

Deep learning-based imaging using single-lens and multi-aperture diffractive optical systems

A. Nikonorov^{1,2}, V. Evdokimova^{1,2}, M. Petrov^{1,2}, P. Yakimov¹,
S. Bibikov¹, Y. Yuzifovich¹, R. Skidanov^{1,2}, N. Kazanskiy^{1,2}

¹ Image Processing Systems Institute, Russian Academy of Sciences,
² Samara National Research University

artniko@gmail.com kazansky@smr.ru

Abstract

The pressure to reduce weight and improve image quality of the imaging devices continues to push research in the area of flat optics with computational image reconstruction. This paper presents a new end-to-end framework applying two convolutional neural networks (CNNs) to reconstruct images captured with multilevel diffractive lenses (MDLs). We show that the patch-wise chromatic blur and image-wise context-aware color highlights, the distortions inherent to MDLs, can be successfully addressed with the suggested reconstruction pipeline. The generative adversarial network (GAN) is first used to remove image-wise color distortion, while a patch-wise network is then used to apply chromatic deblur. The proposed approach produces better image quality improvement than the context-independent color correction with a deconvolution-based chromatic deblur. We also show that the proposed end-to-end reconstruction is equally applicable for single- and multi-aperture MDL-based imaging systems.

1. Introduction

Miniature cameras have always been an object of interest for researchers and consumers [1], [5], [6], [16]. For conventional optics, we need a thicker lens for larger numerical aperture (or resolution) [1]. To compensate the optical aberration inherent to refractive optics, a complex

lens system must be used [2]. To simplify construction, imaging systems based on a single flat lens have recently attracted attention of researchers and the industry. The metalenses are the most popular recent approach to designing flat optics, making it possible to create an optical system with sub-wavelength thickness [3]. While the multilevel diffractive lenses are well known as optical elements since the 19th century, their use for color imaging purposes only started in 2015 [4]. Although metalenses can be made thinner than an MDL, this advantage is so negligible that benefits of diffractive optics (easier to produce, better imaging quality), combined with overall competitive thickness, makes it ultimately a better choice.

Chromatic aberration is a major problem for both MDLs and metalenses. Longitudinal chromatic aberration for flat optics is significantly higher than for conventional systems. Detailed calculations for a typical MDL are shown in section 3, while more information about longitudinal chromatic aberration in metalenses can be found in [1]. For high-chromatism systems, two types of distortions occur: patch-wise image distortions caused by chromatic defocusing and image-wise color distortions caused by the redistribution of energy between diffraction orders.

To remove these distortions and obtain images with acceptable quality, computational post-processing must be applied. In several papers, deconvolution-based approach to compensate chromatic aberrations is described [4], [6], [7]. While papers [6], [7] deal with the chromatic aberration only, [4] and [5] address an additional color correction problem – removing the color shift. Color distortions make

Single frame/First frame of multi-frame

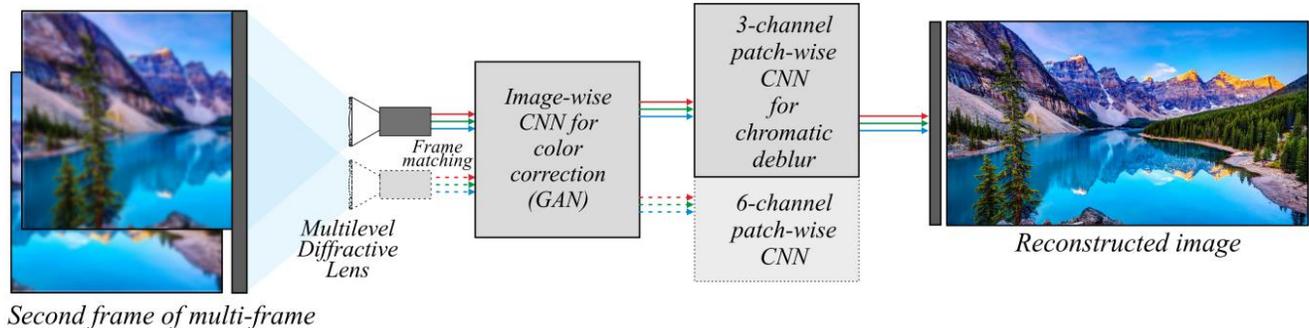


Figure 1: End-to-end deep learning-based image reconstruction pipeline for single and dual MDL.

a significant impact on the quality of an acquired image [4], [5]. This paper shows that for systems with high chromatic aberration, the problem of color distortions is context-sensitive.

The superiority of flat over conventional optics can be even more pronounced in multi-aperture systems. The use of these systems can reduce the dimensions of the optical system [8], [17]. However, for conventional optics, the multi-aperture configuration significantly complicates the structure and increases its weight. Even though theoretical analysis of the performance of several lenses at different wavelengths was performed in [8], one of the first studies of multi-aperture systems based on diffractive optics was carried out later, in [10].

This paper proposes an end-to-end neural network approach to the reconstruction of images captured with systems based on diffractive optics. We propose a two-stage neural network design and show that it successfully solves both the patch-wise chromatic blur problem and image-wise context-aware color distortions. We also show that our reconstruction approach can be applied to multi-aperture systems. A binocular MDL system that we developed and tested showed better quality of the reconstructed image than the monocular one could generate. Figure 1 shows the proposed architecture with two networks: the image-wise network for color correction and the patch-wise network for chromatic deblur.

We benchmark our image reconstruction using PSNR for real images in the same way as in [5], [25].

