# RecRecNet: Rectangling Rectified Wide-Angle Images by Thin-Plate Spline Model and DoF-based Curriculum Learning

Kang Liao† Lang Nie† Chunyu Lin* Zishuo Zheng Yao Zhao
Institute of Information Science, Beijing Jiaotong University
Beijing Key Laboratory of Advanced Information Science and Network
{kang_liao, nielang, cylin, yzhao}@bjtu.edu.cn

# 1. Supplemental Material

## 1.1. Overview

In this document, we provide the following supplementary content:

- Details of justification for rectangling (Section 1.2).
- Details of the dataset construction (Section 1.3).
- More qualitative results of the comparison methods and our approach (Section 1.4).
- More vision perception results (Section 1.5).
- More cross-domain evaluation results (Section 1.6).
- User study (Section 1.7).

## 1.2. Justification for Rectangling

In the main manuscript, we have demonstrated that our rectangling method can significantly facilitate the downstream vision tasks, improving the performance of object detection and semantic segmentation. Besides the benefits of vision tasks, we argue there are other reasons why the rectangling algorithm is worth investigating. First, rectangling allows a visually pleasant structure for humans and a normal format for devices. Previous literature [3, 4] revealed most users prefer rectangular boundaries for publishing, sharing, and printing photos. For example, over 99% images in the tag "panorama" on Flickr (flickr.com) have rectangular boundaries. And human vision system is more sensitive to irregular lines. Moreover, the rectangular image well fits the mainstream display window and screen. Instead, the blank region in a nonrectangular image occupies invalid space, which makes the data storage/compression inefficient.

*Corresponding author. †The first two authors contribute equally.

## 1.3. Details of the Dataset Construction

We construct a rectangling rectification dataset that severing the first dataset in the research region, and we would like to release it to promote the research development. In particular, our dataset is built using the following four steps: **(i) Wide-angle image and rectified image synthesis**. Since it is extremely challenging to collect large-scale paired wide-angle images and their rectified ground truth, we follow the existing distortion rectification approaches [11, 1, 14, 9, 13] to synthesize the dataset. First, the original images are collected from the MS-COCO dataset [10]. We leverage a $4^{th}$ order polynomial model to approximate the radial distortion of the wide-angle image, which is verified to meet most projection models with high accuracy. To be specific, four distortion parameters are randomly generated from the following ranges: $k_1 \in [-1\times10^{-4}, -1\times10^{-8}]$, $k_2 \in [1\times10^{-12}, 1\times10^{-8}]$ or $\in [-1\times10^{-8}, -1\times10^{-12}]$, $k_3 \in [1\times10^{-16}, 1\times10^{-12}]$ or $\in [-1\times10^{-12}, -1\times10^{-16}]$, and $k_4 \in [1\times10^{-20}, 1\times10^{-16}]$ or $\in [-1\times10^{-16}, -1\times10^{-20}]$. Then we perform the distortion rectification on the wide-angle image and obtain the rectified images. **(ii) Rectangling the rectified image**. Formulating an accurate transformation model for rectangling the rectified image is difficult due to its non-linear and non-rigid characteristics. We notice that there is a classical panoramic image rectangling technique He et al. [3] on computer graphics, it makes the stitched image regular by optimizing an energy function with line-preserving mesh deformation. Thus, we perform the same energy function on our rectified image dataset to fit the rectangling rectification task. However, the capability to preserve linear structures in He et al. [3] is limited by line detection. Consequently, some rectangling rectified images have nonnegligible distortions. To overcome this issue, we carefully filter all rectangling results and repeat the selection process three times, resulting in 5,160 training data from 30,000 source images and 500 test data from 2,000 source images. Each manual operation takes around 10s. The size of all

