Optics Letters

Beam profiler network (BPNet): a deep learning approach to mode demultiplexing of Laguerre–Gaussian optical beams

Amit Bekerman,[†] Sahar Froim,[†] Barak Hadad,[†] and Alon Bahabad*

Department of Physical Electronics, School of Electrical Engineering, Fleischman Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel
Corresponding author: alonb@eng.tau.ac.il

The transverse field profile of light has been recognized as a resource for classical and quantum communications for which reliable methods of sorting or demultiplexing spatial optical modes are required. Here we experimentally demonstrate state-of-the-art mode demultiplexing of Laguerre–Gaussian beams according to both their orbital angular momentum and radial topological numbers using a flow of two concatenated deep neural networks. The first network serves as a transfer function from experimentally generated to ideal numerically generated data, while using a unique "histogram weighted loss" function that solves the problem of images with limited significant information. The second network acts as a spatial-modes classifier. Our method uses only the intensity profile of modes or their superposition, making the phase information redundant. © 2019 Optical Society of America

The possibility of employing the spatial degree of photons for communications is gaining interest in recent years due to its unbounded dimensionality [1–4]. A natural basis to span the transverse profile of photons is composed of Laguerre–Gaussian (LG) modes which are characterized with two topological numbers: $l \in \mathbb{Z}$, the orbital index, describing the orbital angular momentum (OAM) in units of $\hbar$ per photon in the beam and $p \in \mathbb{Z}_+$, which is the radial index or radial quantum number. For utilizing LG modes for communications it is essential to be able to perform mode sorting, or demultiplexing on the incoming physical data flow. Essentially, there are two approaches to demultiplexing. The first approach uses (usually complicated) optical setups in which the l and p degrees of freedom are coupled to other degrees of freedom such as the angle of propagation and polarization. Most such methods address either the OAM [5–8] or the radial index [9,10] degrees of freedom, while a recent measuring method handles both degrees [11]. The second approach, which emerged recently, suggests using just a camera to detect the intensity of the incoming light beam and to utilize a deep neural network (DNN) to classify the beam. We note that such works serve as an example of a

more general trend in which deep learning methods are incorporated into optics [12–18]. To date, demonstrated DNN-based demultiplexers addressed solely the OAM degree of light.

Here we present, experimentally, a DNN-based mode sorter for both topological numbers of LG modes, able to classify both the OAM and radial index. Our solution uses two concatenated DNNs. One network is used for mode classification, and it is trained on numerically generated "ideal" images of LG modes and two-mode superpositions. The other network is a calibration network which converts experimentally detected images that suffer from optical aberrations and noise to ideal numerical images, which are then fed to the classifying network.

LG modes are solutions of the paraxial wave equation in cylindrical coordinates. They are given with [9]:

$$\mathrm{LG}_{l,p}(r, \phi, z)$$
$$= \sqrt{\frac{2p!}{\pi(p+|l|)!}} \frac{1}{w_z} \left(\frac{\sqrt{2}r}{w_z} \right)^{|l|} L_p^{|l|} \left(\frac{2r^2}{w_z^2} \right)$$
$$\times \exp\left(-\frac{r^2}{w_z^2} + i \left(\frac{kr^2}{2R_z} + l\phi - (2p+|l|+1)\varphi_g \right) \right), \quad \textbf{(1)}$$

where l and p are the orbital and radial indices, respectively, $L_p^{|l|}$ are the Laguerre polynomials, $w_z = w_0 \sqrt{1 + (z/z_R)^2}$ is the beam waist with w_0 being the waist at $z = 0$, $z_R = (\pi w_0^2)/\lambda$ is the Rayleigh range, $R_z = z(1 + (z_R/z)^2)$ is the radius of curvature, λ is the wavelength, $k = 2\pi/\lambda$ is the wave number, and $\varphi_g = \arctan(z/z_R)$ is the Gouy phase.

This work uses both numerically and experimentally generated LG modes and their superpositions. Experimentally generated data were acquired in a setup consisting of a 532 nm CW laser (Quantum Ventus 532 Solo Laser) which is expanded and collimated before reflecting off a phase-only spatial light modulator (Holoeye Pluto SLM). The phase masks loaded onto the SLM were encoded by extracting the phase of our numerically generated suprimposed modes and then adding a blazed grating to it. The resulting image in the first order of diffraction of the grating is Fourier transformed using a 50 cm focal lens and imaged by a camera (DataRay WinCamD-LCM4). The experimentally generated data are different than the

numerically generated data of ideal modes and their superpositions due to inherent aberrations and noise in the optical system.

