Towards Complete Scene and Regular Shape for Distortion Rectification by Curve-Aware Extrapolation

Kang Liao Chunyu Lin* Yunchao Wei Feng Li Shangrong Yang Yao Zhao
Institute of Information Science, Beijing Jiaotong University
Beijing Key Laboratory of Advanced Information Science and Network
{kang_liao, cylin}@bjtu.edu.cn, wychao1987@gmail.com, {l1feng, sr_yang, yzhao}@bjtu.edu.cn

Abstract

The wide-angle lens gains increasing attention since it can capture a wide field-of-view (FoV) scene. However, the obtained image is contaminated with radial distortion, making the scene not realistic. Previous distortion rectification methods rectify the image in a rectangle or invagination, failing to display the complete content and regular shape simultaneously. In this paper, we rethink the representation of rectification results and present a Rectification OutPainting (ROP) method, aiming to extrapolate the coherent semantics to the blank area and create a wider FoV beyond the original wide-angle lens. To address the specific challenges such as the variable painting region and curve boundary, a rectification module is designed to rectify the image with geometry supervision, and the extrapolated results are generated using a dual conditional expansion strategy. In terms of the spatially discounted correlation, a curve-aware correlation measurement is proposed to focus on the generated region to enforce the local consistency. To our knowledge, we are the first to tackle the challenging rectification via outpainting, and our curve-aware strategy can reach a rectification construction with complete content and regular shape. Extensive experiments well demonstrate the superiority of our ROP over other state-of-the-art solutions.

1. Introduction

A wide-angle lens that can capture the wide field-of-view (FoV) scene, gains increasing applications in recent years. Humans have the natural ability to rectify and understand the distorted scene induced by a wide-angle lens. For computer vision, accomplishing this task requires recovering the realistic geometric distribution based on handcrafted features [6, 8, 29, 14, 1, 26] or deep features [23, 39, 18, 35, 17, 16, 5]. The existing distortion rectification methods generate the rectified image as rectangle or invagination, both of which have the limitations to display the complete content and regular shape simultaneously. For instance, the rectangle construction displays the rectified scene in a regular shape, but it discards the content in the image boundary. Such a representation cannot provide the complete scene and weakens the advantage of the wide-angle lens. Although the invagination construction covers complete contents, it introduces an irregular shape with a visually narrow FoV. These two representations have their strengths and weaknesses. It would be appealing and meaningful if we contribute a representation combined with the complete content and regular shape in the meantime.

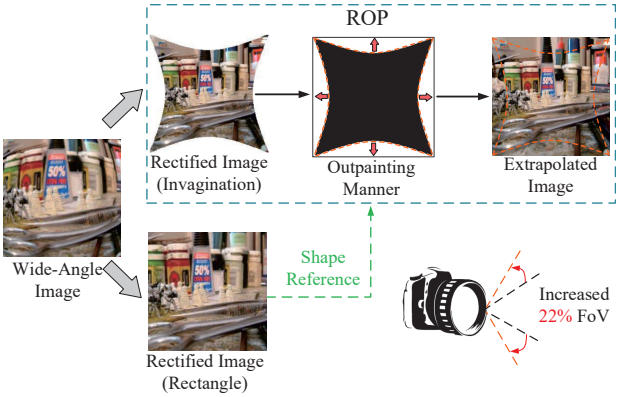

Figure 1. Motivation of the presented Rectification OutPainting (ROP) method. ROP aims to combine the advantages of two traditional rectification constructions: complete content and regular shape, while gaining wider field-of-view (FoV) beyond the original wide-angle lens. Note that the structure of the rectangle rectified image is used as a shape reference for outpainting.

Given an incomplete image, the goal of the image extrapolation is to hallucinate plausible visual contents outside the original boundaries [25, 31, 33, 37, 9, 34]. Different with the actively studied inpainting [40, 41, 38, 12, 36, 13, 42, 43, 44], outpainting receives less attention and is more challenging due to its one-side constraint. In this paper, we rethink the traditional construction of distortion

*Corresponding author

rectification and explore to establish a new representation using outpainting.

As shown in Fig. 1, we propose a Rectification OutPainting (ROP) method. Given a distorted image captured by a wide-angle lens, ROP aims to recover the realistic complete scene in the invagination form and construct a regular shape. However, ROP is challenging due to the following characteristics: (1) *Variable painting regions*. For the rectification result, the blank region surrounding the valid content is different in each image, owing to various degrees of distortions. (2) *Curve boundary*. In contrast to the straight boundary, our ROP extrapolates the coherent semantics and details from the curve boundary. This special structure causes a more complicated spatial correlation between the generated content and the original content.

To be specific, we design a parametric framework to address the aforementioned challenges. First, the distortion rectification module rectifies the input image in the form of invagination with general geometry supervision. Subsequently, the rectified image and the filling mask are fed into an outpainting module to extrapolate the semantically consistent content into the blank region using a dual conditional expansion strategy. Such a strategy considers both the global distribution and expanded content, guiding the network to perceive different filling regions explicitly. Moreover, a curve-aware correlation measurement is presented to enforce the local consistency of the extrapolated content. Extensive experiments demonstrate that our approach can recover the realistic details from the distorted image, with a complete scene and regular shape. We also show that our ROP method enables the captured scene to exhibit a wider FoV beyond the original wide-angle lens, which allows for more accurate scene reasoning and other applications.

In general, our contributions are summarized as follows:

- We are the first to propose a Rectification OutPainting (ROP) method to eliminate the inherent limitations of traditional rectification representations.
- A general geometry supervision and a dual conditional expansion strategy are designed to achieve accurate distortion rectification and image outpainting.
- To address the spatially discounted correlation, a curve-aware correlation measurement is proposed to enforce the local consistency of extrapolation results.

