IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 November 2022, accepted 6 December 2022, date of publication 12 December 2022,
date of current version 16 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228297

RESEARCH ARTICLE

Global Convolutional Neural Networks With Self-Attention for Fisheye Image Rectification

BYUNGHYUN KIM[1], DOHYUN LEE[1], KYEONGYUK MIN[2], (Member, IEEE), JONGWHA CHONG[3], AND INWHEE JOE[1]

[1]Department of Computer Science, Hanyang University, Seoul 04763, South Korea
[2]Department of Electronics Engineering, Hanyang University, Seoul 04763, South Korea
[3]Department of Computer Science, State University of New York Korea, Incheon 21985, South Korea

Corresponding author: Inwhee Joe (iwjoe@hanyang.ac.kr)

ABSTRACT Fisheye images are attracting attention in computer vision such as autonomous vehicles and virtual reality because of the wide field of view (WFoV). However, fisheye images have geometric distortions caused by the refractive index of the lens. Conventional fisheye rectification methods require multiple images to calculate distortion coefficients and lens intrinsic parameters. This means that if the fisheye lens is changed, the same operation will have to be repeated. On the other hand, by using deep learning, images with different distortion coefficients can be rectified. Also, with end-to-end learning, no feature engineering is required. To improve the performance of fisheye image rectification, we propose global convolutional neural networks with self-attention to rectify the fisheye images. The proposed method employs dilated convolutional neural networks (D-CNNs) to enlarge receptive fields, and self-attention to extract the most important features of input images. In this way, the proposed method can extract global features from input images. To better train and evaluate the proposed method, we generate fisheye images from the Place2 dataset with Cartesian and polar coordinates, and label them with original images (ground-truth). we also schedule the learning rate with cosine annealing and use an integrated loss function. The experimental results show that the proposed method achieves an excellent performance in both qualitative and quantitative evaluations.

INDEX TERMS Fisheye image rectification, dilated convolution, self-attention, deep learning.

I. INTRODUCTION

Fisheye images are used in many computer vision tasks [1], [2], [3], [4], [5], [6], [7], [8] because it captures the wide field of view (WFoV). However, the images taken by fisheye lens suffer from sever geometric distortion at the same time. Therefore, fisheye image rectification is absolutely necessary. There are two methods of fisheye image rectification. The first is a pattern based method and the second is a deep learning based method. Most previous fisheye image rectification methods include pattern based methods such as [9], [10], and [11]. In Zhang et al. [9], fisheye images are rectified by planar patterns with a closed-form solution, followed by a non-linear refinement based on the maximum likelihood criterion. Shah et al. [10] employs a non-linear transformation between points in the world coordinate system and their corresponding location on the image plane. The other type of fisheye image rectification methods are deep learning based such as [12], [13], [14], [15], [16], and [17]. They use many distorted images to train rectifica-

FIGURE 1. The proposed method provides the result of rectified images (Bottom row) with different distortion coefficients (Top row).

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Sharif.

tion networks. Networks learn distortion patterns from a large number of images. Generally, deep learning based methods are better than pattern based methods. Most of the deep learning based methods employ convolutional neural networks (CNNs). Networks using CNNs have a lot of parameters, features of each region are reduced through pooling layers. However, when pooling operation is performed, the existing features are lost. In addition, these features are computed over the local neighborhood. It means that repetitive convolutional operations are required to extract overall features of input images.

A mathematical model for fisheye image rectification requires specialized knowledge and complex calculations [17]. Also, since it only calculates the distortion rate for a specific pattern of the fisheye image, it is not possible to rectify the fisheye image of various patterns. On the other hand, the deep learning model can rectify various patterns of fisheye images, and the rectified image can be obtained only by inputting fisheye images without going through complicated calculations.

In this paper, we propose global convolutional neural networks with self-attention to rectify fisheye images. In Fig. 1, the output of our networks is rectified images. The proposed method achieves an end-to-end process from fisheye images to rectified images.

