Providing Real-World Benchmarks for Super-Resolving Fluorescence Microscope Imagery Using Generative Adversarial Networks

1st J. Cooper
Department of Mathematics and Statistics
San José State University
San José, CA, USA
johnclinton.cooper@sjsu.edu

2nd T. B. Issa
Department of Mathematics and Statistics
San José State University
San José, CA, USA
tahirbachar.issa@sjsu.edu

3rd C. Vinegoni
Center for Systems Biology
Harvard University
Boston, MA, USA
cvinegoni@mgh.harvard.edu

4th R. Weissleder
Department of Systems Biology
Harvard University
Boston, MA, USA
rweissleder@mgh.harvard.edu

***Abstract*—Super-resolving and de-noising video-rate acquisition microscope imagery can eliminate the need for practitioners to choose between image acquisition speed and image quality. This task provides a convenient use case for Generative Adversarial Networks (GANs), which have demonstrated impressive pixel-wise reconstruction metrics in microscope image super-resolution tasks. However, the benchmarks which report these metrics typically do so on low-resolution images generated through simple synthetic degradations, which fail to recreate the natural distortions imparted by experimental image acquisition. Because most reported synthetic data generation pipelines rely on the bicubic downsampling of a few high-resolution images, networks trained to correct for this distortion eventually fail downstream when required to super-resolve images containing natural distortions. The literature therefore provides us with an unreliable assessment of GANs' ability to super-resolve microscope imagery in the field. In this work, we present one of the few examples of GANs successfully super-resolving a large cache of experimentally gathered microscope imagery. For our main result, we demonstrate a reliable baseline for the super-resolution task using GAN, in which we obtain a peak-signal-to-noise ratio (PSNR) of 29.21 and a structural similarity index (SSIM) of 0.845 by using all available image pairs and averaging across all sub-datasets. To demonstrate robustness on this task, we present the model with a blind super-resolution task, in which it achieves a PSNR of 25.75 and SSIM of 0.676 after averaging across all sub-datasets. To affirm our results as a reliable baseline, we demonstrate that GANs can fail in the video-rate super-resolution task even when trained on higher-order synthetic degradation pipelines. We confirm this effect by training our model on purely synthetic data, using the pipeline mentioned above, and testing it on a single sub-dataset. In doing so, we observe a -0.06 loss in SSIM and -0.75 loss in PSNR, accompanied by significant quality degradation of the reconstructed images in the form of severe distortion and artifact generation.**

***Index Terms*—GAN, PSNR, SSIM, super-resolution, bicubic downsampling**

I. Introduction

Generative Adversarial Networks (GANs) have shown to be effective tools in generating high-resolution synthetic imagery and super-resolving low-resolution imagery. Specializing in the latter are super-resolution networks such as Super-Resolution Generative Adversarial Networks (SRGan) [9], Enhanced Super-Resolution Generative Adversarial Networks SRGan(ESRGan) [17], and Real Ehanced Super-Resolution Generaative Adversarial Networks [15]. These models are typically applied to the super-resolution of general, un-specialized image data sets. Both of these networks, for example, are trained and tested on the DIV2k [1], FLICKR [10], and OST300 [16] data sets. However, they are more infrequently tasked with the super-resolution of medical imagery, and microscope imagery in particular, seemingly and simply due to a scarcity of data availability. For the handful of instances in which SRGan variants have been applied to super-resolve such specialized images, the authors have been required to make small data sets work for their experiments. Some have done this by applying simple down-sampling techniques (bi-cubic, bi-linear) to high-resolution images to form their low-resolution pairs [2], [4], [5], [14], while others have employed more complex degradation models to mimic the noise distribution of a specific low-resolution imaging source [19]. The data scarcity of high-resolution-low-resolution microscope image pairs in super-resolution tasks, however, persists within this field of research as a major challenge to the interpretability and generalizability of the results produced. In lieu of a sufficient number of ground-truth image pairs, researchers meet a fork in the road: they must either train a model from scratch on limited data, or train and test on synthetically down-sampled data. Both of these choices confuse the interpretation and ap-

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00206

plicability of research outcomes. In this work, we do not claim to introduce any novel improvement to the super-resolution task. Instead, we apply conventional techniques to a relatively large cache of low-high-resolution image pairs. By doing so, we provide one of the few real-world benchmarks for super-resolving microscope imagery using GAN-based methods. Moreover, through our attempt to super-resolve experimentally gathered microscope imagery using a model trained solely on high-order synthetic degradation pipelines, we highlight the problematic nature of current synthetic data generation approaches used to compensate data scarcity.

