Autoencoders for Denoising of Poisson Noise Limited Biological Images

Anonymous Authors[1]

Abstract

Variational Autoencoders (VAE) have been used to denoise images with Gaussian Noise. The noise in the images is assumed to be drawn from a Gaussian distribution. Autoencoders can reduce the dimensionality of large-sized complex data and then reconstruct it back with minimal loss. By tuning the training process and adding a Gaussian noise component the encoder-decoder system can be made into an effective denoiser. Here we present an extension of this concept for non-Gaussian noise, in this case, Poisson noise, typically generated in images where the amount of incident light per pixel is very small. Such low photon number images are generated in the imaging of biological samples under low light excitation. The physical model is captured in the weights of the autoencoder and deploying it for noise reduction yields a good method to reduce complex computation. The literature reviewed presents a large dataset of such images and reviews some traiditional and machince learning based denoising algorithms. We extend the method presented in the literature and apply an autoencoder algorithm to the dataset and present the results. In summary, this method yields an effective method to denoise low light microscopy images and help to speed up the data collection in biological images. A concise summary of the work is also provided at the end.

1. Introduction

1.1. Fluorescence imaging

Fluorescence microscopy is a widespread technique in biomedical and chemistry research, used to image small microscopic samples for study. The typical setup is common and easily available for cheap through commercial solutions. Confocal microscopes and widefield microscopes fall under this category of imaging techniques. Other more expensive techniques include two photon microscopy which promise even smaller levels of background noise in the generated images, promising better research outcomes.

There is, however the problem of inherent noise in the images generated where a detector is involved. The detector can be a Silicon-based Charge Coupled Device (CCD) camera detector, or a Photomultiplier Tube (PMT) or an Avalanche Photodiode (APD). The problem arises because the number of photons received per pixel is rather small compared to traditional photography, 10^2 vs 10^5 per pixel respectively.

The signal is inherently quantised because of quantised nature of light and the assumption of large number statistics per pixel falls through. Regular photography is seen to be dominated by Gaussian Noise when the object is well lit. That does not hold true for microscopy images under specialised illumination conditions. The noise present in this case is Poissonian Noise, and hence it presents a challenge to how we process and analyze the data in such images.

There are some imaging techniques which can be employed in such cases to reduce noise and increase the information that can be obtained from the sample. One way is to increase the illumination power of the laser light or the lamp. This suffers from a big drawback that the biological or chemical samples are both photosensitive and saturable. This means that under strong illumination power it will either receive considerable photodamage or it will not generate a useful photon signal beyond a certain illumination power. This is a common problem in photoemitters in general, that the emitted power cannot be increased linearly with the incident power and drops off as

$$P_{emitted} = \frac{P_{inc}}{P_{inc} + P_{sat}}$$

where P_{inc} is the incident power and P_{sat} is a quantity defined a saturation power. This fundamentally limits how much of a signal can be obtained for some incident illumination power.

Another technique is to increase the illumination and signal collection time, i.e. increase dwell time, exposure length, number of line and frame averages etc. This can also damage the sample and for motile objects in the frame ruin the

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

subsequent images as a messy blur.

This strongly suggests a need for high speed imaging and denoising in low-photon number microscopy. The problem with current denoising system is the assumption of Gaussian Noise in the system. The datasets and models all incorporate this into their processing and hence are unsuitable for cutting-edge scientific research applications in biological and chemical research. Such techniques can also find applications in some areas of physics wherever there is a need of high speed, low-noise images of samples such as applied biophysics and condensed matter physics.

1.2. The Dataset

The authors in (Zhang et al., 2019) have compiled a large dedicated dataset of low photon images in order to facilitate the development of denoising algorithms in Poisson Noise limited cases.

They combine images collected from several biological samples. They have used the 2 Photon Microscopy, Confocal Imaging and Widefield imaging and taken multiple images of the samples. They then account of setup drift during the acquisition and generate the ground truth by taking an average of the obtained data over 50 images taken under same conditions.

This dataset can then be used to build effective models for denoising and machine learning in these settings. Further information about the dataset is in the section related to (Zhang et al., 2019)

1.3. Noise Modeling

To model the noise, a Poisson-Gaussian Noise has been assumed to exist in the system. This has a Poisson component for the shot noisen and a Gaussian component for the thermal and other kinds of noise.