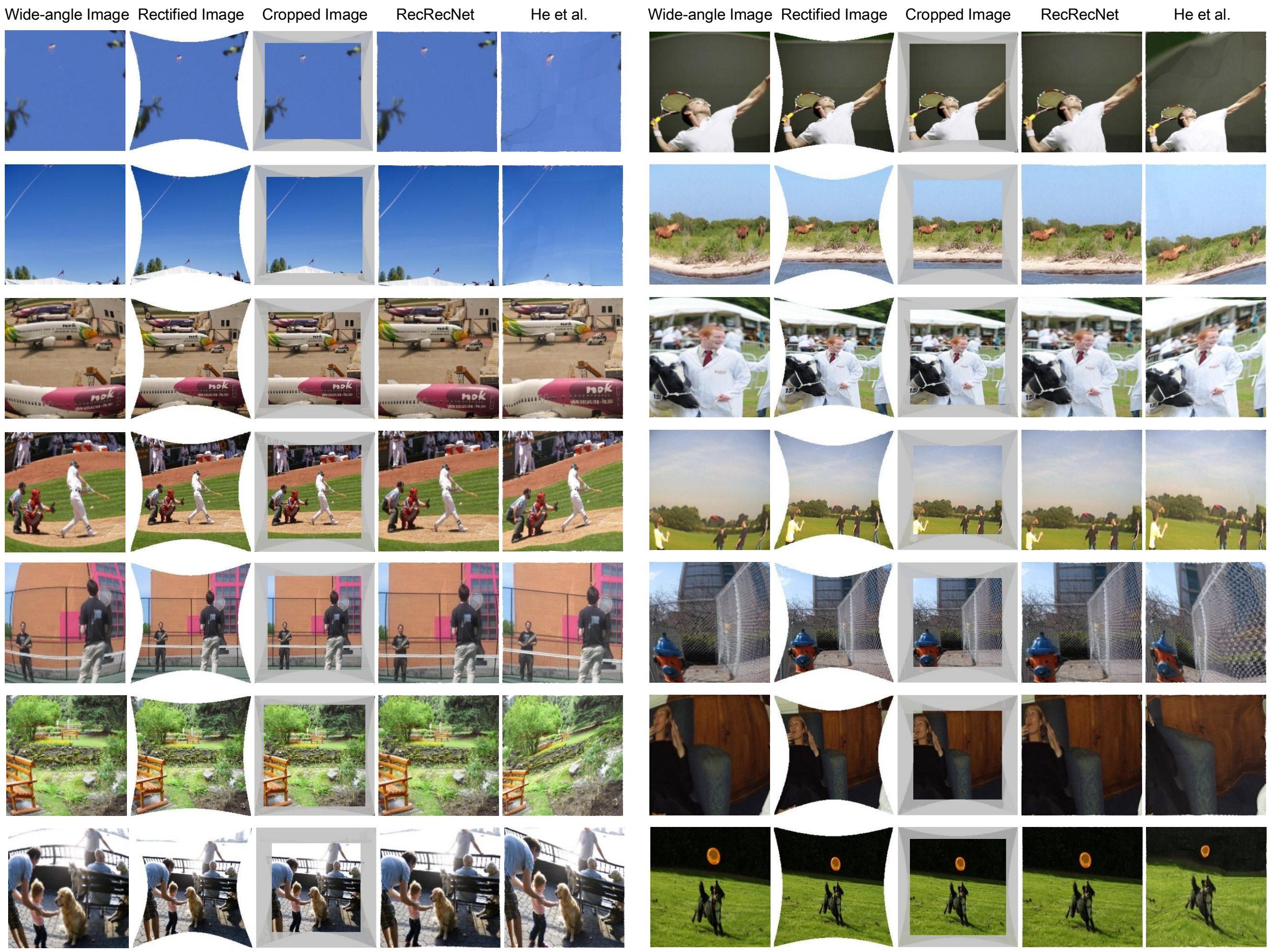


Figure 1. More qualitative results compared to He et al. [3]. We show the wide-angle image, rectified image, cropped rectified image, and the rectangling result of our RecRecNet, as well as the rectangling result by He et al. [3] from left to right.