2. Related works

Metalenses. Recent advances in creating imaging systems with flat optics [3], [1] are built as planar optical components with a single layer of phase shifting nanostructures. To demonstrate planar chromatic-aberration-corrected lenses fabricated from metasurfaces, paper [9] exploits wavelength-dependent phase shift to overcome light dispersion. Alternative ways to create planar optics are reviewed in [3], and [1] contains a comprehensive discussion comparing flat lenses based on diffractive optics and metalenses. Using metasurfaces, planar optics can potentially displace conventional refractive optics in many applications. However, this approach faces multiple difficulties: from theoretical modeling of elements to the design of building-block scattering elements to scaling the fabrication [1].

Diffractive lenses. As shown in [1], a lightweight optical system with a diffractive lens is preferable to a metalens from a practical point of view. Manufacturing of diffractive lens and their use in imaging systems is a subject of several works starting from 2015. While [26] showed one of the first imaging applications of the binary Fresnel lens in a self-powered camera, paper [4] made the idea of using diffractive lens for high resolution imaging mainstream,

with its multilayer Fresnel lens coupled with computational post-processing. Computational reconstruction is necessary to compensate strong chromatic aberration and other distortions affecting images captured by a camera equipped with a single diffractive lens or another similar planar optical element. Papers [6, 7], [25] describe different MDL designs and reconstruction approaches.

Image distortions and the reconstruction. The problem of strong image distortions must be addressed in any attempt to create planar optical elements. In papers reviewed in [3], authors used only optical solutions to eliminate the distortions. In the so-called computational approach, one combines the design of a diffractive lens with the computational reconstruction to compensate the image distortions [4-7]. The main distortion component in diffractive systems is the chromatic aberration, which makes the PSF (point spread function) of the sharpest color channel in the captured image narrower than for two other channels [4], [6]. Typically, the sharpest channel is green and most of reconstruction approaches use this cross-channel prior [5, 6]. The second distortion source is the energy redistribution between diffraction orders. In the next section, we show that this redistribution leads to context-aware distortions in image colors.

Deep learning-based image reconstruction. The first use of DL networks for solving the single image super resolution (SISR) involved the SRCNN [11] network. Then, the VDSR [12] and DRNN [13] networks showed higher image quality. The CNN multi-frame super resolution can significantly improve the quality of the reconstructed image [14]. In [5], it is shown that VDSR, a variation of SISR CNN, can result in a higher quality that exceeds the capabilities of deconvolution-based methods described in papers [4], [6], [7]. However, an end-to-end neural network - based solution was not obtained in [5], [28]; instead, a combination of CNN with deconvolution and color correction based on 3D LUT was used. While this CNN-based approach for removing chromatic blur resulted in better quality than the approach based on the deconvolution, [5] still used a transitional approach that combined neural network with classical correction methods, stopping short of creating a complete end-to-end technique based on neural networks.

In this paper, we present an end-to-end deep learning - based solution for image reconstruction. This end-to-end CNN chain addresses both the local patch-wise chromatic blur and image-wise context-aware color distortions.

Multi-aperture systems. For imaging optical systems, there is a trade-off between viewing angle and linear resolution of the system. Wide-angle systems provide low linear resolution, and long-focus high-resolution systems have a small viewing angle. Recently, a number of multi-sensor and multi-lens systems have emerged, designed to address this tradeoff. Examples of these systems include Dallmeyer Panomera multifocal matrix camera for the video surveillance

market, the patented Nikon four-lens camera [15], and the Light L16 camera [16]. According to the published specifications, Light L16 provides 50-megapixel resolution in combination with a great light sensitivity due to the fact, that it uses 16 sensors. Multi-aperture configurations are used to increase the viewing angle while keeping the spatial resolution. This approach results in super-resolution across several images [17], [18], as well as help with building different-focus systems and post-focusing systems [18]. Despite promising capabilities of these systems, high lens mass of multi-aperture systems remains their significant limitation. Flat diffractive optics - based imaging systems, on the other hand, do not suffer from a weight issue [10]. Flat optics can significantly reduce the weight and the cost of a long-focal distance optical system [5]. Combined with the multi-frame super-resolution [14] approach, image quality can be significantly higher.

3. Optical schematics and image acquisition

This section describes monocular and binocular diffractive-optical systems we built. Figure 2 shows an example of a diffractive lens with a micro-relief on the surface that can replace a set of refractive lenses.

3.1. Single MDL imaging system

For a pure diffractive lens, the focal length shifts almost linearly across the wavelengths and the longitudinal chromatism is comparable to the focal length itself. Using this lens is challenging because only one wavelength is in focus. A more promising approach is the use of so-called harmonic lenses [24], the microrelief height of which is several times higher than the height of the diffraction lens microrelief. With this approach, several wavelengths emerge for which a sharp image can appear in the focal plane.

