Optics Letters

Unpaired data training enables super-resolution confocal microscopy from low-resolution acquisitions

Carlos Trujillo,[1,4] Lauren Thompson,[2] Omar Skalli,[2] and Ana Doblas[3,*]

[1]School of Applied Sciences and Engineering, Universidad EAFIT, Medellin 050027, Colombia
[2]Department of Biological Sciences, The University of Memphis, Memphis, Tennessee 38152, USA
[3]Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, Massachusetts 02747, USA
[4]catrujilla@eafit.edu.co
*adoblas@umassd.edu

Received 26 July 2024; revised 9 September 2024; accepted 12 September 2024; posted 16 September 2024; published 3 October 2024

Supervised deep-learning models have enabled super-resolution imaging in several microscopic imaging modalities, increasing the spatial lateral bandwidth of the original input images beyond the diffraction limit. Despite their success, their practical application poses several challenges in terms of the amount of training data and its quality, requiring the experimental acquisition of large, paired databases to generate an accurate generalized model whose performance remains invariant to unseen data. Cycle-consistent generative adversarial networks (cycleGANs) are unsupervised models for image-to-image translation tasks that are trained on unpaired datasets. This paper introduces a cycleGAN framework specifically designed to increase the lateral resolution limit in confocal microscopy by training a cycleGAN model using low- and high-resolution unpaired confocal images of human glioblastoma cells. Training and testing performances of the cycleGAN model have been assessed by measuring specific metrics such as background standard deviation, peak-to-noise ratio, and a customized frequency content measure. Our cycleGAN model has been evaluated in terms of image fidelity and resolution improvement using a paired dataset, showing superior performance than other reported methods. This work highlights the efficacy and promise of cycleGAN models in tackling super-resolution microscopic imaging without paired training, paving the path for turning home-built low-resolution microscopic systems into low-cost super-resolution instruments by means of unsupervised deep learning.

https://doi.org/10.1364/OL.537713

Deep learning in super-resolution imaging has emerged as a groundbreaking approach over the last decade, leveraging neural networks to enhance image resolution beyond the limits of traditional imaging processing methods using a single image [1]. Deep-learning models can reconstruct high-resolution (HR) images from low-resolution (LR) inputs with remarkable fidelity using sophisticated algorithms and vast amounts of training data. By harnessing the power of convolutional neural networks (CNNs) and generative adversarial networks (GANs), researchers have achieved unprecedented levels of detail and clarity in images, opening new possibilities for analysis, diagnosis, and visualization in various fields such as satellite imagery, photography, and medical and microscopy imaging [2,3]. In particular, deep learning in microscopy has enabled super-resolution imaging by overcoming the diffraction limit and reconstructing super-resolved images (i.e., images whose spatial cutoff frequency is higher than the one limited by the diffraction limit of the conventional microscope) in which we can resolve biological structures at the nanoscale level with improved signal-to-noise ratio [4].

Among the different deep-learning models, supervised and unsupervised training are the two fundamental approaches. Supervised models are those which are trained on a labeled dataset. In other words, each input data is paired with a corresponding target output data. Therefore, the model learns to map inputs to outputs by minimizing the discrepancy between its predictions and the true outputs during the training, being able to generalize the learned mapping and make accurate predictions on unseen data. Supervised models have been reported to transform low-resolution (LR) to super-resolved (SR) images in several imaging modalities, including bright-field microscopy [3], widefield fluorescence microscopy [5], confocal microscopy [5–7], light-sheet microscopy [8], structured illumination microscopy [9], stochastic optical reconstruction microscopy [10], and Photoactivated Localization Microscopy [11].