For a pixel z_i the value registered by the CCD/APD/PMT is assumed to have the expression.

$$z_i = y_i + n_p(y_i) + n_g$$

with y_i being the ground truth, $n_p(y_i)$ being the Poisson Noise and n_g being the independent Gaussian Noise component with a 0 mean.

We can assume a scaling factor of a for the detector to account for 1 photon being absorbed and producing a signal a. This is related to the collection efficiency of the detector. Then, assuming b as the variance of the Gaussian Distribution, we can write

$$(y_i + n_p(y_i))/a \propto P(y_i/a)$$

and

$$n_g \propto N(0, b)$$

Also assuming the distributions are independent we can write the probability distribution of the pixel z_i as

$$p(z_i) = \Sigma_{k=0}^{\infty}(\frac{(y_i/k)^k e^{-y_i/a}}{k!} * \frac{e^{-(z_i - ak)^2/(2b)}}{\sqrt{2\pi b}})$$

The key problem is finding the true value y_i given some pixel value z_i.

There are several algorithms that can be applied to reduce noise by traditional methods. The paper evaluates these algorithms and compares them to some of the state-of-the-art machine learning based algorithms for denoising. They conclude that ML based methods are effective substitutes for traditional methods. They also contend that these trained models can be deployed in real time imaging systems and enhance the data acquisition speeds for cutting edge science applications.

2. Literature Review and critique

2.1. A Poisson-Gaussian Denoising Dataset with Real Fluorescence Microscopy Images

This paper (Zhang et al., 2019) addresses the field of Fluorescence microscopy and the application for deep learning based methods for noise reduction.

Fluorescence microscopy is a valuable tool in the field of biology and is used to image tissue samples at low illuminations. The signals are inherently weak and subject to random Poissonian noise instead of Gaussian noise due to low photon numbers hitting the sensor.

It is impractical to increase the illumination intensity on biological samples due to sensitive nature of the matter and a finite fluorescence saturation rate. Long exposures are also not recommended to avoid long exposure damage to the samples.

Denoising algorithms, hence are a good tool for this kind of data to improve upon the image quality and accelerate the pace of research.

This paper addresses the lack of a large dataset for training and benchmarking when the noise is Poissonian. There are some datasets available for Gaussian noise. Finding and implementing effective denoising algorithms in the low photon count regime will require standard datasets to train the models and evaluate the best performing ones before deployment in real-world labs. This paper provides the first Poisson noise-dominated dataset for denoising model training.

2.1.1. Key Contributions

The authors have used commercial fluorescence microscopes to image real biological samples. In order to obtain Poisson noise dominated images they keep the illumination low enough to yield a noisy image but where features are still discernible.

They image each sample at 20 different fields of view for 50 times, in order to obtain the noisy images with the assumption that the mean of the 50 images would be the ground truth. To obtain different Poisson noise levels from this limited dataset they take multiple averages of the images from the set. Given the high precision repeatability of commercial imaging systems, they can ignore pixel shifting in the captured images. This averaging also helps improve the SNR in the images.

The denoising algorithm used is from Ref. (Foi et al., 2008) to first estimate the parameters of the Poisson-Gaussian noise distribution.

For benchmarking the different kinds of denoising methods available they split the dataset into test and train sets, randomly selecting images for each. The images are split from 512x512 to 4 of 256x256 to save on computational resources. They test a total of 10 different algorithms and find that the deep learning based methods work the best to denoise the images.

The deep learning based models are able to achieve similar denoising times as traditional methods when running on CPU, but running them on the GPU yields upto 1 ms processing time. This can enable real time denoising for video up to 100 frames per second, which would make it a invaluable tool for real time video fluorescence imaging systems. It will enable observation of fast and dynamic processes that aren't clearly visible currently.

2.1.2. Key Criticisms

The key drawback of this dataset is the specific case of applicability. Biological samples imaged are from a limited range of cell samples, including Bovine Pulmonary Artery Endothelial (BPAE), Fixed Mouse Brain tissues, and fixed Zebrafish embryos. This is a highly specific set of data, and any applicability outside of the narrow range of these biological samples would be under question.

1One more drawback is the computational expense of the algorithms even on small image sizes of 512x512. These images need to be split into 4 images of 256x256 in order to be processed means that real time processing will involve another processing layers, and can lead to artefacts when the 4 denoised images are stitched together.