Let us consider the way the images are formed with a single harmonic lens. A focal length of this lens varies with the wavelength according to the following equation [24]:

$$f(\lambda, k) = \frac{m\lambda_0 f_0}{\lambda k}, \quad (1)$$

where m is the multiple of the microrelief height, λ_0 is the base wavelength, λ is the wavelength we estimate the focal length for, k is the number of the current harmonic (for the

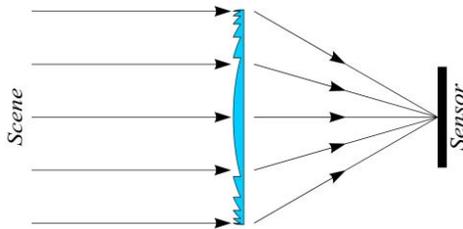


Figure 2: Scheme of an MDL flat lens.

wavelength λ), f_0 is the focal length for the base wavelength λ_0 . The diffraction efficiency of the lens for a given wavelength is described by the equation [24]:

$$e(\lambda, k) = \text{sinc}^2 \left(\frac{m\lambda_0 f_0}{\lambda} - k \right). \quad (2)$$

According to these harmonic lens properties, we can always choose several wavelengths that have exactly the same focal length. Figure 3a shows the dependency graphs of intensity in focus on the z -axis (along the lens optical axis) for a harmonic lens with the following parameters: $m = 10$, $\lambda_0 = 0.65 \mu\text{m}$, $f_0 = 100 \text{ mm}$. The calculation was performed in the interval between $\lambda_{\min} = 0.38 \mu\text{m}$ and $\lambda_{\max} = 0.74 \mu\text{m}$. According to the calculation results, three wavelengths have the same focal length. Figure 3 shows the graphs for $\lambda_1 = 0.406 \mu\text{m}$ – blue, $\lambda_2 = 0.500 \mu\text{m}$ – green, $\lambda_3 = 0.650 \mu\text{m}$ – black. For all these wavelengths, the focal length was $f = 100.15 \text{ mm}$.

However, for other wavelengths of the visible spectrum, the focal point will shift rather widely. The relative change in the optical power D of the harmonic lens will approximately correspond to the equation:

$$\frac{\Delta D}{D} \approx \frac{1}{m}. \quad (3)$$

The change in focal length will also decrease approximately in proportion to f/m . The maximum width of the PSF, W in this case could be determined as follows:

$$W = \frac{\Delta f}{2k}. \quad (4)$$

For the lens described above with $k = 10$, we obtain PSF width $W = 0.5 \text{ mm}$. But at this width the intensity is very low largely creating only background illumination highlight.

Consider the examples of the position and the intensity of focus for different wavelengths. Figure 4a shows the calculated results for a relatively narrow wavelength range. The smooth blue and red curves are envelopes for the intensity level in focus for the 10th and 9th harmonics, respectively. Figure 4b presents experimentally measured results for a 500 nm wavelength.

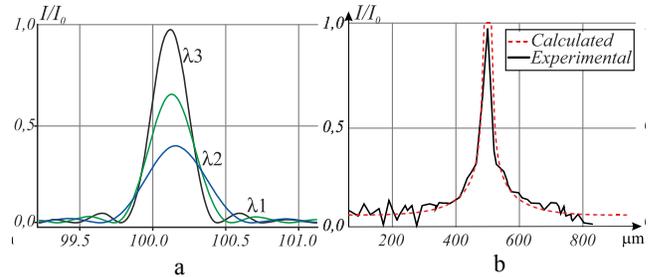


Figure 3: (a) – Focusing intensity plot on the z -axis for a harmonic lens with $m = 10$, $\lambda_0 = 0.65 \mu\text{m}$, $f_0 = 100 \text{ mm}$, $\lambda_1 = 0.406 \mu\text{m}$ – blue, $\lambda_2 = 0.500 \mu\text{m}$ – green, $\lambda_3 = 0.650 \mu\text{m}$ – black. (b) – MDL PSF for white light calculated by the (4) and experimentally estimated (b) – MDL PSF for white light calculated by the (4) and experimentally estimated

As can be seen from Figure 4a, the positions of the foci for a single harmonic lens do vary quite significantly in the range from 90 to 110 mm. Moreover, some wavelengths are represented by two foci. The wavelength of $0.65\mu\text{m}$ falls on the envelope for $k = 10$ (blue line). Wavelengths falling on the envelope for $k = 9$ (red line) have a lower intensity. The magnitude of the longitudinal chromatic aberration for such a lens is about 10 mm. For lens diameter 10mm Δf of 10mm leads to PSF width of about 1mm, or 200 pixels for a 5mkm pitch. However, this large spot will have very low intensity, and will produce only background highlight. Figure 3b shows the calculated PSF for diffractive lens with a diameter of 10 mm and a focal length of 100 mm, and an experimentally measured PSF for the same lens. Figure 3b shows, that a wide highlight spot is present, while PSF width at half maximum is about $30\mu\text{m}$. The color of the described highlights depends on the image context, representing context-aware color distortions, included in general image acquisition model, described in section 3.3.