Despite the prominent successes of supervised deep-learning models, their training and testing can pose several challenges. The amount of training data and its quality are two potential aspects affecting the model's performance. Whereas poor-quality training images affect the performance of the network inference, insufficient training data can easily lead to overfitting, providing a faulty model to generalize unseen data [12–14]. Acquiring accurate and high-quality paired ground-truth images for training can be labor-intensive, requiring an adequate imaging protocol to record experimental imaging from different imaging modalities or under different conditions. Because the

experimental acquisition of large, paired databases is difficult, data augmentation techniques and transfer learning are commonly employed to expand the training dataset. However, data augmentation should be performed cautiously, ensuring that the augmented data adds new information to the training dataset without changing any information associated with the sample's biomolecular information. In summary, the performance of supervised deep-learning models requires a large-paired dataset of high-quality LR and SR images, ensuring models generalize well to unseen data. The challenge of recording large-paired datasets can be overcome by unsupervised models trained on an unpaired dataset, where the LR and HR images depict different fields of view or samples. In 2017, Zhu *et al*. introduced the Cycle-Consistent Generative Adversarial Network (cycleGAN) for image-to-image translation tasks without requiring a paired dataset [15]. The cycleGAN model learns to translate between two domains without requiring direct correspondences between the images in each domain (i.e., circumventing the need for meticulously paired training data) by preserving the sample's structures during the translation process, broadening its applicability to microscopy. As an example, in 2022, Park et al. developed an optimal transport-driven cycleGAN (OT-cycleGAN) to improve the axial resolution in 3D imaging, providing isotropic volumetric confocal and light-sheet microscopy from anisotropic 3D volumes [16]. Although that work improves the axial resolution of both imaging modalities, the lateral resolution improvement was marginal. In this work, we have designed a cycleGAN trained to perform SR confocal microscopy from unpaired LR-recorded confocal images.

The proposed cycleGAN model combines the VGG19 model with custom discriminator and generator architectures to create a robust image-to-image translation system. The VGG19 model, loaded without the top layers and frozen as a fixed feature extractor, helps ensure high-level feature similarity through perceptual loss by focusing on the feature maps produced by the convolutional layers. These feature maps capture essential perceptual characteristics of the input images, ensuring that the generated HR images retain important structural and content details. The discriminator involves a conventional patchGAN with convolutional layers, LeakyReLU activation, and instance normalization. The generator transforms low-resolution (LR) images into super-resolution (HR) images using convolutional layers for downsampling, residual blocks for content preservation, and transposed convolutions for upsampling, incorporating instance normalization and ReLU activations. The composite model integrates the generator, discriminator, and VGG19, training the generator using adversarial, perceptual, and cycle consistency losses to produce realistic images with high-level feature similarity. More details on the implemented cycleGAN model are found in Section 1 of Supplement 1.

The cycleGAN model is trained, validated, and tested using images of U-373MG human glioblastoma cells. Section 2 of Supplement 1 describes the sample protocol used in this case. The experimental confocal fluorescent images were acquired using the Nikon Ti-E confocal microscope at the University of Memphis Integrated Microscopy Center. Twenty-one fields of view of the sample were recorded with a dry 10×/0.3 NA and an oil-immersion 60×/1.4 NA microscope objective lenses to build the unpaired LR and HR datasets. A detailed description of the dataset is provided in Section 3 of Supplement 1. The final dataset contains 1345 LR images and 1345 HR images in each channel (i.e., vimentin-stain and actin-stain images), thereby a total of 5380 (=4 × 1345) images were used to train and validate the model. Both datasets were split by the conventional 80/20 ratio to create the unpaired training and validation datasets.