We employ dilated convolutional neural networks (D-CNNs) [18] to enlarge the receptive field of filters. Previously, in order to enlarge the receptive field, features were reduced through pooling layers and the convolutional operation was repeated. In image reconstruction, there is an auto-encoder consist of an encoder that extracts features through convolutional operations and a decoder that reconstructs images based on features through de-convolutional operations. On the other hand, dilated convolution operation, which can enlarge the receptive field by adding zero padding to the filter, shows good performance [19] without using auto-encoder. Simultaneously, we employ self-attention to find which features of the input image should be more attentive to. Our contributions can be summarized as follows:

- We propose a method for combining CNNs and D-CNNs and applying self-attention to all convolutional layer outputs to rectify fisheye images. And we simplified the model by adjusting the number of parameters in the proposed model.
- We generate fisheye images from Place2 dataset [20]. In the image generation process, the distortion coefficients are set randomly to generate fisheye images with different distortions. These images we generated are used as input to train our networks.
- We schedule the learning rate of the proposed model with the cosine annealing technique, and define and use an integration loss function for efficient fisheye image rectification.

The rest of the paper is organized as follow: Section II introduces previous related works. The generation of fisheye images from original images is detailed in Section III-A. The architecture of proposed method is described in Section III-B. In Section III-C, an integrated loss function applied to our networks is proposed. Finally, Section IV introduces the experimental detail and the qualitative and quantitative evaluations.

II. RELATED WORK

The distortion rectification of the initial fisheye image is an operation of converting the fisheye image into a corrected image through a rectification model to obtain intrinsic parameters of the fisheye lens. Several methods have addressed radial distortion in fisheye lens. These methods obtain the parameters of the camera lens and rectify the fisheye image as a model for radial distortion rectification. Kang [21] predicted parameters using radial distortion made up of parallel straight lines in space through a single image. It uses the minimum vanishing point dispersion constraint to estimate both radial and decentering lens distortion. But it has the disadvantage of locally correcting only the distortion in a specified area. Zhang et al. [9] proposed a method of correcting a single image through a model that calculates parameters by inputting multiple images. This method is widely used for 3D camera calibration. However, there is a disadvantage that images taken from various angles are already required for rectification.

Another distortion rectification method is to use projection. Abidi et al. [22] is a survey approach, using a measuring plate to find out different quadratic surfaces on the x and y axes of the Cartesian coordinates, and then correct the image by projecting onto a plane. However, this method has limitations because it is an approximation rather than finding an ideal coordinate. Melo [23] and Barreto et al. [24] introduced a method of detecting straight lines in an image and using them to correct the image. But these methods are highly dependent on the accuracy of detecting straight lines.

Recently, many image rectification methods based on deep learning have been proposed. Methods using deep learning may have limitations of existing methods, such as requiring multiple images to calibrate a specific fisheye lens. But good performance can be achieved without complex calculations for image rectification. Rong [25] and Yin et al. [26] used convolutional neural networks to rectify image distortion. They predict the distortion coefficients and apply algorithms to rectify images. However, this method requires a separate operation to rectify the image using coefficients. Xue et al. [27] proposed an end-to-end model that predicts a corrected image by receiving a distorted image as an input. This method shows good performance by designing the model structure without image pre-processing. Yin et al. [26] construct a deep CNN model to extract image features and feed the obtained features to a scene parsing network and a distortion parameter estimation network. However, it is not clear whether these obtained scenes can play an important role in the fisheye image rectification. In Yang et al. [28], the predicted parameters are employed to correct strong

distortion that exists in the fisheye image and authors synthesize the corresponding distortion using the original distortion-free image. Their method is excellent, but the rectification for the edge is not perfect.

In this paper, we propose global convolutional neural networks with self-attention to rectify fisheye images. Additionally, we employ dilated convolutional neural networks to enlarge the receptive field of filters and self-attention to find which features of the input image should be more attentive to.

III. PROPOSED METHOD

In this section, we describe the detail of proposed model. The generation of fisheye images from original images is detailed in Section III-A. The architecture of the proposed method is described in Section III-B. In Section III-C, an integrated loss function applied to our networks is proposed.

