Neurodynamics-Driven Coupled Neural P Systems for Multi-Focus Image Fusion Appendix

Bo Li[1,2] Yunkuo Lei[1,2] Tingting Bao[1,2] Hang Yan[1,2]
Yaxian Wang[4] Weiping Fu[1,2] Lingling Zhang[1,2*] Jun Liu[1,3]
[1] School of Computer Science and Technology, Xi'an Jiaotong University
[2] Ministry of Education Key Laboratory of Intelligent Networks and Network Security, China
[3] Shaanxi Province Key Laboratory of Big Data Knowledge Engineering [4] Chang'an University
morvanli@stu.xjtu.edu.cn, zhanglling@mail.xjtu.edu.cn

A. Theoretical Derivation

A.1. Formulation of Continuous Firing Condition

Theorem 4. *The relationship between external input and internal parameters of neurons in the continuous firing state is characterized by the inequality:*

$$I > \frac{\lambda(1-\alpha)(1-\beta)}{(1-\gamma)(1-\beta+\mathrm{sum}(W))} - \mathrm{sum}(W). \tag{1}$$

Proof. We use the firing criterion $U(t)\,(1+V(t)) > T(t)$. Below, we derive a sufficient steady-state condition by letting t go to infinity. Substituting results $U(t)$, $V(t)$, and $T(t)$ obtained by **Theorems 1, 2, and 3**, we have

$$\left(I\frac{1-\alpha^t}{1-\alpha} + \sum_{n=0}^{t-1} K(n)\alpha^{t-n-1}\right)\left(1 + \sum_{n=0}^{t-1} K(n)\beta^{t-n-1}\right) > \lambda\frac{1-\gamma^{t-1}}{1-\gamma}, \tag{2}$$

where $\alpha, \beta, \gamma \in (0,1)$.

We analyze the steady-state continuous-firing regime, such that $P_{kl} \equiv 1$, and obtain $K(n) = \mathrm{sum}(W)$. Using geometric series properties,

$$\sum_{n=0}^{t-1} \alpha^{t-n-1} = \frac{1-\alpha^t}{1-\alpha}, \qquad \sum_{n=0}^{t-1} \beta^{t-n-1} = \frac{1-\beta^t}{1-\beta}. \tag{3}$$

Thus,

$$U(t) = \big[I + \mathrm{sum}(W)\big]\frac{1-\alpha^t}{1-\alpha}, \qquad V(t) = \mathrm{sum}(W)\,\frac{1-\beta^t}{1-\beta}. \tag{4}$$

Substituting into (2) yields

$$\big[I + \mathrm{sum}(W)\big]\frac{1-\alpha^t}{1-\alpha}\left(1 + \mathrm{sum}(W)\,\frac{1-\beta^t}{1-\beta}\right) > \lambda\,\frac{1-\gamma^{t-1}}{1-\gamma}. \tag{5}$$

Expanding the left-hand side,

$$I \cdot \underbrace{\frac{1-\alpha^t}{1-\alpha}\left(1 + \mathrm{sum}(W)\frac{1-\beta^t}{1-\beta}\right)}_{A(t)} + \underbrace{\mathrm{sum}(W)\cdot\frac{1-\alpha^t}{1-\alpha}\left(1 + \mathrm{sum}(W)\frac{1-\beta^t}{1-\beta}\right)}_{B(t)} > \lambda\,\frac{1-\gamma^{t-1}}{1-\gamma}, \tag{6}$$

*Corresponding author.

where $A(t)$ is the coefficient of I, containing α and β, $B(t)$ is the constant term, containing α, β, and $\mathrm{sum}(W)$. Since $1-\beta+\mathrm{sum}(W)>0$ and $\alpha,\beta\in(0,1)$, we have

$$A(t)=\frac{1-\alpha^t}{1-\alpha}\left(1+\mathrm{sum}(W)\frac{1-\beta^t}{1-\beta}\right)>0.$$

Rearranging to isolate I,

$$I\cdot A(t)>\lambda\frac{1-\gamma^{t-1}}{1-\gamma}-B(t),\qquad\Rightarrow\qquad I>\frac{\lambda\frac{1-\gamma^{t-1}}{1-\gamma}-\mathrm{sum}(W)\cdot\frac{1-\alpha^t}{1-\alpha}\left(1+\mathrm{sum}(W)\frac{1-\beta^t}{1-\beta}\right)}{\frac{1-\alpha^t}{1-\alpha}\left(1+\mathrm{sum}(W)\frac{1-\beta^t}{1-\beta}\right)}. \tag{7}$$