Due to the unpaired data learning approach in cycleGANs, the performance of the training procedure was quantified using two metrics based solely on the characteristics of each image: the background standard deviation (BGSTD) and a spatial cutoff frequency (CUF) metric. The BGSTD represents the amount of noise present in an image, and it is measured by the standard deviation of the background intensity. The background region in an image can be segmented from the cellular content by thresholding the image using Otsu's method. In fact, the background region is the remaining portion of the image after thresholding. The BGSTD value is the standard deviation of this background region. This metric has been chosen to assess if the model introduces additional noise. The CUF metric quantifies the compact support of the image spectrum, providing a measure of the spatial cutoff frequency in the images. Figure S1 in Section 4 of Supplement 1 shows the spectrum of a pair of LR/HR images and the corresponding CUF metric measurement with a smaller CUF value for LR images. The cutoff frequency marks the circular boundary of the compact support circular region. The compact support consists of the subset of spatial frequencies that carry non-zero frequency components of the image. The CUF is defined as the spatial frequency (in pixels) that encloses 95% of the total energy of the image. The CUF value is normalized by the image dimensions. Plots of the BGSTD and CUF metrics for both training and validation datasets (Fig. S2 in Section 5 of Supplement 1) demonstrate the robustness of the implemented training model. Although the model has been trained up to 100 epochs, both metrics converge after epoch 40 for both the training and validation datasets. The CUF and BGSTD metrics converge at 0.22 and 0.13 a.u., respectively.

Although the BGSTD and the CUF metrics quantify the evolution of the training procedure, they do not objectively measure the accurate conversion from the LR to the HR domains. We have measured three additional metrics for a small, paired testing dataset of 68 images. This paired dataset was compiled by imaging the same two cells for both microscope objective lenses (i.e., LR and HR configuration). We followed the same experimental protocol as the one used for the unpaired dataset. After appropriately scaling the LR images, a FoV matching algorithm was implemented using image registration to geometrically align the LR image with the HR image using OpenCV built-in functions [17]. The three additional metrics computed for this paired dataset are: (1) the Mean Square Error (MSE), (2) the Peak Signal-to-Noise Ratio (PSNR), and (3) the Structural Similarity Index Metric (SSIM). The trained model converges (see Fig. S3 in Supplement 1) after epoch 20 in terms of these three metrics, achieving a low MSE value ($\sim 10^{-4}$), high PSNR value (~25 dB), and a high SSIM value (~0.7 a.u.) on average for images in the paired dataset.

Figure 1 illustrates the performance of the trained cycleGAN model's LR-HR generator on the paired LR/HR dataset. We have reported the SSIM and CUF values in Fig. 1. To quantitatively assess the model's effectiveness, we have included the estimated finals obtained using conventional imaging processing methods and reported deep-learning models. Specifically, we have applied the super-resolution convolutional neural network (SRCNN) [18], regular bicubic interpolation, Neighbor Embedding with Locally Linear Embedding (NE + LLE) [19], Anchored Neighborhood Regression (ANR) [20], Image

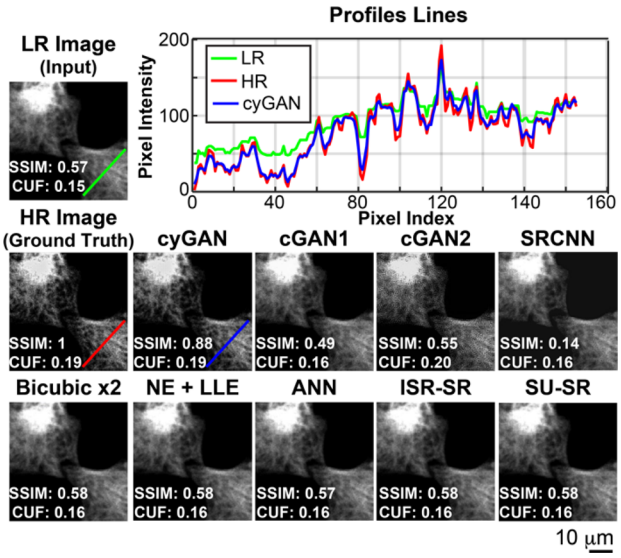

Fig. 1. Comparison of the cycleGAN model (cyGAN) versus other reported super-resolution methods using a paired testing image.