2.2. Physics-based Iterative Projection Complex Neural Network for Phase Retrieval in Lensless Microscopy Imaging

This paper (Zhang et al., 2021) introduces a technique to obtain image phase in lensless microscopy imaging. Lensless microscopy imaging is a popular technique in biological science and physics. The basic setup consists of a coherent light source (typically fiber coupled laser), a semi-transparent sample slide and a CMOS detector to detect the image at the end. This technique can capture the intensity of the laser transmitted through the sample and then view it as images.

The CMOS sensor is not able to capture the phase of the images, only transmitted intensity. This makes for incomplete information retrieval from the image. Phase retrieval is an important problem in the fields of computational microscopy, X-Ray crystallography, holography and others.

2.2.1. Key Contributions

The key contribution of the paper is an interpretable, unsupervised and physics based neural network that is complex valued. The mathematics of the the Complex Fourier transform is embedded into the complex weights of the neural network. This allows it to handle the task of Fourier transform and convolution with the transmission function easily.

The paper unfolds the process of alternative projections for loss function minimization into a neural network. The flow of the network mimics the process of alternate projections.

This incorporates the inherent physics of the problem, while making it interpretable after each stage of the projection.

They are able to train the network without labelled data. This strategy simplifies the network inference.

2.2.2. Key Criticisms

The main drawback of the paper is the lack of reliance on any ground truth data for phase retrieval. This is a significant concern as a loss-minimization mathematical approach will only yield itself to a very narrow range of applications.

Another issue is the lack of testable code and datasets to compare other approaches to. The authors have submitted a very small number of images with superior performance, but there is a possibility of the images being cherry picked. The availability of the datasets is a requirement for reproducibility and benchmarking against other methods for the phase retrieval problem.

2.3. Unsupervised Instance Segmentation in Microscopy Images via Panoptic Domain Adaptation and Task Re-weighting

In this paper (Liu et al., 2020), the authors use fluorescence microscopy images to create a Cycle Consistency Panoptic Domain Adaptive Mask R-CNN (CyC-PDAM) architecture for unsupervised nuclei segmentation in histopathology images. They present a nuclei inpainting mechanism to remove auxiliary generated objects in their images. They also design a semantic branch with a domain discriminator to achieve panoptic level domain adaptation. Finally, in order to dynamically add trade-off weights for task-specific loss functions, they propose a task re-weighing mechanism.

Nuclei instance segmentation is important for pathologists to diagnose cancers according to cell counts, the structure of each nucleus and nuclei spatial distribution. Current methods of supervised learning based instance segmentation need large scale data and expert annotation. Removing the dependence on annotations is needed to reduce workload and improve processing times.

2.3.1. Key Contributions

There is a lack of unsupervised domain adaptation methods for instance segmentation. Current methods ignore the domain shift in terms of background/foreground, spatial object distribution and other semantic changes. Moreover they optimize multiple loss functions simultaneously. If feature extraction fails to yield domain invariant features, they tend to bias towards the source which was used to train them.

This paper proposes a simple mechanism to inpaint nuclei and remove auxiliary nuclei.

Their task re-weighing mechanism down-weighs the task if the predictions are source biased, and up-weighed if the obtained features are hard to differentiate.

They conduct experiments on 3 publicy available datasets of histopathology images by unsupervised domain adaptation of a fluorescence imaging dataset since fluorescence imaging data is easier to find labeled.

They are able to obtain state-of-the-art performance compared to other unsupervised domain adaptation methods for nuclei segmentation.

2.3.2. Key Criticisms

One of the major drawbacks is the limited number of organs whose images form the histopathology datsset. The organs being liver, breast,kidney, prostate, bladder, colon and stomach means that the applicability of the methods to other organs will be under question. They also have to segment images into 256x256 patches, which might end up ignoring some larger features in the dataset.

3. Implementation, Evaluation and Results

3.1. Setup Information

We perform all experiments and training on a Ryzen 4950 Mobile CPU, paired with a NVidia RTX2060 Mobile GPU with 6 GB of VRAM, and the total system RAM is 16 GB. The operating system is Windows 11. We also take a subset of the total dataset (4000 images out of 12000) in order to keep the training and testing methods within practical bounds. The benchmark used is code provided with the paper and we use our method to extend their model to an autoencoder architecture. We run their code to understand how their model works and then write our own models and train them on the same dataset as their model. This provides the necessary means of comparison between their and our methods and augments the image loss as a means of gauging the effectiveness of the model.