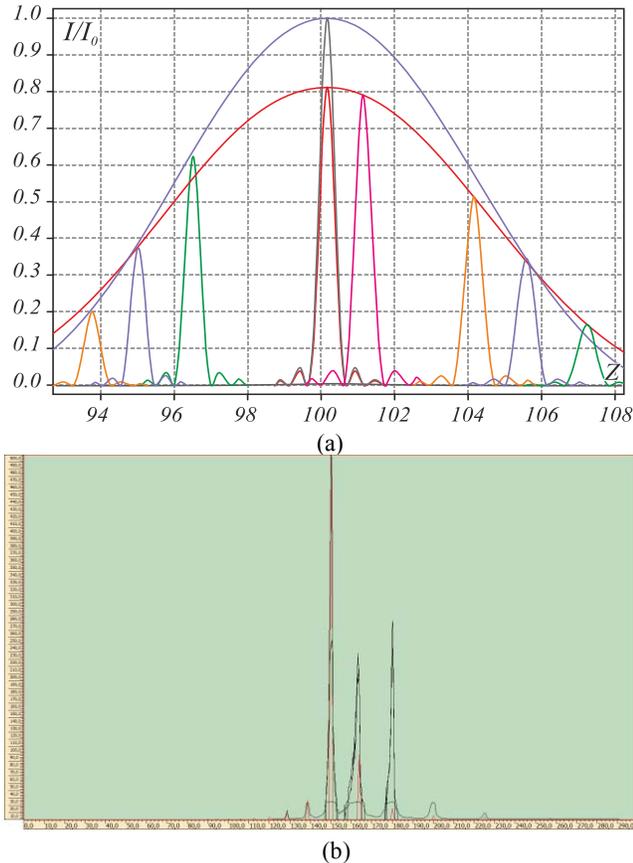


Figure 4: (a) The dependency graphs of intensity in focus on the z coordinate for a harmonic lens with the following calculation parameters: $m = 10$, $\lambda_0 = 0.65\mu\text{m}$, $f_0 = 100\text{mm}$, $\lambda_1 = 0.65\mu\text{m}$ – black, $\lambda_2 = 0.675\mu\text{m}$ – green, $\lambda_3 = 0.685\mu\text{m}$ – blue, $\lambda_4 = 0.695\mu\text{m}$ – orange, $\lambda_5 = 0.715\mu\text{m}$ – purple; (b) experimental intensity distributions along the z axis (solid line) versus calculated (dashed line) for a wavelength of 500 nm.

3.2. A multi-aperture system

While described above single-lens system significantly outperforms classical optics in terms of its weight, it is significantly inferior in image quality due to distortions. The use of multi-lens configurations can improve the image quality. In [2], a method is shown to improve the image quality by using MDLs with different base wavelengths and focusing in the red, green and blue parts of the spectrum. In this paper, we show that the use of several lenses of the same spectral range can also significantly increase the resolution of the system.

Figure 5 shows an optical system with two harmonic lenses combined into one system, with each forming its own image on each individual photosensitive matrix. The presence of several images taken from different points will increase final image resolution. There is also an opportunity to produce a multispectral camera using the same principle, if it is formed from harmonic lenses that focus different base wavelengths.

Technical simplicity of manufacturing of an array of harmonic lenses allows to significantly increase the number of simultaneously capturing matrices in the system and significantly improve the resolution without increasing the focal length.

3.3. Image acquisition by diffractive lens

Images taken with harmonic lenses suffer from special types of image distortions because its PSF strongly depends on the wavelength, leading to strong chromatic aberration. As shown in 3.1, for some wavelengths the problem manifests itself as chromatic blur, applied locally in each pixel. For other wavelengths it may lead to color highlights, applied to a large part of the image. The area of the distortion is determined by the PSF width (4), and its intensity is determined by the efficiency expression (2). These distortions are inherent to various optical systems with a strong longitudinal chromatic aberration. Thus, image acquisition could be formulated as follows:

$$p_{RGB}(\mathbf{x}) = \mathbf{B}_{RGB} \otimes p_{RGB}^0(\mathbf{x}) + H(p_{RGB}^0, p_{RGB}^0(\mathbf{x}), \mathbf{x}) + \eta, \quad (5)$$

where $p_{RGB}(\mathbf{x})$ is one of the red, green, or blue color channels of the image acquired with a diffractive lens, $p_{RGB}^0(\mathbf{x})$ is original image, \mathbf{B}_{RGB} is the PSF for each color channel; $H(p_{RGB}^0, p_{RGB}^0(\mathbf{x}), \mathbf{x})$ is the context-aware color distortion, and η is additive noise. As shown in Section 3.1, the context-aware color distortion $H(p_{RGB}^0, p_{RGB}^0(\mathbf{x}), \mathbf{x})$ is specific for MDL and depends not only on image color in given point $p_{RGB}^0(\mathbf{x})$, but on the entire image p_{RGB}^0 , so an overall image context affects the additional color distortion in any given point, which is like light color flare slightly changing over the entire image.

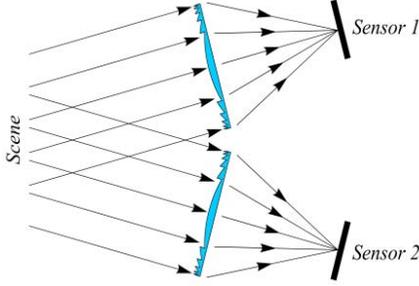


Figure 5: Multi-aperture optical scheme.

Previous works on image reconstruction in MDL use the simplification of (5). Works [7], [8] consider only first blurring term of (5). Color correction of color highlights is considered in [4], [5]. However, these papers use constant model of color distortions and use context-independent 3D LUT color correction [5]. Here, we propose two CNNs: one for context-aware color correction and another one for local correction of chromatic blur.

4. Deep learning-based image reconstruction for diffractive optical systems

In this work, we use a generative adversarial network (GAN) to address image-wise context-aware color distortion, the second term in (5). After applying an image-wise CNN, we use a patch-wise modified single image super-resolution CNN, which removes the chromatic blur, the first term in (5). Our image reconstruction pipeline of two sequentially connected neural networks is shown in Figure 1.

4.1. Image-wise color correction

It is imperative to address context-aware distortion (5) in all optical systems that have a high level of chromatic aberrations. In early works, the conventional, context-independent, techniques of the color correction were used. However, the experiments described in section 6 show that conventional methods are not good enough for color correction of images captured by MDL. We use GAN to address the context-aware distortions. GAN CNNs have been shown to successfully solve problems similar to the considered image reconstruction: a solution of removing distortions for underwater photography is proposed in [19]. Paper [20] describes successful cases of applying GAN to the white point estimation, which is close to our problem of context-aware color correction. As an adaptive filter for removing of context-aware distortions inherent to MDL, we use the GAN image-to-image translation with adaptive loss function, implemented by the GAN’s discriminator part.