Super-Resolution Via Sparse Representation (ISR-SR) [21], and Scale-Up using sparse representations (SU-SR) [22]. Additionally, we have implemented two common cGAN models reported in the literature. Both cGAN models have the same generator and discriminator sub-models as our cycleGAN model. The first cGAN model (cGAN1) was trained on a simulation dataset. The simulated LR and HR confocal images were created by convolving experimental confocal images from the Kinome Atlas in the Cell Imaging Library with two Point Spread Functions corresponding to the two experimental conditions. Because experimental confocal images are affected by Poisson noise, the simulated LR and HR images were degraded using a Poisson-distributed noise, setting the SNR to 45 dB. The paired dataset in the second trained cGAN (cGAN2) model consisted of experimental HR images and their corresponding degraded LR images as Ref. [6] describes. In summary, the degradation model to generate the LR dataset firstly applies a Gaussian blur with a standard deviation value of 1 pixel. Then, we reduced the resolution by downsampling the degraded HR images by a factor of 2. Finally, the down-sampled images are then up-sampled back to their original dimensions. Whereas supervised methods should outperform unsupervised ones, the performance of both cGAN models is poorer than the proposed cycleGAN model because they were trained with cells of a totally different structure. The results (i.e., predicted image and CUF and SSIM values) in Fig. 1 demonstrate the superior performance of the proposed model. Whereas the CUF metric reflects changes in spatial frequency spectra, its value may not perfectly align with a subjective visual assessment of the resolution improvement. In fact, the CUF metric is highly sensitive to noise, leading to higher CUF values for the cGAN1 model since it produces noisier images.

Figure 2 illustrates an example where the performance of cycleGAN model is poorer (SSIM value = 0.53) compared to Fig. 1, showing significant discrepancies between the ground-truth image and the model's prediction. Although the attained SSIM value for the proposed cycleGAN model is not ideal, it is still superior to the performance of the other methods as Table 1 reports. This non-ideal result may be related to the

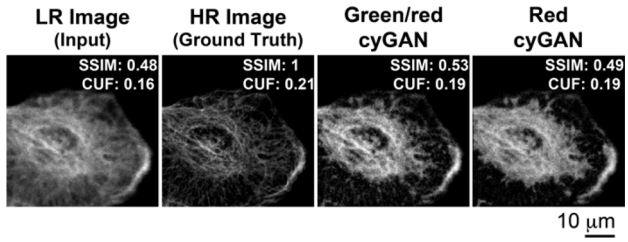

Fig. 2. Example where the cycleGAN model performs poorly.

Table 1. Quantitative Evaluation of the Models' Performance for the FoV Shown in Fig. 2

	SSIM (a.u.)	CUF (a.u.)
LR	0.48	0.16
HR	1.00	0.21
cyGAN	0.53	0.19
ISR-SR	0.49	0.16
NE + LLE	0.49	0.16
SU-SR	0.49	0.16
Bicubic	0.48	0.16
ANR	0.47	0.16
cGAN1	0.45	0.20
cGAN2	0.40	0.17
SRCNN	0.14	0.15

presence of blur in the input LR image due to the reduced optical sectioning capability of the LR objective lens compared to the one provided by the HR objective lens. Nonetheless, as Fig. 2 shows, the proposed cycleGAN model performs better when it is trained with the two-channel vimentin-stain and actin-stain dataset (green/red cyGAN in the figure) than the one trained exclusively on the vimentin-stain dataset (red cyGAN in the figure).

To fully assess the performance of the proposed cycleGAN model compared to all reported methods, we have performed a statistical analysis on the SSIM and the CUF values for 12 paired HR/LR images, which were randomly selected from the paired testing dataset. The results are presented in Table 2. The HR results, being the ground truth data, serve as the benchmark. Again, our results show that the proposed cyGAN model significantly outperforms the other methods, achieving an SSIM maximum value of 0.96, an SSIM mean value of 0.73, and a CUF mean value of 0.208. The lowest-performing method in terms of SSIM mean value is the SRCNN, with a value of 0.17. The CUF metric shows a similar trend, with the cyGAN model achieving the highest CUF mean value of 0.208, followed closely by the cGAN2 model at 0.196, while the other methods, such as NE + LLE and ANR, show significantly lower performance with average CUF values of 0.157 and 0.160, respectively. This table highlights the superiority of the cyGAN model in both SSIM and CU metrics, indicating its effectiveness in predicting high spatial frequencies and generating super-resolved images from LR images.