images is $256 \times 256$. **(iii) Cross-domain validation**. In addition to the synthesized dataset, we collect 300 rectified wide-angle image results from the state-of-the-art rectification methods [8, 13]. 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. **(iv) DoF-based curriculum dataset**. To relieve the challenge of the structure approximation in the rectangling task, we proposed a Degree of Freedom (DoF)-based curriculum learning. Specifically, three curriculum stages are leveraged to inspire our RecRecNet, namely, from similarity transformation (4-DoF) to homography transformation (8-DoF), and to rectangling transformation. Thus, we also construct two datasets (4-DoF dataset and 8-DoF dataset) to provide the basic transformation knowledge, in which each dataset contains 5,000 image pairs. For the transformation synthesis, we randomly perturb four corners of the original image and warp it to the target image following the previous work [2].

### 1.4. More Qualitative Comparison Results

As shown in Figure 1, we exhibit more qualitative comparison results. Our RecRecNet is capable of rectangling the rectified wide-angle image in various scenes. We can observe the deformed boundary is straightened and the image content is rearranged to keep undistorted by RecRecNet, contributing a win-win rectification representation for the wide-angle image. By contrast, previous methods fail to trade off the image boundary and image content in the rectified image. Their results usually show incomplete content or distorted distributions. For example, some rectangling results produced by He et al. [3] rotate the original scene and twist the object due to their line-preserving mesh deformation.

### 1.5. More Vision Perception Results

As illustrated in Figure 2, we show more vision perception results, including the object detection and seman-


Figure 2. More detection and segmentation results by Mask R-CNN [5]. We show the rectified wide-angle image and its vision perception result, and our rectangling result and its vision perception result, from top to bottom.

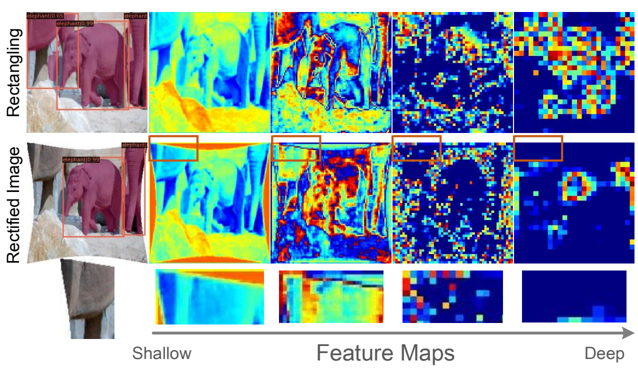


Figure 3. Failure case of Mask R-CNN [5] in regards to the missing perception. We find that the deformed boundary can introduce new features to the original feature maps. As a result, the network cannot recognize the elephant near the deformed image boundary as the original features are blinded in the deep feature maps.

tic segmentation results, which are derived by the popular perception model Mask R-CNN [5]. As we can observe, the learning model is misled by the deformed structure in the original rectified image, leading to wrong perception and missing perception results. As a benefit of the proposed flexible TPS transformation module and DoF-based curriculum learning, our RecRecNet can significantly help the downstream vision tasks by eliminating the deformed boundary issue. And the perception performance is recovered especially near the image boundary. For the performance degradation in rectified images, one straightforward reason is that the perception model has never seen the data with a deformed image boundary during training, and the domain gap between test data and training data confuses it on rectified images. Nevertheless, we would like to explore deeper reasons and disclose why the deformed geometry makes the vision perception deformed.

A sample is shown in Figure 3 to visualize why the deformed image boundary makes the perception deformed. As described in the main manuscript, we found the deformed boundary can introduce *new features* onto the feature maps, which (i) form new semantics (leads to the wrong perception) or (ii) cause blind spots for original features (leads to missing perception). The noticeable line artifacts and curve artifacts can be observed at the outermost boundaries and deformed boundaries respectively in shallow feature maps. In particular, the line artifacts are essentially generated by zero padding. Zero padding is a fundamental component of prominent CNNs architectures, it serves to maintain the size of the feature maps across the network by adding zero values to the feature map. When a convolutional kernel extracts the feature at the boundary with zero padding, the sharp transitions between the zero values and the original content are wrongly identified as an edge.

