

Ultra-Robust Imaging Restoration of Intrinsic Deterioration in Graded-Index Imaging Systems Enabled by Classified-Cascaded Convolutional Neural Networks—Supplementary Material

Zaipeng Duan,^{a,b,c,†} Yang Yang,^{a,b,†} Ruiqi Zhou,^{a,b} Jie Ma,^c Jiong Xiao,^{a,b} Zihang Liu,^{a,b} Feifei Hao,^{a,b} Jinwei Zeng,^{a,b,*} Jian Wang^{a,b,*}

^aWuhan National Laboratory for Optoelectronics and School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

^bOptics Valley Laboratory, Wuhan 430074, Hubei, China

^cNational Key Laboratory of Science and Technology on Multispectral Information Processing and school of artificial intelligence and automation, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[†]These authors contributed equally to this work.

* Corresponding authors: zengjinwei@hust.edu.cn, jwang@hust.edu.cn

1. Imaging principle of GRIN waveguide

Graded-index (GRIN) waveguide is a light-transmitting medium characterized by a graded refractive index [1]. In this paper, we have chosen the widely used GRIN lens as the central component of our experiments. Its refractive index exhibits a radial distribution proportional to the square of the radius, while it remains constant along the axial direction. Consequently, GRIN lenses can effectively transmit nearly axial meridional rays to achieve imaging. These rays follow a sinusoidal path and periodically converge, significantly reducing intermodal dispersion [2]. Here, we will provide a brief derivation of this phenomenon.

Here, we take the optical axis as the z-axis. Since only meridional rays are considered, their trajectory always lies in a plane passing through the optical axis. Due to symmetry considerations, we only need to consider the radial direction, denoted as the r-axis. Therefore, the refractive index distribution of the GRIN lens can be expressed as:

$$n(r, z) = n_0 \left(1 - \frac{A}{2} r^2 \right) \quad (S1)$$

Where n_0 is the refractive index at the center of the GRIN, and A is the refractive index distribution coefficient. This is based on the ray equation from the generalized law of refraction [3]:

$$\frac{d}{ds} \left(n \frac{d\vec{r}}{ds} \right) = \nabla n \quad (S2)$$

Where \vec{r} represents the coordinates of the ray trajectory, and it is given by:

$$\vec{r} = (r(z), z) \quad (S3)$$

To describe the trajectory of light within the GRIN lens, it is necessary to find the relationship between $r(z)$ and z . Here, we introduce the paraxial approximation:

$$ds \approx dz \quad (S4)$$

Substituting equation (S2) and simplifying accordingly, we obtain:

$$\nabla n = \frac{d}{ds} \left(n \frac{d\vec{r}}{ds} \right) \approx \frac{d}{dz} \left(n \frac{d\vec{r}}{dz} \right) = \frac{dn}{dz} \cdot \frac{d\vec{r}}{dz} + n \frac{d^2 \vec{r}}{dz^2} \quad (S5)$$

Where, as obtained from equation (S1):

$$\begin{cases} \nabla n = -n_0 A(r, 0) \\ \frac{dn}{dz} = 0 \end{cases} \quad (S6)$$

Substituting into equation (S5), we have [4]:

$$-n_0 A \cdot (r, 0) = n_0 \left(1 - \frac{A}{2} r^2 \right) \cdot \left(\frac{d^2 r}{dz^2}, 0 \right) \quad (S7)$$

Simplifying, we get:

$$\frac{d^2 r}{dz^2} + A \cdot r = 0 \quad (S8)$$

The above is the trajectory equation for the on-axis meridional ray in a GRIN lens. For a point source located at a distance D along the optical axis of the GRIN lens, with an outgoing ray at an angle θ , the initial conditions are given by:

$$\begin{cases} r(0) = D\theta \\ \frac{dr}{dz}(0) = \theta \end{cases} \quad (S9)$$

Substituting equation (S9) as initial conditions into equation (S8) and solving, we obtain the relationship between $r(z)$ and z as follows:

$$r(z) = D\theta \cdot \cos(\sqrt{A}z) + \frac{\theta}{\sqrt{A}} \cdot \sin(\sqrt{A}z) \quad (S10)$$