Finally, we predicted the SR image from a full-FoV LR acquisition for both the red and green channels in Fig. 3. The green and red channels were independently input into the network, generating independent outputs which are later combined. In this validation example, the model enhances the CUF, increasing from 0.16 to 0.21 a.u. in the red channel and from 0.15 to 0.20 a.u. in the green channel. Additionally, the BGSTD

Table 2. Comparative Analysis of the Proposed CycleGAN Model and Other Reported Methods for 12 Random Images of the Paired Testing Dataset

	SSIM (a.u.)				CUF (a.u.)			
	Mean	STD	Max	Min	Mean	STD	Max	Min
LR	0.50	0.16	0.77	0.26	0.170	0.010	0.179	0.154
HR	1.0	0.0	1.0	1.0	0.220	0.008	0.222	0.142
cyGAN	0.73	0.19	0.96	0.39	0.208	0.006	0.221	0.152
cGAN1	0.56	0.12	0.75	0.35	0.196	0.002	0.198	0.193
cGAN2	0.53	0.15	0.74	0.23	0.162	0.007	0.170	0.149
ANR	0.51	0.16	0.77	0.26	0.160	0.005	0.165	0.150
Bicubic	0.51	0.16	0.77	0.26	0.157	0.008	0.166	0.141
ISR-SR	0.51	0.16	0.77	0.26	0.157	0.008	0.166	0.141
NE + LLE	0.51	0.16	0.77	0.26	0.157	0.008	0.166	0.141
SU-SR	0.51	0.16	0.77	0.26	0.157	0.008	0.166	0.141
SRCNN	0.17	0.09	0.35	0.07	0.155	0.005	0.162	0.147

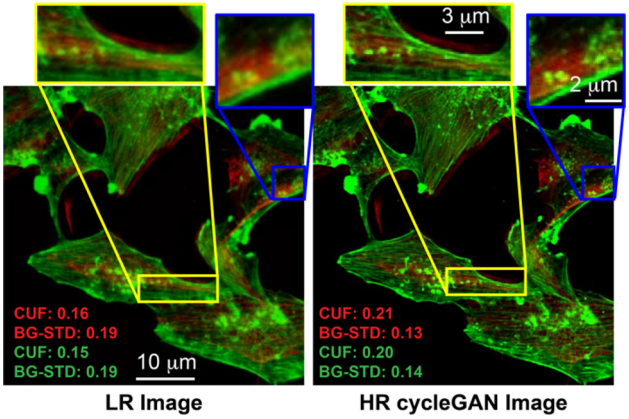

Fig. 3. Evaluation of the cycleGAN model on an unpaired LR image, demonstrating enhanced visualization of sample details.

decreased, from 0.19 to 0.14 a.u. in the green channel and from 0.19 to 0.13 a.u. in the red channel. The zoom-in regions in Fig. 3 confirm the HR image shows improvements in the individual visualization of invadopodias, which are structures that cancer cells use to invade tissues. Whereas all invadopodias seem to be colocalized in the LR image, showing an orange-greenish appearance, we can only identify particular structures that are colocalized between both channels. The proposed model also provides an improvement in the colocalization of the stress fibers shown in the yellow inset, differentiating between actin and vimentin cytoskeletal systems. These results validate our model's capability to effectively transform LR confocal acquisitions, enhancing spatial frequency content, reducing background noise, and providing improved visualization of the imaged sample's details.

