

DYNAMIC FOCUS MECHANISM-BASED DUAL-DOMAIN RECONSTRUCTION NETWORK FOR ACCELERATED MRI

Zhongxian Wang¹, Zhiwen Wang¹, Zhongzhou Zhang¹, Ziyuan Yang¹, Maosong Ran¹, Hui Yu¹, Zhenyang Yu¹, Yi Zhang^{2,1}, Senior Member, IEEE

¹College of Computer Science, Sichuan University, Chengdu, P. R. China

² School of Cyber Science and Engineering, Sichuan University, Chengdu, P. R. China

ABSTRACT

The integration of information from both k -space and image domains has become a topic of growing interest in accelerated Magnetic Resonance Imaging (MRI) using dual-domain reconstruction networks. However, many existing methods treat sampled and unsampled measurements equally in the k -space domain, leading to progressive feature shift and suboptimal reconstruction performance. Additionally, networks with single-scale receptive field in the image domain are insufficient for fully exploring global-local information and lack channel-wise feature interaction to preserve structural details. To address these issues, we propose a novel dual-domain reconstruction network named DFRNet, which incorporates a dynamic focus mechanism. In the k -space domain, we introduce an area normalization module that dynamically eases feature shift and predicts reliable k -space measurements. In the image domain, our proposed dynamical attention module with channel-wise gating mechanism extracts rich global-local features for detail recovery. Quantitative and qualitative experiments show that our proposed DFRNet achieves competitive performance to several state-of-the-art methods while being more lightweight.

Index Terms— dynamic focus, MRI reconstruction, multi-scale feature aggregation

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) is a widely used non-invasive medical imaging technique for high-resolution imaging of human tissues. However, the acquisition of high-quality MR images can be time-consuming, which can cause patient discomfort and increase the cost for hospitals.

To accelerate the MRI scan, compressed sensing (CS) [1] has been introduced to reduce the sampling number in k -space [2–4]. However, the huge computational cost and laborious parameter selection hinder the clinical applications. In recent years, deep learning (DL)-based reconstruction methods [5, 6] have received increasing attention due to their superior inference speed and reconstruction performance com-

pared to CS-based methods, under the premise that large-scale training data is available [7–9].

In current DL-based methods, the dual-domain approaches achieve a state-of-the-art (SOTA) performance due to the rich information from both image and k -space domains [7]. On the other side, some normalization techniques [10, 11] are widely used to eliminate the vanishing gradient and ease the training of dual-domain networks [12]. However, existing normalization methods are not perfectly suitable for MRI reconstruction. Specifically, applying existing normalization for the undersampled measurements in k -space may result in severe mean and variance shift of features due to the existence of zero-filled values. In the image domain, previous networks based on convolutional neural networks (CNNs) or Transformers use isolated features from local or global receptive field, which lacks adaptive channel-wise feature interaction and limits the network ability in recovering fine-grained structural details [13, 14].

In this paper, inspired by [12, 15], we propose a novel dynamic focus mechanism-based dual-domain parallel reconstruction network (DFRNet) for relieving the limitations mentioned above. Our proposed dynamic focus mechanism includes two main components, an area normalization (AN) layer for k -space domain and a dynamic attention module for image domain. Particularly, AN performs normalization separately for the sampled and the unsampled measurements in k -space domain. It focuses on the sampled areas, dynamically estimates mask probability distribution of reconstructed k -space data for next AN operation and progressively recovers the unsampled measurements. In the image domain, a novel dynamic attention with multi-scale receptive fields and channel-wise gating mechanism, called MSGA, is proposed. MSGA uses convolution kernels with different sizes and dilation rates to encourage the network to focus on the multi-scale features from both local and global ranges. The gating mechanism of channel dimension introduces adaptive weight calculation for different spatial features to improve the representation for better detailed reconstruction performance.

