# Innovating Real Fisheye Image Correction with Dual Diffusion Architecture

Shangrong Yang, Chunyu Lin*, Kang Liao, Yao Zhao
Institute of Information Science, Beijing Jiaotong University
Beijing Key Laboratory of Advanced Information Science and Network, Beijing, 100044, China
{sr_yang, cylin, kang_liao, yzhao}@bjtu.edu.cn

## 1. Supplemental Material

### 1.1. Overview

To fully illustrate our experiments, we offer additional experimental details and results. The following is a list of our supplemental materials:

- Detailed algorithms of training and testing strategies in dual diffusion architecture (DDA) (Section 1.2).
- More subjective comparison between our proposed method and the state-of-the-art methods on synthetic fisheye images correction (Section 1.3).
- Evaluating the quality of detail in synthetic fisheye correction results through a representative subset of comparison methods (Section 1.4).
- Additional real fisheye image correction results and more comprehensive analysis of detail comparisons. (Section 1.5).
- Testing the performance of different correction results on downstream tasks. (Section 1.6).

### 1.2. Detailed Algorithms of DDA

To provide a detailed explanation of our dual diffusion architecture (DDA), we have included two corresponding algorithm tables (Algorithm 1 and Algorithm 2). The training strategy details are outlined in Algorithm 1. We utilize the one-pass network (OPN) to predict distortion flows $W$ in both real and synthetic fisheye images, and perform pre-correction ($S_p$ and $R_p$) to generate new conditions. Subsequently, we randomly sample a noise ground truth $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and add it to both the synthetic image ground truth $S_{gt}$ and the real fisheye image $R_f$ to generate corresponding noisy data ($\bar{S}_{gt}$ and $\bar{R}_f$). Our conditional diffusion module (CDM) takes $\bar{S}_{gt}$ as input, and uses the new condition $y_n$ (concatenated by $S_p$ and $R_p$) to predict the noise. The UDM directly takes $\bar{R}_f$ as input to predict the noise. Finally, by supervising these two noises, our DDA is able to optimize and train.

**Algorithm 1** Training dual diffusion architecture

Requirement: $C_\theta$, $U_\theta$, $OPN_\theta$ (conditional and unconditional diffusion module, one-pass network), $S_f$ & $S_{gt}$: paired synthetic images, $R_f$: real fisheye image
1: repeat
2: $R_p \leftarrow OPN_\theta(R_f)$, $S_p \leftarrow OPN_\theta(S_f)$
3: $t \sim Uniform\{1, ..., T\}$
4: $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Perform Gradient descent according
$$\bigtriangledown_\theta ||2\epsilon - C_\theta(\sqrt{\bar{\alpha}_t}S_{gt} + \sqrt{1-\bar{\alpha}_t}\epsilon, \bar{\alpha}_t, y_n) - U_\theta(\sqrt{\bar{\alpha}_t}R_f + \sqrt{1-\bar{\alpha}_t}\epsilon, \sqrt{\bar{\alpha}_t}S_{gt} + \sqrt{1-\bar{\alpha}_t}\epsilon, \bar{\alpha}_t)||_1$$
6: until converged

**Algorithm 2** Correct the synthetic and real fisheye image

1: $R_p \leftarrow OPN_\theta(R_f)$, $S_p \leftarrow OPN_\theta(S_f)$ ▷ One-pass correction scheme
2: $R_{c(T)}, S_{c(T)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
3: for $t = T, ..., 1$, do
4: $t \sim Uniform\{1, ..., T\}$
5: $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $z = 0$
6: $R_{c(t-1)} = \frac{1}{\sqrt{\alpha_t}}(R_{c(t)} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}C_\theta(\tilde{c}_t, \bar{\alpha}_t, y_n))$
$S_{c(t-1)} = \frac{1}{\sqrt{\alpha_t}}(S_{c(t)} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}C_\theta(\tilde{c}_t, \bar{\alpha}_t, y_n))$ ▷ Inference correction scheme
7: end for
8: return $R_p$, $S_p$, $R_{c(0)}$, $S_{c(0)}$

The testing strategy is illustrated in Algorithm 2. Benefiting from our DDA, we provide two optional test methods (one-pass correction and inference correction). For one-pass correction, we use the OPN obtained from our trained DDA to directly predict the distortion flow $W$ of the fisheye image (real or synthetic). We then use $W$ to warp the fisheye image and quickly obtain the correction result.

For inference correction, we optimize the network using

*Corresponding author: cylin@bjtu.edu.cn

only OPN and CDM to predict the noise. We also use the OPN to predict the distortion flow $W$ and pre-correct the synthetic or real fish images. Then we randomly sample a noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ as initial image. The initial image and the pre-corrected result are fed into the CDM to predict the noise for denoising. By repeatedly predicting the noise and recalculating new images, we can obtain high-quality results.