3.2. Autoencoder Architecture and Implementation Details

To implement and run the autoencoder we had to limit ourselves to a fairly simple representation of the autoencoder. This was kep so due to the fact that the computational power is limited on the system and other problems as discussed later in the paper.

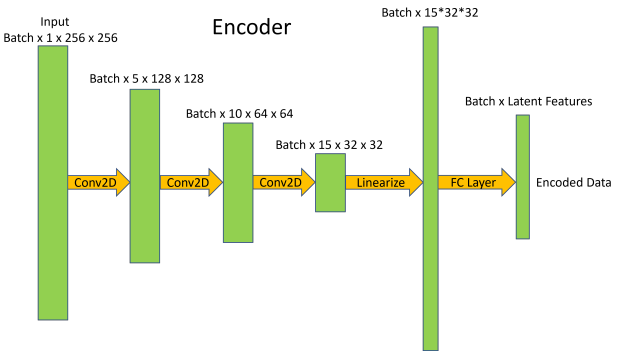

Figure 1. Architecture of the Encoder for denoising

We have three convolutional layers and one fully connected layer with a latent features size of 256. This was near the limit of the system and hence we could not increase the complexity of the system any more than the one presented here.

The total size of the dataset is 12000 images. We trained and evaluated our system on a subset of 4000 images. This was due to resource constraints on the system.

The evaluation method for testing is to calculate two quantities, the Peak Signal to Noise Ratio (PSNR) in the predicted image, and the Mean Square Error (MSE) loss. For a pre-

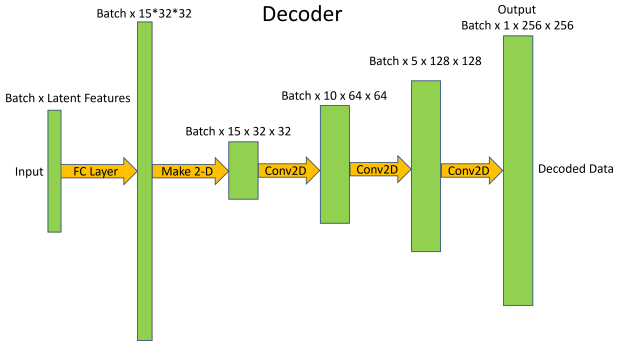

Figure 2. Architecture of the Decoder for denoising

dicted image I and ground truth J the MSE loss is described as

$$l(I,J) = L = l_1, l_2, ..{l_N}^T, l_n = (I_n - J_n)^2$$

and $MSE = sum(L)$. the PSNR is defined as

$$PSNR(I,J) = 10 * log(\frac{MAX_I}{MSE})$$

.

Then we track the evolution of the train and test PSNR and MSE as a function of epochs and keep evaluating the model on the data.

Since this is an autoencoder based denoiser model, we cannot define a quantity such as accuracy and only track other indicators such as the loss and PSNR with the training and testing epoch. Lowering the MSE and increasing the PSNR with each epoch shows that the model is getting better with each training.

The dataset is randomly split into train and test and the dataloaders provided with the sample code are used to run the experiments, training and testing. The loss function used is the MSE loss as defined previously and the PSNR is used to quantify how close the images are to each other.

3.3. Problems and Challenges faced during Project

We faced several issues during this project, ranging from trivial ones to serious ones. Some were simple enough to be able to be resolved with simple Google searches and StackOverFlow browsing and for some help and support from the Teaching Assistants was required. The key decision was to run and train the model locally on the laptop PC, which led to a lot of these issues. Since much of applied ML is done locally to save on compute costs, this seemed like a practical decision at the start, but led to a lot of issues down the line.

The first and foremost was when trying to run the code provided by the authors of the paper itself. Installing Pytorch itself took a couple of days since it demands the installation of CUDA. Finding and installing all of the required software and packages before Pytorch took time. Then it took time to find the versions compatible with the system, CUDA and the paper itself.

Trying to execute their model then took some time. Understanding their code and trying to run it had a few issues. One thing that was totally unfamiliar was how VSCode does not run python files properly when the Run button is pressed. Asking for help from the TA during the in-person office hours helped to recognise this issue and get a workaround by running it directly via the command line. Each epoch takes a few minutes to train even on a subset of the dataset, which means that training the model for 400 epochs as done in the paper is not viable.