Different datasets, each with a different number of superimposed modes, were created. We mark the datasets with DB_N^{type}, where $N = 1, 2, 3$, is the number of superimposed modes and type $\in$ {num, exp} stands for numerically or experimentally generated dataset. Each member of the datasets was realized according to $\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathrm{LG}_{l_n, p_n}$, where l_n and p_n are the orbital and radial indices that are being used in a particular superposition. All images are set to be of the size 224×224 pixels with 256 values per pixel in the range of [–1, 1] while, for the experimental datasets, the average pixel value (between all images) was set to 0, and the variance was set to 1. (We note that experimentally acquired images are first obtained at a size of 500×500 with pixel values in the range of [0,255].) Initially, each of the datasets $DB_N^{num}, N = 1, 2, 3$ contains 36^N members (as the range of the p and l topological numbers we apply is 0–5 each), while DB_2^{num} contains DB_1^{num}, and DB_3^{num} contains both DB_2^{num} and DB_1^{num}. For training a DNN, very large datasets are required. For this reason, the datasets are "augmented" with new members generated from the old ones. Specifically, for DB_1^{num} and DB_2^{num}, we augment the basic training dataset of images by adding 70 image variations per image in the basic set through beam rotations (rotation angles were uniformly distributed over $[0, 2\pi]$ with 1 deg $\simeq 17$ mrad resolution), beam shifts (uniformly distributed over [0,16] pixel shifts in both x and y coordinates), and the addition of Gaussian noise with mean $\mu = 0$ and variance $\sigma^2 = 0.2$. This amounts to overall $\sim$2,000 and $\sim$90,000 samples for DB_1^{num} and DB_2^{num}, respectively. In DB_3, the number of augmentations per unique combination was 2, leading to a total dataset size of $\sim$90,000. Similarly, DB_1^{exp} and DB_2^{exp} were generated and numerically augmented 100 fold by applying similar rotations and random noise (with no beam shifts).

Our solution to the beam profiler network (BPNet) consists of two concatenated networks, trained separately: a calibration network and a classifier network. Both networks were created using keras API [19].

The calibration network, based on U-Net's architecture [20], converts LG beam images taken in the lab (both single-mode and superpositions) into ideal images of the same LG beams without changing the overall size of the image. The output of the network for each image, which is also the label during training, is an image of the same LG mode (or superposition of the LG modes), albeit an ideal one, created using a simulation.

Since large parts of each image in the dataset are dark, a simple Mean-Square-Error (MSE) or Mean-Absolute-Error (MAE) loss functions cannot allow the calibration network to converge properly; as with such loss functions, the network converges to a poor local minimum, predicting only dark images. To solve this convergence issue, we introduce here a new type of loss function, called "Histogram Weighted Loss" (HWL). This loss gives higher significance to pixel values that are less common in the image, since they are the ones carrying important information in sparse images. To implement this cost function we first calculate the histogram of each image and modify the calculation of the regular MAE loss during training, by multiplying the loss of each pixel by 1 minus

the pixel probability in the image (which is determined by the histogram). In this case, a wrong prediction of a less common pixel will have a higher cost. The HWL is given by the following equation:

$$\mathrm{HWL} = \frac{1}{N \times M} \sum_{j=1}^{M} \sum_{i=1}^{N} (1 - \mathrm{prob}_{i,j})^{\gamma} |y_{i,j} - y_{i,j}^p|, \quad \textbf{(2)}$$

where N is the number of pixels in an image, M is the number of images in a given batch, $y_{i,j}$ is the value of pixel i in the jth (target) image in the batch, $y_{i,j}^p$ is the prediction for the value of the same pixel generated by the network, $\mathrm{prob}_{i,j}$ is the probability for the said pixel value in the target image (extracted from the histogram), and $\gamma = 4$ is a hyper-parameter which we found by trial and error to produce the best results. HWL is somewhat related to the concept of focal loss [21] in which whole images are given different significance for a specific loss calculation.

The classifier network classifies images fed by the calibration network according to the values of the index numbers of the modes composing the images. The classifier network is trained using simulated images of LG beams (both single-mode and superpositions). This network is based on Mobilenet V2's architecture [22], where the last fully connected layer outputs 36 labels $((p, l) = (0, 0), (1, 0), \ldots(5, 5))$. We refer to the set of these 36 labels as the "modes vector." Each label output is set in the range [0,1], indicating the probability for successful detection of a specific mode. We decide on a mode being detected if the appropriate value in the modes vector is higher than 0.5. The input to this network is a numerically calculated image of an LG beam, or a superposition of such beams, and its output is a labeling of the different modes it contains.