### 1.3. More Subjective Comparison

For a more comprehensive comparison, we have included additional subjective comparison results of synthetic fisheye image correction in Figure 1. As can be observed, the images corrected using Blind [1] and DCCNN [2] still exhibit some distortion. Blind selectively only corrects the center area during training, and DCCNN has a restricted quantization interval, hence their correction on large distortion images is not effective. DeepCalib [3] performs better than Blind and DCCNN for a more reasonable spherical fisheye model. However, this new model causes some over-rectification and boundary loss. The rectification results of DDM [4] and PCN [5] have no boundary loss. Although they significantly improve correction quality through a multi-stage correction scheme, some artifacts appear in the results. In contrast, our dual diffusion architecture combines the advantages of DDPMs and GANs. It isolates structure correction from content reconstruction, which improves the quality of corrected images. Our results, particularly the inference result, show more precise structures and greater sharpness with no artifacts. It is worth mentioning that our method offers two correction schemes, both of which outperform the comparison method.

### 1.4. Evaluating on Representative Subset

Due to the limited resolution of the images, it can be challenging to distinguish the texture differences between the correction results of some effective methods, such as DeepCalib, DDM, and PCN. Therefore, we have enlarged the local area of correction results in Figure 2 to enable easier discernment of the texture details. As can be observed, our rectification results exhibit clearer texture when compared with the mainstream methods. In particular, the texture of our inference results is superior to other approaches. Additionally, we have used the Sobel operator [7] to detect correction result edges in Figure 3. The edge of DeepCalib has obvious bending, while DDM and PCN exhibit some edge confusion. In contrast, our edge is clean, demonstrating that our results are more precise in both structure and texture than the comparison methods.

### 1.5. More Results on Real Fisheye Correction

We have included more testing results on real fisheye images in Figure 4-6 to demonstrate the practicability of our dual diffusion architecture. Since the distortion of real fisheye images is greater than that of synthetic images, Blind and DCCNN, which achieve incomplete correction on synthetic images, are also incomplete for real fisheye images. DDM and PCN can achieve better structure correction, but the created artifacts are more noticeable since they were unable to learn the distribution of real fisheye images during training. The results of DeepCalib have no artifacts, but it still exhibits over-rectification and boundary loss in real fisheye correction. In contrast, our method achieves complete correction, with no artifacts, over-rectification, or boundary loss in our real correction results. In particular, due to multiple iterative recalculations, our inference results are far more visually appealing than comparison results.

### 1.6. Comparison on Downstream Tasks

To evaluate the efficacy of our correction method on semantic segmentation, we leveraged a mainstream method [6] to perform semantic segmentation on the correction results of different methods. The result of semantic segmentation is shown in Figure 7. It is evident that the semantic segmentation algorithm's efficacy has increased with distortion reduction. However, the comparison approaches cannot provide appropriate semantic segmentation due to the remaining distortion and artifacts. In contrast, our method successfully removes distortion and artifacts, resulting in an ideal segmentation effect. As there are no labels for real fisheye images, we computed quantitative results on synthetic images. We used the semantic segmentation results of the ground truth as labels to calculate the pixel accuracy (Pixel Acc.) and mean intersection-over-union (mIoU) of each correction method, as shown in Table 1. Compared with other methods, our results obtain the best performance, further demonstrating that our corrected images have superior quality.

Table 1. Performance comparision on semantic segmentation

| Comparison | | Metric | |
|---|---|---|---|
| Methods | Type | mIoU | Pixel Acc. |
| Blind | Regression | 0.1666 | 68.46% |
| DCCNN | Regression | 0.1963 | 73.87% |
| DeepCalib | Regression | 0.3135 | 81.91% |
| DDM | Generation | 0.3557 | 84.62% |
| PCN | Generation | 0.3781 | 89.30% |
| Ours (one-pass) | Generation | 0.3575 | 85.93% |
| Ours (inference) | Generation | **0.4519** | **92.13%** |

## References

[1] X. Li, B. Zhang, Pedro V. Sander, and J. Liao. Blind geometric distortion correction on images through deep learning. In

