

Исследование метаобучения для нейросетевой реконструкции изображений в дифракционно-оптических системах¹

В.В. Евдокимова^{*,**}, В.Д. Рябов^{*}, А.П. Алексеев^{*,**}, А.В. Никоноров^{*,**}

^{*} Самарский национальный исследовательский университет им. академика С. П. Королева, Самара,
Россия

^{**} Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника»
РАН, Самара, Россия

Поступила в редколлегию 14.07.2023

Аннотация—В настоящее время выполняется все больше исследовательских проектов в области создания систем формирования изображения на основе плоской оптики. Основным недостатком плоских оптических систем является значительное снижение качества получаемых изображений из-за хроматических аберраций. В статье представлены результаты исследования одношагового метаобучения для повышения качества реконструкции изображений в дифракционно-оптических системах.

КЛЮЧЕВЫЕ СЛОВА: многоуровневые дифракционные линзы, нейросетевая реконструкция изображений, метаобучение, обучение по нулевой выборке.

DOI: 10.53921/18195822_2023_23_3_335

1. ВВЕДЕНИЕ

В настоящее время актуальным становится некоторое переосмысление методов глубокого обучения и искусственных нейронных сетей (ИНС), популярность которых растёт с 2012 года. Один из наиболее актуальных вопросов - снижение требований к объёмам выборки для обучения ИНС, а также создания универсальных нейросетевых моделей, обученных на большом объёме данных для определенного круга задач, и допускающих адаптацию к новой задаче без дообучения на дополнительном наборе данных.

Такие подходы к обучению носят название подходов обучения по малым (Few Shot Learning) или нулевым обучающим выборкам (Zero Shot Learning). Эти подходы позволяют решать задачи без огромных наборов данных, специфичных для конкретной задачи. Применение подобных подходов актуально как для классических для ИНС задач классификации, так и для более специфичных, таких, как задачи вычислительной фотографии или сверхразрешения.

Задача реконструкции изображений как в однолинзовых так и в мультиапертурных дифракционно-оптических системах представляет собой задачу, схожую с задачей сверхразрешения на основе одного изображения (single image super resolution, SISR), или сверхразрешения на основе последовательности кадров (multi frame super resolution) [1].

Современные нейросетевые решения задачи сверхразрешения обеспечивают хорошее визуальное качество изображений высокого разрешения и высокие значения пикового отношения сигнала к шуму (PSNR) и индекса структурного сходства (SSIM) [2, 3]. Однако большинство из существующих подходов используют известную модель деградации, основанную на

¹ Работа выполнена в рамках Гранта РНФ 22-19-00364 в теоретической части, Государственного задания ФНИЦ «Кристаллография и фотоника» РАН в части экспериментального исследования.

бикубической интерполяции изображения низкого разрешения или на размытии изображений различными ядрами [4, 5, 6, 7, 8]. Модели деградации для изображений реального мира неизвестны или сложны для моделирования [9], что приводит к появлению различных артефактов при обработке реальных изображений. В работах [9, 10, 11] предложены современные решения для преодоления проблемы артефактов для задачи сверхразрешения. В работе [11] предлагается адаптация универсальных нейросетевых моделей для конкретного случая деградации изображения, основанная на процедуре метаобучения по нулевой обучающей выборке (MZSR). В [10] предложены архитектура и процедура обучения сети, позволяющие получить точное представление деградации для конкретных изображений. Для моделирования деградаций изображений реального мира в [9] предложена процедура моделирования деградаций высоких порядков на основе простых моделей, таких как размытие изображения, изменение размера и др. Артефакты, возникающие при реконструкции реальных изображений в дифракционных оптических системах связаны с различиями в условиях съемки обучающего набора данных, использованного для обучения нейросетевой модели реконструкции, и изображений реальных сцен. Экспозиция и динамический диапазон на изображениях реальных сцен значительно отличаются от результата съемки с экрана, которая проводилась в условиях средней освещенности в 100 люкс. Кроме того, причинами артефактов реконструкции могут быть усиление (gain) или светочувствительность (ISO) камеры и сжатие с потерями, исследуемые в работах [12, 13].