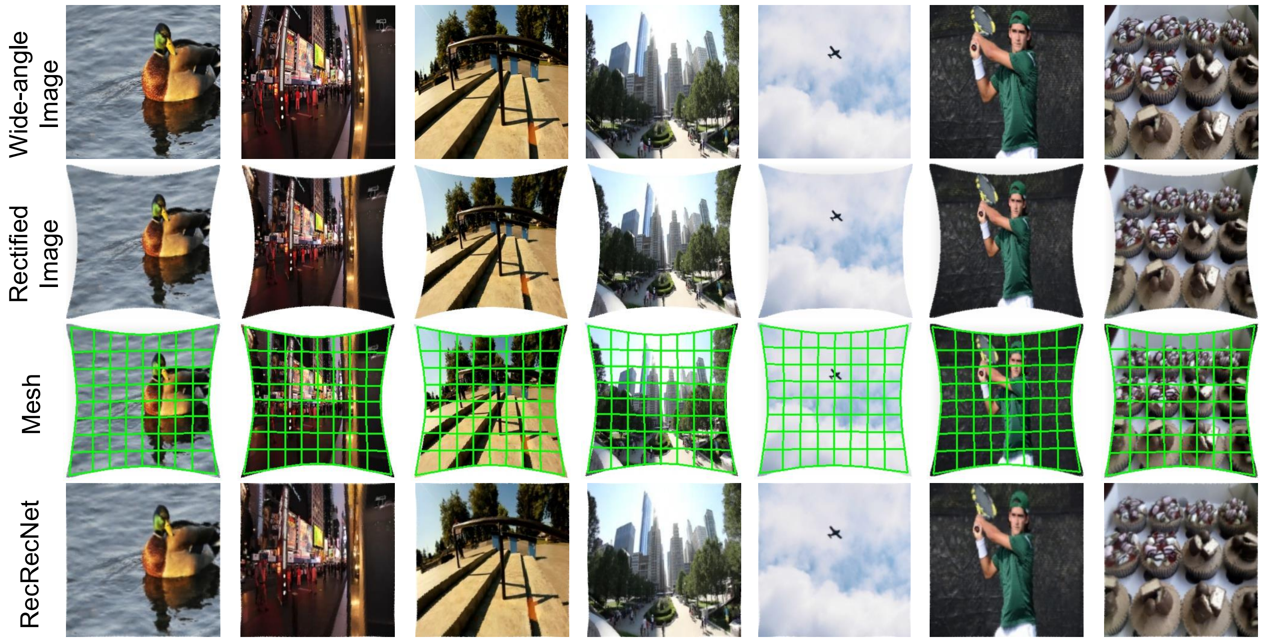


Figure 4. More results for the cross-domain evaluation. We show the wide-angle image, rectified image, predicted mesh, and rectangling result of our RecRecNet from top to bottom.

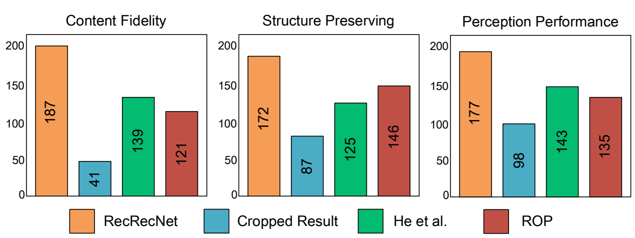


Figure 5. User study for the rectangling rectification results.

Compared to the regular image, the rectified image allows a faster extension of the edge effect. Thus, more new features will be involved in the inference process. We call it a dual-edge extension effect.