II. Related Work

With paired-image scarcity being such a persistent problem in microscope image super-resolution, current research seems to obtain their low-resolution pairs in one of five ways: applying simple downsampling to high-resolution images [2], [4], [5], [14], [27], [28], [33], employing sophisticated generative networks or degradation pipelines [19], [22], [31], [34], experimentally collecting them [7], [26], [29], [30], [32], or using a combination of experimental and generated low-resolution images [20].

Perhaps the most closely related to our work is that of [32], which introduces a feedback (recurrent) module to the GAN's generator that reduces network overhead and produces iterative losses as features are cycled through the module. A novel Fourier frequency transform loss is also introduced to be used alongside the typical perceptual, generator, and discriminator losses. Their training strategy consists of a progressive super-resolution task in which the model is trained to resolve low-resolution 256×256 images to high-resolution images 1024×1024 for the first dataset. This same model is then trained to resolve low-resolution 512×512 images of the second dataset into high-resolution 2048×2048 images. Both datasets comprised 901 image pairs in total. Importantly, they demonstrate that their method is able to outperform ESRGan [17] across all super-resolution tasks. We note that changes made to the architecture and augmentation pipeline of ESRGan to produce RESRGan produce significant gains in perceptual quality metrics [15], [17].

Others eschew experimentally gathered paired images in favor of synthetic degradation pipelines. In certain instances, these efforts have demonstrated that one can use a relatively simple degradation pipeline, such as additive Gaussian noise, to generate low-resolution data for the super-resolution task and ultimately obtain higher signal-to-noise ratios than even the ground-truths [34]. The quality of these resolved images have also been proven to correlate closely to expert-analysis [22], [34]. Others employ physics-inspired generative pipelines, such FluoGAN, to achieve expert-approved super-resolved images or high-throughput enhancement [19], [20], [31].

Despite these efforts to overcome the paired-image problem, innovative methods nonetheless require the results produced from super-resolution benchmarks using experimentally-gathered paired-image data to demonstrate the scope of their successes. Current literature attempts to benchmark these results, and does so with impressive performance metrics, but frequently only to the extent that their methods may be applied to the super-resolution of synthetically-degraded images. Even if these methods are tested on real-world low-resolution data, there often isn't a way to benchmark the success of the proposed method without also doing the very thing which the method set out to preclude - laboriously obtaining low-resolution pairs. This even applies to successful efforts, such as [34], which produce images with high signal-to-noise-ratio but also oversmoothing.

The results of many interesting architectural changes or training methodologies are undermined, at least temporarily, by their dependence on simple degradation pipelines. One method, for example, applies parameter interpolation between a natural-image pre-trained ESRGan and a microscope-image trained ESRGan to produce impressive super-resolution performance metrics [4] without re-training their model. Yet their training and testing pipeline use bi-cubic downsampled data, which measures the ability of the network to resolve bi-cubic distortions. Even within the pre-trained network itself, an ESRGan, bi-cubic down-sampling is used in the pre-train pipeline degradation pipeline [17].

Another method proposes small architectural changes to SRGan in order to reduce checkerboard-pixelation patterns in generated images, overfitting, and computation time. [2], and succeeds in doing so. But by training and testing on a large microscope image database, The Human Protein Atlas [23], the network trains on bicubic down-sampled low-resolution images. In this case, the data itself sets a too-easy standard for super-resolution tasks. Another method improves SRGGAn's GAN and discriminator loss to improve robustness, but still reports on down-sampled test sets [14].

[19], one of the aforementioned models which uses physics inspired down-sampling, employs an SRGan to train an instantaneous, on-chip super-resolver for microscope imagery. However, due to the explored image acquisition method, low-resolution counterparts for each high-resolution image were unable to be. They therefore simulated low-resolution images from high-resolution images based on the parameters of their low-resolution acquisition system, using a simulating model based on [18]. They prove that such a method can work on both simulated and non-simulated low-resolution images. Their results demonstrate relatively, but not exceedingly, high performance metrics which seem to be characteristic of testing on experimentally gathered low-resolution images. That is, this work demonstrates less impressive results when its methods are actually tested on real-world low-resolution images. It seems that models trained and tested on experimentally-gathered low-resolution images may perform worse than models synthetically generated data [5].