A. BASICS OF IMAGE DISTORTION

The Cartesian coordinates consist of x and y axes. In contrast, the polar coordinates consist of the distance r away from the origin and the angle θ. Fisheye images can be generated through radial distortion. The origin of the image coordinates are moved to the center, and then by reflecting the actual pixel distance, world coordinates are obtained. In Fig. 2, we can obtain equations (1) and (2).

$$r^2 = x^2 + y^2 \tag{1}$$

$$\tan\theta = \frac{y}{x} \tag{2}$$

$$\begin{cases} x &= r\cos\theta \\ y &= r\sin\theta \end{cases} \tag{3}$$

The above equation (3) shows the relationship between Cartesian and polar coordinates.

$$r_d = r_u(1 + k_1 r_u^2 + k_2 r_u^4 + k_3 r_u^6) \tag{4}$$

where r_d is distortion radius, r_u is normal radius, and k_i is i-th polynomial coefficient.

The image can be radially distorted using equation (4) above. A distorted image can be obtained by converting the polar coordinates into the Cartesian coordinates through the distortion radius.

Usually k_3 has little effect on radial distortion, so we only use k_1 and k_2, and we distort the image by setting k_1 and k_2 to random values and k_3 to 0.

B. NETWORK ARCHITECTURE

The structure of our networks is shown in Fig. 5 and consists of the following three layers: base layer, dilated convolutional layer, and self-attention layer. Our networks are supervised, and it uses three layers below to extract global features and implements end-to-end fisheye image rectification from fisheye images to rectified images. D-CNNs use enlarged filters to expand the receptive field. In the convolution operation, the extended receptive field has less dimensional loss, so it is possible to extract global features. The self-attention extracts the most important features and residual mapping improves the training efficiency of networks. We conducted ablation studies on the proposed method to justify the combination of each layer of our networks affects the performance. The details of each layer are introduced in the following section.

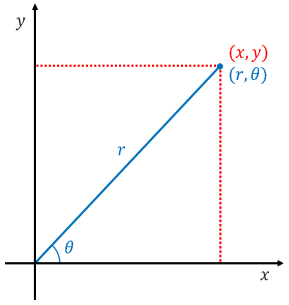

FIGURE 2. Relationship between Cartesian and polar coordinates.

1) BASE LAYER

The base layer is designed to extract features from the input for fisheye image rectification. Many deep learning studies suggest that CNNs trained with large amounts of data perform well in various computer vision tasks, such as image classification and object detection. So we employ CNNs for our base layer. Five base layers are used to extract features from the image. The number of filters for the convolutional layer is 16, 32, 64, 128, and 3. The output of the dilated convolutional layer is used as an input to the base layer. Also, self-attention is connected to the output of each convolutional layer to extract global features.

2) DILATED CONVOLUTIONAL LAYER

Dilated convolutional neural networks (D-CNNs) represent extension of CNNs using enlarged filters. A way to improve the performance of networks using CNNs is to enlarge the receptive field. However, in traditional CNNs, enlarging the receptive field increases the computational amount of the networks. So in general, the filter size is reduced and the depth of networks is increased by stacking layers. In contrast, D-CNNs use enlarged filter termed as dilated convolutional filter with dilation rate. The dilation rate is a spacing between the values in a filter. The dilated convolutional filter is shown in Fig. 3, where (a) is a traditional convolutional 3×3 filter, (b) shows that 3×3 dilated convolutional filter with a dilation rate of 2 in two dimensions will have the same field of view as a 5×5 filter. Similarly, (c) is a dilated convolutional filter when the dilation rate is 3. By adjusting the dilation rate, the receptive field can be enlarged with the same amount of computation as a 3×3 convolutional filter. Also, it shows better performance than auto-encoder [19]. The number of dilated convolutional layers and the number of filters are the same as the base layer.