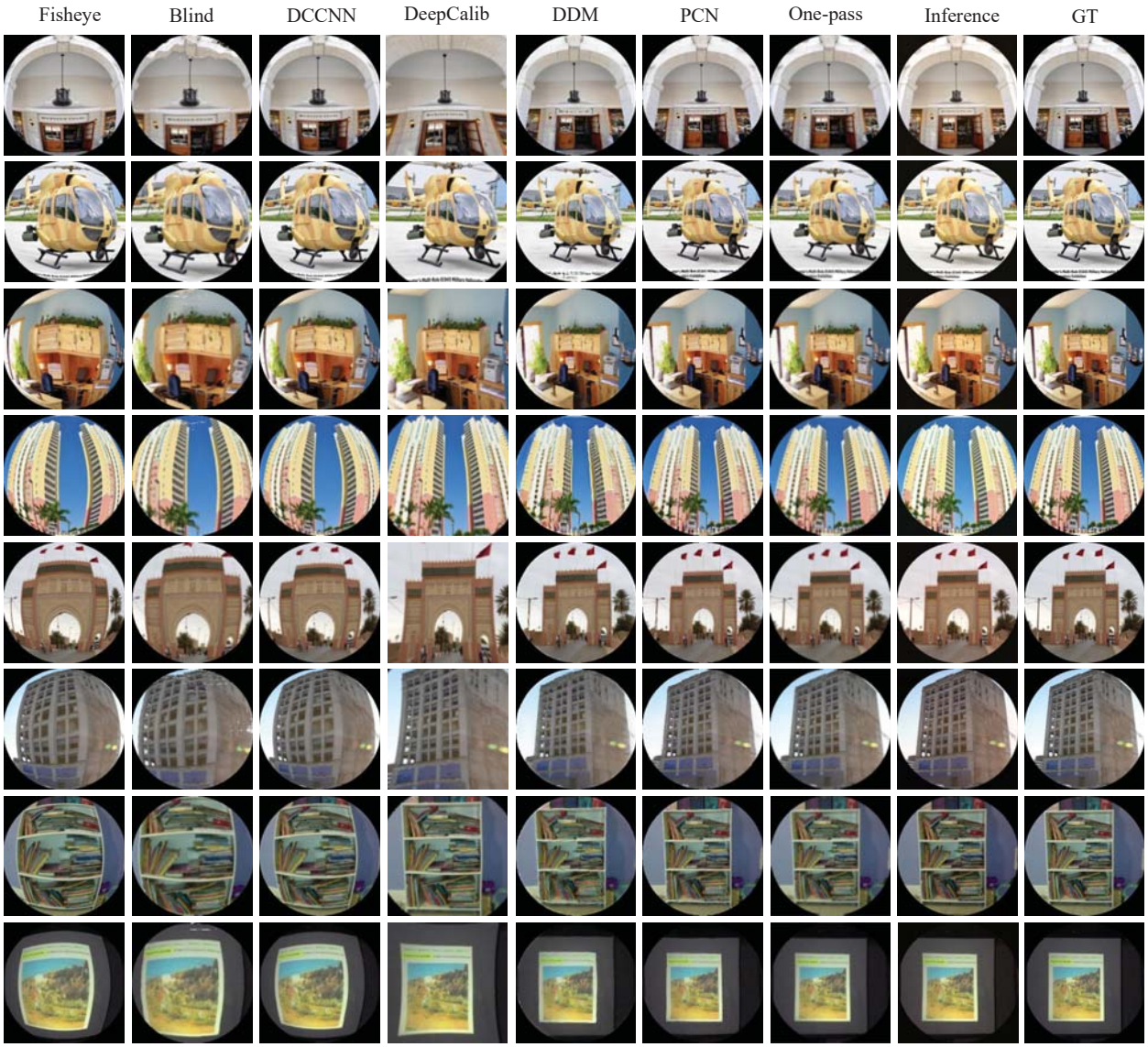


Figure 1. We conducted full comparisons between our methods and the state-of-the-art methods for synthetic image rectification. The state-of-the-art methods include Blind [1], DCCNN [2], DeepCalib [3], DDM [4], and PCN [5].

*CVPR*, pages 4855–4864, 2019. 2, 3, 6

[2] J. Rong, S. Huang, Z. Shang, and X. Ying. Radial lens distortion correction using convolutional neural networks trained with synthesized images. In *ACCV*, 2016. 2, 3, 6

[3] O. Bogdan, V. Eckstein, F. Rameau, and J. Bazin. Deepcalib: a deep learning approach for automatic intrinsic calibration of wide field-of-view cameras. In *CVMP*, 2018. 2, 3, 4, 5, 6

[4] K. Liao, C. Lin, Y. Zhao, and M. Xu. Model-free distortion rectification framework bridged by distortion distribution map. *IEEE Transactions on Image Processing*, 29:3707–3718, 2020. 2, 3, 4, 6

[5] S. Yang, C. Lin, K. Liao, C. Zhang, and Y. Zhao. Progressively complementary network for fisheye image rectification using appearance flow. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6344–6353, 2021. 2, 3, 4, 5, 6

[6] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, 127:302–321, 2018. 2, 9

[7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification and scene analysis*, volume 3. Wiley New York, 1973. 2