Within the current literature, there are a few instances of model improvements on large paired-image data sets. One instance, however, is demonstrated in [7], which applies GAN to super-resolve video-rate microscope images. Notably, [7] applies their network to a relatively large set (5600 for

training) of experimentally gathered, high resolution confocal microscope images. They obtain one-hundred experimentally gathered low-resolution (video-rate) images per high-resolution image, thus allowing a faithful testing of the ability of deep super-resolution networks on excessively noisy real-world microscopy data. In doing so, they demonstrate impressive and reliable results for benchmarking super-resolution tasks on real-world microscope imagery.

III. Data

To illustrate the range of our model we employ four data sets to work with which we call MITO, MPM, MEM, and NUC, and stand for "mitochondria", "multiphoton mitochondria", "membrane", and "nuclei", respectively, comprising an extension of the data presented in [7]. MITO, MEM, and NUC are acquired through single photon confocal scans, and MPM is acquired via multiphoton scans. Together they comprise a total of 25,452 images of 252 high-resolution images and 25,200 low-resolution images. Individually they contribute a somewhat imbalanced number of images to the overall set, as seen in Table I.

TABLE I
Composition of (paired) data for training

Data	Train	Validation	Test	Total
MITO	3600	1100	300	5000
MPM	3600	1100	2400	7100
MEM	3600	1100	1900	6600
NUC	3600	1100	1800	6500

Both high and low resolution images were obtained from an Olympus FV1000 system employing a confocal microscope [7]. High-resolution images are acquired at a 512×512 resolution using a unidirectional scan with a frame rate of 0.08 frames per second, while low-resolution images are acquired at a 256×256 resolution using a bi-directional scan with a frame rate of 20 frames per second [7]. Both low and high resolution images are affected by various sources of noise, including system-induced noise, scattering, and auto-fluorescence signal noise, with the prior experiencing higher rates of distortion. However shot noise and dark noise present as the major sources of quality degradation in the low resolution image acquisitions [7]. Their effects are assumed to comprise a significant proportion of the noise seen in the low-resolution images [7], as seen Fig. 1.

IV. Methods

A. RESRGan

RESRGan, or Real Enhanced Super Resolution GAN, descends from the Super-Resolution GAN family [9], [15], [17]. It uses a "residual in residual dense block" generator [17], a a U-Net discriminator [12], [13], [15] with ten layers, spectral normalization, and skip connections. It also employs relativistic [8] and perceptual loss, as well as perception-PSNR based network interpolation which balances perceptual and pixel-wise image fidelity [15]. During its training, RESRGan introduces a complex degradation process in two stages. In the first stage, several forms of degradation are applied to the image; namely, isotropic and anisotropic Gaussian blurring, 2D sinc filtering, bicubic, bilinear, area resizing, Gaussian noise, poisson noise, color noise, gray noise, and JPEG compression . In the second stage, called "higher-order degradation", a second layer of blurring, resizing, noise, and JPEG compression and 2D sinc filtering are applied. This complex degradation process is meant to produce a blind super-resolver for real-world imagery in RESRGan [17]. Unsurprisingly, it demonstrates distinct visual and metric-based improvements over its predecessor. We refrain from explicating all modifications to the original GAN architecture [23] which RESRGan makes, as this model's deployment is well documented in super-resolution tasks and does not stray from the basic workflow of GAN [23].

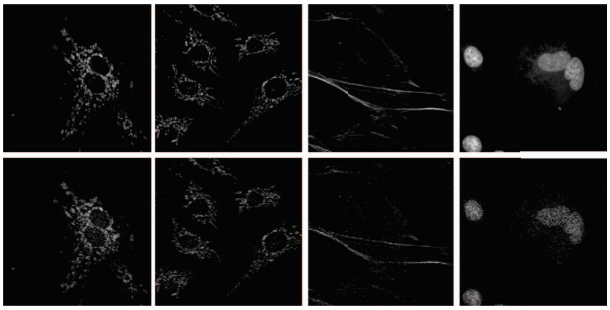

Fig. 1. Examples of low-resolution-high-resolution paired images from the working data set. From left-to-right: MITO, MPM, MEM, NUC. Best viewed on a computer.

B. Perceptual Image Quality Metrics

The use of perceptual image quality metrics play a crucial role in assessing generated image quality for the super-resolution GAN family [15], [17]. In particular, ESRGan and RESRGan use the Perceptual Index [3], a linear combination of Ma's score [25] and the NIQE score [11], to judge the naturalness of generated images. They propose using this score to balance the pixel-wise quality metrics which tend to produce unnatural artifacts in generated images. We omit the use of the Perceptual Index used in favor of two separate image quality scores, the Naturalness Image Quality Evaluator (NIQE) [11] and the Frechet Inception Distance (FID) [6]. We note that in using these scores to assess microscopy images, we depart from the natural image distribution which both metrics derive from. The scales of our scores are therefore different from those of [17]; in fact, they are much higher, suggesting poorer image quality overall. However, we notice that NIQE and FID seem to be highly correlated with more faithful reproductions of ground-truth images. We therefore use them to guide our hyper parameter selection, as suggested in [17] and [15].