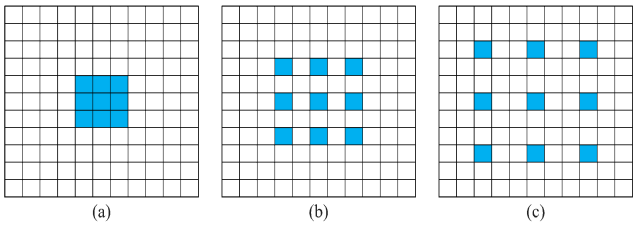

FIGURE 3. 3 × 3 dilated convolutional filters according to the dilation rates. (From left to right, the dilation rates are 1, 2, and 3.)

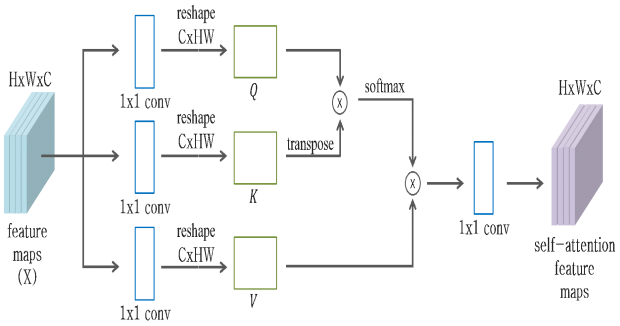

FIGURE 4. Illustration of Self-attention layer applied to features.

3) SELF-ATTENTION LAYER

Recently, attention mechanisms are the most popular in models that need to capture global dependencies [29]. In particular, self-attention [30] calculates attention score in single image by attending to all positions within the same image. The self-attention is shown in Fig. 4. Given an input features X; three different sets of features: queries Q, keys K, and values V are calculated using a linear transformation:

$$Q = W^Q X + b^Q \tag{5}$$
$$K = W^K X + b^K \tag{6}$$
$$V = W^V X + b^V \tag{7}$$

Then the self-attention weight is calculated by a scaled dot product between Q and K, the scaling factor d is the dimension of the vectors in Q and K. The self-attention layer uses this attention weight and value of matrix V to compute its output (attention score).

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d}})V \tag{8}$$

C. LOSS FUNCTION

Our loss function is defined as:

$$L = \alpha L_2 + (1 - \alpha) L_{SSIM} \tag{9}$$

First, we use the L_2 loss, which is the sum of the squares of the difference between the input and ground-truth.

$$L_2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{10}$$

where y_i and $\hat{y}_i$ are the ground-truth and the predicted value from networks respectively. However, when the networks is trained with L_2 loss, it has splotchy artifacts.

To compensate for splotchy artifacts, L_{SSIM} loss is jointed. The L_{SSIM} is used to produce visually pleasing images. $SSIM$ [31] is defined as:

$$SSIM(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{11}$$

where μ_x and μ_y are photometric measure of the luminous intensity per unit area of light travelling in a given direction, and σ_x and σ_y are difference in luminance or color that makes an object (or its representation in an image or display) distinguishable. σ_{xy} is cross-covariance for x, y, and C_1, and C_2 are constant values.

$$L_{SSIM} = 1 - SSIM(x, y) \tag{12}$$

We empirically set coefficient α to 0.85. The training goal is to minimize this loss function (see equation (9)).

IV. EXPERIMENTS

A. FISHEYE IMAGE GENERATION FOR TRAINING

A crucial problem remains in training the proposed networks which require fisheye and real images. To generate fisheye images, we randomly set k_1, k_2 (see equation (4)). The range is limited from 0.01 to 0.2. Then we use the calculated distortion radius to transform the pixels of the image from the polar coordinates to the Cartesian coordinates to obtain fisheye images. The results of fisheye image generation are shown in Fig. 6.