The GAN architecture consists of a generator G and a discriminator D . Figure 6 illustrates GAN architecture, built to remove context-aware color distortions. The generator takes a captured three-channel RGB image with the color

distortion as input and generates a color corrected image Y . Then, the discriminator D classifies whether Y is a ground-truth image, or it was generated by G . Finally, an output of D is passed to G to learn to the desired color correction.

In this work, we use Pix2Pix [21] architecture of GAN for image-wise color correction task. This architecture is based on U-Net [22] as the generator and PatchGAN [27] as the discriminator.

4.2. Patch-wise chromatic deblur

In this paper, we considered two patch-wise networks to remove chromatic blur for images captured by a single-lens optical system: the 3-channel modification of VDSR [12] and DRRN [13]. In [5], a VDSR-inspired CNN with grey-edge penalty was successfully applied to solving the same task, but it took only one channel as an input. The 3-channel modification of VDSR takes an entire distorted RGB image as an input and outputs an RGB deblurred image.

In contrast to VDSR, DRRN has both residual and recursive blocks. We utilize DRRN architecture with 9 residual units and 1 recursive block, creating a 20 layers-deep DRRN. The mean squared error loss function is used for DRRN training. For both VDSR and DRRN we use a parametric rectified linear unit (PReLU) activation function.

For both CNNs we used the loss function with cross-channel prior:

$$L(\hat{p}, p^0, p_G^0, \mathbf{w}) = \frac{1}{N} \sum_{j=1}^N d(\hat{p}^{(j)} - p^{0(j)}) + \lambda \frac{1}{N} \sum_{j=1}^N \|p_G^{0(j)} \nabla \hat{p}^{(j)} - \hat{p}^{(j)} \nabla p_G^{0(j)}\|_2 + \eta \|\mathbf{w}\|_2, \quad (6)$$

where N is the number of images in one training minibatch, j is the index of an image in the minibatch, $p^{0(j)}$ is the ground truth RGB image, $\hat{p}^{(j)}$ is the reconstructed RGB image, $p_G^{0(j)}$ is the green channel of the reference image, \mathbf{w} is weights of CNN. The first term of the loss function minimizes the distance (6) between the ground truth image and the network output image. The second term implements

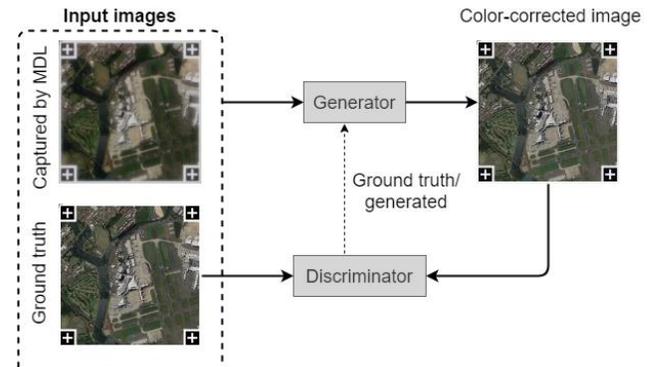


Figure 6: GAN-based correction of context-aware color highlight

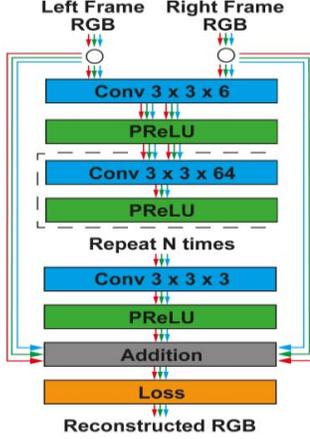


Figure 7: 6-channel patch-wise CNN for multi-aperture imaging the penalty for deviation of the gradient from the green channel of the ground truth image. The last term applies constraints on the network weights, following the weight decay rule.

4.3. Image reconstruction for dual-lens system

For a dual-lens system, we use a combination of the Pix2Pix GAN for color correction and a 6-channel modification of VDSR, which takes two blurred RGB images and outputs a reconstructed RGB image. Figure 7 illustrates the architecture of our 6-channel VDSR.

In the last layer of the VDSR (the addition layer), we add the output of the previous layer to a linear combination of input images:

$$\begin{aligned} \text{Addition} &= Y_{prev} + X_{in}, \\ X_{in} &= (1-\gamma)X_{left} + \gamma X_{right}, \end{aligned} \quad (7)$$

where Y_{prev} is the output of the previous layer, X_{left} is the RGB image produced by the left lens, X_{right} is the RGB image produced by the right lens, γ is the training parameter.

5. Implementation and training

5.1. Optical setup and image capturing

We have designed two optical systems. The first one is a single-lens MDL with a focal length of 150 mm and a lens diameter of 50 mm. The fabrication technique is based on direct laser writing [29] and is similar to the one described in [5]. The second one is a multi-aperture system that has two identical lenses. These two lenses were fixed at 40 cm distance between them and captured the same scene. In both cases, the lens focusing is adjusted manually. In both cases, for training and validation we captured images projected on the screen by 3LCD Epson projector at a distance of 6 meters.