C. Training Scheme

We train two models total. For the first, we employ a RESRGan model, pre-trained on natural image data sets, to fine-tune on the full MITO data set. We then use this newly fine-tuned model to separately finetune on the rest of the data sets. That is, the individual fine-tuning processes for NUC, MPM, and MEM begin fresh from the model fine-tuned on MITO. Our second model is used to test the performance of a model trained only on synthetic data. For this model, we train on the high-resolution images of the MITO data set, using the complex synthetic data generation pipeline of RESRGan to create low-resolution ground truths. We do not perform any few-shot testing for this model.

Respectively, we label them RESRGan-Base (RESRGan-B) and RESRGan-Synthetic (RSRGan-S). For the base model, we take the approach of [7] in testing the capability of the model on few-shot tasks. We do so by selecting random subsets of each training data set consisting of 200, 400, and 800 paired images. This idea is illustrated in Fig 2. Note how this process spawns successive permutative training schemes across all few-shot and full-shot tasks and data sets.

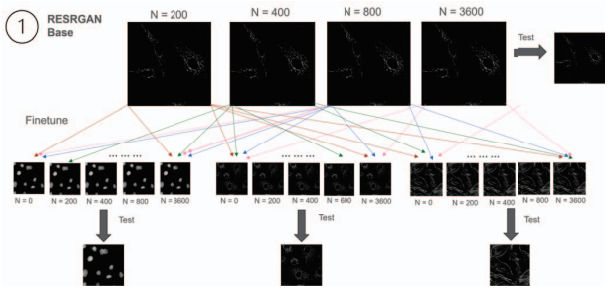

Fig. 2. Successive Fine-Tuning Explicated

This differs from that of [7], as we test the robustness on 200 paired images for all data sets. Moreover, we decrease the baseline number of total paired training images from 5600 to 3600 for the full model, and increase our validation size from 400 to 1100. All models are trained on a maximum of 10000 iterations, with an early stopping of 500 iterations for low-to-medium learning rates and 1000 iterations (as a warm up) for high learning rates. We use no data augmentation techniques. The overall process of our training strategy is outlined in Fig. 3.

D. Objective

In the previous section, we identified RESRGan-B and RESRGan-S as the two models which we trained on purely experimental and purely synthetic data, respectively. The purpose of the first experiment is to continue expanding baselines for GAN-based microscope image super-resolution tasks, by grounding their results in large sets of experimental image-pairs. While this benchmark is specific to video-rate laser scanning confocal microscope imaging, it's able to

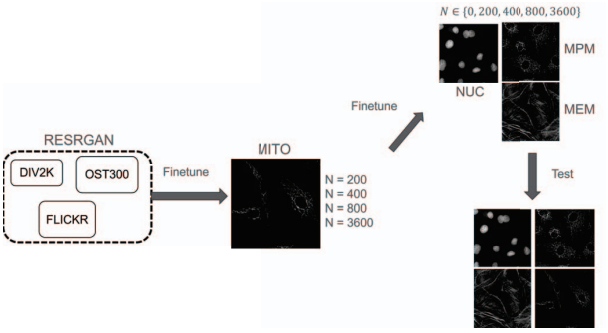

Fig. 3. Training Strategy Overview

set standards for GAN-based super-resolution specific to (1) the severity of noise induced by video-rate acquisitions, (2) the effect of natural image model pre-training, and (3) the amount of experimental data necessary for successful super-resolution. Our hope is to aid practitioners in benchmarking novel methods that move us away from a dependence on painstakingly acquired low-resolution data. In this way, we can alleviate researchers, pursuing things such as unsupervised synthetic data generation, of the burden of having to re-train and benchmark their models on experimental data that doesn't exist.

The purpose of the small experiment using RESRGan-S is to demonstrate how despite training on the complex and compounded synthetic data degradation pipeline of RESRGan, which was built as a blind super-resolver for generalized noise [15], we're still unable to capture the noise distribution of video-rate imaging. While several works have proposed well-formulated physics-grounded degradation or generative pipelines [19], [20], [31], we continue to see a large number of efforts to validate on bicubic downsampled data. We hope to provide evidence here that this, in fact, does not generalize.