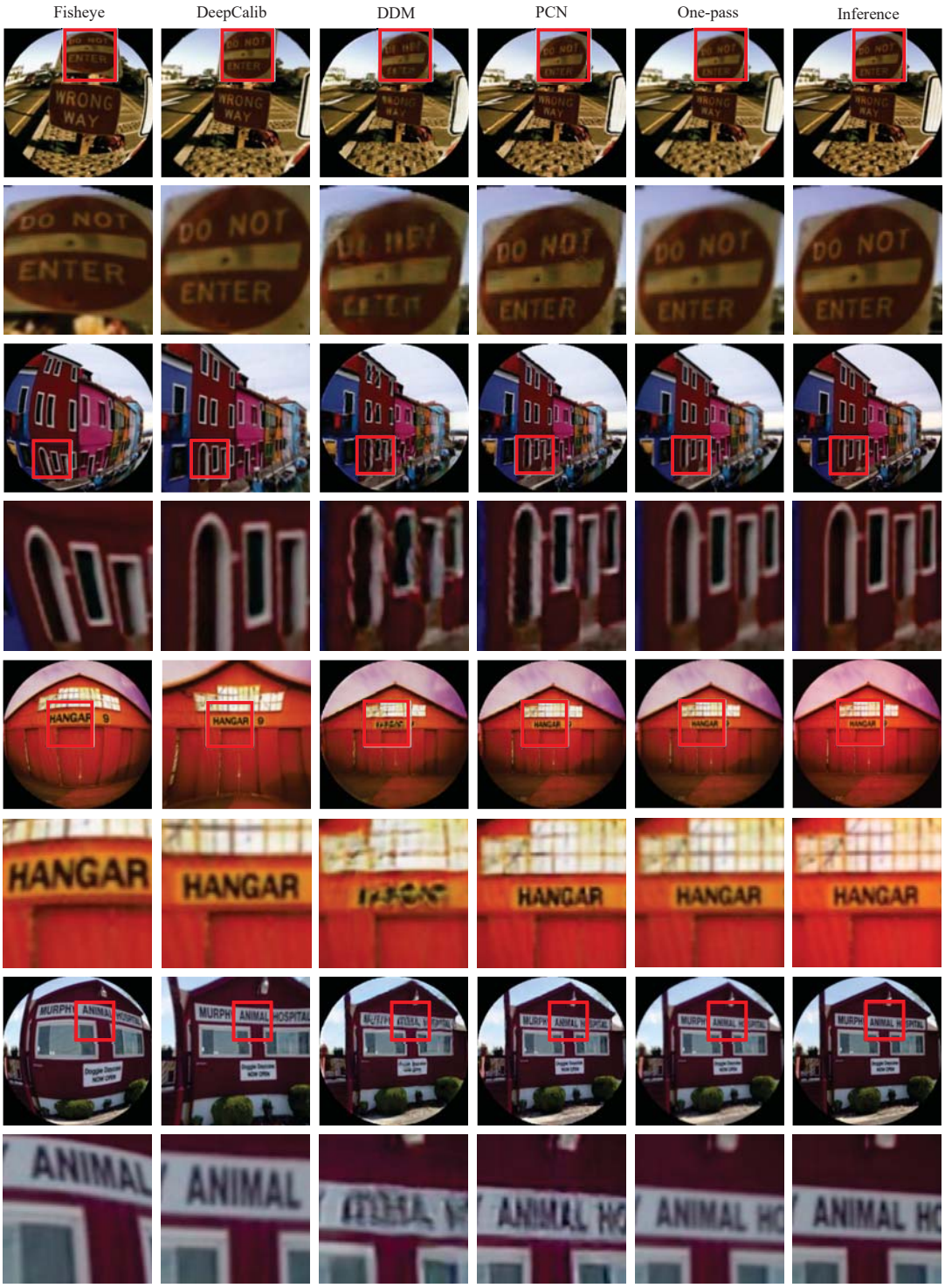


Figure 2. Evaluating on a representative subset of comparison methods, which include DeepCalib [3], DDM [4], and PCN [5]. We enlarge the local region (marked by red boxes) to compare the image texture.