The images acquired by the camera with diffractive lens have projective distortions. To measure the quality of the implemented reconstruction, we have to align the captured images with the corresponding originals. First, we add cross-like markers on the corners of original image before capturing it by a diffractive lens. After threshold processing, the center points of these markers are calculated both in original and captured images using contour analysis and template matching. Applying inverse projective transform, we obtain matched captured images. To make this transformation more accurate, we use iteration matching with PSNR minimization between matched images following the approach described in [25].

5.2. The dataset

We assembled two datasets using single-lens and dual-lens MDL systems. The first single-lens dataset has 350 pairs of diffraction and ground truth images of the size 1024×1024 pixels. The second dual-lens dataset includes 148 1024×1024 diffraction images captured by the left and the right lenses and corresponding ground truth images.

The datasets were then randomly divided into training and testing samples. For single-lens and dual-lens datasets, the training samples include 321 and 119 images respectively, and 29 images were used for testing in both cases. To train the image-wise network, we combined left and right training images for a dual-lens system. In this case, the training dataset includes 642 images. The datasets are available online at:

<https://1drv.ms/f/s!AiQGcNBKHALepTCPCRIEasXf92cO>

5.3. Data preparation

To train patch-wise networks, training images are split into 51×51 patches with the stride of 51. In addition, data augmentation is performed on the training images. In particular, we rotate the original images by 90° , 180° , 360° , flip them and add Gaussian noise, creating 15 additional augmented versions for each original patch.

For a dual-lens system, we perform image matching and perspective distortion correction during preprocessing. We use the left image as base and transform the right one to match it. We combined two RGB images into a 6-channel image before training and testing stages, as shown in Figure 7.

5.4. Training the networks

Pix2pix GAN training is performed with full-size training images using minibatch stochastic gradient descent and the Adam solver [23], with a learning rate of 0.0002, and Adam momentum parameters set to 0.5 and 0.999.

To train DRRN, we used the Adam optimizer and set the mini-batch size to 32, learning rate to 10^{-4} , weight decay to 10^{-6} . The gradients are clipped to $[-\theta/\gamma, \theta/\gamma]$, where γ is the current learning rate and $\theta = 0.01$ is the gradient clipping parameter.

For VDSR training, we also utilized the Adam optimizer and set the mini-batch size to 64. The network depth, initial learning rate, weight decay, and edge penalty are set to 18, 10^{-4} , 10^{-4} , and 10^{-2} , respectively. The learning rate is halved every 10 epochs.

6. Results

In this section, we compare the results of our reconstruction pipeline with the deconvolution-based reconstruction [25]. In previous papers authors used synthesized images [7] or cropped the central part of an image to estimate the reconstruction quality [5]. This simplification was done due to complex nature of image-wise distortions, generating too low PSNR values for entire images. Proposed in this paper end-to-end approach correctly processes image-wise distortions, finally allowing us to estimate PSNR values for the entire RGB image.

Table 1 shows PSNR values for our reconstruction technique against the values obtained with deconvolution-based reconstruction from [6], [25]. More details and comparison of deconvolution-based reconstruction are provided in [5] and [25]. We included mean values for input distorted images, color corrected images, and reconstructed images after removing chromatic blur. For validation, we averaged mean values across a diverse set of 29 images, which includes natural scenes and remote sensing images, as shown in Figure 9.

For a single-lens system, the combination of Pix2Pix and 3-channel VDSR cannot adequately deblur the image. Therefore, we propose to use DRRN, which achieved the best

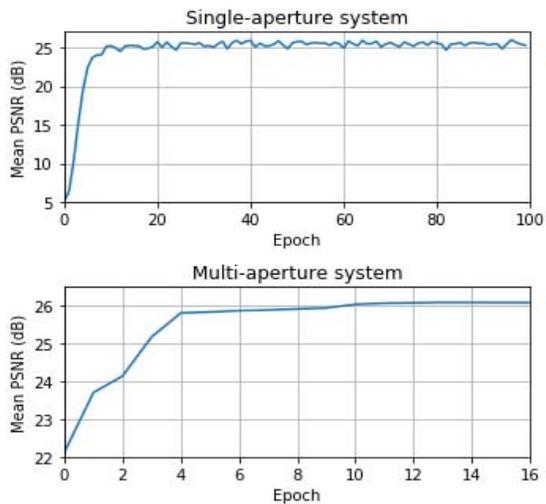


Figure 8: Validation for patch-wise CNN.

performance after 96 training epochs. Figure 8 illustrates the validation process for our DRRN and a 6-channel VDSR. In contrast to a 3-channel VDSR, a 6-channel VDSR achieves almost the same reconstruction quality at epoch 12 as DRRN at epoch 96, showing much better convergence.

Reconstructed images in Figure 9, shows good visual quality. Line *c* shows that the patch-wise CNN makes a color-corrected by GAN image sharper. Our CNNs were trained using a dataset of images captured indoor on fixed distance, but this pipeline is also applicable for outdoor scenes, captured from different distances, two examples are shown in line *d*. Line *d* in Figure 9 also shows the processing artifacts produced by the image-wise CNN. The artifacts sometimes occur when our pipeline is applied to real scene images. A possible cause of these artifacts is illumination irregularity.

As we showed, our approach reconstructs an image obtained by a diffractive single- or dual-lens optical system outperforming other methods. Our method can also be applied to images taken with a multi-aperture MDLs setup, showing additional quality gains.