B. EXPERIMENTAL DETAILS

We resize the images of Place2 dataset [20] to 256 × 256, and then we generate fisheye images (Section III-A) and label them as real images. We use AdamW [32] optimizer to train our networks and schedule the learning rate with cosine annealing [33] (see equation (13)). The initial learning rate is set to 0.001.

$$LR_t = LR^i_{min} + \frac{1}{2}(LR^i_{max} - LR^i_{min})(1 + \cos(\frac{T_{cur}}{T_i}\pi)) \tag{13}$$

where LR_t is the current learning rate calculated by cosine annealing, and LR^i_{max} and LR^i_{min} are the maximum and minimum values of the learning rate. T_{cur} is the current epoch in cycle (decay steps), and T_i is the cycle to perform cosine annealing. The learning rate of the proposed method is shown in Fig. 8.

Batch size is set as 16 and total number of training data is set as 60K for Place2 dataset. We apply the Rectified Linear Unit (ReLU) activation function to all base layers. Following [34], we apply the LeakyReLU activation function to dilated convolutional layers. In addition, the workstation configuration used in the experiment is shown in Table 1.

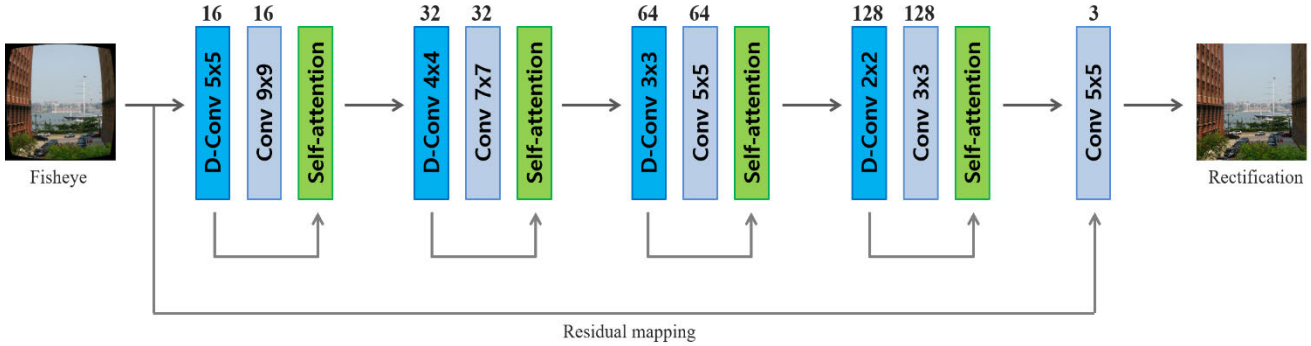

FIGURE 5. The architecture of the proposed method. The entire networks consist of three layers: D-CNNs, CNNs, and Self-attention layers. D-CNNs use enlarged filters to expand the receptive field. The filter size of CNNs matches the receptive field range of D-CNNs. In addition, self-attention is applied to the output of CNNs to extract global features, and residual mapping is applied to optimize our networks.

FIGURE 6. Fisheye image generation on real image (From the left, the values of k_1 are 0.0372, 0.1813, 0.0126; The values of k_2 are 0.0772, 0.1852, and 0.0432).

TABLE 1. Workstation configuration.

Hardware / Software	Specification
CPU	AMD Ryzen 5 5600X
GPU	GeForce RTX 3070
RAM	DDR4 64GB
Language	Python 3.7
Framework	TensorFlow 2.4.0

C. QUALITATIVE EVALUATION

To demonstrate the effectiveness of the proposed method, we provide a qualitative comparison with recent works in terms of visual performance, as shown Fig. 7. Chao et al. [12] advance self-supervised learning strategies to rectify fisheye images. This GAN-based model learns pixel-level distortion flows with unique cross rotation. However, artifacts are easily found in the rectified image that is the output of the model. Li et al. [13] propose general framework to rectify different types of geometric distortion from distorted images. They estimate the distortion parameters, and use it to generate rectified images. Their proposed framework is not as robust as other fisheye image rectification methods because it aims to rectify various types of distortion as well as radial distortion of fisheye images. Yang et al. [14] propose a parallel complementary structure to rectify fisheye images. This complementary mechanism is a method of correcting features through an encoder that reduces the degree of distortion by successive convolution and pooling operations and a decoder that is a flow estimation. This method shows comparable performance to our method. As shown in Fig. 9, our method has better quality in the local region.