In the steady-state limit $t\to\infty$ (so $\alpha^t,\beta^t,\gamma^{t-1}\to 0$),

$$A(\infty)=\frac{1}{1-\alpha}\left(1+\mathrm{sum}(W)\frac{1}{1-\beta}\right),\qquad B(\infty)=\mathrm{sum}(W)\cdot\frac{1}{1-\alpha}\left(1+\mathrm{sum}(W)\frac{1}{1-\beta}\right). \tag{8}$$

Hence,

$$I>\frac{\lambda\frac{1}{1-\gamma}-\mathrm{sum}(W)\cdot\frac{1}{1-\alpha}\left(1+\mathrm{sum}(W)\frac{1}{1-\beta}\right)}{\frac{1}{1-\alpha}\left(1+\mathrm{sum}(W)\frac{1}{1-\beta}\right)}. \tag{9}$$

Multiplying numerator and denominator by $(1-\alpha)(1-\beta)$ gives

$$I>\frac{\lambda\frac{(1-\alpha)(1-\beta)}{1-\gamma}-\mathrm{sum}(W)\big((1-\beta)+\mathrm{sum}(W)\big)}{(1-\beta)+\mathrm{sum}(W)}. \tag{10}$$

Therefore, the final steady-state condition is

$$I>\frac{\lambda(1-\alpha)(1-\beta)}{(1-\gamma)(1-\beta+\mathrm{sum}(W))}-\mathrm{sum}(W). \tag{11}$$

Theorem 4 is proved. □

A.2. Detailed CNP Neuron Parameter Configurations

This subsection details the parameter configuration process of CNP neurons based on Corollary 1.

Corollary 1. *Let I denote the input at an arbitrary pixel position (i,j) in the image $\mathcal{Y}\in[0,1]^{H\times W}$. To prevent uncontrolled sustained firing of the neuron at that position σ_{ij}, the input must satisfy*

$$I\le\frac{\lambda(1-\alpha)(1-\beta)}{(1-\gamma)(1-\beta+\mathrm{sum}(W))}-\mathrm{sum}(W). \tag{12}$$

Within the constraints of Eq. (12), we perform model parameterization.

Feeding input unit. The parameter u, representing the consumption value of the feed input unit, is difficult to compute directly. Instead, we introduce the exponential decay coefficient α. A larger α results in a narrower distribution, indicating an inverse relationship between α and the standard deviation of the input signal. We find that the logarithmic function of the reciprocal of the image standard deviation (std) effectively captures this relationship:

$$\alpha=1-\exp\left(-\log_{10}\left(\frac{1}{\mathrm{std}(\mathcal{Y})}\right)\right). \tag{13}$$

Here, $\mathrm{std}(\mathcal{Y})$ denotes the standard deviation of the normalized image, and α adaptively decreases with increasing contrast, enabling the neuron to adjust its decay rate according to local intensity variations.

Since $\mathcal{Y}$ is normalized, $\text{std}(\mathcal{Y}) \leq 0.5$. Thus, we obtain the consumed u in neuron σ_{ij} at moment t:

$$u = \exp\left(-\log_{10}\left(\frac{1}{\text{std}(\mathcal{Y})}\right)\right) U_{ij}(t). \tag{14}$$

Linking input unit. The parameter v denotes the consumption value of the linking input unit, which physically represents a coupled complement [5]. Accordingly,

$$v = \left(1 - \exp\left(-\log_{10}\left(\frac{1}{\text{std}(\mathcal{Y})}\right)\right)\right) V_{ij}(t). \tag{15}$$

Threshold weight. The parameter λ indicates the weight of the dynamic threshold. Following [1], it is calculated as

$$\lambda = \exp\left(-\log_{10}\left(\frac{1}{\text{std}(\mathcal{Y})}\right)\right) + \frac{\max(\mathcal{Y})}{o'(\mathcal{Y})}, \tag{16}$$

where $\max(\cdot)$ denotes the maximum value operation, and $o'(\cdot)$ stands for the Otsu threshold operation.