Table 1 – Mean PSNR values for different reconstruction algorithms (dB)

Pipeline	Captured	Color corrected image	Deblurred image
Deconvolution [6], [25] (single-aperture)	17.21	17.32	18.41
Pix2pix+ 3-channel VDSR (single-aperture)	17.21	25.51	25.59
Pix2Pix + DRRN (single-aperture)	17.21	25.51	25.96
Pix2Pix + 6-channel VDSR (multi-aperture)	17.21 (left), 17.19 (right)	25.41 (left), 25.08 (right)	26.07

7. Conclusion

We propose an end-to-end deep learning-based approach to image reconstruction and applied it to images captured by single and dual MDLs imaging systems. The reconstruction quality was estimated as an average PSNR value applied to a wide image dataset, and it reached 26 dB. In contrast to previous works [5] and [25], where PSNR was estimated for synthetic tests or for the central part of an image only, we estimate the quality for the entire RGB image. Not only we now use an improved, more realistic PSNR measurement, but we were also able to achieve a higher measured quality according to this stricter metric.

The proposed approach relies on the general model of patch-wise and image-wise distortions, inherent to the optical systems with high longitudinal chromatic aberration. We conjecture that our reconstruction approach will also be useful for other flat optical systems, based on

MDLs or metalenses. The reconstruction pipeline was trained using an indoor imaging setup, but it shows good results when applied to outdoor images too. Nevertheless, the spatial and depth-dependence of distortions, and illumination variations need to be further investigated.

Finally, we publicly released a set of images captured by our single and dual MDL imaging systems to allow researchers to train and validate image reconstruction methods for flat lens imaging –

<https://1drv.ms/f/s!AiQGeNBKHALepTCPCRlEasXf92cO>.

Acknowledgement. Methods and algorithms were developed with the support of RFBR grants (19-29-01235-mk, 16-29-11744-ofi-m, № 16-29-09528-ofi-m, № 17-29-03112-ofi-m, № 18-07-01390-A), optics and experiments – in the framework of state assignment of the IPSI RAS - a branch of the Federal Scientific-Research Center "Crystallography and Photonics" of the RAS (agreement № 007-Г3/Ч3363/26).

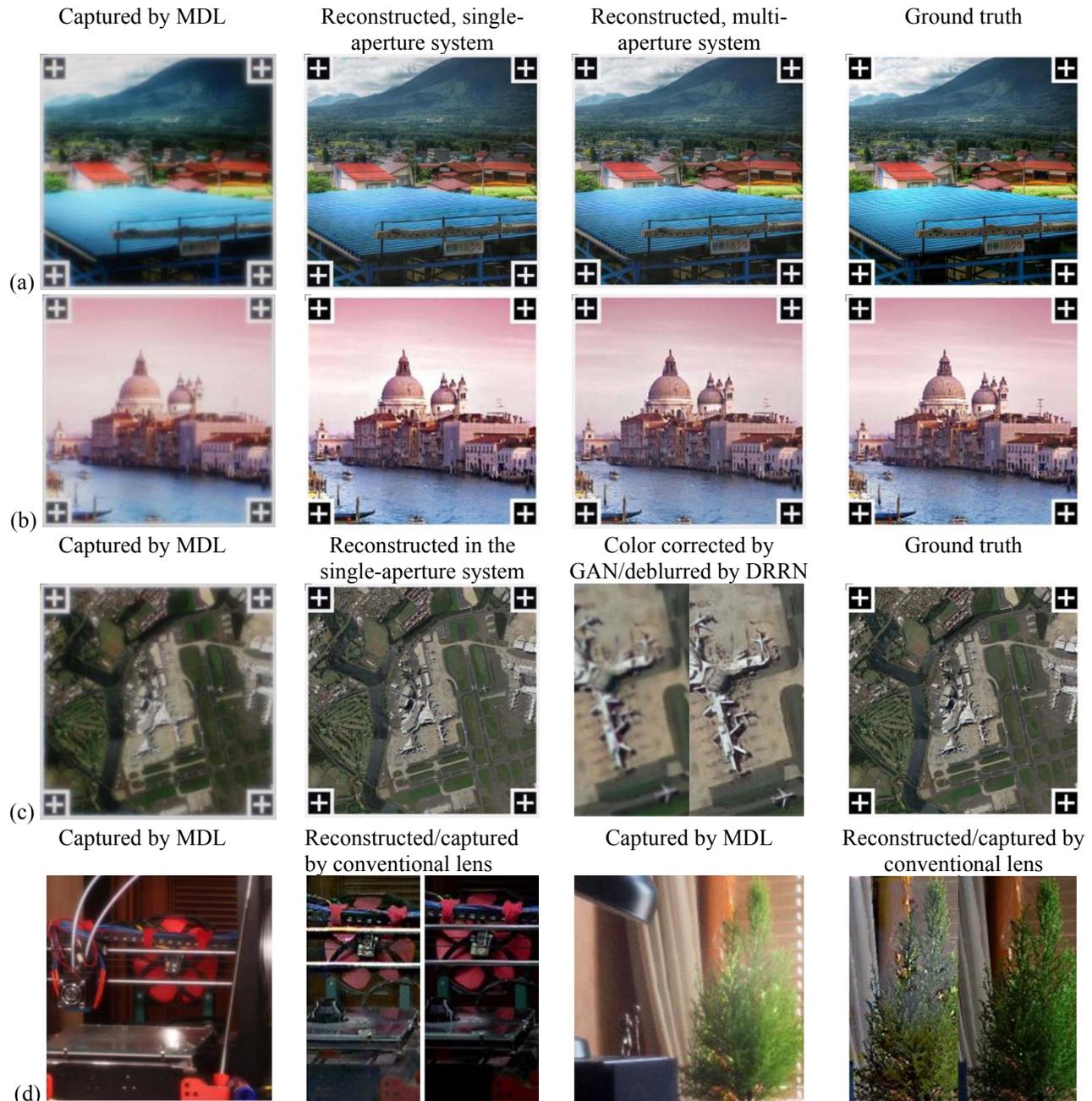


Figure 9: (a)(b)(c) Test image examples; (d) Examples of images with reconstruction artifacts.