2. Related Works

Distortion Rectification: Suffering from the severe distortion induced by a wide-angle lens, the distortion rectification task plays a vital role in computer vision. Previous methods solve the rectification problem mainly based on the hand-crafted features [6, 8, 29, 14, 1, 26] or deep features [23, 39, 18, 35, 17, 16, 5]. For example, Kang et al. [14] leveraged the consistency of pairwise tracked point features in a sequence to self-calibrate the catadioptric camera. The curve-based methods [6, 8, 29, 1] predicted the distortion parameters in terms of the detected distorted lines and removed the distortion using the estimated parameters. On the other hand, learning methods expanded the distortion rectification with deep features. For instance, the convolutional neural networks are used to extract the semantic features of distorted images and predict the distortion parameters [23, 39, 35, 5]. To achieve the blind rectification, the generation-based methods [18, 17, 16] employed the encoder-decoder architecture to rectify the distorted image, enables a flexible paradigm of parameter-free rectification.

Despite the encouraging performance achieved by the above methods, the construction of rectification results has the limitations of displaying a complete content and regular shape simultaneously. In this work, we would like to draw attention from the precise rectification performance to the few concerned representations of rectification results.

Image Completion: Image completion targets reconstructing missing parts in damaged images, of which the image inpainting technique gains well explored. Prior works in inpainting can be classified into traditional methods [2, 4, 3, 27] and learning methods [40, 38, 12, 36, 42, 43, 44]. Image outpainting is more challenging than inpainting due to the one-side constraint. For the first time, Sabini et al. [25] achieved the deep learning-based outpainting with the powerful generation ability of the generative adversarial networks. The feature expansion module and content prediction module are proposed to improve the outpainting performance gradually in [33]. Teterwak et al. [31] conditioned the discriminator with pre-trained features from InceptionV3 network, which enables the outpainting image to match the ground truth in the semantics space. Guo et al. [9] designed a spiral generative network to conduct the image extrapolation following the perception fashion of humans.

Thanks to the above methods' promising efforts, we get a meaningful inspiration and investigate to utilize the outpainting technique to construct a novel representation for the distortion rectification result.

3. Methodology

3.1. Problem Formulation

Given a distorted image $I_d \in \mathbb{R}^{h\times w\times 3}$, our rectification outpainting (ROP) aims to generate a rectified and extrapolated image $I_{re} \in \mathbb{R}^{h\times w\times 3}$. To be specific, the distortion rectification module takes I_d as the input and outputs a rectified image $I_r \in \mathbb{R}^{h\times w\times 3}$ and a filling binary mask $M \in \mathbb{R}^{h\times w\times 1}$. Then, the outpainting module receives the I_r and M, generating I_{re} with the complete content and regular shape.

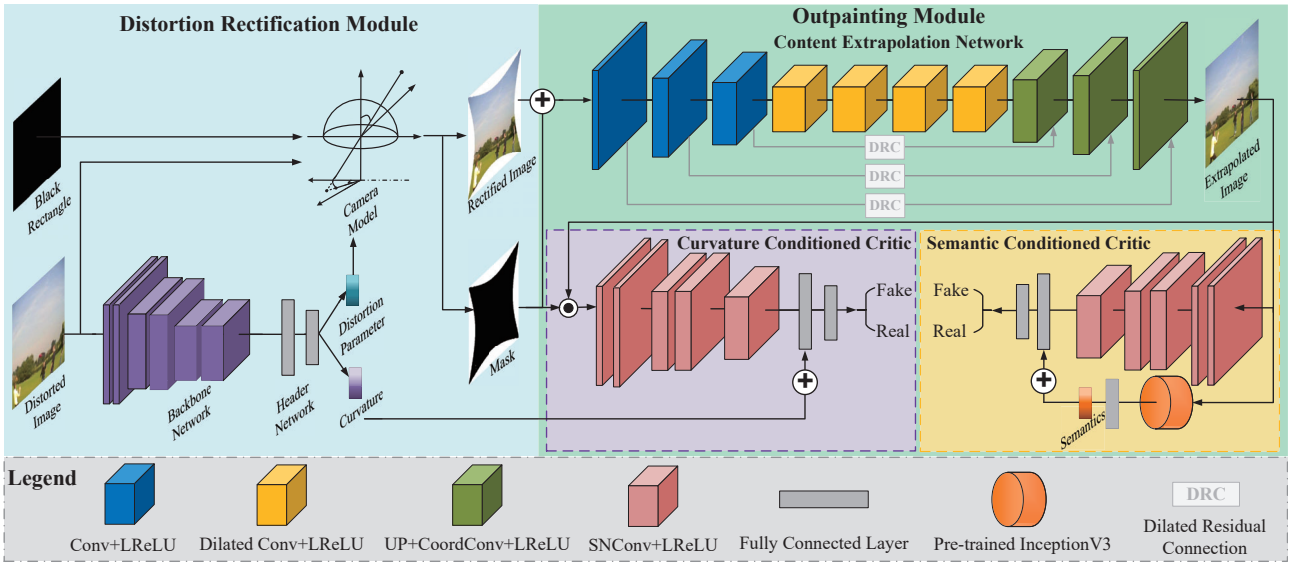

Figure 2. An overview of our rectification outpainting (ROP) framework. In this work, the crucial concept: curvature, can bridge the relationship between the distortion rectification module and the outpainting module.

3.2. Framework Design

3.2.1 Distortion Rectification Module

Our framework comprises a distortion rectification module (DRM) and an outpainting module (OM), as shown in Fig. 2. DRM rectifies the distorted image in the form of invagination and generates a filling mask, which is supervised by the proposed general geometry rectification.