Dynamic threshold unit. The parameter τ refers to the consumption value of the dynamic threshold unit. We calculate γ strictly based on the theoretical analysis. Note that we denote the synaptic matrix as

$$W = \begin{bmatrix} 0.5 & 1 & 0.5 \\ 1 & 0 & 1 \\ 0.5 & 1 & 0.5 \end{bmatrix},$$

where the elements satisfy the neighborhood weight definition: (1) $W_{22} = 0$ (center neuron self-inhibition), $W_{ij} = \frac{1}{(i-2)^2+(j-2)^2}$ for $i \neq 2$ or $j \neq 2$ ($1 \leq i, j \leq 3$). Then, replacing the calculated values of α, β, and λ in Eq. (12), we obtain γ:

$$\gamma \geq 1 - \frac{\lambda(1-\alpha)(1-\beta)}{(I + \text{sum}(W))(1 - \beta + \text{sum}(W))}. \tag{17}$$

We further obtain the consumed τ in neuron σ_{ij} at moment t:

$$\tau \leq \frac{\lambda(1-\alpha)(1-\beta)}{(I + \text{sum}(W))(1 - \beta + \text{sum}(W))} T_{ij}(t). \tag{18}$$

In summary, these parameter settings strictly satisfy Corollary 1 and ensure the proper functioning of CNP neurons.

B. Extended Dealing with Multiple Inputs

In this section, we will perform source image sequence fusion, which requires more than two source images. In this case, we can fuse these source images sequentially. As shown in Fig. 1, the model is first used to generate the first decision map (DM_1), and Eq. (19) is used to obtain the intermediate result.

$$\text{F} = \text{A} \times \text{DM} + \text{B} \times (1 - \text{DM}), \tag{19}$$

where A and B denote the source multi-focus images, and F is the fused image.

Then, the intermediate result with another source image is input into the model to generate the second decision map (DM_2), which is again fused using Eq. (19) to obtain the final fused image. In this manner, ND-CNPFuse is theoretically capable of fusing any number of multi-focus image inputs.

C. More Experiment Results

C.1. More Qualitative Fusion Results

We further show more qualitative comparison fusion results from Lytro [4], MFI-WHU [9], and Real-MFF [10] datasets. The purpose is to fully indicate the superiority of our proposed method. Specifically, as follows:

Fig. 2 shows the fusion results of 10 different fusion methods on "Lytro-5", we can observe that almost all methods produce acceptable fusion results. However, PADCDTNP performs imperfectly in detail, as shown by the small red rectangular boxes,

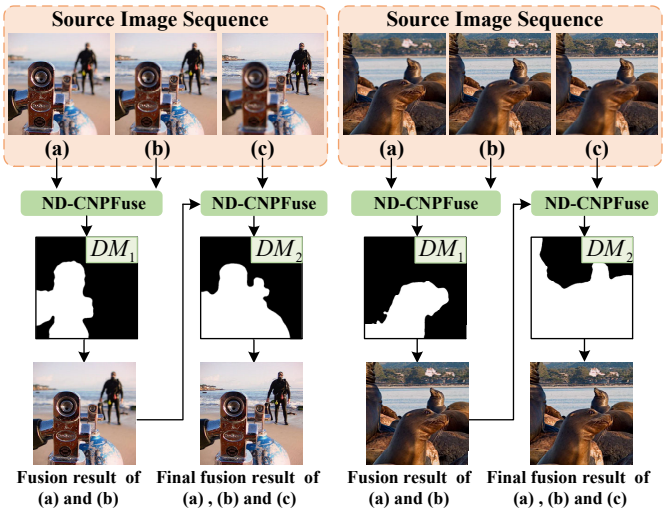

Figure 1. ND-CNPFuse fuses the sequence of images from the Lytro dataset.