D. QUANTITATIVE EVALUATION

To verify the robustness of the proposed method, we quantitatively compare it with other fisheye image rectification methods, including: Chao et al. [12], Li et al. [13], and Yang et al. [14]. Specifically, we select Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Map (SSIM) as evaluation metrics. Our quantitative comparison results are shown in Table 2.

As shown in the Table 2, the proposed method is superior to other methods in PSNR and SSIM. Chao [12] and Li et al. [13] show poor performance in PSNR and SSIM, and Yang et al. [14] shows comparable performance to the proposed method.

E. ABLATION STUDY

We analyze the effectiveness of base layers, dilated convolutional layers, and self-attention layers of the proposed method. First, we remove layers to confirm the need for dilated convolutional layers (w/o dilated convolutional layers). The removed dilated convolutional layers are replaced with base layers. Secondly, we determine how extracting global features affects the performance of the networks by removing self-attention layers (w/o self-attention layers). Additionally, we check whether the networks optimization is affected when residual mapping layers are removed (w/o residual-mapping layers). As a result of Table 3, the w/o self-attention layers shows the lowest performance, and the w/o dilated convolutional layers with the highest performance among them also shows lower performance

FIGURE 7. Qualitative comparison of the proposed method and other methods. From left to right are the fisheye image, Chao et al. [12], Li et al. [13], Yang et al. [14], the output of the proposed method, the ground truth. The results of this experiment are superior to those of related studies.

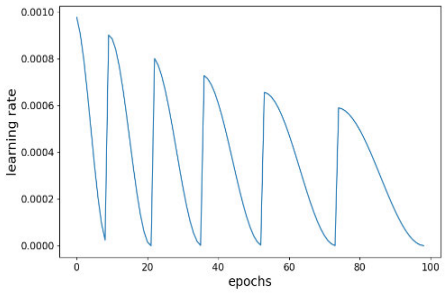

FIGURE 8. Cosine annealing with an initial learning rate of 0.001, decay steps of 10. For each annealing, the initial learning rate is decreased and the period is increased.

than the networks including the entire layers. Therefore, it can be found that each layer plays an important role in improving the performance of the proposed networks.

TABLE 2. Quantitative comparison (PSNR and SSIM) between the proposed method and other methods on our test dataset. The number of test data is set to 20K for the Place2 dataset. As a result of testing with PSNR and SSIM, the proposed method shows the best performance.

Method	PSNR	SSIM
Chao [12]	16.68	0.5493
Li [13]	17.18	0.6220
Yang [14]	26.46	0.8405
Ours	**27.93**	**0.8658**

F. COMPARISON OF MODEL PARAMETERS

We compare the number of parameters of the proposed model and comparison models to demonstrate the advantages of the proposed model.

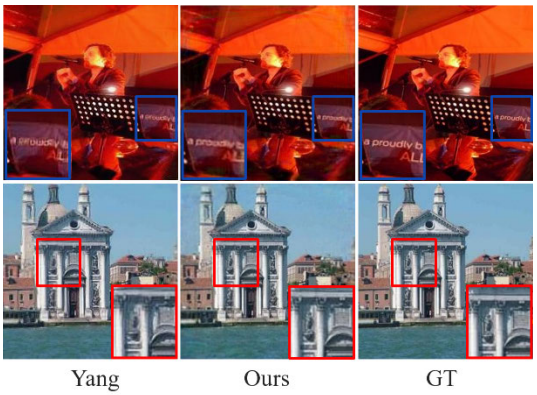

FIGURE 9. A detailed comparison of Yang [14] and our method.

TABLE 3. Ablation study of the proposed method.

Method	PSNR	SSIM
w/o dilated convolutional layers	23.93	0.7274
w/o self-attention layers	21.40	0.5910
w/o residual mapping layers	22.88	0.6537
Full	**27.93**	**0.8658**

TABLE 4. Comparison of model parameters.