V. Results

Here, we report the results for RESRGan-B and RESRGan-S, the training methodologies of which are described in the previous section. We note that RESRGan-B is the best performing model, with RESRGan-S experiencing significant performance degradation.

A. RESRGan-B

For all results in this section, we refer to RESRGan-B as the model fine-tuned on the MITO dataset, i.e. the base model. If RESRGan-B is followed by a number, this means that the base model was fine-tuned on NUM $\times$ 100 paired training images from the respective data set. RESRGan-BFULL refers to the base model fine-tuned on all paired training images from the respective data set.

Note that when RESRGan is present under another data sets test performance results, that means we are testing the base model's zero-shot capabilities on that data set. That is, we

test the MITO-trained model on a completely unseen data set without any extra fine-tuning. In the special case where we fine-tune the model on a subset of the MITO data, we label this RESRGan followed by a number, e.g. RESRGan-2. Also note that all fine-tuning was conducted with identical hyper parameters to those of the base MITO-tuned model. Namely, fine-tuning was conducted uniformly with learning rate 5×10^{-4} and cosine-annealing scheduling, patch size 128×128, loss weights $\lambda = 1 \times 10^{-1}$ and $\eta = 1$, and a 10,000 iteration runtime.

Overall, the performance trends are somewhat surprising. We notice either marginal improvements in performance when using more training data, or even slight degradation. In most cases, we see that moving from a zero-shot setting to two-hundred training pairs notably improves performance metrics. However, there is ambiguity in performance gains between using four hundred training pairs and the full data set. Although SSIM seems to converge towards a stable result with access to the full data set, we continue to see fluctuations in PSNR.

Generated images were judged not only on PSNR or SSIM, but on perceptual quality metrics. Namely, we use FID and NIQE to fine-tune the hyperparameter selection process, and show that the incorporation of these metrics does, in fact, seem to lead to more visually faithful reproductions of high-resolution ground-truths. For example, the model with a learning rate of 5e-4, a 64×64 patch size, and weighted loss, which was validated on SSIM, obtained the highest PSNR among all models. However, with a NIQE equal to 8.12 ± 0.044, it seemed to produce overly smooth or blurry images with unwanted artifacts, or images which lacked small but important details as seen in Fig. 4. Overall, we observe this trend for models which produce high PSNR/SSIM as well as high FID/NIQE.

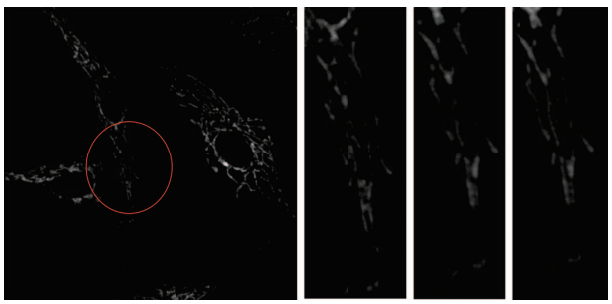

Fig. 4. From left-to-right: (1)-(2) the ground-truth image and the area which is being viewed, (3) the cropped area from the model with the highest PSNR/SSIM, (4) the cropped area from our best model

The best model's performance on the MITO test set, after being fine-tuned on only MITO, is seen in II. We see in Fig. 5 the production of nearly-identical images.

For MPM, it's notable that while performance seems to improve with the amount of training data, it only does so marginally. Moreover, the model doesn't obtain a maximum PSNR when trained on the full data set. We suspect that

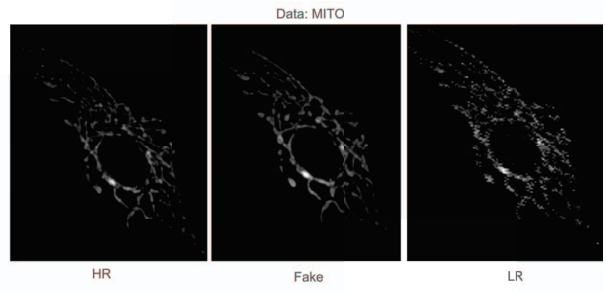

Fig. 5. RESRGan-B on MITO

the structural similarity shared between MPM and MITO contributes to this "plateauing" effect in MPM. Despite these performance metrics, from a perceptual perspective the images produced are nearing excellent, aside from a very small amount of blurring in the generated image. Figure 6 shows a comparison between ground-truth, low-resolution, and generated image. Looking closely, one can see both blur and structural degradation in areas containing the most fine-grained features.