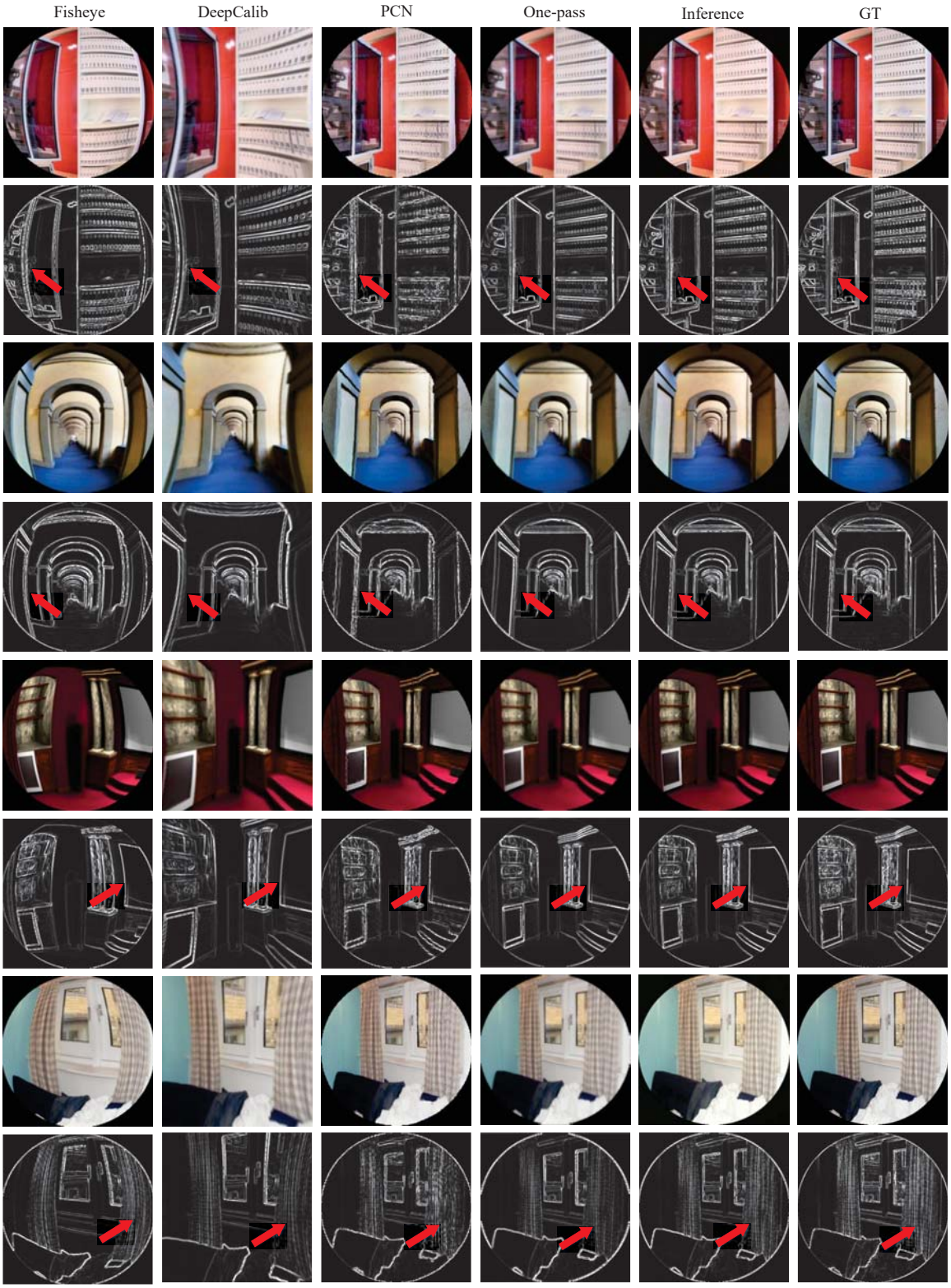


Figure 3. Evaluating on a representative subset of comparison methods, which include DeepCalib [3] and PCN [5]. We highlight the structural differences (marked by red arrows).