The classifier network was first tested in a stand-alone configuration, by being fed directly by numerically generated and augmented datasets $DB_N^{num}, N = 1, 2, 3$. We split our three datasets into training (85%) and validation (15%) sets. The validation scored a perfect 100% success rate in all three cases. This shows that the network architecture we used was adapted effectively enough so as to learn even small datasets (DB_1) and datasets without a lot of repetition in the samples (DB_3). It is notable that when noisy, but otherwise undistorted, images are supplied to the classifier network, it exhibits very high performance. However, when testing the classifier network on experimental data that were measured in the lab, the inherent aberrations (in any optical system) degrade the performance considerably (see, e.g., Ref. [7]). At this point, we can choose between two strategies. One option is training the classifier network on experimental data (a strategy that was adapted for example in Ref. [6]) which has two problems—the performance would rely on the amount of distortions (aberrations) in the optical system, and the solution is applicable to a specific optical setup. The second strategy that we chose is using a calibration network which lets us use a high-performance classifier network that, in principle, could work with any setup-specific calibration network. The performance of the whole system thus becomes dependent mostly on the quality of the calibration network whose performance depends, in turn, on the quality of the optical system. If the quality of the optical setup is high enough, the classifier network can be used as a stand-alone.

The next stage was training the calibration network and then testing the whole BPNet (calibration+classifier). For this

Table 1. Results for Different Training Configurations[a]

Test-Set	Calibration network DB	Classification network DB	Results
Single-Mode	DB_1^{exp}	DB_1^{num}	100%
Single-Mode	DB_1^{exp}	DB_2^{num}	96.39%
Single-Mode	DB_1^{exp}	DB_3^{num}	86.39%
2-Mode Superposition	DB_2^{exp}	DB_2^{num}	91.30%
2-Mode Superposition	DB_2^{exp}	DB_3^{num}	63.45%

^{a}"*DB*" stands for the training set being used as detailed in the text. Note that the modes range from $((p, l) = (0, 0), (1, 0), \ldots (5, 5))$.

purpose, we divided DB_1^{exp} and DB_2^{exp} into a training set (72%), a validation set (18%), and to a test set (10%). The whole network was tested in several different training configurations as described in Table 1. The main conclusion from these results is that we were able to get state-of-the-art real-world single-mode detection and two-mode superposition demultiplexing. It is noteworthy that the single-mode training yields perfect performance although the dataset for training was relatively small. The relatively low performance for demultiplexing two-modes superpositions when the classifier network was trained with three-modes superpositions is attributed to the small level of augmentation in DB_3^{num}.

A few examples of successful demultiplexing-detection by the BPNet for two-modes superpositions are shown in Fig. 1. In this figure, we show the phase masks that were applied to the SLM in the experimental setup, the input to the network which is the image captured by the camera, the prediction of the calibrat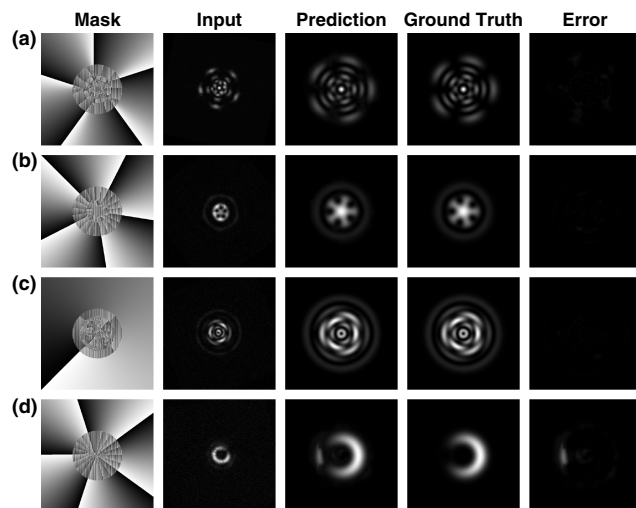ion network which is fed to the classification network, the ground truth for that network, as well as the error which is the absolute difference between the ground truth and the prediction. Cases (a)–(c) show a negligible error. An interesting case is shown in Fig. 1(d) in which the calibration network added an artifact to the prediction, but the resulting classification was still correct. This shows that the classifier network has some degree of robustness to artifacts introduced by the calibration network.