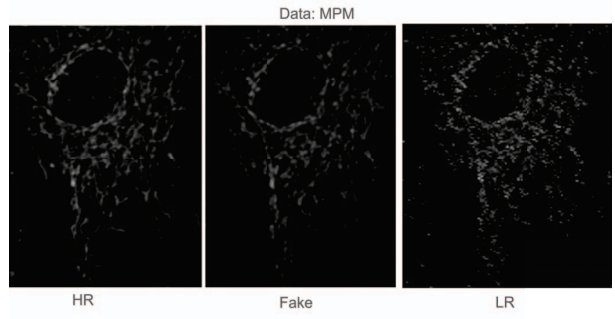

Fig. 6. RESRGan-B on MPM

We also notice an interesting trend in the performance on the NUC data set. It demonstrates (1) the largest content performance gains between the zero-shot performance and the fully pre-trained performance, and (2) the highest content performance among all fine-tuning episodes. However the fully fine-tuned model suffer in terms of perceptual quality, outputting a NIQE of 9.75. This is in contrast to, for example, the fully fine-tuned MEM model which outputs a NIQE of 5.42, but seems to suffer from lower content metrics as seen in IV. Note the image generation fidelity of MEM in Fig. 7.

However, as seen in Fig. 8, there are clear problems with the images generated on the fully fine-tuned NUC model. Specifically, the model seems to identify features as noise, and almost seems to be randomly guessing where the features are.

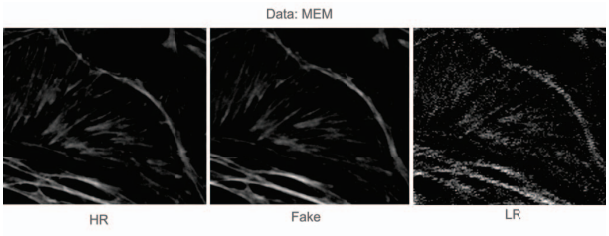

Fig. 7. RESRGan-B on MEM

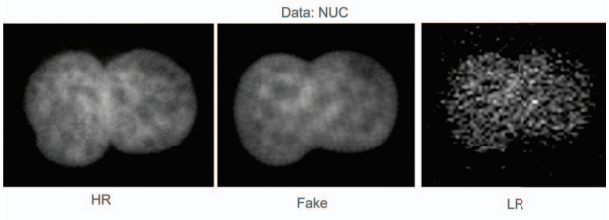

Fig. 8. RESRGan-B on nuc

TABLE II
RESRGAN-B TEST PERFORMANCE ON MITO ACROSS FINE-TUNING EPISODES

DATA	MODEL	SSIM	PSNR
MITO	RESRGAN-2	0.860±0.002	27.87±0.119
	RESRGAN-4	0.876±0.001	29.21±0.100
	RESRGAN-8	0.864±0.001	28.49±0.099
	RESRGAN-B	0.887±0.001	29.47±0.124

TABLE III
RESRGAN-B TEST PERFORMANCE ON MPM ACROSS FINE-TUNING EPISODES

DATA	MODEL	SSIM	PSNR
MPM	RESRGAN-B	0.853±0.001	28.26±0.019
	RESRGAN-B2	0.856±0.001	28.33±0.018
	RESRGAN-B4	0.862±0.001	28.47±0.018
	RESRGAN-B8	0.861±0.001	28.55±0.018
	RESRGAN-BFULL	0.863±0.001	28.50± 0.018

TABLE IV
RESRGAN-B TEST PERFORMANCE ON MEM ACROSS FINE-TUNING EPISODES

DATA	MODEL	SSIM	PSNR
MEM	RESRGAN-B	0.705±0.001	26.17±0.034
	RESRGAN-B2	0.728±0.001	26.95±0.039
	RESRGAN-B4	0.743±0.001	28.05±0.032
	RESRGAN-B8	0.745±0.001	28.01±0.026
	RESRGAN-BFULL	0.752±0.001	28.13±0.031

TABLE V
RESRGAN-B TEST PERFORMANCE ON NUC ACROSS FINE-TUNING EPISODES

DATA	MODEL	SSIM	PSNR
NUC	RESRGAN-B	0.676±0.001	22.83±0.037
	RESRGAN-B2	0.867±0.001	30.73±0.045
	RESRGAN-B4	0.875±0.001	31.42±0.041
	RESRGAN-B8	0.877±0.001	30.67±0.047
	RESRGAN-BFULL	0.876±0.001	30.74±0.047

B. RESRGan-S

The impressiveness of RESRGan is captured by its RRDB network, its powerful and relativistic discriminator, and its pre-activation perceptual loss. Yet it also introduces a complex data synthesizing pipeline used to facilitate the blind super-resolution of distorted real-world images [15]. By using limited paired-image data sets to train RESRGan we essentially, but almost necessarily, limit the model's power in super-resolving noisy images.