Detailed Parameter Estimation: Considering that the general polynomial camera model [15] is widely used in the approximation of wide-angle lens, we formulate the relationship between the projection manner and distortion parameters as follows.

$$r(\theta) = \sum_{i=1}^{N} k_i \theta^{2i-1}, N = 1, 2, 3, \cdots, \tag{1}$$

where r is the distance between the principal point and the pixels in the image, and θ represents the angle between the incident ray and the optical axis of wide-angle lens.

To rectify the distorted image, we design a learning model consisting of a backbone network and a header network. In particular, the backbone network extracts the general representation of the distortion context in the form of the high-level semantic features using stacked convolutional layers. We pre-train the backbone network on ImageNet dataset [7] and fine-tune it on our distorted image dataset. Here, the classical networks such as the VGG and ResNet (removed the fully connected layers) can be exploited in the architecture. The header network, composed of fully connected layers, combines the general representation of the distorted image and constructs a feature vector to estimate the distortion parameters. Then, the pixel in the distorted image can be warped and rectified by the estimated parameters. In addition, we rectify a black rectangle using the same parameters to provide a filling mask.

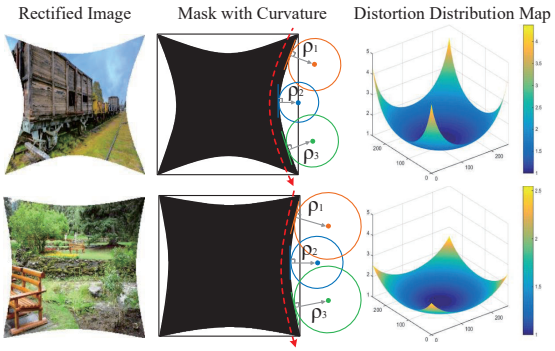

Figure 3. Explicit geometry relation between the rectified image and distortion feature bridged by the curvature.

General Geometry Supervision: Using the above network, we can obtain the distortion parameters and rectify the distorted image. However, accurately regressing these parameters is still difficult due to the implicit relationship between the learning target and the image's visual feature. Thus, we present a general geometry supervision to constrain the explicit structure of the rectified image.

As illustrated in Fig. 3, we show two examples of rectified images, filling masks marked with osculating circles and curvatures, and 3D distortion distribution maps (DDM). Intuitively, the greater the degree of invagination in a rectified image, the greater the curvature of the rectified boundary, and the greater the distortion in its corresponding 3D

DDM. For a boundary curve $f(x)$ of rectified image or filling mask, the curvature ρ of point $(t, f(t))$ can be described by the radius r_{os} of an osculating circle O_{os}

$$\rho = \frac{1}{r_{os}} = \frac{|f''(t)|}{(1 + (f'(t))^2)^{\frac{3}{2}}}. \quad (2)$$

3D distortion distribution map denoted as $\mathbf{D}$, shows the distortion degree and distortion distribution in a distorted image, which is formed based on the distortion level $\mathcal{D}$ [18]:

$$\mathbf{D}(x, y) = \{x, y, \mathcal{D}(x, y)\}, 0 \leq x \leq W, 0 \leq y \leq H, \quad (3)$$

where W and H indicate the width and height of distorted image. Then, the description of $\mathcal{D}$ can be given as follows:

$$\mathcal{D}(x, y) = \frac{x'}{x} = \frac{y'}{y} = k_1 + k_2 l + k_3 l^2 + \cdots, \quad (4)$$

where (x', y') is the coordinate of the corresponding pixel in rectified image, l represents the Euclidean distance between the pixel (x, y) in a distorted image and the principal point. Assuming a pixel (x_b, y_b) located at the boundary in a distorted image is rectified to a pixel (x'_b, y'_b) located at the curve boundary, then the distortion level $\mathcal{D}(x_b, y_b)$ that represents the ratio relation of coordinates also indicates the deviation degree of this curve boundary, i.e., curvature. Consequently, a mathematical relation can be derived: $\mathcal{D} \propto \rho$.

From the above equations and Fig. 3, we can construct an intuitive relationship between the structural appearance r_{os} of the rectified image and the parametric distortion model $\mathcal{D}$ of a distorted image, bridged by the curvature ρ. Thus, the rectification process of a distorted image can be constrained in terms of the global structure. In the implementation, we define a maximum curvature ρ_{max} for the curve boundary in the rectified image as another label. Besides the detailed distortion parameter supervision, we also teach the learning model to focus on the general prior knowledge of distortion using the ρ_{max} estimation. As a result, our distortion rectification module gains a more comprehensive perception of distortion and achieves more accurate rectification performance. Most important of all, the estimated curvature, also implying the area of extrapolated content in the filling mask, effectively guides the subsequent outpainting module.

3.2.2 Outpainting Module

Outpainting module consists of a content extrapolation network, a curvature conditional critic network, and a semantic conditional critic network. The last two parts are dubbed as a dual conditional expansion paradigm.

Content Extrapolation Network: The outpainting can be regarded as an image-to-image transformation task, and thus we apply the encoder-decoder architecture. Different from the classical U-Net [24], our content extrapolation network G has the following specific designs. First, to capture the long-range spatial information with a large receptive field, we leverage the kernels of size 7×7 in the first hierarchy of the encoder part. Moreover, the dilated convolutional group with the dilation rate of 2, 4, 8, 16 is used at the end of the encoder. Second, due to some coarse-rectified results, we implement the Coordinate Convolution (CoordConv) [20] in each convolutional layer of the decoder to help recover the accurate distribution. Finally, a dilated residual connection (DRC) is designed to replace the skip connection. Particularly, the DRC module that consists of three residual blocks with increasing dilated rates, takes various receptive fields of inputs to selectively adopt the useful content from early layers to the decoder. Thus, we can relieve the issues such as introducing the blurred content caused by common skip connection operations and pre-paint the feature maps from the encoder.