Our main contributions can be summarized as follows:
(1) we propose a novel area normalization layer in k -space

Yi Zhang is the corresponding author

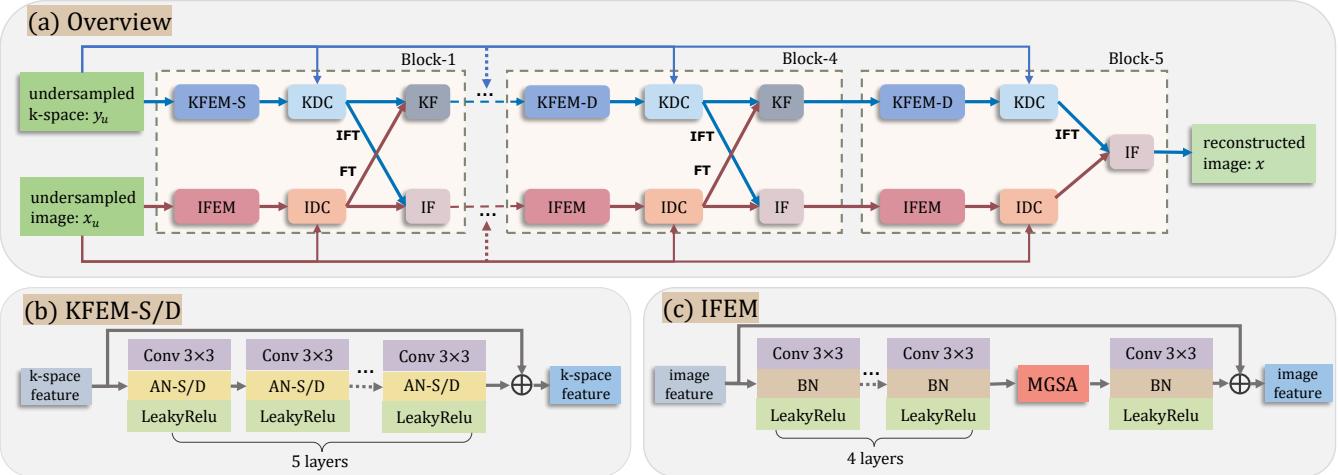


Fig. 1. The overall architecture of the proposed DFRNet. (a) The backbone of the network. (b) the details of the KFEM module. (c) the details of the IFEM module. BN: batch normalization. AN: area normalization. 'S' and 'D' represent the shallow and the deep versions, respectively and the details are given in Section 2.2. KFEM: k -space feature extraction module. 'S' and 'D' represent using different types of AN modules. IFEM: image feature extraction module. KDC: k -space data consistency. IDC: image data consistency. FT: Fourier Transform. IFT: inverse Fourier Transform. KF: k -space data fusion layer. IF: image data fusion layer. MSGA: multi-scaled gating attention module and the details are given in Section 2.3.

network to alleviate the mean and variance shift of features in k -space; (2) we design a dynamic attention mechanism to capture the local and global features from multi-scale receptive fields and improve the channel-wise feature extraction; (3) DFRNet with dynamic focus mechanism is proposed for CS-MRI reconstruction, whose effectiveness and efficiency are experimentally validated.

2. METHODS

2.1. The Overall Network Architecture

The overall architecture of DFRNet is shown in Fig. 1(a). We choose MD-Recon-Net [7] as the backbone to build the proposed DFRNet. As a lightweight network, only five blocks are adopted. The undersampled measurement y_u and the zero-filling reconstruction x_u are fed into the network to predict the full-sampled reconstruction x . The first three blocks of the network share the same structure. In the fourth and fifth blocks, k -space feature extraction module-deep (KFEM-D) replaces k -space feature extraction module-shallow (KFEM-S) for k -space feature extraction. The details of KFEM-S/D are given in Section 2.2. The fifth block removes the k -space data fusion (KF) module to only fuse the features in image domain and generate the final reconstruction. As illustrated in Fig. 1(a), each block has two parallel branches, and the features extracted from both k -space and image domains are interacted via the fusion module in a cascaded manner. The structures of k -space and image feature extraction module (KFEM and IFEM) are demonstrated in Fig. 1(b) and (c). The IFEM module includes an extra MSGA block to model

the spatial and channel relationships while two types of area normalization blocks (AN-S/D) in KFEM-S/D module replace the traditional batch normalization (BN) module to relieve the feature shift in k -space domain. Similar to [3], data consistency modules in dual domains (KDC and IDC) are adopted, which constrain the reconstruction close to the original measurements. KF and IF are the fusion modules in k -space and image domains, respectively, which are formulated as: $F = \frac{1}{1+\gamma}f_1 + \frac{\gamma}{1+\gamma}f_2$, where f_1 and f_2 denote the inputs from two branches and γ is a learnable balancing coefficient.