The issue in using synthetically generated data for our purposes lies in the possibly large distance between the noise distribution of un-specialized "real-world" data and the photon noise present in our microscopy data. Because RESRGan proposes a generalizable blind super-resolver whose power lies in its higher-order degradation method, we test to see if these degradations are able to capture the distribution of photon noise that's particular to video-rate microscope imagery acquisiton. Because RESRGan includes photon noise in its degradation process, we were curious to see if more severe degradations of the high-resolution images were capable of capturing the subsetted noise distribution present in the original low-resolution images. To this end, we omit the use of paired images, and only use synthetically degraded high-resolution images for low-resolution image pairings. We both validate and test the model on the original, true low-resolution images.

For this experiment, we only extend the tests to the MITO data set, as the results observed were much less performant than in other models. As seen in Figure 9, the model trained only on synthetically degraded ground-truths fails when tested on the original low-resolution images.

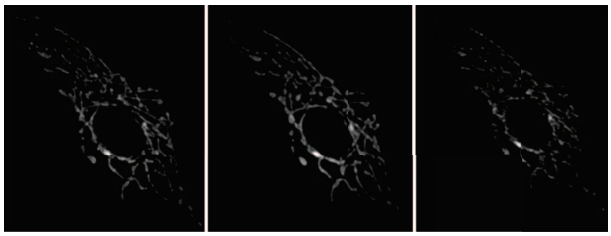

Fig. 9. Cropped ground-truth (left) compare to RESRGan-B (middle) and RESRGan-S (right)

We speculate that although photon noise is added to the ground-truths to generate synthetic images, that either this distribution of photon noise does not approximate that which is present in the low-resolution images, or that the conjunction of various types of degradations fails to super-set the noise present in the low-resolution images. Although the performance metrics in Table VI does not seem all too poor, the image itself, as well as the perceptual metrics indicate otherwise. We observe a NIQE of 7,14 but a FID of 104.91, the latter of which is roughly 30 points higher than most of our other models.

TABLE VI
TEST PERFORMANCE ON MITO (SYNTHETIC) VERSUS REAL

DATA	MODEL	SSIM	PSNR
MITO	RESRGAN-S	0.83±0.002	28.72±0.109
MITO	RESRGAN-B	0.887±0.001	29.47±0.124

VI. CONCLUSION

We use a state-of-the-art super-resolution network to train and test on a large cache of experimentally-gathered fluorescence microscope images, providing a benchmark for future improvements to super-resolution tasks in fluorescence microscope imaging.

We test whether the complex degradation process employed by RESRGan, and utilized in blind super-resolution of natural images, captures the shot noise of video-rate micro- scope images. We demonstrate that such a network fails to resolve experimentally-gathered low-resolution images even when the degradation pipeline used is significantly more complex than bi-cubic or bi-linear down-sampling.

We justify the need for either (1) bypassing paired-image training completely, (2) using degradation pipelines which simulate instrument noise, or (3) using larger experimentally-gathered paired data sets when reporting results.

REFERENCES

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 1122–1131, 2017.

[2] Vibhu Bhatia and Yatender Kumar. Attaining real-time super-resolution for microscopic images using gan, 2020.

[3] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. CoRR, abs/1711.06077, 2017.

[4] Manna Dai, Gao Xiao, Lance Fiondella, Ming Shao, and Yu Shrike Zhang. Deep learning-enabled resolution-enhancement in mini- and regular microscopy for biomedical imaging. Sensors and Actuators A: Physical, 331:112928, 2021.

[5] Orkun Furat, Donal P. Finegan, Zhenzhen Yang, Tom Kirstein, Kandler Smith, and Volker Schmidt. Super-resolving microscopy images of li-ion electrodes for fine-feature quantification using generative adversarial networks. npj Computational Materials, 8(1), April 2022.

[6] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, G unter Klambauer, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a nash equilibrium. CoRR, abs/1706.08500, 2017.

[7] T. B. Issa, C. Vinegoni, A. Shaw, P. F. Feruglio, R. Weissleder, and D. Uminsky. Video-rate acquisition fluorescence microscopy via generative adversarial networks. In 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), pages 569–576, 2020.

[8] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard GAN. CoRR, abs/1807.00734, 2018.