Examples of some unsuccessful results are shown in Fig. 2. In these cases, it is clear that the images captured (using an automated procedure) by the camera are distorted and clouded. Even though the results were misclassified for these cases, we can appreciate that the calibration network locked on to some features in the input images. Explicitly, we can observe that in all images [except Fig. 2(a)], even though the ground truth and the captured image look similar, some additional artifacts were introduced to the image and, thus, the input was misclassified. In Fig. 2(a), the classification network adds another mode to the label of the image, due to a deformation in the outer ring. In Fig. 2(b), the calibration network converts the input image into a completely different superposition of modes. In Fig. 2(c), a spurious ring appears again, but this time, instead of adding another mode, it simply increases the OAM index for the first mode and the radial index for the second mode. In Fig. 2(d), the input looks similar to the ground truth image, but the calibration network (probably due to added noise) predicts an image with a closed inner ring, therefore leading to the omission of one of the superimposed modes. Still, it is noteworthy that, in two of the cases, the predicted images were almost identical to the ground truth images.
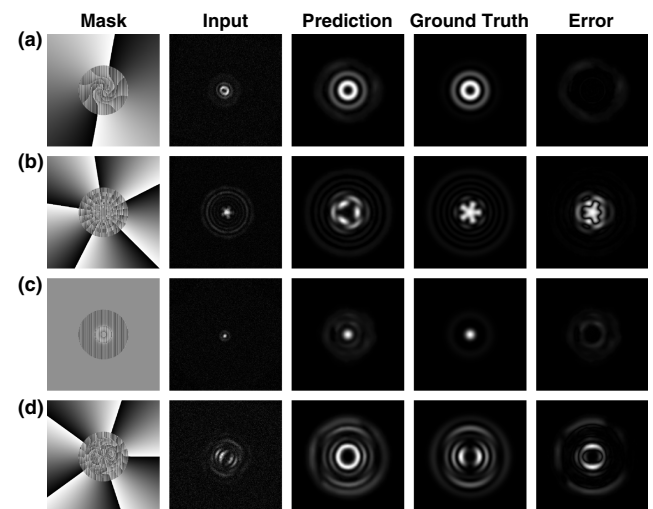

Fig. 1. Successful classification of two-mode superpositions. Some examples of successful calibration and classification of modes using the whole BPNet flow. The "Mask" column shows phase masks that were loaded onto the SLM. The "Input" column shows images captured by the camera. The "Prediction" column shows the output of the calibration network, while the "Ground Truth" column shows the projected output for a perfect calibration. The "Error" column shows the absolute difference between the predicted image and the ground truth image. The superpositions $((p_1, l_1), (p_2, l_2))$ demonstrated are (a) $((4,0),(2,5))$, (b) $((1,5),(0,0))$, (c) $((5,1),(0,4))$, and (d) $((0,4),(0,5))$.

Fig. 2. Unsuccessful calibration and classification of two-mode superpositions. Some examples of unsuccessful calibration and classification of the whole BPNet flow. The "Mask" column shows phase masks that where loaded into the SLM. The "Input" column shows images captured by the camera. The "Prediction" column shows the output of the calibration network, while the "Ground Truth" column shows the projected output for a perfect calibration. The "Error" column shows the absolute difference between the predicted image and the ground truth image. The superpositions demonstrated and the unsuccessful predictions are (a) $((1, 2), (2, 2)) \rightarrow ((1, 2), (2, 2), (3, 2))$, (b) $((5,5),(0,0)) \rightarrow ((5,5),(0,2))$, (c) $((0,0),(2,0)) \rightarrow ((1,0),(2,1))$, and (d) $((4, 5), (2, 3)) \rightarrow ((4, 5))$.
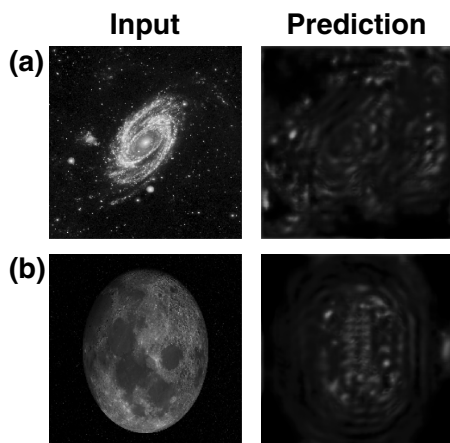
Input   **Prediction**

Fig. 3. Random inputs. Some examples of random inputs to the calibration network and their predictions.

Finally, it can be suggested that the calibration network did not actually learn to transform its input images to undistorted images (undoing the optical aberrations in the experimental setup), but that it simply learned a mapping between the input and output images. To refute this argument we fed a few random images to the calibration network (see Fig. 3), where we observe that the output of the calibration network would not simply map any image with some features to a multiplexed mode, but instead it does learn some kind of a transfer function (albeit, it might be restricted to work correctly with the LG modes fed to the network).