Curvature Conditional Critic: In the adversarial learning strategy, a critic or discriminator network is often utilized to judge if the generated result is real or fake, improving the reasonable approximation of data distribution. For the outpainting task that hallucinates plausible semantics outside the original boundaries, it requires a more discerning critic than that of simple image-to-image translation tasks. As mentioned above, our outpainting differs from conventional outpainting, especially in the variable filling regions and curve boundaries. In Section 3.2.1, we built a intuitive relationship between the degree of invagination in the rectified image and the curvature of the rectified boundary. The degree of invagination also shows a positive correlation with the area of the filling region. Thus, we can guide the network to explicitly perceive the variable filling regions of rectified images using this prior knowledge. Specifically, we design a critic network D^c and condition it with the maximum curvature of the rectified boundary,

$$\min_G \max_{D^c} \mathcal{L}_{adv1}(G, D^c) = \mathbb{E}_{y \sim data(y)} \log D^c((y \odot M)|\rho_{max}) + \mathbb{E}_{x \sim data(x)} \log (1 - D^c((G(x) \odot M)|\rho_{max})). \quad (5)$$

Recall that ρ_{max} implies the extrapolated area and M indicates the filling mask. We denote the data distributions as $x \sim data(x)$ and $y \sim data(y)$, where x and y represent the rectified image and the ground truth extrapolated result. During the adversarial learning, we encourage the curvature conditioned critic to pay more attention to the extrapolated results with the larger filling region. Thus, the outpainting module gains an evident cognition regarding the variable filling regions and produces better expansion images.

Semantic Conditional Critic: Inspired by the conditional projection adversarial works [21, 31], we then present a semantic conditioned expansion scheme. Unlike the above

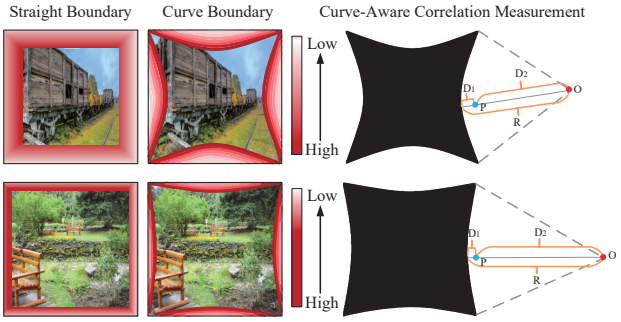

Figure 4. Curve-aware correlation measurement strategy. At the first two columns, we show different spatially discounted correlations in conventional outpainting and our outpainting. The spatial correlation (D_1) of a filling pixel P can be calculated by the subtraction of the radius R and its distance (D_2) to center (O), based on a circle including the curve boundary as an arc.

works, our main aim is to enable an authoritative reference feature from a pre-trained learning model to the outpainting critic network D^s. In detail, the backbone network of InceptionV3 model I [30] is added into our adversarial learning strategy, which is pre-trained on ImageNet [7] and provides the high-level semantic features of the input image. These reference features are further combined with the features extracted by the original convolutional layers D^s_c in critic network, to discriminate if the image is derived from generator G or not, using fully connected layers D^s_f:

$$\min_G \max_{D^s} \mathcal{L}_{adv2}(G, D^s) = \mathbb{E}_{y\sim data(y)} \log D^s_f(D^s_c(y) + I(y)) + \mathbb{E}_{x\sim data(x)} \log\left(1 - D^s_f(D^s_c(G(x)) + I(G(x))\right), \tag{6}$$

Getting the well-represented reference feature from the pre-trained model, our outpainting critic network achieves more comprehensive discrimination, impelling the generator network to hallucinate more realistic extrapolated results.

3.2.3 Curve-Aware Correlation Measurement

Previous outpainting methods reason the generation region based on a spatially discounted reconstruction. In general, the pixel far away from the boundary has less relation to original content. Concretely, we can measure this spatial correlation with a weighted mask as M^w:

$$M^w(i,j) = \lambda^{dis(i,j,\mathcal{B})}, \ \ 0 < \lambda < 1, \tag{7}$$

where $dis(\cdot)$ calculates the distance from pixel (i,j) to its nearest boundary $\mathcal{B}$. In contrast to the straight boundary case, our outpainting needs to create the content from the curve boundary, showing a more complicated spatial correlation between the generated content and the original content. Thus, we design a curve-aware correlation strategy to measure the spatially discounted reconstruction. As illustrated in Fig. 4, we show two examples of how to calculate the distance between a filling pixel and its adjacent curve boundary. Supposed that the curve boundary of rectified image derives from an arc $\mathcal{A}$ of a circle $\mathcal{O}$, and then we can compute the center (x_c, y_c) and radius R using any three points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ on $\mathcal{A}$:

$$\begin{cases} x_c = (bf - ec)/(bd - ea); \\ y_c = (dc - af)/(bd - ea); \\ R = \sqrt{(x_c - x_1)^2 + (y_c - y_1)^2}, \end{cases} \tag{8}$$

where the above coefficients can be obtained by:

$$\begin{cases} a = 2(x_2 - x_1); \\ b = 2(y_2 - y_1); \\ c = x_2^2 + y_2^2 - x_1^2 - y_1^2; \\ d = 2(x_3 - x_2); \\ e = 2(y_3 - y_2); \\ f = x_3^2 + y_3^2 - x_2^2 - y_2^2. \end{cases} \tag{9}$$

Subsequently, the distance from a filling pixel (x_f, y_f) to its nearest curve boundary $\mathcal{A}_f$ can be given by:

$$dis(x_f, y_f, \mathcal{A}_f) = ||R - \sqrt{(x_f - x_c)^2 + (y_f - y_c)^2}||_1. \tag{10}$$

With the proposed curve-aware correlation measurement, we enable the outpainting module to extrapolate the consistent semantics from the curve boundary, achieving more reasonable and coherent completed reconstruction.