2. МЕТАОБУЧЕНИЕ ДЛЯ ДИФРАКЦИОННО-ОПТИЧЕСКИХ СИСТЕМ

Успешное использование сквозного нейросетевого подхода к реконструкции изображений позволяет получить систему компьютерного зрения, имеющую практическое применение, в работе [12] описан спутник CubeSat с камерой на основе МДЛ, впервые запущенный в космос. Однако, как показано в [12], существует ряд проблем при реконструкции изображений, причиной которых является различия в условиях съемки реальных изображений от обучающей выборки, такие как различия в экспозиции и коэффициента усиления камеры и артефактов реконструкции. В данной работе описывается подход на основе метаобучения, независимого от модели деградации (model-agnostic metalearning).

Подход на основе метаобучения, не зависящего от модели деградации (MAML) был успешно применен для задачи сверхразрешения по нулевой или малой выборке в работе [11]. Данный подход включает 3 этапа:

1) Обучение на достаточно большом наборе данных высокого качества с известной моделью деградации. На данном этапе определяются значения весов начальной модели W_T .

2) Мелкошаговое метаобучение. Дообучение начальной модели W_T на N наборах данных, сформированных с использованием модели деградации с различными параметрами. При одновременном дообучении на N наборах данных проводится усреднение N траекторий градиентного спуска, что позволяет определить «среднюю» точку W_M в пространстве параметров модели для N задач.

3) Этап метатестирования с использованием подхода к сверхразрешению по нулевой выборке (ZSSR), в рамках которого модель W_M дообучается на одном изображении низкого разрешения (input LR), модель деградации которого неизвестна. Поскольку для данного изображения нет эталона, то дообучение проводится на следующей паре: применяется понижение разрешения исходного изображения низкого разрешения (input LR) и сформированное изображение подается на вход сети, а исходное изображение низкого разрешения (input LR) считается эталонным и используется при расчете функции потерь. Спустя несколько шагов градиентного спуска обучение останавливается. Количество шагов определяется экспериментальным путем. В результате дообучения сформирована финальная модель W_k , на вход которой по-

дается исходное изображение низкого разрешения (input LR). Финальная модель генерирует изображение высокого разрешения для изображения низкого разрешения с неизвестной моделью деградации. Подробнее данный процесс описан в работе [11].

В данной работе подход на основе метаобучения успешно используется для решения задачи реконструкции изображений, полученных в дифракционно-оптических системах. Однако в данной работе вместо метаобучения с усреднением градиентов функций потерь на каждом шаге [11] используется усреднение параметров модели в один шаг [14], а также в качестве эталона для реального кадра используется улучшенное экспертом Adobe Photoshop изображение. Предлагаемый в данной работе алгоритм включает 3 этапа:

1) Обучение на исходном наборе данных. Перед обучением проводится аугментация обучающей выборки моделированием ISO-шума и изменением экспозиции. Процедура аугментации описана в работе [15]. Выбор начальной точки данного подхода W_T в пространстве параметров модели проводится на основе критерия min-FEL, предложенного в работе [15].

2) Крупношаговое метаобучение. Последовательное дообучение начальной модели W_T на N наборах данных, сформированных с использованием различных типов деградации, характерных для дифракционно-оптических систем [15]: увеличение экспозиции, уменьшение экспозиции, моделирование ISO-шума с различными параметрами, сжатие изображений. В отличие от оригинальной работы, где проводилось усреднение траекторий градиентного спуска, в данной работе предлагается усреднять веса N моделей, полученных в результате дообучения начальной модели W_T на N наборах данных:

$$W_M = \frac{1}{N} \sum_{i=1}^N W_i. \quad (1)$$

Обоснование данного шага описано в разделе 3. Выбор лучшей точки в пространстве параметров модели при дообучении начальной модели W_T на N наборах данных осуществляется на основе критерия min-FEL [15].

3) Этап метатестирования или дообучения на нулевой выборке. На данном этапе модель W_M дообучается на одном реальном изображении с неизвестной моделью деградации, эталон для которого сформировал эксперт в программе Adobe Photoshop. На рисунке 1 представлен пример реального изображения, используемого на этапе метатестирования, и его результат обработки в Adobe Photoshop, используемый как эталон при обучении. На рисунке 1а черной рамкой выделен фрагмент изображения, используемый для расчета FEL. Количество эпох обучения определяется экспериментальным путем. Модель W_k , сформированная в результате дообучения на изображении реальной сцены используется для обработки новых изображений реальных сцен.