The dataset also consists of a large number of images, which severely limits how complex a model can be made to run. One very frequent error was that the GPU was out of memory, with the program trying to assign 9 or 10 Gigabytes of memory when only 6 Gigabytes is physically present. This caused us to reduce our model's complexity and restrict how many epochs we could train it for. Having fully connected (FC) layers adds a large number of parameters for image sizes of 256X256, and such large FC layers cannot fit in the GPU memory. Typically models are being trained for 400-500 epochs in the papers that were read and considered, but that number is impractical given the quick prototyping required in this particular application.

4. Results and Discussion

During training, the model is able to reduce its loss to a saturation point. The loss then stabilizes and does not reduce any further.

The plot of the MSE Loss with the epochs is shown below.

This is a worse performance than the DNCNN (Zhang et al., 2017) based model code as provided with the paper.

To present a comparison with the DNCNN method presented in the paper we present some denoised images as follows.

The DNCNN model works better than our model and is able to present some denoised images along with a reduction in the loss function.

This is as expected since the autoencoder is among the simplest models that are deployed. More complex FC layers cannot be accommodated into the memory of the GPU and hence it is difficult to make it better. More sophisticated models are expected to perform better on specific datasets and hence this result is well within the realm of expectations.

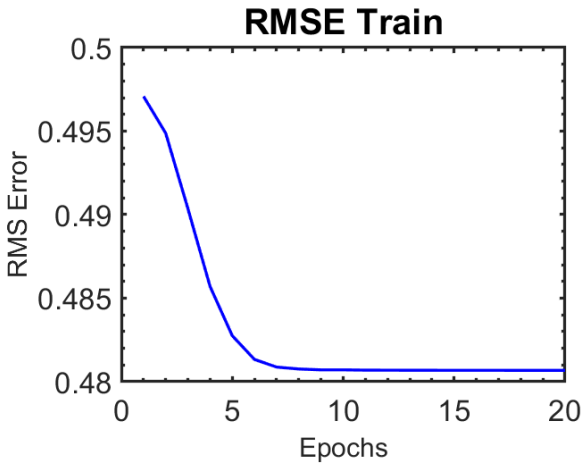

Figure 3. Train Loss for the model

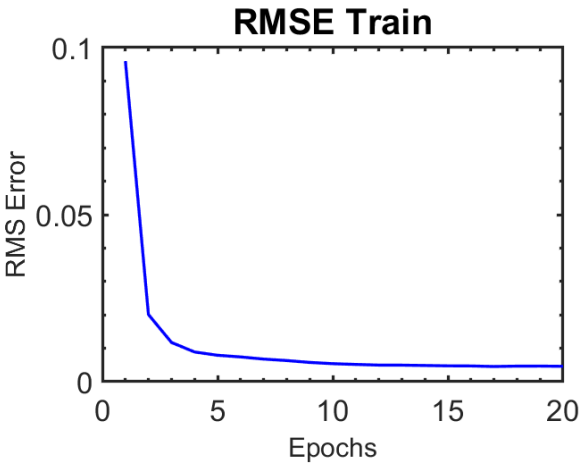

Figure 5. Train Loss for the DNCNN model

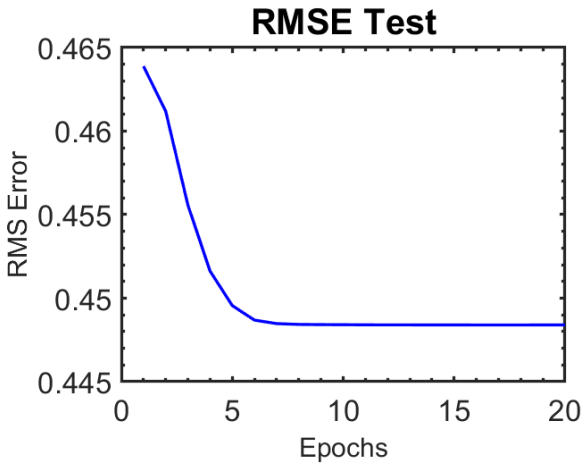

Figure 4. Test Loss for the model

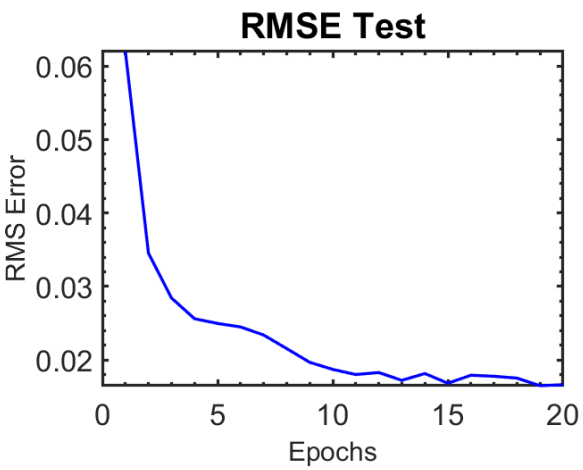

Figure 6. Test Loss for the DNCNN model

5. Summary