3.3. Training Loss

The overall framework is separately optimized for the distortion rectification module (DRM) and the outpainting module (OM). First, DRM is designed based on a parameter supervision $\mathcal{L}_{ps}$ and a curvature supervision $\mathcal{L}_{cs}$ by:

$$\mathcal{L}_{DRM} = \mathcal{L}_{ps} + \mathcal{L}_{cs}, \tag{11}$$

where we use the smooth $\mathcal{L}_1$ [22] to calculate the value of $\mathcal{L}_{ps}$ and $\mathcal{L}_{cs}$, which can alleviate the exploding gradient problem during the training process.

Subsequently, OM extrapolates new contents with the dual conditional expansion and curve-aware correlation measurement, of which the parameters can be learned by

$$\mathcal{L}_{OM} = \mathcal{L}_{dc} + \mathcal{L}_{ca} + \lambda_{adv1}\mathcal{L}_{adv1} + \lambda_{adv2}\mathcal{L}_{adv2}, \tag{12}$$

where $\mathcal{L}_{adv1}$ and $\mathcal{L}_{adv2}$ are introduced in Eq. 5 and 6. λ_{adv1} and λ_{adv1} are the balance factors. $\mathcal{L}_{dc}$ computes the difference between the outpainting image $\hat{I}_{rop}$ and ground truth I_{rop} on the feature maps $\phi_{i,j}$, obtained from the j-th convolution before the i-th maxpooling layer in the VGG19 [28]:

$$\mathcal{L}_{dc} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} ||\phi_{i,j}(I_{rop})_{x,y} - \phi_{i,j}(\hat{I}_{rop})_{x,y}||_2. \tag{13}$$

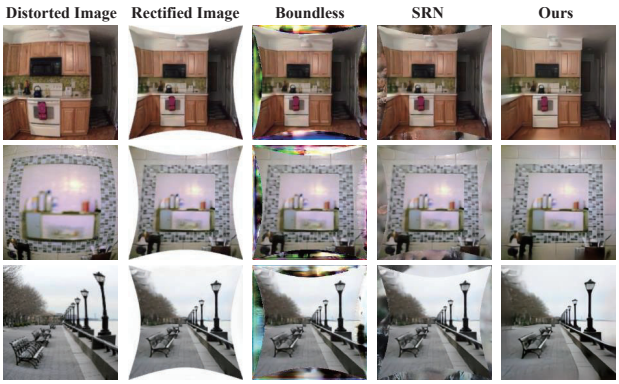

Figure 5. Visual comparisons of the outpainting methods: Boundless [31] and SRN [33] and our curve-aware outpainting.

The $\mathcal{L}_{ca}$ focuses on the optimization of extrapolated contents. It contains a curve-aware correlation measurement to weigh the spatially discounted reconstruction.

$$\mathcal{L}_{ca} = \mathcal{L}_{dc}(I_{rop} \odot M, \hat{I}_{rop} \odot M) \odot M^{w}, \qquad (14)$$

where M^{w} can be calculated by Eq. 7 and Eq. 10.

4. Experiment

4.1. Implementation Details

To train our framework, we establish a standard dataset using original images from MS-COCO [19]. To be specific, our dataset consists of the distorted image, rectified image, filling mask, maximum curvature of mask boundary, and extrapolated image. Considering the general polynomial camera model [15] is widely used for the approximation of wide-angle lens, we construct our dataset based on this model. The transformation of coordinates is described in Eq. 1. Following previous distortion rectification methods [39, 18, 17], we leverage the fourth order polynomial model to synthesize the distorted image, which is sufficient for the approximation of most application scenarios.

For the outpainting module, we add the instance normalization [32] after each convolutional layer, except for the first layer of the generator network, because it can reduce the artifacts in generated results [31]. During training, Adam optimizer with learning rates 5×10^{-4} and 10^{-4} are adopted for the distortion rectification module, the generator network and critic network in the outpainting module. The batch size of training is set to be 4. The input image and output results are linearly clipped within the range $[-1, 1]$. For the hyper-parameters of training loss, We empirically set $\lambda_{adv1} = 0.04$ and $\lambda_{adv2} = 0.02$.

4.2. Image Outpainting Results

We found that previous outpainting methods are hard to be applied in the case of rectified shapes. For example, NSIO [37] does not support the mask-based outpainting. Boundless [31] and SRN [33] require the filling region to be a rectangle. These constraints limit the practical application of the outpainting technique, leading to inferior and seamed extrapolation results as shown in Fig. 5. By contrast, some methods [13, 40, 41, 38, 12, 44] in the inpainting region can complete the image with irregular masks flexibly. Thus, for a fair comparison, we mainly compare our approach with these inpainting methods.

Quantitative Evaluation: Following previous methods, we use three metrics to evaluate the comparison methods: the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM), and Fréchet Inception Distance (FID) that describes the perceptual quality of the generated results [11]. We compare our approach with the GL [13], DeepFill_v1 (DF_v1) [40], DeepFill_v2 (DF_v2) [41], RK [12], and HiFill [38]. All the methods are used to conduct the outpainting on the test dataset including 1,000 images. Then we compute three evaluation metrics using the difference between each extrapolated image and the ground truth image. For a comprehensive evaluation, we split the test dataset into three categories: easy, moderate, and hard, based on the area of the blank region in the rectified image. As listed in Table 1, our approach outperforms the comparison methods in all evaluation metrics.