3. ОБОСНОВАНИЕ КРУПНОШАГОВОГО МЕТАОБУЧЕНИЯ

Согласно [18] мелкошаговое метаобучение состоит из следующей последовательности шагов. Расчет оценки параметров для каждого набора данных T_i при $i = \overline{1, N}$:

$$W'_i = W - \alpha \nabla_W L_{T_i}(f_W). \quad (2)$$

Расчет градиентного шага метаобучения для оценки параметров выглядит следующим образом:

$$W \leftarrow W - \beta \nabla_W \sum_{i=1}^N L_{T_i}(f_{W'}). \quad (3)$$

Согласно [18] gjcktljdfntkmyjcnm (2)–(3) приводит к вычислению Гессиана. Также в [18] экспериментально показано, что аппроксимация Гессиана первым порядком возможно и приводит

к незначительному снижению точности. Согласно [18], для сетей с ReLU активацией это может быть обосновано тем, что для ReLU активаций вторая производная почти всегда близка к нулю [19].

При таком упрощении (3) примет вид:

$$W \leftarrow W - \beta \nabla_W \sum_{i=1}^N L_{T_i}(f_W). \quad (4)$$

В предположении, что градиентные шаги (4) выполняются в течение K итераций финальная оценка параметров примет вид:

$$W_K = W - \beta \sum_{i=1}^N \sum_{k=1}^K \nabla_W L_{T_i}(f_{W_k}). \quad (5)$$

С другой стороны, для каждого набора данных финальная оценка параметров, полученная за K шагов примет вид:

$$W_{i,k} = W - \alpha \sum_{k=1}^K \nabla_W L_{T_i}(f_{W_k}). \quad (6)$$

Представляет интерес выражение оценки (5) через среднее значение оценок (6).

Используя выражение

$$\alpha \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \nabla_W L_{T_i}(f_{W_k}) = W - \frac{1}{N} \sum_{i=1}^N W_{i,k}. \quad (7)$$

можно выразить второе слагаемое в (4) как

$$W_k = \left(1 - \frac{\beta}{\alpha}\right) W + \frac{\beta}{\alpha} \sum_{i=1}^N W_{i,k}. \quad (8)$$

Случай при $\frac{\beta}{\alpha} = \frac{1}{N}$ приводит (8) к виду (1), используемому в настоящей работе в качестве основной формулы для оценки крупношагового метаобучения.

4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В данной работе проводится обучение сети U-Net, архитектура которой представлена в работе [15].

Преимуществом усреднения моделей, а не траекторий градиентного спуска является то, что при появлении нового способа деградации изображений новая «средняя» модель W_M формируется с минимальными вычислительными затратами. Достаточно дообучить модель W_T на одном наборе данных, сформированном на основе новой модели деградации и добавив новую модель W_{N+1} в уже существующий стек из N моделей. При усреднении траекторий градиентного спуска второй этап необходимо повторить полностью, дообучая модель на всем стеке из $N + 1$ моделей. На рисунке 2 представлены графики обучения сети в рамках первого этапа алгоритма. График FEL для фрагмента реального изображения представлен, начиная с 30 эпохи обучения, поскольку из-за проблем с масштабированием может быть немного затруднительно увидеть общую тенденцию: на первых эпохах обучения уровень ложных контуров значительно выше, чем на последующих, что не позволяет наглядно представить график на интересующем временном периоде, где FEL достигает минимальных значений. По критерию min-FEL значения весов модели были зафиксированы на 140 эпохе, где FEL достиг значения 0,61%. Среднее значение PSNR на валидационной выборке составило 25,67 дБ.