2.2. Area Normalization

To vary the representations of sampled and unsampled measurements, we propose AN module for the sampled areas and global interpolation is adopted for the unsampled areas to relieve the feature shift. Let $K \in \mathbb{R}^{B \times C \times H \times W}$ be the input k -space features, $k_{b,c,h,w}$ be the basic element of K , and $K_{b,c} \in \mathbb{R}^{H \times W}$ be a channel-and-batch subset of K at spatial scale. B , C , H and W are batch size, number of channels, height and width, respectively. On the one hand, given an initial undersampling mask, $K_{b,c} \in \mathbb{R}^{H \times W}$ can be divided into the following areas: $A_{b,c}^1 \cup A_{b,c}^2 \cup \dots \cup A_{b,c}^m \cup \dots \cup A_{b,c}^M$, where m denotes a phase coding line in Cartesian sampling or a spoke line in radial sampling and M is the total number of sampling lines. Then, the mean $\mu_{b,c}^m$ and standard deviation $\sigma_{b,c}^m$ of each area ($A_{b,c}^m$) can be calculated by:

$$\mu_{b,c}^m = \frac{\sum_{k_{b,c,h,w} \in A_{b,c}^m} k_{b,c,h,w}}{|A_{b,c}^m|} \quad (1)$$

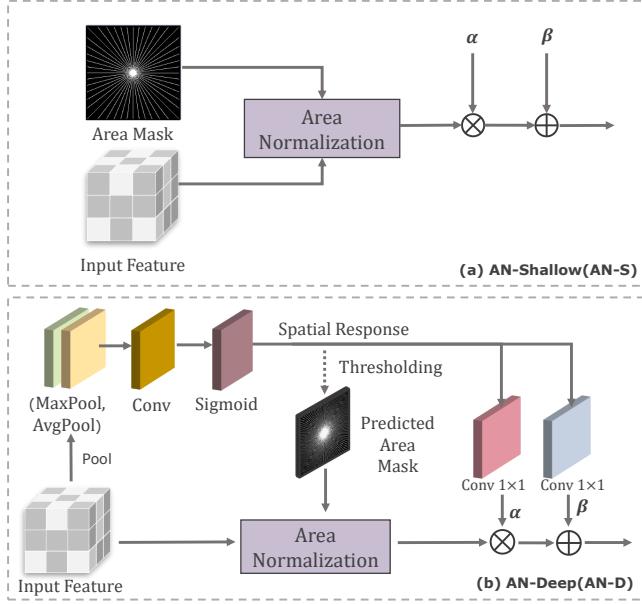


Fig. 2. The details of area normalization (AN). Area mask is the same as undersampling mask used in DC layers.

and

$$\sigma_{b,c}^m = \sqrt{\frac{1}{|A_{b,c}^m|} \sum_{k_{b,c,h,w} \in A_{b,c}^m} (k_{b,c,h,w} - \mu_{b,c}^m)^2 + \epsilon}, \quad (2)$$

where ϵ is a trivial constant to avoid being divided by zero. Each sampled area $A_{b,c}^m$ in k -space processed by AN is denoted as $\hat{A}_{b,c}^m$. On the other hand, the initial unsampled measurements are interpolated along the sampled points in the global-average manner to be normalized. $\hat{A}_{b,c}^m$ and the normalized features from the unsampled measurements are integrated to get the area-normalized feature $\hat{K}_{b,c}$. As shown in Fig. 2(a), AN-Shallow (AN-S) takes k -space features and the initial area mask as input. α and β are two learnable affine parameters. AN-S is used in the first three layers of DFRNet to relieve the feature shift in each module of k -space gradually, when the initial undersampled mask is more accurate in the shallow layers of DFRNet. As the network goes deeper, the missing parts of k -space data are progressively recovered, leading to inaccuracy initial undersampled mask. AN-Deep (AN-D) module is developed to generate pseudo undersampled mask and focus on the sampled areas dynamically. The AN-D is revealed in Fig. 2(b). Given the input k -space features, a global spatial response is produced using two pooling layers and a 3×3 convolution with a sigmoid function. The pseudo undersampled mask stems from the selection of global spatial response with 0.8 as the threshold. Unlike AN-S, the affine parameters here are learned from the spatial response via 1×1 convolution to get more global representation. In this way, the fusion of the sampled and unsampled regions can be