Qualitative Evaluation: In Fig. 6, we show the results derived from compared methods and our approach. Concretely, the results generated by DeepFill_v1 [40], and DeepFill_v2 [41] suffer from poor extrapolated contents with unrelated objects and noises. Although more visually pleasing results are produced by RK [12], HiFill [38], and ProFill [44], the fracture of semantics occurs especially in the boundary, as the spatial discounted correlation increases. Instead, our approach achieves the best extrapolation performance regarding the visual appearance and semantic consistency, demonstrating that the designed dual conditional expansion and curve-aware correlation measurement meet the special painting paradigm in ROP.

4.3. Distortion Rectification Results

Although our main aim is to extrapolate the rectified image, we also evaluate our rectification algorithm with the state-of-the-art methods. From the Fig. 7, we can observe that compared with Li [16] and DeepCalib [5], our rectification results yield significant improvements on both PSNR and SSIM, since our approach guides the learning model to build a general geometry perception on the distorted image. Instead, such prior knowledge is ignored in most previous distortion rectification methods. In addition, we achieve a competitive performance with Liao [18], while this method is more complicated and it requires three types of networks to conduct the rectification. In terms of the visual appearance, our method performs more robust than

Table 1. Quantitative evaluation of the extrapolation results obtained by different methods. **Red** text indicates the best and blue text indicates the second-best performing method.

	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓
Comparison Methods	*Easy*			*Moderate*			*Hard*		
GL (ToG'17) [13]	13.52	0.37	62.15	13.06	0.35	88.41	11.63	0.26	100.78
DeepFill_v1 (CVPR'18) [40]	16.93	0.59	49.33	16.21	0.52	80.92	14.03	0.35	90.29
DeepFill_v2 (ICCV'19) [41]	18.39	0.64	41.93	17.32	0.62	71.41	15.02	0.41	82.16
RK (ECCV'20) [12]	20.43	0.68	37.85	17.51	0.64	62.22	16.13	0.45	67.31
HiFill (CVPR'20) [38]	20.82	0.68	32.59	18.52	0.58	64.78	17.50	0.47	65.92
Ours	**23.40**	**0.74**	**21.48**	**22.08**	**0.71**	**51.88**	**19.41**	**0.61**	**53.19**

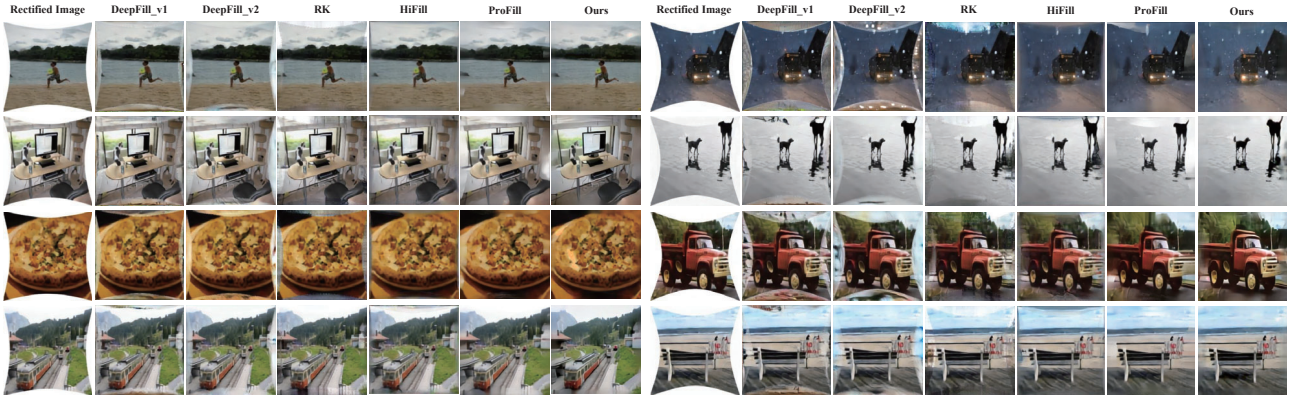

Figure 6. Qualitative comparison on the test dataset. For each case, we show the rectified image, and the extrapolated results derived by DeepFill_v1 [40], DeepFill_v2 [41], RK [12], HiFill [38], ProFill [44] and our approach, from left to right.

traditional feature-based method [1] in real-world setting, in which the distorted images are captured by the popular wide-angle lens such as Opteka 6.5mm Lens, iZugar MKX22 Lens, and GoPro.

4.4. Why Curvature Guidance Is Necessary

In our method, the curvature of the boundary in the rectified image plays a vital role. For distortion rectification, besides the detailed supervision provided by the distortion parameters, the general supervision introduced by the curvature teaches neural networks to learn the global structure of the rectified image. In addition, the curvature of a boundary in the rectified image also implies the area of the filling region, which can guide the critic network to judge the different extrapolation results. For deep analysis, we explore the importance of curvature guidance in ROP as follows.