In summary, this study evaluates the performance of an unsupervised cycleGAN model to generate super-resolved confocal images from LR images, overcoming the experimental challenge of acquiring a paired dataset by using an unpaired dataset. The proposed cycleGAN model has effectively learned to translate between low- and high-resolution images without the need for precise spatially matched pairs. The performance of the cycleGAN model is consistently superior to the results from reported processing and deep-learning imaging methods. The proposed cycleGAN model is well poised to overcome current instrumentation limitations by improving the resolution capability of our laboratory devices using advanced deep-learning methods. For example, after training the cycleGAN model with more generalized data and applying transfer learning approaches, LR confocal images from a laboratory-built optical imaging system can be computationally upgraded to HR commercial confocal microscopes by means of unsupervised deep learning without any funding investment. Future work should explore optimizing the cycleGAN model for a wider range of imaging conditions, incorporating additional types of noise and artifacts, and extending the learning-based model to other imaging modalities, including cross-modality imaging.

Funding. National Science Foundation (2404769, 2042563); Universidad EAFIT (Fundamental Sciences area, Vicerrectoría de Ciencia, Tecnología e Innovacion).

Acknowledgment. The authors thank Shashwat Patra for his contribution to the initial implementation of the unpaired model. We also acknowledge the Integrated Microscopy Center (IMC) staff at the Univ. of Memphis for providing support and advice during the confocal imaging.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [23].

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. C. Dong, C. C. Loy, K. He, *et al.*, IEEE Trans. Pattern Anal. Mach. Intell. **38**, 295 (2016).
2. X. Yuan, J. Shi, and L. Gu, Expert Syst. Appl. **169**, 114417 (2021).
3. L. Tian, B. Hunt, M. A. L. Bell, *et al.*, Lasers Surg. Med. **53**, 748 (2021).
4. C. Qiao, D. Li, Y. Guo, *et al.*, Nat. Methods **18**, 194 (2021).
5. H. Wang, Y. Rivenson, Y. Jin, *et al.*, Nat. Methods **16**, 103 (2019).
6. X. Li, J. Dong, B. Li, *et al.*, in *IEEE International Conference on Computational Photography* (2020).
7. W. Wang, B. Wu, B. Zhang, *et al.*, Opt. Lett. **46**, 4932 (2021).
8. H. Zhang, C. Fang, X. Xie, *et al.*, Biomed. Opt. Express **10**, 1044 (2019).
9. L. Jin, B. Liu, F. Zhao, *et al.*, Nat. Commun. **11**, 1934 (2020).
10. E. Nehme, L. E. Weiss, T. Michaeli, *et al.*, Optica **5**, 458 (2018).
11. W. Ouyang, A. Aristov, M. Lelek, *et al.*, Nat. Biotechnol. **36**, 460 (2018).
12. S. Shalev-Shwartz and S. Ben-David, in *Understanding Machine Learning: From Theory to Algorithms*, 1st ed. (Cambridge University Press, 2014).
13. I. Goodfellow, Y. Bengio, and A. Courville, in *Deep Learning*, 1st ed. (MIT Press, 2016).
14. X. Wang, in *Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science*, L Leal-Taixé and S. Roth, eds., Vol. 11133, pp. 63–79.
15. J.-Y. Zhu, T. Park, P. Isola, *et al.*, in *2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017), pp. 2242–2251.
16. H. Park, M. Na, B. Kim, *et al.*, Nat. Commun. **13**, 3297 (2022).
17. B. Zitová and J. Flusser, Image Vis. Comput. **21**, 977 (2003).
18. X. Ji, Y. Lu, and L. Guo, in *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)* (IEEE, 2016), pp. 626–630.
19. S. T. Roweis and L. K. Saul, Science **290**, 2323 (2000).
20. R. Timofte, V. De, and L. Van Gool, in *2013 IEEE International Conference on Computer Vision* (IEEE, 2013), pp. 1920–1927.
21. J. Yang, J. Wright, T. S. Huang, *et al.*, IEEE Trans. on Image Process. **19**, 2861 (2010).
22. R. Zeyde, M. Elad, and M. Protter, Curves and Surfaces **6920**, 711 (2012).
23. C. Trujillo, L. Thompson, O. Skalli, *et al.*, "CycleGAN for Confocal Microscopy Image Translation," GitHub (2024) [retrieved 30 September 2024]https://github.com/catrujilla/cycleGAN-SR_confocal.