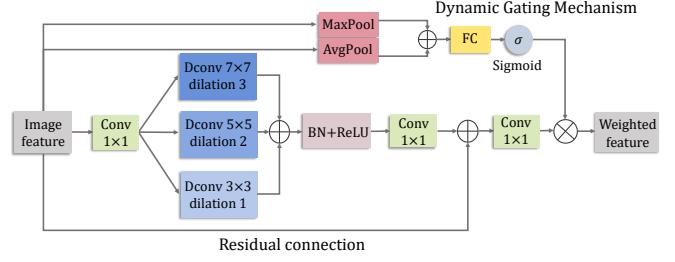


Fig. 3. The structure of MSGA. Dconv: Depth-wise convolution. FC: Fully connected layer

enhanced naturally.

2.3. Multi-Scale Gated Attention

To efficiently utilize the local-global information, we propose MSGA with mixed receptive fields, which is implemented with parallel multi-scale convolution kernels with different dilation rates [16]. The details of MSGA is illustrated in Fig. 3. Three different kernels, including 3×3 with dilation rate 1, 5×5 with dilation rate 2 and 7×7 with dilation rate 3, are employed to provide diverse features in image domain. For spatial attention part of MSGA, we designed a dynamic channel-wise gated mechanism, which consists of two global pooling layers, one fully connected layer and one sigmoid function. It is helpful to dynamically calculate the channel-wise relationships, which are ignored in spatial information fusion. Additionally, it improves the significance of the spatial features in image domain by dynamically emphasizing the non-trivial ones.

3. EXPERIMENTS AND RESULTS

3.1. Implementation Details

The public brain MR raw data set Calgary Campinas [17], which comes from the clinic MR scanner (Discovery MR750; General Electric (GE) Healthcare, Waukesha, WI), are used to validate our proposed model. The training set contains 4524 slices from 25 subjects, and the testing set contains 1700 slices from 10 subjects. The acquisition matrix size is set to 256×256 .

We use Adam ($\beta_1 = 0.9, \beta_2 = 0.999$) as the optimizer. The learning rate is set to 0.001 and gradually decays to 0 in 200 epochs with a cosine annealing strategy. The proposed DFRNet is implemented by PyTorch with a NVIDIA RTX 3090 GPU. To evaluate the performance of the proposed DFRNet, three SOTA methods, including UNet [8], MD-Recon-Net [7], and SwinMR [9] are compared. UNet is a typical MR reconstruction network. MD-Recon-Net employs a parallel dual-domain architecture, but ignores the characteristic differences in spatial and k -space domains. SwinMR is built based on the popular Swin-Transformer [18]. Peak signal to noise ratio (PSNR), and structural similarity index

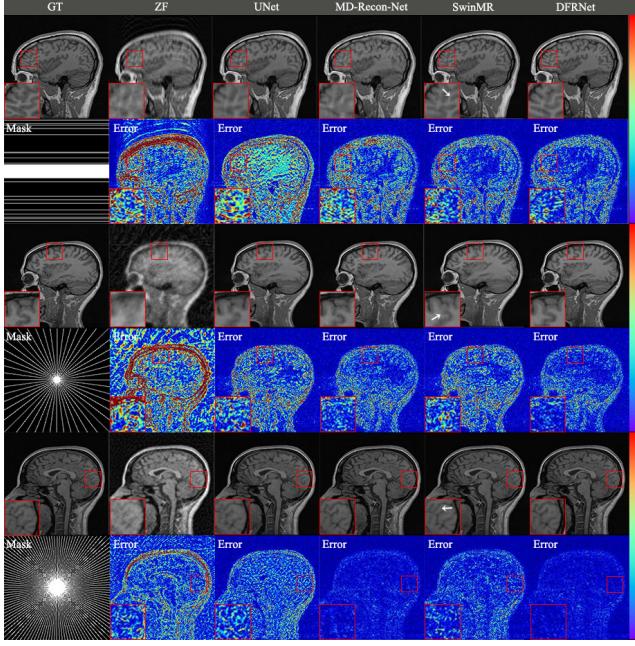


Fig. 4. Visual results of different methods with different undersampling patterns and sampling rates.