The above equation describes the trajectory of the on-axis meridional ray in a GRIN lens, taking the form of a sinusoidal curve. Clearly, this trajectory is dependent on the incident conditions and the length of the GRIN lens, denoted by L . The on-axis meridional ray will converge at a distance D' along the optical axis:

$$D' = -\frac{x(L)}{x'(L)} = \frac{D \cdot \cos(\sqrt{A}L) + \frac{1}{\sqrt{A}} \cdot \sin(\sqrt{A}L)}{D\sqrt{A} \cdot \sin(\sqrt{A}L) - \cos(\sqrt{A}L)} \quad (11)$$

In this case, D' is independent of the angle of the incident light, indicating that all on-axis meridional rays converge at the same point. This implies that a GRIN lens can achieve ideal imaging for on-axis meridional rays.

It is worth mentioning that Equation (S10) is a periodic function, and the length of one period corresponds to the pitch (P) value of the GRIN waveguide:

$$P = \frac{2\pi}{\sqrt{A}} \quad (S12)$$

The imaging behavior of a GRIN waveguide depends on the number of pitches it encompasses, thus being closely related to its length.

2. Aberration analysis of GRIN waveguide

As discussed earlier, GRIN lenses can effectively converge on-axis meridional rays to achieve imaging. Rays that do not satisfy this condition, including non-axial and non-meridional rays, introduce aberrations to the GRIN imaging system. As shown in Fig. S1, ZEMAX simulation results indicate that when the aperture of the GRIN lens is large, the trajectories of rays that do not satisfy the on-axis condition are compressed axially, meaning their P values are relatively small, resulting in convergence at different points along the axis. Moreover, due to the inability to provide analytical

solutions for the trajectory equation under these conditions, this leads to complex higher-order aberrations.

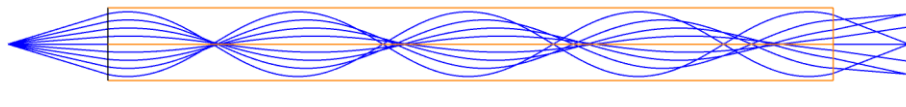


Fig. S1. Simulation of trajectories for non-axial rays in a GRIN lens.

It is evident that, under the influence of non-axial rays, the received image on the ideal image plane will be plagued. Furthermore, for non-meridional rays, their trajectories and aberrations are more complex, and further exploration can be found in the work of Merchand E. It's worth noting that the two mentioned aberrations are closely related to the length of the GRIN lens. Specifically, both aberrations accumulate as the light travels through the lens, meaning that the more pitches the GRIN lens contains, the greater the accumulated aberrations. Additionally, the distribution of aberrations is associated with the length of the lens, making it challenging to provide a more precise quantitative analysis.

Length mismatch in GRIN waveguides is another significant source of aberrations, introducing unpredictable distortions to imaging and even altering the reality of the images formed. Two main reasons contribute to the length mismatch in GRIN waveguides. Firstly, due to the presence of material dispersion, the refractive index of the GRIN waveguide is wavelength-dependent. Consequently, the effective length of the GRIN changes significantly when illuminated with different wavelengths or white light sources. The second source arises from mechanical errors and potential wear. While these factors can be controlled to a certain extent, they become non-negligible, especially in the case of small-aperture GRIN endoscopes.

The numerical aperture ($N.A.$) of a GRIN lens is solely dependent on the refractive index distribution and is expressed as follows [5]:

$$N.A. = \sqrt{n_0^2 - n^2(R)} \quad (S13)$$

Here, R represents the half-radius of the GRIN lens. GRIN endoscopes often demand a smaller aperture and a larger $N.A.$ Considering equation (1), reducing the value of R and increasing the $N.A.$ requires a larger refractive index constant, denoted as A in equation (S12). Consequently, this leads to a decrease in the P value of the GRIN lens [6]. Therefore, minor length mismatches result in more variations in the number of pitches, amplifying the impact of length mismatch in GRIN endoscopes.