In Fig. 8, we show the performance improvements supported by the curvature guidance in the distortion rectification module (top) and outpainting module (bottom). Two prevalent backbone networks, VGG16 and ResNet50 are employed respectively in the distortion rectification module, to eliminate the impact of network architecture on performance. Obviously, the error of distortion estimation gains significant decreases with the curvature guidance in both two backbone cases, while the module performs poorly and tends to the model overfitting without the curvature

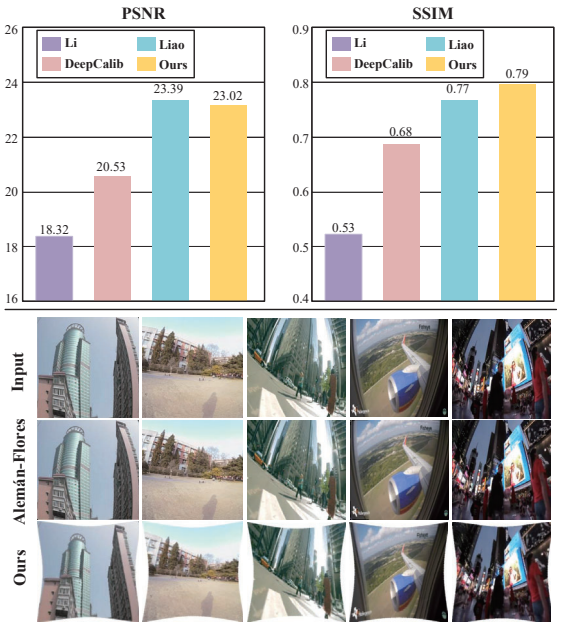

Figure 7. Distortion rectification results. We show the quantitative evaluation with learning-based methods (top) and visual comparison with traditional feature-based methods (bottom).

guidance. For the outpainting module, the learning process is boosted by the curvature guidance and achieves smaller content extrapolation errors. We can conclude that the cur-

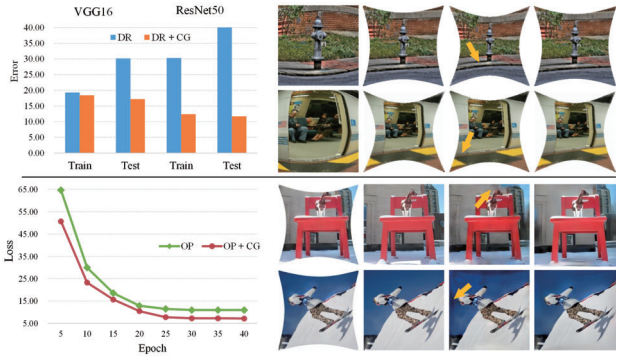

Figure 8. Improvements provided by the curvature guidance (CG) in distortion rectification module (top) and outpainting module (bottom). For visual comparisons, we show the input, ground truth, generated results, and generated results with curvature guidance. The arrows highlight the unsatisfactory parts.

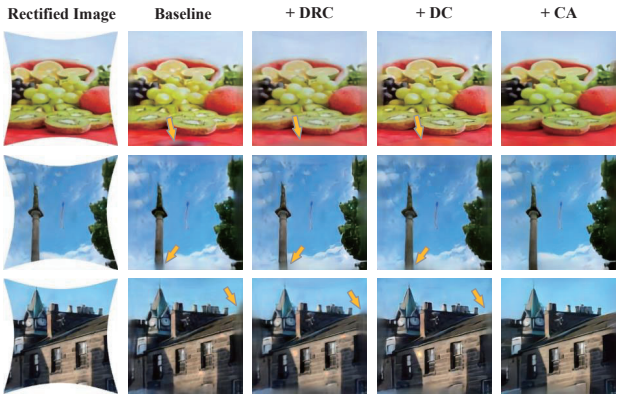

Figure 9. Visual ablation comparison on our outpainting approach. The arrows highlight the artifacts.

vature guidance bridges the gap between the implicit learning target and explicit visual representation in ROP.

4.5. Ablation Study

To validate the effectiveness of different components in our approach, we conduct an ablation study as follows. First, we implement the learning model as a baseline without the dilated residual connection (DRC), dual conditional expansion mechanism (DC), and curve-aware outpainting strategy (CA). Then, we gradually add these components to show different extrapolation performance. The complete approach with the above components achieves the best visual appearance as depicted in Fig. 9. Using the dilated residual connection, the burdens of the decoder network can be relieved since the introduced feature maps from the encoder network are pre-painted under multiple receptive fields. With the dual conditional expansion, the model can generate smoother contents and details. Furthermore, the curve-aware correlation measurement enables our model to harmonize color and semantics consistency, constructing a realistic and seamless scene.

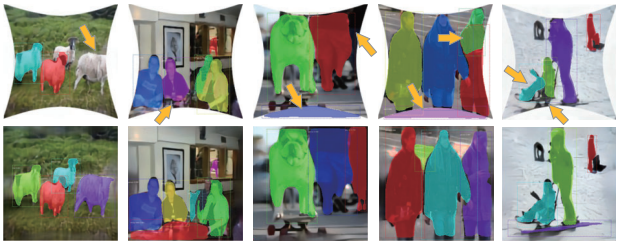

Figure 10. Object detection and semantic segmentation results of the rectified images (top) and the extrapolated images (bottom). The arrows highlight the missing or wrong detection parts.

4.6. Benefits for Scene Reasoning

The proposed ROP method constructs a nearly perfect representation for rectification results and extrapolates the original scene with wider FoV. Thus, our method can help downstream vision tasks such as object detection and semantic segmentation, which is crucial to scene analysis and motion prediction. As shown in Fig. 10, the detection and segmentation results are derived from Mask R-CNN [10]. We can notice some missing and wrong detection parts in the rectified image (top), especially in the boundary, since the outer objects cannot be captured completely within a fixed FoV. Moreover, the curve boundary of the rectified image would mislead the scene understanding. For example, the blank region surrounded by the curve boundary is recognized as a bed at the third and fourth columns. By contrast, our approach can exhibit the complete scene in a regular construction and create semantically coherent objects beyond the original FoV. Consequently, the ROP boosts the more accurate performance of scene perception, improving the integrity of the object in image boundary.