- We present an autoencoder-based model for denoising Poisson Loss limited biological images.
- We download the dataset provided in the literature reviewed and use it for the models.
- We run testing and training on the DNCNN model provided in the paper. We use it to benchmark the autoencoder model.
- We train and test the autoencoder model and observe the results are not as good as the DNCNN model
- The reason for this performance deficiency is the large size of the fully connected layers in the model. They cannot be fit into the GPU memory
- We are restricted to fairly simple models for training and testing.
- The DNCNN model has no fully connected layers and hence a lower number of parameters. It can fit and train in the limited GPU memory.
- To fully train and run the autoencoder effectively, more compute power and time is needed.

References

Foi, A., Trimeche, M., Katkovnik, V., and Egiazarian, K. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.

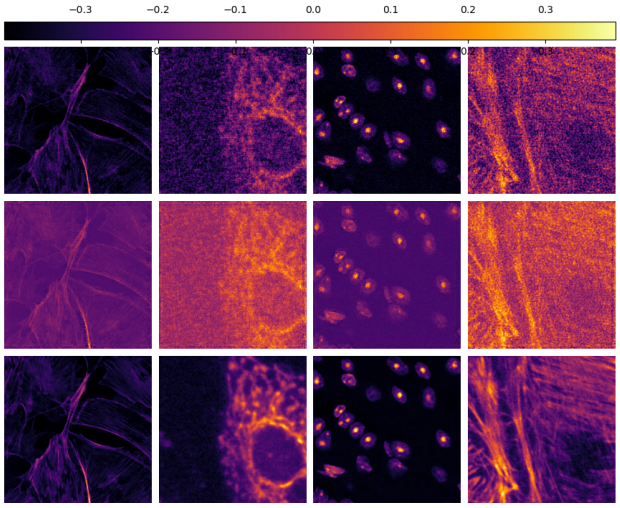

Figure 7. Test images for the DNCNN model from the paper. Top Row: Noisy test images from dataset. Middle row: Denoised Predictions from DNCNN model. Bottom row: Ground truth images

Liu, D., Zhang, D., Song, Y., Zhang, F., O'Donnell, L., Huang, H., Chen, M., and Cai, W. Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4243–4252, 2020.

Zhang, F., Liu, X., Guo, C., Lin, S., Jiang, J., and Ji, X. Physics-based iterative projection complex neural network for phase retrieval in lensless microscopy imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10523–10531, 2021.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. doi: 10.1109/TIP.2017.2662206.

Zhang, Y., Zhu, Y., Nichols, E., Wang, Q., Zhang, S., Smith, C., and Howard, S. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11710–11718, 2019.

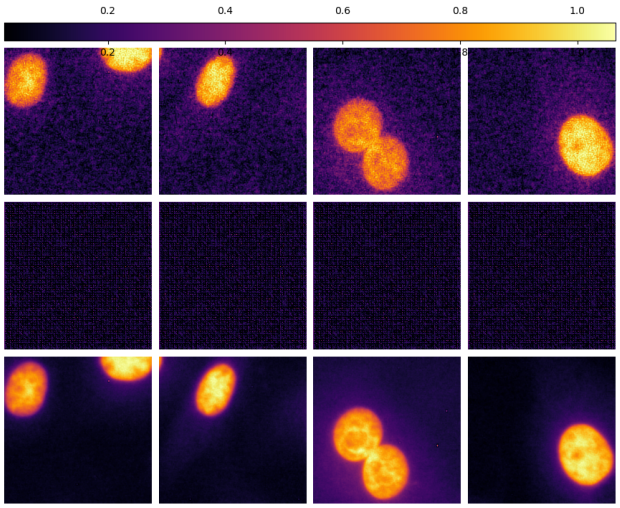

Figure 8. Test images for our trained AE model. Top Row: Noisy test images from dataset. Middle row: Denoised Predictions from DNCNN model. Bottom row: Ground truth images