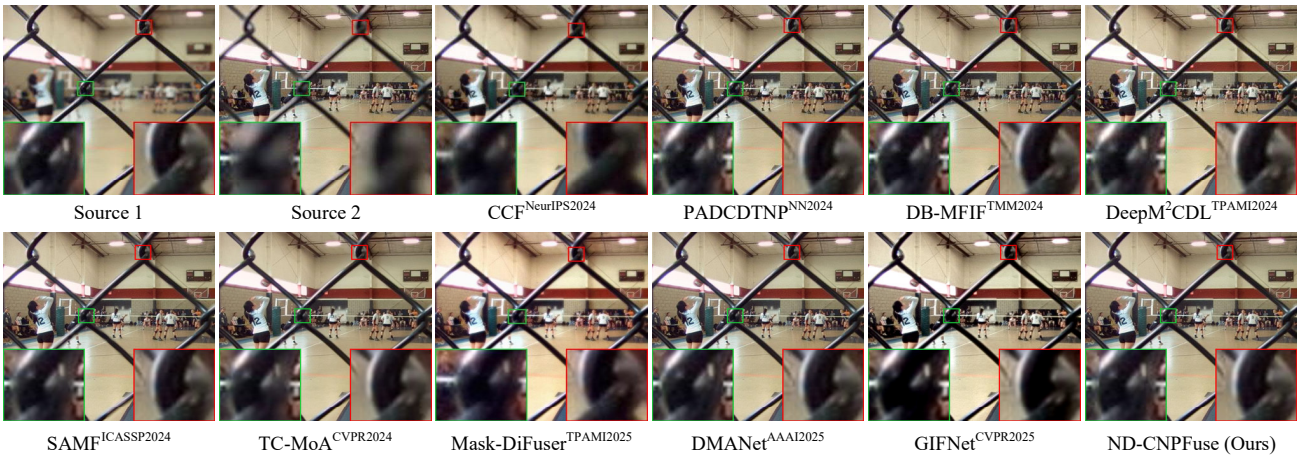

Figure 2. Qualitative comparison obtained for "Lytro-5" in the Lytro dataset, with red and green demarcated areas zoomed in 5 times for easy observation

and it incorrectly detects some regions as unfocused regions. In addition, the fused images of CCF, DeepM^{2}CDL, TC-MoA, Mask-DiFuser, and GIFNet all suffer from some degree of distortion, which is mainly due to end-to-end methods losing more information about the source images during the fusion process. In contrast, our proposed ND-CNPFuse performs better in retaining the source image information and also shows superior performance in detail detection.

Fig. 3 presents the fusion results of various fusion methods on "MFI-WHU-120". From Fig. 3, we can see that all fusion algorithms accomplish the fusion task. However, CCF and DB-MFIF exhibit poor performance in capturing details, as indicated by the enlarged red and green regions, indicating their inability to detect focused areas accurately. In contrast, SAMF surpasses previous methods in detail but sacrifices some gradient information due to post-processing, leading to excessive smoothness between focused and unfocused boundaries. Furthermore, we can also conclude that there is significant distortion in the images fused by Mask-DiFuser and GIFNet. The remaining fusion methods yield satisfactory results, at least in terms of visual performance.

Fig. 4 illustrates the fusion results of various fusion methods on "Real-MFF-27". It can be seen that ND-CNPFuse and DMANet effectively minimize errors caused by the lack of precise alignment. Conversely, the fusion results of PADCDTNP and SAMF show biases in focus property detection near boundaries, leading to slight blurring visual effects. The results of DeepM^{2}CDL, Mask-DiFuser, and GIFNet contain obvious chromatic distortion. To sum up, we can see that our ND-CNPFuse outperforms other methods in most cases, showing good stability in MFIF. Our ND-CNPFuse maintains more competitive fusion performance among all the methods.