To conclude, we have realized a novel method for spatial-mode demultiplexing, relating to the two topological numbers characterizing LG modes using a flow of two concatenated DDNs: a calibration network (transferring from experimentally acquired images in the lab to "ideal" images) and a classifier network. We have shown that our classifier is able to demux up to three superimposed spatial modes with perfect accuracy, while we demonstrated that the whole flow exhibits state-of-the-art performance for detecting two-mode superpositions acquired in the lab. In this Letter, we demonstrated de-multiplexing of two modes experimentally and three modes numerically. In principle, it is possible to extend the number of modes that can be de-multiplexed in order to be on par with recent state-of-the-art figures (e.g., 12 modes as communicated in Ref. [23]). An important ingredient in this Letter is the introduction of the HWL which helps handle sparse images where most pixels do not carry information. This loss function might be relevant to other fields that encounter sparse images such as medical imaging and astrophysics.

†These authors contributed equally to this Letter.

REFERENCES

1. J. Wang, J.-Y. Yang, I. M. Fazal, N. Ahmed, Y. Yan, H. Huang, Y. Ren, Y. Yue, S. Dolinar, M. Tur, and A. E. Willner, Nat. Photonics **6**, 488 (2012).
2. E. Nagali, F. Sciarrino, F. De Martini, L. Marrucci, B. Piccirillo, E. Karimi, and E. Santamato, Phys. Rev. Lett. **103**, 013601 (2009).
3. N. Bozinovic, Y. Yue, Y. Ren, M. Tur, P. Kristensen, H. Huang, A. E. Willner, and S. Ramachandran, Science **340**, 1545 (2013).
4. M. Krenn, M. Huber, R. Fickler, R. Lapkiewicz, S. Ramelow, and A. Zeilinger, Proc. Natl. Acad. Sci. USA **111**, 6243 (2014).
5. S. Lightman, G. Hurvitz, R. Gvishi, and A. Arie, Optica **4**, 605 (2017).
6. T. Doster and A. T. Watnik, Appl. Opt. **56**, 3386 (2017).
7. S. Lohani, E. M. Knutson, M. O'Donnell, S. D. Huver, and R. T. Glasser, Appl. Opt. **57**, 4180 (2018).
8. A. Ruelas, S. Lopez-Aguayo, and J. C. Gutiérrez-Vega, J. Opt. **18**, 095605 (2016).
9. X. Gu, M. Krenn, M. Erhard, and A. Zeilinger, Phys. Rev. Lett. **120**, 103601 (2018).
10. Y. Zhou, M. Mirhosseini, D. Fu, J. Zhao, S. M. H. Rafsanjani, A. E. Willner, and R. W. Boyd, Phys. Rev. Lett. **119**, 263602 (2017).
11. F. Bouchard, N. H. Valencia, F. Brandt, R. Fickler, M. Huber, and M. Malik, Opt. Express **26**, 31925 (2018).
12. J. Jiang and J. A. Fan, "Global optimization of dielectric meta-surfaces using a physics-driven neural network," arXiv:1906.04157 (2019).
13. J. Jiang, D. Sell, S. Hoyer, J. Hickey, J. Yang, and J. A. Fan, "Data-driven metasurface discovery," arXiv:1811.12436 (2018).
14. Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, "Deep learning approach based on dimensionality reduction for designing electro-magnetic nanostructures," arXiv:1902.03865 (2019).
15. Y. Kiarashinejad, S. Abdollahramezani, M. Zandehshahvar, O. Hemmatyar, and A. Adibi, "Deep learning reveals underlying physics of light-matter interactions in nanophotonic devices," arXiv:1905.06889 (2019).
16. R. Trivedi, L. Su, J. Lu, M. F. Schubert, and J. Vuckovic, "Data-driven acceleration of photonic simulations," arXiv:1902.00090 (2019).
17. Y. Li, Y. Xue, and L. Tian, Optica **5**, 1181 (2018).
18. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, Optica **5**, 803 (2018).
19. F. Chollet et al., "Keras," 2015, https://keras.io.
20. O. Ronneberger, P. Fischer, and T. Brox, in International Conference on Medical Image Computing and Computer-assisted Intervention (Springer, 2015), pp. 234–241.
21. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, in Proceedings of the IEEE International Conference on Computer Vision (2017), pp. 2980–2988.
22. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018), pp. 4510–4520.
23. K. Ingerslev, P. Gregg, M. Galili, F. Da Ros, H. Hu, F. Bao, M. A. U. Castaneda, P. Kristensen, A. Rubano, L. Marrucci, K. Rottwitt, T. Morioka, S. Ramachandran, and L. K. Oxenløwe, Opt. Express **26**, 20225 (2018).