На рисунке 3 приведены результаты реконструкции реального изображения, не участвовавшего в процессе обучения, после трех этапов алгоритма. На рисунке 3б приведен результат

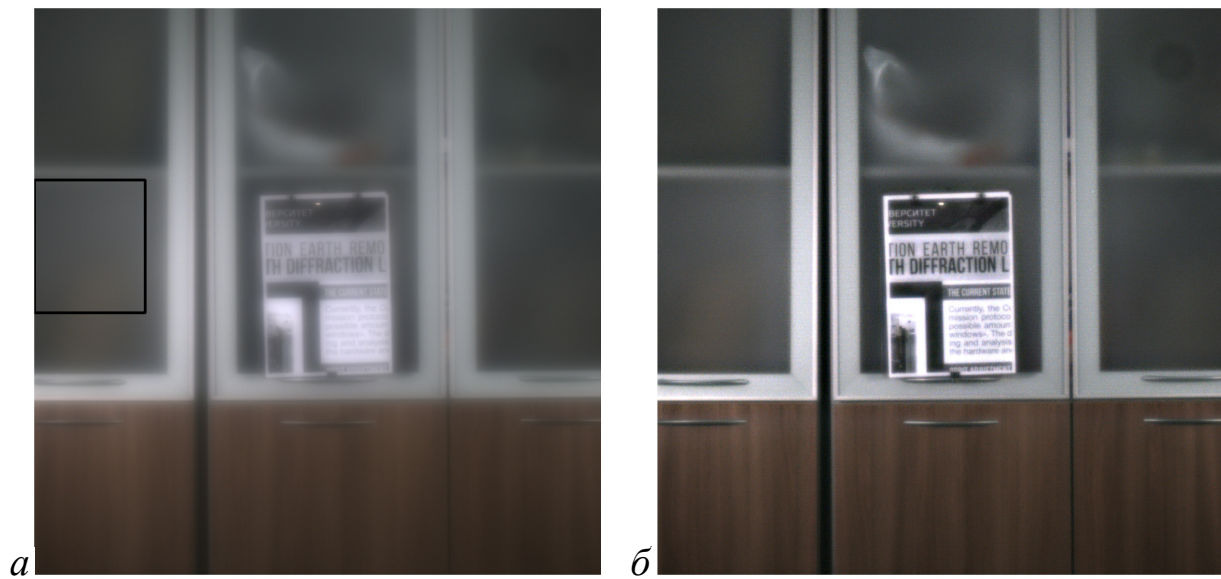


Рис. 1. Изображение реальной сцены, используемое на 3 этапе алгоритма при дообучении на нулевой выборке. а) Изображение, снятое на основе диракционно-оптической системы, с выделенным фрагментом для расчета FEL б) Эталонное изображение, обработанное средствами Adobe Photoshop.

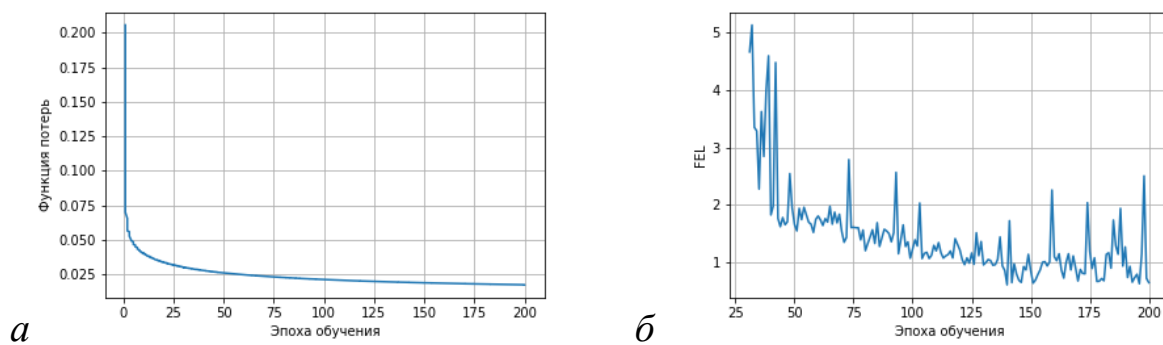


Рис. 2. Графики обучения сети на 1 этапе алгоритма а) функция потерь б) FEL на фрагменте реального изображения.