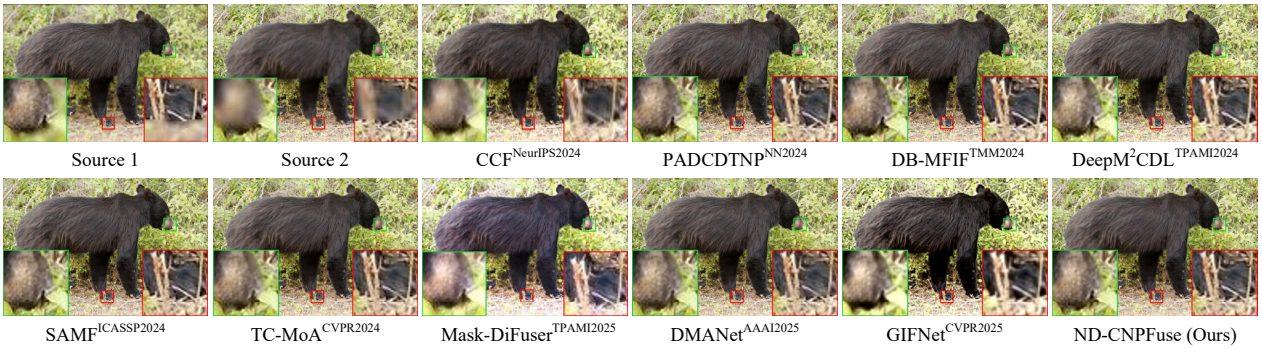

Figure 3. Qualitative comparison obtained for "MFI-WHU-120" in the MFI-WHU dataset, with red and green demarcated areas zoomed in 5 times for easy observation.

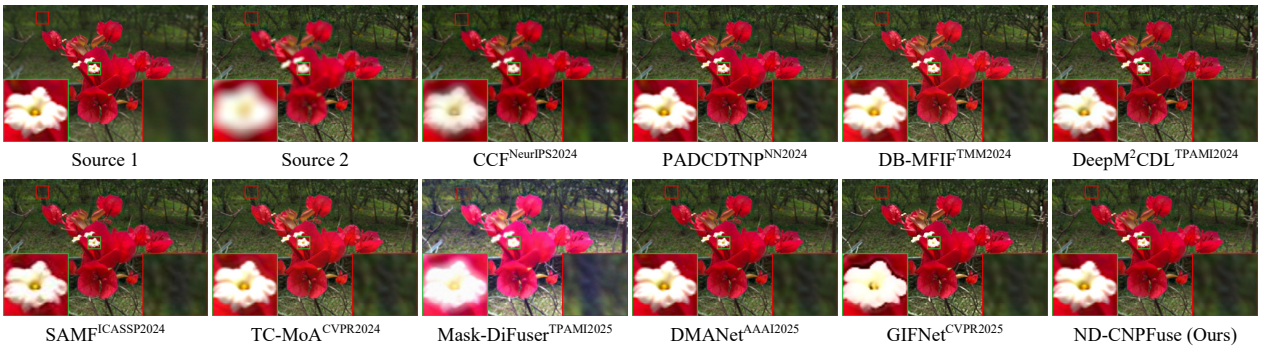

Figure 4. Qualitative comparison obtained for "Real-MFF-27" in the Real-MFF dataset, with red and green demarcated areas zoomed in 5 times for easy observation.

C.2. Downstream Salient Object Detection

Clear images adequately presenting scene information are essential for various computer vision tasks. To evaluate the effectiveness of MFIF in downstream tasks, we adopt a representative salient object detection algorithm, U^2Net [6], for the evaluation. In Fig. 5, we show the saliency maps obtained from fused images generated by different algorithms after U^2Net. We can find that: (1) The quality of the fused images directly affects the results of salient object detection. (2) Our fused images produced the best saliency maps, indicating that our ND-CNPFuse generates better-quality fused images.

C.3. Comparison with the LIF Model and Dedicated MFIF Methods

To further validate the superiority of our proposed spiking neuron ND-CNP model, we compared it with the classical spiking neuron LIF model [2]. Both were evaluated under an identical fusion pipeline, differing only in the neural model used. Tab. 1 summarizes the quantitative results of the two spiking neuron models across four datasets. We can see that the LIF model performs significantly worse than ND-CNP across all metrics. Specifically, averaged over the four datasets, the six metrics decline by 36.01%, 44.60%, 21.36%, 19.02%, 38.56%, and 28.34%. This can be attributed to the simplified biological properties of LIF, which ignore the neuronal coupling and dendritic structure of real neurons. This simplification limits the spatio-temporal dynamics of neurons, thereby reducing their feature expressive power. In contrast, our proposed ND-CNP model is more biologically plausible and exhibits richer dynamics.

We further evaluated our proposed ND-CNPFuse against three specialized MFIF methods (ZMFF [3], MDLSR-RFM [7], and MCCSR-Net [11]) across four datasets. As shown in Tab. 2, our proposed ND-CNPFuse remains competitive.

C.4. Analysis of SML Under Noisy Conditions

To further analyze the role of SML within the overall framework, we evaluated its performance under typical Gaussian noise inputs. Tab. 3 presents the experimental results. We can find that the decision maps generated after applying SML exhibit better accuracy, while the fusion image quality also improves, indicating that SML plays a positive role in the framework.