5. Conclusion

In this paper, we rethink the traditional representation of distortion rectification and present a rectification outpainting (ROP) method. Given a distorted image, ROP aims to recover the geometric distribution of the realistic scene and construct the results with complete content and regular shape. Moreover, we can obtain a wider field-of-view scene beyond the original wide-angle lens, which extends an appealing and promising application for video surveillance and autonomous driving. To address the specific challenges, we propose a parametric framework with a general geometry supervision and a dual conditional expansion strategy. Considering the spatially discounted correlation, a curve-aware correlation measurement is designed to enforce the local consistency of extrapolated details. In future work, we plan to explore the outpainting scheme based on the rectangle rectification construction.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (No.62172032, No.61772066).

References

[1] Miguel Alemánflores, Luis Alvarez, Luis Gomez, and Daniel Santanacedrés. Automatic lens distortion correction using one-parameter division models. *Image Processing on Line*, 4, 2014. 1, 2, 7

[2] Coloma Ballester, Marcelo Bertalmio, Vicent Caselles, Guillermo Sapiro, and Joan Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE transactions on image processing*, 10(8):1200–1211, 2001. 2

[3] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28(3):24, 2009. 2

[4] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, 2000. 2

[5] 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, 6

[6] Christian Bräuer-Burchardt and Klaus Voss. A new algorithm to correct fish-eye and strong wide-angle lens distortion from single images. volume 1, pages 225–228, 2001. 1, 2

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 3, 5

[8] Christopher Geyer and Kostas Daniilidis. Paracatadioptric camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:687–695, 2002. 1, 2

[9] Dongsheng Guo, Hongzhi Liu, Haoru Zhao, Yunhao Cheng, Qingwei Song, Zhaorui Gu, Haiyong Zheng, and Bing Zheng. Spiral generative network for image extrapolation. In *European Conference on Computer Vision*, 2020. 1, 2

[10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017. 8

[11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017. 6

[12] Liu Hongyu, Jiang Bin, Song Yibing, Huang Wei, and Yang Chao. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. In *Proceedings of the European Conference on Computer Vision*, 2020. 1, 2, 6, 7

[13] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017. 1, 6, 7

[14] Sing Bing Kang. Catadioptric self-calibration. In *IEEE International Conference on Computer Vision*, 2000. 1, 2

[15] Juho Kannala and Sami 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. 3, 6

[16] Xiaoyu Li, Bo Zhang, Pedro V Sander, and Jing Liao. Blind geometric distortion correction on images through deep learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4855–4864, 2019. 1, 2, 6

[17] K. Liao, C. Lin, Y. Zhao, and M. 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, 2, 6

[18] Kang Liao, Chunyu Lin, Yao Zhao, and Mai Xu. Model-free distortion rectification framework bridged by distortion distribution map. *IEEE Transactions on Image Processing*, 29:3707–3718, 2020. 1, 2, 4, 6

[19] 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, 2014. 6

[20] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. In *Advances in Neural Information Processing Systems*, pages 9605–9616, 2018. 4

[21] Takeru Miyato and Masanori Koyama. cGANs with projection discriminator. In *International Conference on Learning Representations*, 2018. 4

[22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015. 5

[23] 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, 2016. 1, 2

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015. 4

[25] Mark Sabini and Gili Rusak. Painting outside the box: Image outpainting with gans. *arXiv preprint arXiv:1808.08483*, 2018. 1, 2

[26] YiChang Shih, Wei-Sheng Lai, and Chia-Kai Liang. Distortion-free wide-angle portraits on camera phones. *ACM Transactions on Graphics (TOG)*, 38:1 – 12, 2019. 1, 2

[27] Denis Simakov, Yaron Caspi, Eli Shechtman, and Michal Irani. Summarizing visual data using bidirectional similarity. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2

[28] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5

[29] Rahul Swaminathan and Shree K. Nayar. Nonmetric calibration of wide-angle lenses and polycameras. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:1172–1178, 2000. 1, 2

[30] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016. 5

[31] Piotr Teterwak, Aaron Sarna, Dilip Krishnan, Aaron Maschinot, David Belanger, Ce Liu, and William T Freeman. Boundless: Generative adversarial networks for image extension. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10521–10530, 2019. 1, 2, 4, 6

[32] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6924–6932, 2017. 6

[33] Yi Wang, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Wide-context semantic image extrapolation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1399–1408, 2019. 1, 2, 6

[34] Yaxiong Wang, Yunchao Wei, Xueming Qian, Li Zhu, and Yi Yang. Sketch-guided scenery image outpainting. *IEEE Transactions on Image Processing*, 30:2643–2655, 2021. 1

[35] Zhucun Xue, Nan Xue, Gui-Song Xia, and Weiming Shen. Learning to calibrate straight lines for fisheye image rectification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1643–1651, 2019. 1, 2

[36] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1, 2

[37] Zongxin Yang, Jian Dong, Ping Liu, Yi Yang, and Shuicheng Yan. Very long natural scenery image prediction by outpainting. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10561–10570, 2019. 1, 6

[38] Zili Yi, Qiang Tang, Shekoofeh Azizi, Daesik Jang, and Zhan Xu. Contextual residual aggregation for ultra high-resolution image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7508–7517, 2020. 1, 2, 6, 7

[39] 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 *European Conference on Computer Vision*, pages 469–484, 2018. 1, 2, 6

[40] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018. 1, 2, 6, 7

[41] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4471–4480, 2019. 1, 6, 7

[42] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4471–4480, 2019. 1, 2

[43] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Learning pyramid-context encoder network for high-quality image inpainting. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1486–1494, 2019. 1, 2

[44] Yu Zeng, Zhe Lin, Jimei Yang, Jianming Zhang, Eli Shechtman, and Huchuan Lu. High-resolution image inpainting with iterative confidence feedback and guided upsampling. In *European Conference on Computer Vision*, pages 1–17, 2020. 1, 2, 6, 7