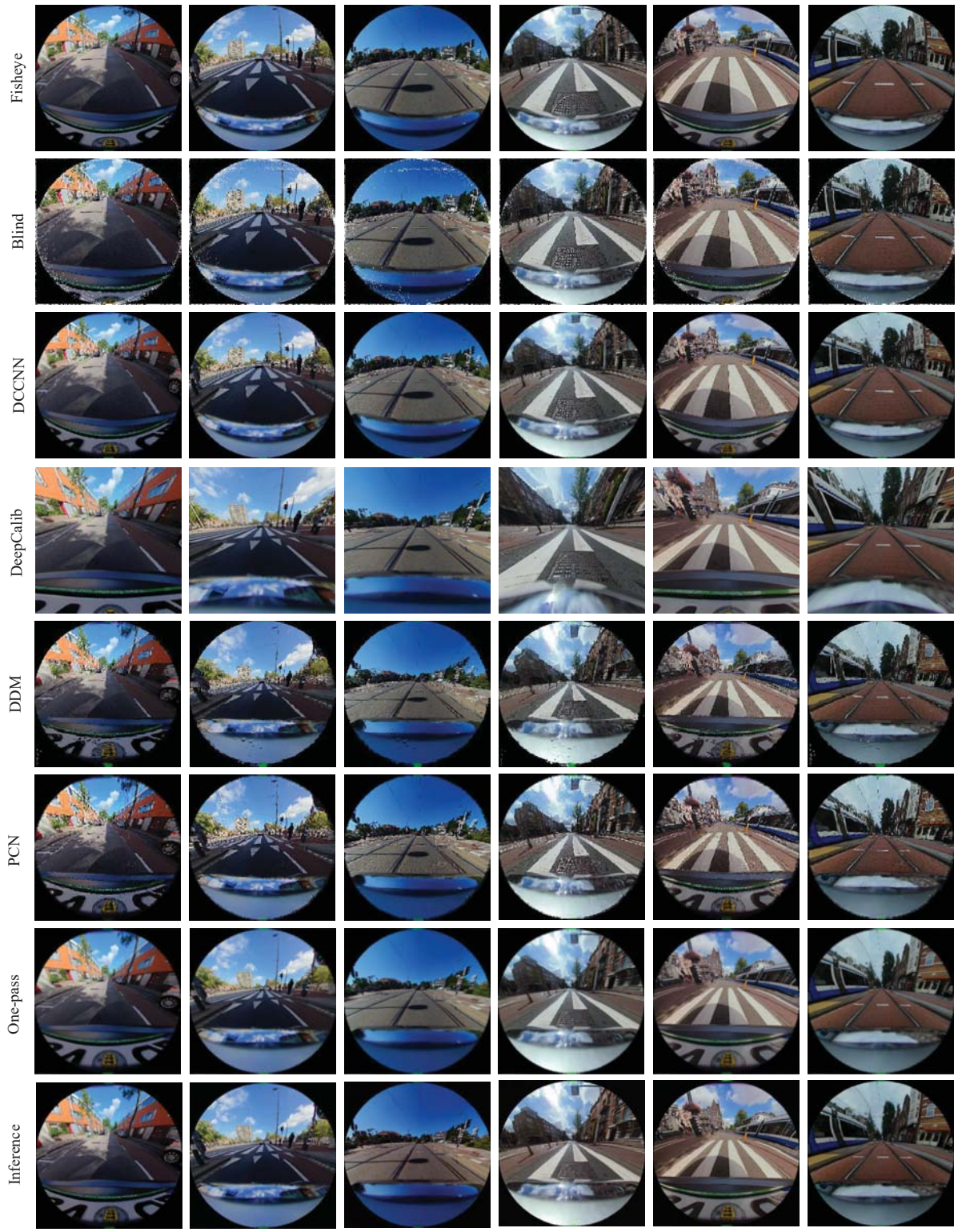


Figure 4. Full comparisons of our methods with the state-of-the-art methods in real fisheye rectification. The state-of-the-art methods include Blind [1], DCCNN [2], DeepCalib [3], DDM [4], and PCN [5]. Our results exhibit precise structure and clearer texture.