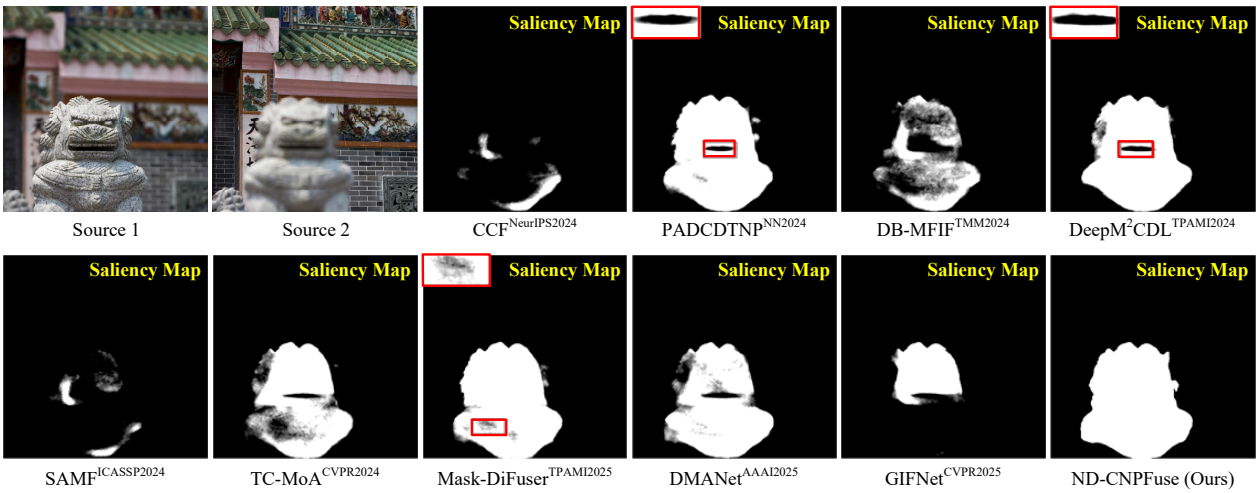

Figure 5. Salient object detection results for various fused images using U^2Net. **Since existing MFIF datasets lack annotation information (e.g., bounding boxes),** we cannot compute quantitative metrics (e.g., IoU or mAP). We adopt the same subjective evaluation as Y-Net [8] and MCCSR-Net [11] to assess result quality.

Table 1. Quantitative evaluation results of the LIF model and ND-CNP model on four MFIF datasets.

	Multi-focus Image Fusion on Lytro						Multi-focus Image Fusion on MFFW					
	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑
LIF model [2]	0.5707	0.4312	0.7005	22.0771	0.5611	0.6085	0.5409	0.3733	0.6304	20.1507	0.5304	0.5308
ND-CNP model (Ours)	**0.7621**	**0.5967**	**0.8541**	**26.9904**	**0.8466**	**0.8092**	**0.7399**	**0.5434**	**0.8362**	**23.7154**	**0.7294**	**0.7291**
	Multi-focus Image Fusion on MFI-WHU						**Multi-focus Image Fusion on Real-MFF**					
	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑
LIF model [2]	0.4697	0.4371	0.7309	22.8100	0.5334	0.6709	0.7009	0.4412	0.8255	29.8396	0.7503	0.7008
ND-CNP model (Ours)	**0.7346**	**0.6268**	**0.8400**	**27.7140**	**0.7611**	**0.8265**	**0.8207**	**0.6666**	**0.9569**	**34.2024**	**0.9237**	**0.8398**

Table 2. Quantitative results compared with three specialized MFIF methods. **Bold** and underlined indicate the highest and second-highest.