3. More details about the cascaded network

Here are the specific details of the cascaded network. In Fig. S2(a), the improvement of ResNeSt50 [7] over ResNet is illustrated, namely, the Split-Attention mechanism. Split-Attention is a distinctive feature of the ResNeSt model, designed to enhance the network's ability to learn interrelations among features. In this mechanism, the network segments channels into several groups, treating each group as a distinct subset. Within each group, an attention mechanism is implemented, enabling the network to assign varying weights to different groups of channels. This process allows for more selective capturing of interrelations among input features. By integrating channel grouping with attention mechanisms, "Split-Attention" promotes interaction between various channel groups, empowering the network to amalgamate information from diverse channel subsets and thereby gain a deeper understanding of inter-feature relationships.

Moreover, this design mitigates channel competition, fostering a more synergistic approach to processing information from different channels, as opposed to having them compete during the learning process. In essence, the innovation of "Split-Attention" lies in its incorporation of channel

grouping and attention mechanisms, significantly boosting the network's capacity to discern the interrelation of input features, which in turn enhances overall model performance. This attribute is especially advantageous for visual tasks, as it allows the model to more flexibly and effectively learn complex relationships among different channels in images. In Fig. S2(b), our enhancement to the U-Net model [8], namely the spatial attention mechanism. The channel attention mechanism in SENet (Squeeze-and-Excitation Network) is operationalized through the integration of the Squeeze-and-Excitation (SE) module. This process begins with the module aggregating global information for features on each channel via global average pooling.

Following this, a compact fully connected network is deployed to learn and assign weights to each channel based on this globally aggregated information, thereby highlighting the significance of individual channels. Once these channel-specific weights are determined, they are applied to modulate the original features, resulting in a feature representation that prioritizes more critical channels. This channel attention mechanism empowers the SENet to dynamically modify the weights of each channel during the learning process, significantly enhancing the network's capacity to discern features across different channels, which ultimately leads to improved model performance. Such a design not only augments the flexibility of the network but also enables it to more adeptly conform to the characteristics of the input data.

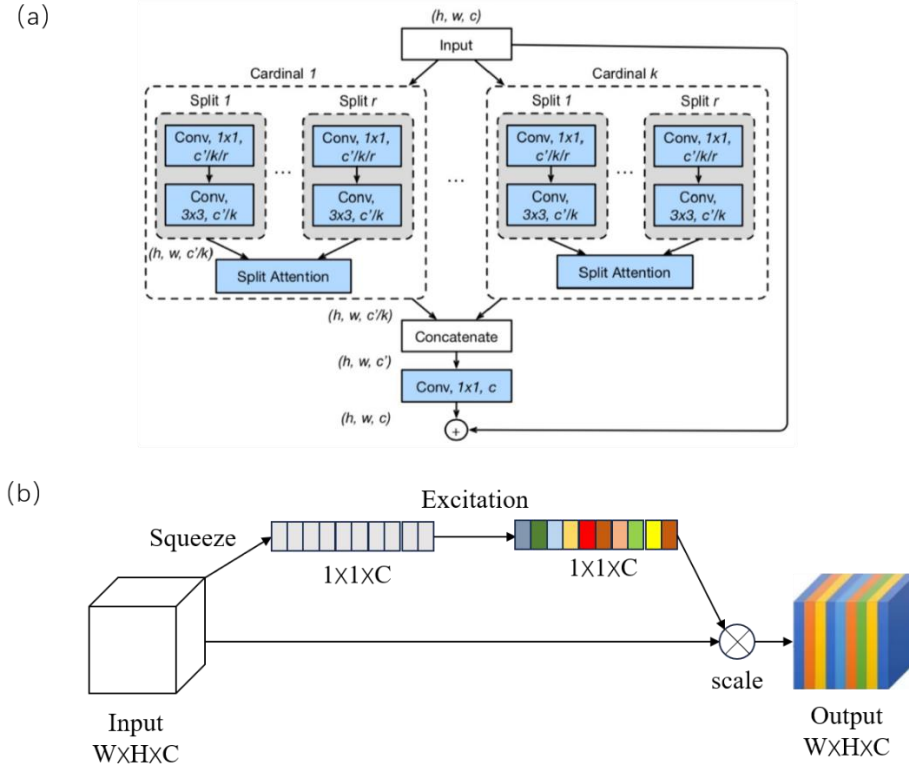














































Fig. S2 (a) The Structure of the Split-Attention mechanism. (b) The Structure of the spatial attention mechanism.