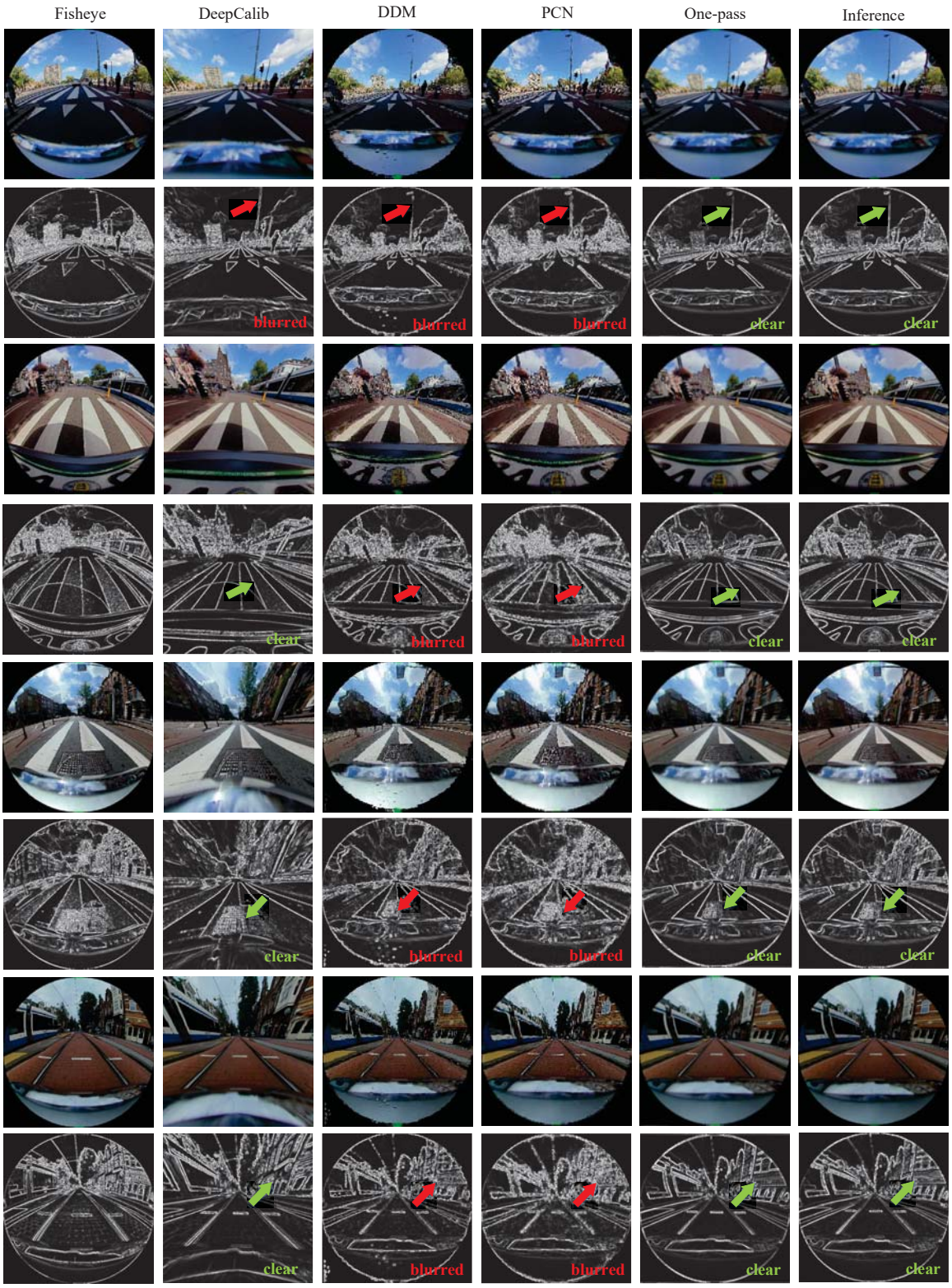


Figure 5. Edge detection on correction results. We utilize arrows to indicate the areas that require attention (green arrows represent clear structure and red arrows represent blurred structure). Our rectification results demonstrate a clearer structure.

