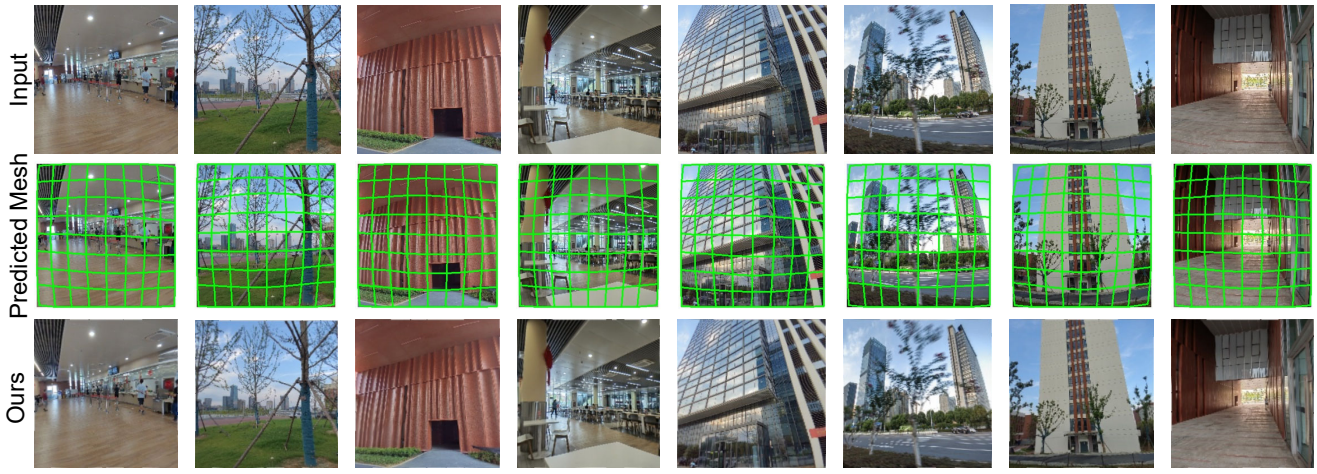

Figure 3. More results on the real-world photos from different lenses.

residual distortion left by its rectification stage, which the subsequent rectangling module cannot resolve without introducing content warping.

A.5. More Generalization Evaluation Results

As detailed in the main text, we test our method on distorted images captured by diverse real-world wide-angle lenses, including the GoPro Hero 10, Opteka 6.5mm Fish-

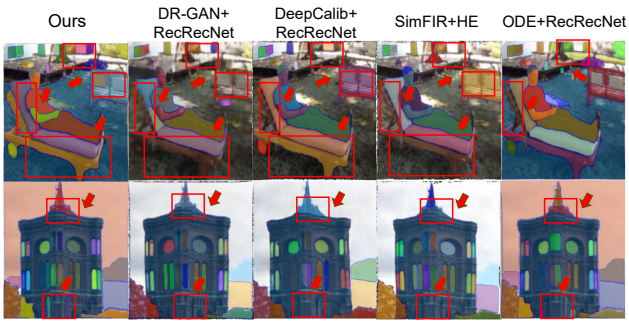

Figure 4. ConBo-Net enhances downstream vision perception tasks, such as image segmentation, as demonstrated in[5]. The arrows indicate instances where the two-stage approach fails to produce accurate results.

Figure 5. We compared ConBo-Net with other two-stage pipelines in terms of object detection performance. ConBo-Net achieves higher confidence scores in detecting objects like pizza and trains, showing its better detection performance.

eye, and SAMSUNG 10mm F3.5. As shown in Fig. 3, our approach robustly generalizes to complex outdoor scenes (*e.g.*, urban landscapes and natural environments) under varying distortion patterns while preserving rectangular image boundaries critical for perspective coherence. For instance, architectural structures (*e.g.*, building facades), linear objects (*e.g.*, streetlights), and roads are accurately restored to their intrinsic straight geometries, eliminating curvature artifacts. This results in processed images with significantly improved geometric fidelity, closely resembling undistorted real-world perspectives.

A.6. Downstream Tasks Performance

To evaluate the practical impact of our method on downstream vision tasks, we combine ConBo-Net with the Segment Anything Model (SAM) [5] for image segmentation and You Only Look Once (YOLO) [3] for object detection. Comparative experiments show that our method significantly improves segmentation and detection accuracy, especially in capturing fine structural details like thin edges and occlusion boundaries.