[9] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. CoRR, abs/1609.04802, 2016.

[10] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July 2017.

[11] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters, 20(3):209–212, 2013.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. CoRR, abs/1505.04597, 2015.

[13] Edgar Schönfeld, Bernt Schiele, and Anna Khoreva. A u-net based discriminator for generative adversarial networks. CoRR, abs/2002.12655, 2020.

[14] Uddeshya Upadhyay and Suyash P. Awate. Robust super-resolution gan, with manifold-based and perception loss. CoRR, abs/1903.06920, 2019.

[15] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data, 2021.

[16] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. CoRR, abs/1804.02815, 2018.

[17] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. ESRGAN: enhanced super-resolution generative adversarial networks. CoRR, abs/1809.00219, 2018. IEEE Transactions on Medical Imaging, 37(6):1310–1321, 2018.

[18] Hao Zhang, Chunyu Fang, Xinlin Xie, Yicong Yang, Wei Mei, Di Jin, and Peng Fei. High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network. Biomed. Opt. Express, 10(3):1044–1063, March 2019.

[19] Hao Zhang, Tingting Zhu, Xiongchao Chen, Lanxin Zhu, Di Jin, and Peng Fei. Super-resolution generative adversarial network (SRGAN) enabled on-chip contact microscopy. Journal of Physics D Applied Physics, 54(39):394005, September 2021.

[20] Wenjian Wang, Huaying Wang, Shaokai Yang, Xiaolei Zhang, Xue Wang, Jieyu Wang, Jialiang Lei, Zijian Zhang, Zhao Dong, Resolution enhancement in microscopic imaging based on generative adversarial network with unpaired data, Optics Communications, Volume 503, 127454, 2022.

[21] X. Li et al., "Unsupervised content-preserving transformation for optical microscopy," Light: Science & Applications, vol. 10, no. 1, p. 44, Mar. 2021, doi: https://doi.org/10.1038/s41377-021-00484-y.

[22] Mayeul Cachia, Vasiliki Stergiopoulou, Luca Calatroni, Sébastien Schaub, Laure Blanc-Féraud. Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN). Inverse Problems, 2023, 39 (5), pp.054006. ⟨10.1088/1361-6420/acc889⟩. ⟨hal-03790156v2⟩

[23] Human Protein Atlas. Human Protein Atlas Image Classification — Kaggle., October 3, 2018

[24] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sher- jil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[25] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution, 2016.

[26] Zafran Hussain Shah, Marcel Müller, Tung-Cheng Wang, Philip Maurice Scheidig, Axel Schneider, Mark Schüttpelz, Thomas Huser, and Wolfram Schenck. Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images, Photon. Res., May 2021.

[27] Kang, M., Cha, E., Kang, E., Ye, J., Her, N., Oh, J., Nam, D., Kim, M. & Yang, S. Accuracy improvement of quantification information using super-resolution with convolutional neural network for microscopy im-

ages. *Biomedical Signal Processing And Control*. **58** pp. 101846 (2020), https://www.sciencedirect.com/science/article/pii/S1746809420300021

[28] Chen, R., Tang, X., Zhao, Y. et al. Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. Nat Commun 14, 2854 (2023). https://doi.org/10.1038/s41467-023-38452-2

[29] Qiao, C., Li, D., Guo, Y. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods 18, 194–202 (2021). https://doi.org/10.1038/s41592-020-01048-5

[30] Huang, B., Li, J., Yao, B. et al. Enhancing image resolution of confocal fluorescence microscopy with deep learning. PhotoniX 4, 2 (2023). https://doi.org/10.1186/s43074-022-00077-x

[31] Baoyuan Zhang, Xuefeng Sun, Jialuo Mai, and Weibo Wang, "Deep learning-enhanced fluorescence microscopy via confocal physical imaging model," Opt. Express 31, 19048-19064 (2023)

[32] X. Li et al., "Fast confocal microscopy imaging based on deep learning," 2020 IEEE International Conference on Computational Photography (ICCP), St. Louis, MO, USA, 2020, pp. 1-12, doi: 10.1109/ICCP48838.2020.9105215.

[33] Ayas, S., Ekinci, M. Microscopic image super resolution using deep convolutional neural networks. Multimed Tools Appl 79, 15397–15415 (2020). https://doi.org/10.1007/s11042-019-7397-7

[34] Fang, L., Monroe, F., Novak, S.W. et al. Deep learning-based point-scanning super-resolution imaging. Nat Methods 18, 406–416 (2021). https://doi.org/10.1038/s41592-021-01080-z