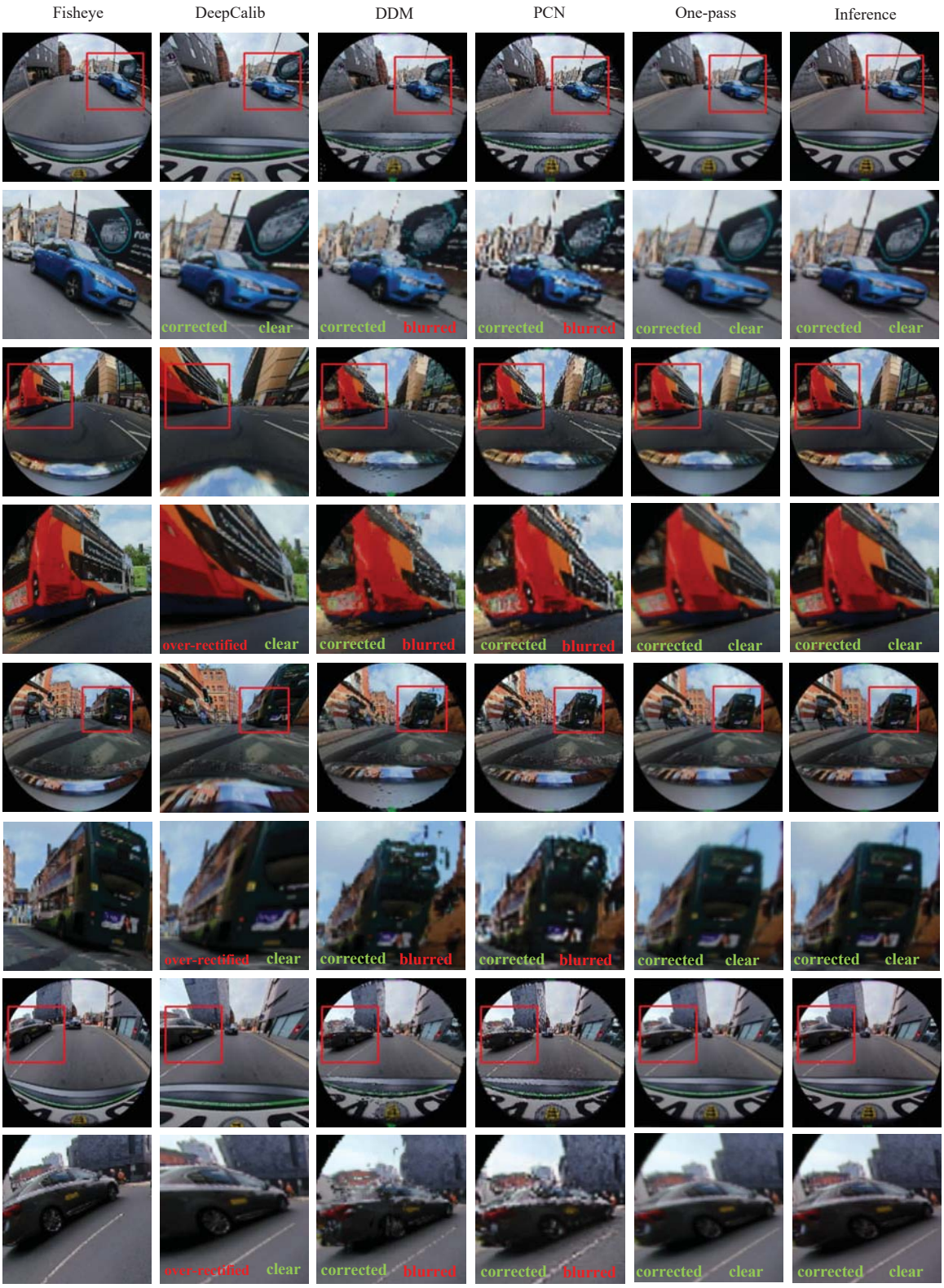


Figure 6. To provide a more intuitive comparison, we enlarged the local area marked by red boxes. Our approaches, which ensure the integrity of the boundaries, result in precise corrected structures and clear content for correcting real fisheye images.

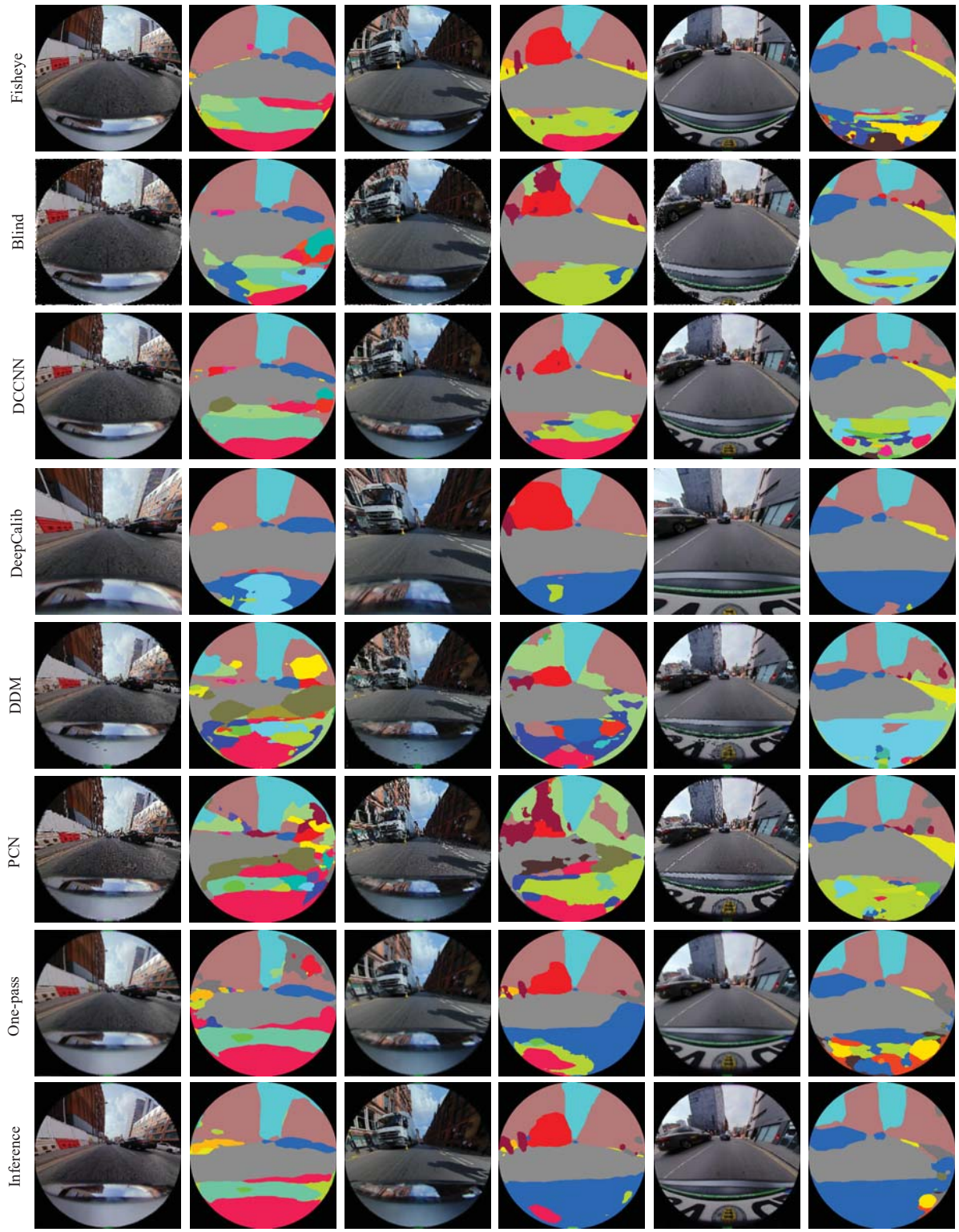


Figure 7. Semantic segmentation results on corrected images. Our corrected images demonstrate improved performance in semantic segmentation. Specifically, the semantic segmentation method [6] achieves more accurate segmentation of each object, providing further evidence of the higher quality of our correction results.