4. More CNN recovery results

Here, we provide additional restoration results below shown as Fig. S3. The conclusions align with the descriptions in the main text, as both methods exhibit comparable performance in restoring real images, while the single CNN performs significantly worse in restoring virtual images compared

to the classified-cascaded CNN. Therefore, Fig. S4 presents more comparative images of their performance in restoring virtual images.

Target	Lens1	Lens2	Lens3	Lens4	Lens5	Lens6	Lens7	
								Original Imaging
Single CNN								Restoration
PSNR	33.5635	33.8275	31.9867	29.7746	23.9905	24.0534	30.5639	
SSIM	0.9279	0.9469	0.9061	0.9020	0.7867	0.7783	0.9214	
Classified-Cascaded CNN								Restoration
PSNR	33.4530	33.3431	31.7792	30.6875	31.9179	33.2614	31.8737	
SSIM	0.9292	0.9372	0.9076	0.8969	0.9144	0.9240	0.9235	

Target	Lens1	Lens2	Lens3	Lens4	Lens5	Lens6	Lens7	
								Original Imaging
Single CNN								Restoration
PSNR	31.4842	34.1131	32.5029	28.3451	26.9041	26.7722	31.9595	
SSIM	0.8938	0.9507	0.9052	0.8775	0.7612	0.7487	0.9066	
Classified-Cascaded CNN								Restoration
PSNR	32.0665	34.4218	33.0502	29.0521	30.1930	34.4997	33.3323	
SSIM	0.9191	0.9523	0.9063	0.8694	0.9113	0.9199	0.9205	























Target	Lens1	Lens2	Lens3	Lens4	Lens5	Lens6	Lens7	
								Original Imaging
Single CNN								Restoration
PSNR	29.9991	33.4437	28.1345	30.4139	19.4012	19.1965	30.7949	
SSIM	0.8950	0.9438	0.8459	0.9003	0.5604	0.5564	0.9042	
Classified-Cascaded CNN								Restoration
PSNR	30.3677	33.6877	28.7517	30.2747	32.7352	31.4675	31.3395	
SSIM	0.9095	0.9493	0.8622	0.9116	0.9441	0.9215	0.9135	

Fig. S3. Restoration results for imaging random-length GRIN waveguides using Single CNN and Classified-Cascaded CNN.





























Target							
Original							
Single CNN							
PSNR	20.7806	22.9151	24.3818	22.1089	21.9757	23.8161	23.5721
SSIM	0.6393	0.7429	0.7429	0.6343	0.5027	0.6321	0.5998
Classified-Cascaded CNN							
PSNR	31.4090	31.6373	28.1345	30.9251	30.5936	32.7193	32.6945
SSIM	0.8385	0.8552	0.8552	0.8117	0.8627	0.8602	0.8477

Fig. S4. Comparison of restoration results for virtual images using Single CNN and Classified-Cascaded CNN.

References

1. Stewart G . “Optical waveguide theory”. Chapman and Hall, 1983.
2. Butylkin V S , Shalyaev M F . “Excitation of stimulated Raman scattering in graded-index fiber waveguides by an arbitrary Gaussian beam”. Soviet Journal of Quantum Electronics, 1982, 12(11):1505-1507.
3. Ghatak, Bak , and K. Thyagarajan . “Contemporary Optics”. Plenum*, 1978.
4. C Gómez-Reino, Llares, Jesús. “Optical path integrals in gradient-index media”. Journal of the Optical Society of America A, 1987, 4(8):1337-1341.
5. Acosta E , Gomez-Reino C , Llares J . “Effective radius and numerical aperture of GRIN lenses with revolution symmetry”. Applied Optics, 1987, 26(15):2952-2955.

6. Khular E , Kumar A , Ghatak A K , et al. "Effect of the refractive index dip on the propagation characteristics of step index and graded index fibers". *Optics Communications*, 1977, 23(2):263-267.
7. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE, 2016), pp. 770–778.
8. N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds., *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*, Lecture Notes in Computer Science (Springer International Publishing, 2015), Vol. 9351.