measure (SSIM) are adopted as the quantitative metrics. The number of parameters (Params) is used to assess the model scale.

3.2. Comparisons with The State-Of-The-Art Methods

Three different sampling strategies, including Cartesian sampling with 20% rate and radial sampling with 10% and 25% sampling rate are utilized to verify the reconstruction performance. Visual results and error maps are shown in Fig. 4. It is noticed that our proposed DFRNet achieves overall lowest error residuals in all the cases, thus leading to the best visual effect, especially for detail recovery of brain tissue. As shown in the magnified regions marked by the red boxes, MD-Recon-Net and UNet tend to oversmooth the cerebral gyri and sulci, while SwinMR generates incorrect brain soft tissues indicated by the white arrows. Quantitative results are listed in Table 1. We can see that our DFRNet outperforms other methods. On the other side, the amount of parameters of our DFRNet is much less than UNet and SwinMR and has same order of magnitude to MD-Recon-Net.

3.3. Ablation Study

In this part, ablation experiments are conducted to verify the improvement of reconstruction performance from AN and MSGA modules. All experiments are performed on the data with 20% Cartesian sampling. As shown in Fig. 5(a), two proposed modules bring a progressive improvement to the imaging quality. Fig. 5(b) shows the improvements

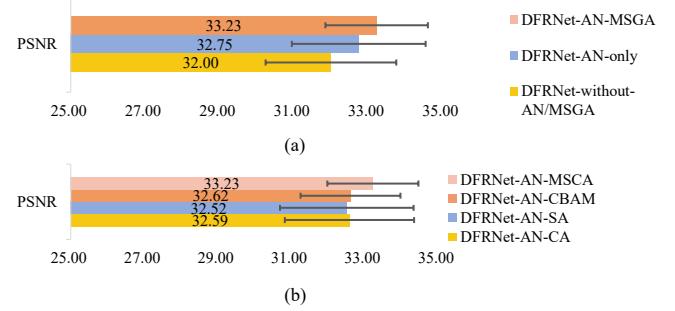


Fig. 5. The quantitative results of ablation study.(a) The effect of AN and MSGA modules.(b) Comparisons with different attention modules.

Table 1. PSNR, SSIM and Params values of different methods with different sampling patterns and sampling rates

	Cartesian			Radial	Params
	20%	10%	25%		
Zero-filling	24.62±1.35 0.69±0.06	24.18±1.43 0.61±0.05	28.84±1.59 0.79±0.03	/	
UNet [8]	30.31±1.95 0.88±0.03	29.61±2.38 0.86±0.04	34.71±2.25 0.93±0.02	7.76M	
MD-Recon-Net [7]	32.00±1.38 0.90±0.02	32.92±1.50 0.91±0.02	39.85±1.18 0.97±0.01	0.29M	
SwinMR [9]	32.14±1.70 0.90±0.02	31.16±2.38 0.87±0.04	37.31±2.14 0.94±0.01	11.40M	
DFRNet	33.23±1.76 0.92±0.02	34.19±2.05 0.92±0.02	40.84±1.56 0.97±0.01	0.42M	

by adding different attention modules, including channel attention (CA) [19], spatial attention (SA) [20], convolutional block attention module (CBAM) [20] and the proposed MSCA. It can be observed that benefiting from the multi-scale features and dynamic gating mechanism, our proposed MSGA module achieves the highest PSNR scores.

4. CONCLUSION

In this work, we developed a lightweight dynamic focus mechanism-based parallel dual-domain network for accelerated MRI reconstruction. Experiments show that our method can effectively reduce the artifacts and achieve better reconstruction with different sampling patterns and rates than several SOTA methods.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by CALGARY-CAMPINAS. Ethical approval was not required as confirmed

by the license attached with the open access data.

6. CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

7. ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 62271335; in part by the Sichuan Science and Technology Program under Grant 2021JDJQ0024; and in part by the Sichuan University “From 0 to 1” Innovative Research Program under Grant 2022SCUH0016.

8. REFERENCES

- [1] L. David, “Donoho. compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] P. Kellman, F. H. Epstein, and E. R. Mcveigh, “Adaptive sensitivity encoding incorporating temporal filtering (tsense).,” *Magnetic Resonance in Medicine*, vol. 45, no. 5, pp. 846–852, 2010.
- [3] S. Ravishankar and Y. Bresler, “Mr image reconstruction from highly undersampled k-space data by dictionary learning.,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028, 2011.
- [4] Jose Caballero, Anthony N. Price, Daniel Rueckert, and Joseph V. Hajnal, “Dictionary learning and time sparsity for dynamic mr data reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 33, no. 4, pp. 979–994, 2014.
- [5] Y. Yang, J. Sun, Huibin Li, and Z. Xu, “Admm-csnet: A deep learning approach for image compressive sensing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 521–538, 2020.
- [6] S. Wang, Z. Su, L. Ying, P. Xi, and L. Dong, “Accelerating magnetic resonance imaging via deep learning,” in *2016 IEEE 13th International Symposium on Biomedical Imaging*. 2016, pp. 514–517, IEEE.
- [7] M.S. Ran et al., “Md-recon-net: A parallel dual-domain convolutional neural network for compressed sensing mri,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 120–135, 2020.
- [8] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, and Tullie Murrell, “fastmri: An open dataset and benchmarks for accelerated mri,” *CoRR*, vol. abs/1811.08839, 2018.
- [9] J. Huang, Y. Fang, Y. Wu, H. Wu, and G. Yang, “Swin transformer for fast mri,” *Neurocomputing*, vol. 493, pp. 281–304, 2022.
- [10] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton, “Layer Normalization,” *CoRR*, vol. abs/1607.06450, 2016.
- [11] Sergey Ioffe and Christian Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on Machine Learning*. 2015, vol. 37 of *JMLR Workshop and Conference Proceedings*, pp. 448–456, PMLR.
- [12] T. Yu, Z. Guo, X. Jin, S. Wu, and S. Liu, “Region normalization for image inpainting,” in *The Thirty-Fourth Conference on Artificial Intelligence*. 2020, pp. 12733–12740, AAAI Press.
- [13] Zimian Wei, Hengyue Pan, Xin Niu, and Dongsheng Li, “ConvFormer: Closing the Gap Between CNN and Vision Transformers,” *CoRR*, vol. abs/2209.07738, 2022.
- [14] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H. Li, Shan Liu, and Ge Li, “Structureflow: Image inpainting via structure-aware appearance flow,” in *International Conference on Computer Vision*. 2019, pp. 181–190, IEEE.
- [15] X. Ding, X. Zhang, Y. Zhou, J. Han, G. Ding, and J. Sun, “Scaling up your kernels to 31x31: Revisiting large kernel design in cnns,” *CoRR*, vol. abs/2203.06717, 2022.
- [16] Mingxing Tan and Quoc V. Le, “Mixconv: Mixed depthwise convolutional kernels,” in *30th British Machine Vision Conference*. 2019, p. 74, BMVA Press.
- [17] Roberto Souza, Oeslle Lucena, Julia Garrafa, et al., “An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement,” *NeuroImage*, vol. 170, pp. 482–494, 2018.
- [18] Ze Liu, Han Hu, and Yutong Lin, “Swin transformer V2: scaling up capacity and resolution,” in *Conference on Computer Vision and Pattern Recognition*. 2022, pp. 11999–12009, IEEE.
- [19] Jie Hu, Li Shen, and Gang Sun, “Squeeze-and-excitation networks,” in *Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7132–7141, IEEE.
- [20] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, “CBAM: convolutional block attention module,” in *Computer Vision 15th European Conference*. 2018, pp. 3–19, Springer.