As illustrated in Fig. 4, in the segmentation task, the two-stage pipeline (DR-GAN [7] → RecRecNet [8]) exhibits critical limitations—highlighted by red arrows—including fragmented segmentation in high-curvature regions and blurred boundaries around deformable objects. For example, the two-stage method erroneously merges the backrest of the chair with the background. In contrast, our method achieves crisper segmentation coherence through distortion-aware feature adaptation. As shown in Fig. 5, ConBo-Net shows higher detection rates than other two-stage methods, such as DR-GAN+RecRecNet, PCN+HE, SimFIR+HE, and ODE+RecRecNet in YOLO evaluation. In the pizza detection example, ConBo-Net achieves higher confidence scores (0.80 and 0.72) than other methods, indicating more accurate identification of pizza objects. Similarly, in train detection, ConBo-Net attains a confidence score of 0.49, outperforming other approaches. ConBo-Net enhances image quality, which contributes to better object detection performance.

These results underscore our method's dual strength: it not only mitigates distortion-induced segmentation errors but also captures semantically critical regions (*e.g.*, functional object parts) that cascaded approaches often miss. By maintaining geometric fidelity and structural integrity, our approach significantly improves the interpretability and utility of segmentation outputs across diverse real-world scenarios.

References

[1] Hao Feng, Wendi Wang, Jiajun Deng, Wengang Zhou, Li Li, and Houqiang Li. Simfir: A simple framework for fisheye image rectification with self-supervised representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12418–12427, 2023. 2

[2] Kaiming He, Huiwen Chang, and Jian Sun. Rectangling panoramic images via warping. *ACM Transactions on Graphics*, page 1–10, 2013. 2

[3] Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Laughing, tkianai, Adam Hogan, lorenzomammana, yxNONG, AlexWang1900, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, Francisco Ingham, Frederik, Guilhen, Hatovix, Jake Poznanski, Jiacong Fang, Lijun Yu, changyu98, Mingyu Wang, Naman Gupta, Osama Akhtar, PetrDvoracek, and Prashant Rai. ultralytics/yolov5: v3.1 – bug fixes and performance improvements, 2020. 4

[4] J. Kannala and S.S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340, 2006. 2

[5] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023. 4

[6] Kyu-Yul Lee and Jae-Young Sim. Warping residual based image stitching for large parallax. In *2020 IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), pages 8195–8203, 2020. 1

[7] Kang Liao, Chunyu Lin, Yao Zhao, and Moncef Gabbouj. Dr-gan: Automatic radial distortion rectification using conditional gan in real-time. *IEEE Transactions on Circuits and Systems for Video Technology*, page 725–733, 2020. 2, 4

[8] Kang Liao, Lang Nie, Chunyu Lin, Zishuo Zheng, and Yao Zhao. Recrecnet: Rectangling rectified wide-angle images by thin-plate spline model and dof-based curriculum learning. *arXiv preprint arXiv:2301.01661*, 2023. 2, 4

[9] Jiangpeng Rong, Shiyao Huang, Zeyu Shang, and Xianghua Ying. Radial lens distortion correction using convolutional neural networks trained with synthesized images. In *Asian Conference on Computer Vision*, 2016. 1, 2

[10] Xueyang Wang, Zhixin Zheng, Jiandong Shao, Yule Duan, and Liang-Jian Deng. Adaptive rectangular convolution for remote sensing pansharpening, 2025. 1

[11] Zhucun Xue, Nan Xue, Gui-Song Xia, and Weiming Shen. Learning to calibrate straight lines for fisheye image rectification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[12] Lichun Yang, Bin Tian, Tianyin Zhang, Jiu Yong, null, and Jianwu Dang. Image rectangling network based on reparameterized transformer and assisted learning. *Scientific Reports*, page 6981, 2024. 1

[13] Shangrong Yang, Chunyu Lin, Kang Liao, Chunjie Zhang, and Yao Zhao. Progressively complementary network for fisheye image rectification using appearance flow. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[14] Xu-Cheng Yin, Xinchao Wang, Jun Yu, Maojun Zhang, Pascal Fua, and Dacheng Tao. Fisheyerecnet: A multi-context collaborative deep network for fisheye image rectification. *Cornell University - arXiv,Cornell University - arXiv*, 2018. 1, 2