реконструкции моделью, обученной с использованием только 1 и 3 этапов алгоритма (трансферное обучение), то есть без этапа метаобучения. Также на восстановленных изображениях приведены значения FEL, рассчитанные по трем фрагментам, выделенным ограничивающей рамкой. Видно, что подход на основе одношагового метаобучения и усреднения моделей позволяет значительно сократить количество артефактов реконструкции по сравнению с трансферным обучением. Среднее значение FEL по трем фрагментам восстановленного изображения моделью, обученной без метаобучения, составило 32,1%, что соответствует визуально высокому уровню артефактов реконструкции. При использовании подхода на основе метаобучения среднее значение FEL по трем фрагментам восстановленного изображения снизилось до 7,3%. Предложенный в данной работе подход позволяет снизить влияние условий съемки изображений на результат реконструкции для дифракционно-оптических систем, которые в ранних работах [12, 13, 16, 17] приводили к артефактам реконструкции на изображениях реальных сцен и препятствовали использованию разрабатываемых алгоритмов реконструкции в прикладных задачах. Использование метрики качества PSNR в работах [12, 13, 16, 17] не обеспечивало минимизацию уровня артефактов по сравнению с используемой в данной работе метрикой FEL.

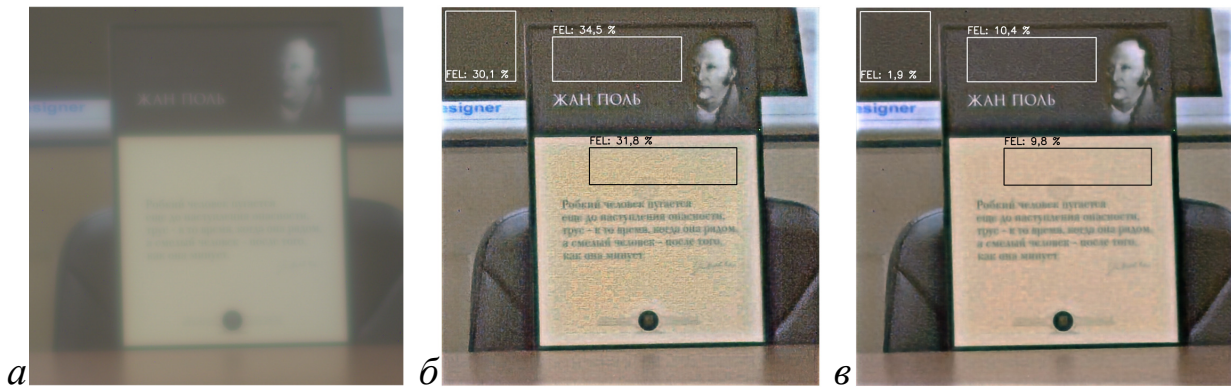


Рис. 3. Результаты реконструкции изображения реальной сцены, не участвовавшего в процессе обучения а) Изображение, снятое на основе дифракционно-оптической системы. б) Результат реконструкции изображения моделью, обученной без второго этапа метаобучения, но с использованием трансферного обучения на 3 этапе на основе модели, полученной на 1 этапе (1 и 3 этапы алгоритма) в) Результат реконструкции изображения моделью, обученной трехэтапным алгоритмом на основе одношагового метаобучения, предложенного в данной работе (1, 2, 3 этапы алгоритма).

5. ЗАКЛЮЧЕНИЕ

Использование подхода на основе одношагового метаобучения с усреднением моделей и обучения по нулевой выборке для сквозной нейросетевой реконструкции изображений в дифракционно-оптических системах позволил в данной работе снизить влияние различий в условиях съемки обучающего набора данных и реальных кадров, таких как различия в экспозиции и коэффициента усиления камеры. На восстановленных изображениях на основе описанного трехэтапного подхода наблюдается значительное сокращение количества артефактов реконструкции по сравнению с использованием трансферного обучения. Среднее значение FEL, рассчитанное по трем фрагментам восстановленного изображения, не участвовавшего в процессе обучения, удалось снизить с 32,1% (без второго этапа алгоритма) до 7,3% при использовании подхода на основе одношагового метаобучения. Преимуществом усреднения моделей, а не траекторий градиентного спуска является то, что при появлении нового способа деградации изображений новая «средняя» модель формируется с минимальными вычислительными

затратами по сравнению с оригинальным подходом. Более глубокое исследование описанного подхода на основе одношагового метаобучения для гибридных и мультиапертурных дифракционно-оптических систем является предметом дальнейшей работы.