	Multi-focus Image Fusion on Lytro						Multi-focus Image Fusion on MFFW					
	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑
ZMFF [InfFus2023]	0.7031	0.5523	0.8395	26.7154	0.7853	0.7412	0.6655	0.3219	0.8148	23.6539	0.6469	0.6749
MDLSR-RFM [InfFus2024]	0.7621	0.5938	0.8404	26.3823	0.8460	0.8089	0.7306	0.5377	0.8199	23.6392	0.7171	0.7234
MCCSR-Net [InfFus2025]	**0.7623**	0.5960	0.8402	26.8821	0.8456	0.8084	0.6924	0.3546	0.8138	23.4489	0.6898	0.7233
ND-CNPFuse (Ours)	0.7621	**0.5967**	**0.8541**	**26.9904**	**0.8466**	**0.8092**	**0.7399**	**0.5434**	**0.8362**	**23.7154**	**0.7294**	**0.7291**
	Multi-focus Image Fusion on MFI-WHU						**Multi-focus Image Fusion on Real-MFF**					
	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑	Qabf↑	FMIw↑	SSIM↑	PSNR↑	QP↑	QCB↑
ZMFF [InfFusS2023]	0.6321	0.3811	0.8208	26.7501	0.6632	0.7093	0.7582	0.5081	0.9464	32.2025	0.8343	0.7567
MDLSR-RFM [InfFus2024]	0.7340	0.6290	0.8301	26.8097	0.7584	0.8261	0.8201	**0.6735**	0.9510	33.6161	0.9228	0.8363
MCCSR-Net [InfFus2025]	0.7300	**0.6315**	0.8301	26.8087	0.7586	0.8254	0.8205	0.6708	0.9510	33.8236	0.9232	0.8332
ND-CNPFuse (Ours)	**0.7346**	0.6268	**0.8400**	**27.7140**	**0.7611**	**0.8265**	**0.8207**	0.6666	**0.9569**	**34.2024**	**0.9237**	**0.8398**

Table 3. Noise generalization with/without SML on the MFFW dataset, evaluated on SSIM and PSNR metrics.

Configuration	Method	Gaussian noise (std $\sigma = 15$)	No noise
✓ SML	ND-CNPFuse (Ours)	0.830 (-0.7%) / 23.597 (-0.5%)	**0.836 / 23.715**
✗ SML	ND-CNPFuse (Ours)	0.814 (-2.2%) / 22.885 (-3.6%)	0.832 / 23.709

References

[1] Xiangyu Deng, Chunman Yan, and Yide Ma. Pcnn mechanism and its parameter settings. *IEEE Transactions on Neural Networks and Learning Systems*, 31(2):488–501, 2019. 3

[2] Wei Fang, Yanqi Chen, Jianhao Ding, Zhaofei Yu, Timothée Masquelier, Ding Chen, Liwei Huang, Huihui Zhou, Guoqi Li, and Yonghong Tian. Spikingjelly: An open-source machine learning infrastructure platform for spike-based intelligence. *Science Advances*, 9(40):eadi1480, 2023. 5, 6

[3] Xingyu Hu, Junjun Jiang, Xianming Liu, and Jiayi Ma. Zmff: Zero-shot multi-focus image fusion. *Information Fusion*, 92:127–138, 2023. 5

[4] Mansour Nejati, Shadrokh Samavi, and Shahram Shirani. Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion*, 25:72–84, 2015. 3

[5] Hong Peng and Jun Wang. Coupled neural P systems. *IEEE Transactions on Neural Networks and Learning Systems*, 30(6):1672–1682, 2019. 3

[6] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognition*, 106:107404, 2020. 5

[7] Jiwei Wang, Huaijing Qu, Zhisheng Zhang, and Ming Xie. New insights into multi-focus image fusion: A fusion method based on multi-dictionary linear sparse representation and region fusion model. *Information Fusion*, 105:102230, 2024. 5

[8] Zeyu Wang, Xiongfei Li, Libo Zhao, Haoran Duan, Shidong Wang, Hao Liu, and Xiaoli Zhang. When multi-focus image fusion networks meet traditional edge-preservation technology. *International Journal of Computer Vision*, 131(10):2529–2552, 2023. 6

[9] Hao Zhang, Zhuliang Le, Zhenfeng Shao, Han Xu, and Jiayi Ma. Mff-gan: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion. *Information Fusion*, 66:40–53, 2021. 3

[10] Juncheng Zhang, Qingmin Liao, Shaojun Liu, Haoyu Ma, Wenming Yang, and Jing-Hao Xue. Real-mff: A large realistic multi-focus image dataset with ground truth. *Pattern Recognition Letters*, 138:370–377, 2020. 3

[11] Kecheng Zheng, Juan Cheng, and Yu Liu. Unfolding coupled convolutional sparse representation for multi-focus image fusion. *Information Fusion*, 118:102974, 2025. 5, 6