References

- [1] S. Banerji, M. Meem, A. Majumder, F.G. Vasquez, B. Sensale-Rodriguez, and R. Menon. Imaging with flat optics: metalenses or diffractive lenses? *Optica*, 6(6):805-810, 2019.
- [2] P. Wang, N. Mohammad, and R. Menon. Chromatic-aberration-corrected diffractive lenses for ultra-broadband focusing. *Scientific Reports*, 6:21545, 2016.
- [3] P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, and R. Devlin. Recent advances in planar optics: from plasmonic to dielectric metasurfaces. *Optica*, 4(1): 139-152, 2017.
- [4] A. Nikonorov, R. Skidanov, V. Fursov, M. Petrov, S. Bibikov, and Y. Yuzifovich. Fresnel Lens Imaging with Post-Capture Image Processing. *CVPRW-2015:33-41*, 2015.
- [5] A. Nikonorov, M. Petrov, S. Bibikov, P. Yakimov, V. Kutikova, Y. Yuzifovich, A. Morozov, R. Skidanov, and N. Kazanskiy. Toward Ultralightweight Remote Sensing With Harmonic Lenses and Convolutional Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9):3338-3348, 2018.
- [6] Y. Peng, Q. Fu, H. Amata, Sh. Su, F. Heide, and W. Heidrich. Computational imaging using lightweight diffractive-refractive optics. *Optics Express*, 23(24):31393-31407, 2015.
- [7] T. Sun, Y. Peng, and W. Heidrich. Revisiting cross-channel information transfer for chromatic aberration correction. *CVPR-2017:3248-3256*, 2017.
- [8] G. Carles, G. Muyo, N. Bustin, A. Wood, and A.R. Harvey. Compact multi-aperture imaging with high angular resolution. *Journal of the Optical Society of America A*, 32(3):411-419, 2015.
- [9] F. Aieta, M.A. Kats, P. Genevet, and F. Capasso. Multiwavelength achromatic metasurfaces by dispersive phase compensation. *Science*, 347(6228):1342-1345, 2015.
- [10] A. Nikonorov, R. Skidanov, V. Kutikova, M. Petrov, A. Alekseyev, S. Bibikov, and N. Kazanskiy. Towards multi-aperture imaging using diffractive lens. *SPIE Proceedings:11146*, 2019.
- [11] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295-307, 2015.
- [12] J. Kim, J. Lee, and K. Lee. Accurate image super-resolution using very deep convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2016):1646-1654*, 2016.
- [13] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017):3147-3155*, 2017.
- [14] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia. Detail-revealing Deep Video Super-resolution. *IEEE International Conference on Computer Vision (ICCV-2017): 4482-4490*, 2017.
- [15] M. Ohara, Y. Seto, T. Kanamori, Y. Banba. Patent № JP2013081156A, 09.04.2013.
- [16] F.E. Sahin, R. Laroia. Light L16 Computational Camera. *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, OSA Technical Digest:P.JTu5A.20, 2017.
- [17] J. Tanida. Multi-aperture optics as a universal platform for computational imaging. *Optical Review*, 23(5):859-864, 2016.
- [18] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (TOMBO): concept and experimental verification. *Applied Optics*, 40(11):1806-1813, 2001.
- [19] X. Yu, Y. Qu, M. Hong. Underwater-GAN: Underwater Image Restoration via Conditional Generative Adversarial Network. *ICPR 2018*, 11188:66-75, 2019.
- [20] P. Das, A. Baslamisli, Y. Liu, S. Karaoglu, and T. Gevers. Color constancy by GANs: an experimental survey. Preprint arXiv:1812.03085v1, 2018.
- [21] P. Isola, J.-Y. Zhu, T. Zhou, and A. Efros. Image-to-image translation with conditional adversarial networks. Preprint arXiv:1611.07004v3, 2018.
- [22] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention:234-241*, 2015.
- [23] D. P. Kingma, and J. Ba. Adam: A method for stochastic optimization. arXiv, arXiv:1412.6980, 2014.
- [24] D.W. Sweeney, and G.E. Sommargren. Harmonic diffractive lenses. *Applied Optics*, 34(14):2469-2475, 1995.
- [25] A. Nikonorov, M. Petrov, S. Bibikov, Y. Yuzifovich, P. Yakimov, N. Kazanskiy, R. Skidanov, and V. Fursov. Comparative Evaluation of Deblurring Techniques for Fresnel Lens Computational Imaging. *ICPR 2016:775-780*, 2016.
- [26] S.K. Nayar, D.C. Sims, and M. Fridberg. Towards Self-Powered Cameras. *Proceeding ICCP*, 1-10, 2015.
- [27] C. Li, and M. Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. *European Conference on Computer Vision (ECCV-2016)*.
- [28] A. Nikonorov, M. Petrov, S. Bibikov, V. Kutikova, A. Morozov, N. Kazanskiy. Image restoration in diffractive optical systems using deep learning and deconvolution. *Computer Optics*; 41(6): 875-887, 2017
- [29] V.P. Korolkov, R.K. Nasyrov, R.V. Shimansky. Zone boundary optimization for direct laser scripture of continuous diffractive optical elements. *Applied optics*, 41(1):53-62, 2006.