Method	Total params
Chao [12]	16,011,522
Li [13]	19,337,857
Yang [14]	35,637,458
Ours	**747,923**

As shown in the Table 4, the number of parameters of the proposed model is 21 times less than Chao et al. [12] and 25 times less than Li et al. [13]. Moreover, the number of parameters of the proposed model is 47 times less than Yang et al. [14]. Therefore, the proposed model is simplified and can be used in the environment of low computing power.

V. CONCLUSION

In this paper, we propose global convolutional neural networks with self-attention to rectify fisheye images. Additionally, we employ dilated convolutional neural networks to have the same computation as traditional convolutional neural networks and to enlarge the receptive field of filters, and we apply self-attention to global features, which allows the input image to interact with each other and finds features in which input should pay more attention. To better train and evaluate the proposed method, we generate fisheye images from the Place2 dataset with Cartesian and polar coordinates and label them with original images, also we schedule the learning rate with cosine annealing and use an integrated loss function. To demonstrate the robustness of the proposed method, qualitative and quantitative evaluations with other related methods are performed. In addition, an ablation study is conducted to confirm whether each layer of the proposed method contributes to networks performance improvement. As a result of the experiment, the proposed method shows excellent performance both visually and numerically.

REFERENCES

[1] M. Bertozzi, A. Broggi, and A. Fascioli, "Vision-based intelligent vehicles: State of the art and perspectives," *Robot. Auto. Syst.*, vol. 32, no. 1, pp. 1–16, 2000.

[2] J. Huang, "6-DOF VR videos with a single 360-camera," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2017, pp. 37–44.

[3] Y. Gao, C. Lin, Y. Zhao, X. Wang, S. Wei, and Q. Huang, "3-D surround view for advanced driver assistance systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 320–328, Jan. 2018.

[4] H. Lee, S. Song, and S. Jo, "3D reconstruction using a sparse laser scanner and a single camera for outdoor autonomous vehicle," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 629–634.

[5] J. M. Alvarez, "Road scene segmentation from a single image," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 376–389.

[6] X. Li, C. Ma, B. Wu, Z. He, and M.-H. Yang, "Target-aware deep tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1369–1378.

[7] W. Bao, W.-S. Lai, X. Zhang, Z. Gao, and M.-H. Yang, "MEMC-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 933–948, Mar. 2021.

[8] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.

[9] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 1, Sep. 1999, pp. 666–673.

[10] S. Shah and J. K. Aggarwal, "Intrinsic parameter calibration procedure for a (high-distortion) fish-eye lens camera with distortion model and accuracy estimation," *Pattern Recognit.*, vol. 29, no. 11, pp. 1775–1788, Nov. 1996.

[11] X. Ying and H. Zha, "Identical projective geometric properties of central catadioptric line images and sphere images with applications to calibration," *Int. J. Comput. Vis.*, vol. 78, no. 1, pp. 89–105, 2008.

[12] C.-H. Chao, P.-L. Hsu, H.-Y. Lee, and Y.-C.-F. Wang, "Self-supervised deep learning for fisheye image rectification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 2248–2252.

[13] X. Li, B. Zhang, P. V. Sander, and J. Liao, "Blind geometric distortion correction on images through deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4855–4864.

[14] S. Yang, C. Lin, K. Liao, C. Zhang, and Y. Zhao, "Progressively complementary network for fisheye image rectification using appearance flow," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6348–6357.

[15] K. Liao, C. Lin, and Y. Zhao, "A deep ordinal distortion estimation approach for distortion rectification," *IEEE Trans. Image Process.*, vol. 30, pp. 3362–3375, 2021.

[16] Z. Xue, N. Xue, G.-S. Xia, and W. Shen, "Learning to calibrate straight lines for fisheye image rectification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1643–1651.