Recent works also demonstrated that the zero padding can unintentionally leak the positional information in CNNs [6, 7, 12]. In the case of the rectified image with a deformed boundary, we reason that the original positional information could be disturbed by the introduced new features. As a result, the perception model cannot build an accurate spatial distribution and fails to understand the relationships among adjacent semantics. Furthermore, it is interesting to find that the new semantic features introduced by the deformed boundary are jointly understood with their near features. In other words, the neural networks perceive the object relation and the scene context.

For the missing perception case, if the features near the boundary are non-salient, the newly introduced features can cover them and generate blind spots for the perception model. As a result, the network cannot recognize the elephant near the deformed image boundary. We argue that such an effect can raise in most image transformation cases such as image warping, in which the boundaries are deformed to trade off different requirements on image content.

### 1.6. More Cross-Domain Evaluation Results

As mentioned in Section 1.3, we collect 300 rectified wide-angle image results from the state-of-the-art rectification methods [8, 13] to conduct the cross-domain evaluation. Their results are derived from other types of synthesized datasets and real-world wide-angle lenses with different camera models. Figure 4 shows the cross-domain evaluation results. While the new data domain has never been seen, our RecRecNet can achieve a promising generalization ability by learning a flexible TPS transformation. And we can observe that the rectangling results display reasonable global distributions and visually pleasing local details. Moreover, the predicted mesh locates at the most spatial range of the rectified image, demonstrating the effectiveness of the learned rectangling transformation.

### 1.7. User Study

The aim to yield the rectangular images is mainly for the visual sense and vision perception, thus we conduct a user study to evaluate the rectangling methods based on three aspects: content fidelity, structure-preserving, and perception performance. In particular, content fidelity denotes the measurement of the image content, such as the objects and background. Structure-preserving requires evaluation, especially for the rectangling image boundary, namely, if it is straight or not. Perception performance means the intuitive object detection and semantic segmentation results. Subsequently, we collect 200 samples in random order, and 10 volunteers with vision expertise are required to vote on the results under different contexts. As shown in Figure 5, RecRecNet achieves the highest votes in three tests and shows a superior capacity for scene faithfulness.

## References


[1] Oleksandr Bogdan, Viktor Eckstein, Francois Rameau, and Jean-Charles Bazin. Deepcalib: a deep learning approach for automatic intrinsic calibration of wide field-of-view cameras. In *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*, pages 1–10, 2018. 1

[2] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation. *arXiv preprint arXiv:1606.03798*, 2016. 2

[3] Kaiming He, Huiwen Chang, and Jian Sun. Rectangling panoramic images via warping. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013. 1, 2

[4] Kaiming He and et al. Content-aware rotation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 553–560, 2013. 1

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 3

[6] Md Amirul Islam, Sen Jia, and Neil DB Bruce. How much position information do convolutional neural networks encode? In *International Conference on Learning Representations*, 2019. 4

[7] Osman Semih Kayhan and Jan C van Gemert. On translation invariance in cnns: Convolutional layers can exploit absolute spatial location. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14274–14285, 2020. 4

[8] Kang Liao, Chunyu Lin, Lixin Liao, Yao Zhao, and Weiyao Lin. Multi-level curriculum for training a distortion-aware barrel distortion rectification model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4389–4398, 2021. 2, 4

[9] 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*, 30(3):725–733, 2020. 1

[10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 1

[11] 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*, pages 35–49. Springer, 2016. 1

[12] Rui Xu, Xintao Wang, Kai Chen, Bolei Zhou, and Chen Change Loy. Positional encoding as spatial inductive bias in gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13569–13578, 2021. 4

[13] Shangrong Yang, Chunyu Lin, Kang Liao, Chunjie Zhang, and Yao Zhao. Progressively complementary network for fisheye image rectification using appearance flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6348–6357, June 2021. 1, 2, 4

[14] Xiaoqing Yin, Xinchao Wang, Jun Yu, Maojun Zhang, Pascal Fua, and Dacheng Tao. Fisheyerecnet: A multi-context collaborative deep network for fisheye image rectification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 1