СПИСОК ЛИТЕРАТУРЫ

1. Choi J. H., Kim J. H., Lee J. S. SRZoo: An integrated repository for super-resolution using deep learning. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020.
2. Zhang Yu. Image super-resolution using very deep residual channel attention networks. Proceedings of the European Conference on Computer Vision (ECCV), 2018.
3. Wang Z., Chen J., Hoi S. C. H. Deep Learning for Image Super-resolution: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
4. Kim J., Lee J. K., Lee K. M. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
5. Zhang Y. Residual dense network for image super-resolution. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2472–2481.
6. Dong C. Learning a deep convolutional network for image super-resolution. European conference on computer vision. Springer, Cham, 2014, pp. 184–199.
7. Zhang Y. Image super-resolution using very deep residual channel attention networks. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 286–301.
8. Zhang K. Zuo W., Zhang L. Learning a single convolutional super-resolution network for multiple degradations. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 3262–3271.
9. Wang X. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021, pp. 1904–1914.
10. Wang L. Unsupervised degradation representation learning for blind super-resolution. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10581–10590.
11. Soh J. W., Cho S., Cho N. I. Meta-transfer learning for zero-shot super-resolution. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3516–3525.
12. Ivliev N., Evdokimova V., Podlipnov V., Petrov M., Ganchevskaya S., Tkachenko I., Abrameshin D., Yuzifovich Y., Nikonorov A., Skidanov R., Kazanskiy N., Soifer V. First Earth-Imaging CubeSat with Harmonic Diffractive Lens. Remote Sensing, 2022, vol. 14, № 9, pp. 2230.
13. Evdokimova V., Petrov M., Klyueva M., Zybin E., Kosianchuk V., Mishchenko I., Novikov V., Selvesiuk N., Ershov E., Ivliev N., Skidanov R., Kazanskiy N., Nikonorov A. Deep learning-based video stream reconstruction in mass production diffractive optical systems. Computer Optics, 2021, vol. 45, № 1, pp. 130–141.
14. Finn C., Abbeel P., Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Book Model-agnostic meta-learning for fast adaptation of deep networks. Ed. Sydney, NSW, Australia: JMLR.org, 2017. - p. 1126–1135.
15. Evdokimova V.V., Podlipnov V.V., Ivliev N.A., Petrov M.V., Ganchevskaya S.V., Fursov V.A., Yuzifovich Y., Stepanenko S.O., Kazanskiy N.L., Nikonorov A.V. Hybrid RefractiveDiffractive Lens with Reduced Chromatic and Geometric Aberrations and Learned Image Reconstruction. Sensors, 2023, vol. 23, pp. 415.
16. Никоноров А.В., Петров М.В., Бибииков С.А., Кутикова В.В., Морозов А.А., Казанский Н.Л. Реконструкция изображений в дифракционно-оптических системах на основе сверточных нейронных сетей и обратной свертки. Компьютерная оптика, 2017, Т. 41, № 6, С. 875–887.

17. Nikonorov A.V., Petrov M.V., Bibikov S.A., Yakimov P.Y., Kutikova V.V., Yuzifovich Y.Y., Morozov A.A., Skidanov R.V., Kazanskiy N.L. Toward Ultralightweight Remote Sensing With Harmonic Lenses and Convolutional Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, vol. 11, no. 9, pp. 3338–3348.
18. Finn C., Abbeel P., Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. *ArXiv*, abs/1703.03400.
19. Goodfellow I. J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples. *International Conference on Learning Representations (ICLR)*, 2015.

Study of meta-learning for neural network image reconstruction in diffraction optical systems

V.V. Evdokimova, V.D. Ryabov, A.P. Alekseev, A.V. Nikonorov

Currently, more and more research projects are being carried out in the field of creating imaging systems based on flat optics. The main disadvantage of flat optical systems is a significant decrease in the quality of the images obtained due to chromatic aberrations. The article shows how it is possible to cope with such a deterioration in quality using a meta-learning approach without depending on the model.

KEYWORDS: multilevel diffraction lenses, neural network image reconstruction, meta-training, zero-sampling training.