[17] S. Sha, Z. Wen, X. Lin, Q. Luo, and C. Kong, "Fisheye image calibration and super-resolution method based on deep learning," in *Proc. 4th Int. Conf. Image Graph. Process.*, Jan. 2021, pp. 102–108.

[18] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[19] L. C. Chen, G. Papandreou, and I. Kokkinos, "DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Jun. 2016.

[20] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, Jul. 2018.

[21] S. B. Kang, "Semiautomatic methods for recovering radial distortion parameters from a single image," Cambridge Res. Lab., Cambridge, U.K., Tech. Rep. CRL 97/3, May 1997.
[22] M. A. Abidi, R. O. Eason, and R. C. Gonzalez, "Camera calibration in robot vision," in *Proc. 4th Scandinavian Conf. Image Analysis*, 1985.
[23] R. Melo, "Unsupervised intrinsic calibration from a single frame using a 'plumb-line' approach," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 537–544.
[24] J. P. Barreto, J. Roquette, P. Sturm, and F. Fonseca, "Automatic camera calibration applied to medical endoscopy," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 1–11.
[25] J. Rong, "Radial lens distortion correction using convolutional neural networks trained with synthesized images," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016.
[26] X. Yin, "FishEyeRecNet: A multi-context collaborative deep network for fisheye image rectification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 1–16.
[27] Z.-C. Xue, N. Xue, and G.-S. Xia, "Fisheye distortion rectification from deep straight lines," 2020, *arXiv:2003.11386*.
[28] S. Yang, C. Lin, K. Liao, Y. Zhao, and M. Liu, "Unsupervised fisheye image correction through bidirectional loss with geometric prior," *J. Vis. Commun. Image Represent.*, vol. 66, Jan. 2020, Art. no. 102692.
[29] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
[30] H. Zhang, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.
[31] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[32] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
[33] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 2016, *arXiv:1608.03983*.
[34] X. Zhang, Y. Zou, and W. Shi, "Dilated convolution neural network with LeakyReLU for environmental sound classification," in *Proc. 22nd Int. Conf. Digit. Signal Process. (DSP)*, Aug. 2017, pp. 1–5.

BYUNGHYUN KIM received the B.S. degree in computer engineering from Seokyeong University, Seoul, South Korea, in 2020. He is currently pursuing the M.S. degree in software engineering with Hanyang University, Seoul. His research interests include artificial intelligence, machine learning, deep learning, computer vision, and natural language processing.

DOHYUN LEE received the B.S. degree in computer engineering from Seokyeong University, Seoul, in 2020. He is currently pursuing the M.S. degree in software engineering with Hanyang University, Seoul, South Korea. His research interests include artificial intelligence, machine learning, deep learning, computer vision, and bioinformatics.

KYEONGYUK MIN (Member, IEEE) received the B.S. degree in physics from Korea University, Seoul, South Korea, in 1992, and the M.S. and Ph.D. degrees in electronics engineering from Hanyang University, Seoul, in 1996 and 2010, respectively. His current research interests include digital cinema, implementation of FMCW radar, and hardware design of real-time H.264 encoder/decoder and JPEG2000 encoder/decoder.

JONGWHA CHONG received the B.S. and M.S. degrees in electronics engineering from Hanyang University, Seoul, South Korea, in 1975 and 1979, respectively, and the Ph.D. degree in electronics and communication engineering from Waseda University, Tokyo, Japan, in 1981. From 1981 to 2017, he served as a Professor at the Department of Electronics Engineering, Hanyang University. Since 2018, he has been a professor with the Department of Computer Science, State University of New York Korea, Incheon, South Korea. His current research interests include SoC design methodology (including memory centric design and physical design automation of 3D ICs), indoor wireless communication SoC designs for ranging and location, and video systems.

INWHEE JOE received the B.S. and M.S. degrees in electronics engineering from Hanyang University, Seoul, South Korea, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1998. Since 2002, he has been a Faculty Member with the Department of Computer Science, Hanyang University. His current research interests include the Internet of Things, cellular systems, wireless powered communication networks, embedded systems, network security, machine learning, and performance evaluation.