Project Proposal: Machine Learning Analysis of X-ray Photon Fluctuation Spectroscopy Experiments (SLAC/LCLS)

Author: Sathya Chitturi
Advisors: Prof. Mike Dunne, Dr. Youssef Nashed, Dr. Daniel Ratner

Problem Description / Overview

X-ray Photon Fluctuation Spectroscopy is a newly developed (at LCLS) coherent X-ray imaging technique used to probe ultrafast materials dynamics. Data from this experiment comes in the form of 2D arrays (90x90 pixels) which amounts to single-photons hitting a detector (See Figure 1). Each frame represents a noisy measurement of the intrinsic photon map.

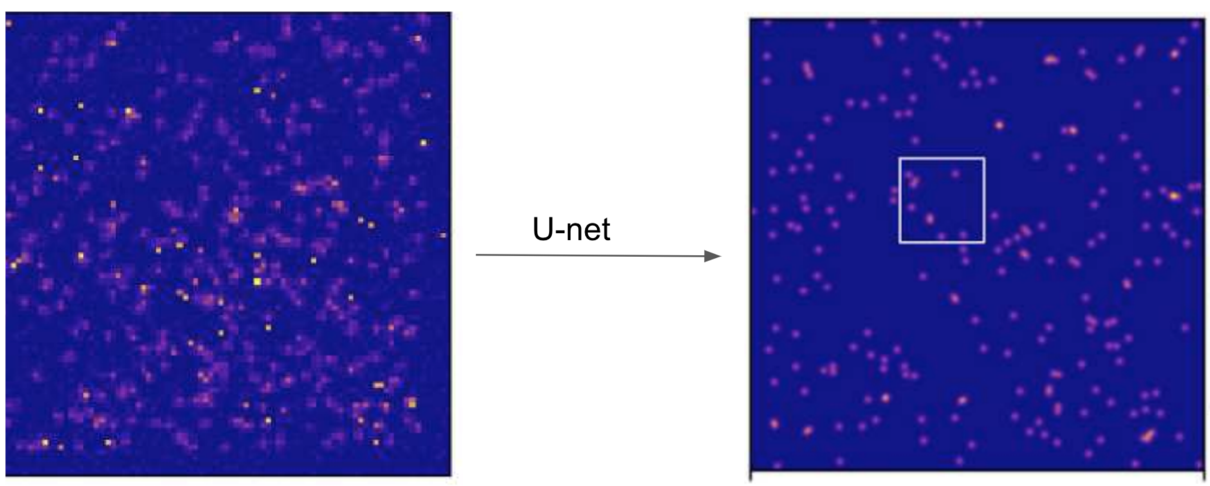

Figure 1. An example of a noisy photon image and the corresponding denoised photon map.

The photon map is a discrete distribution over photon counts. More explicitly, each pixel records the number of photons that hit it during the measurement. The underlying probability distribution for the data is known to follow a negative binomial distribution, parameterized by scalars known as the contrast and the mean photon count. The physical insight from these measurements is the contrast and is extracted by accurately fitting the negative binomial distribution. Here it is worth noting that this measurement is an exquisitely sensitive statistical technique which requires tens of thousands of shots to obtain a good estimate for the contrast.

The goal of the analysis pipeline is to recover photon maps from measurements – in essence, this task is a denoising procedure. This task is challenging for a number of reasons:

1. One reason is that the non-linear transformation kernel is unknown. It likely that the data has multiple unknown convolutional blurring kernels for the charge clouds. These can vary in size between 1 pixel and 25 pixels.
2. There is an underlying read noise distribution which is unknown.
3. There are no ground truth photon maps (i.e. this task is essentially unsupervised).

Current Progress

- We have developed a forward model of the process which represents the charge smearing and noise. This forward model, however, is not so accurate. Based on this forward model, we simulated a large training dataset in order to train a denoising CNN U-net.
- We find that while the U-net performs well on simulated data, the generalization to real data does not work as expected (due to data shift between training and testing distributions).

Related Work

The forward model that we use is based on an implementation given in [1]. We note, also, that there are other methods to perform this task including the Droplet Algorithm and the Droplet Least Squares Algorithm [2,3]. However, these algorithms are relatively slow and are ill-suited to our data which exhibits higher noise and more charge smearing. It is worth noting that the Droplet algorithm is an unsupervised algorithm.

1. **Burdet, N. G., Esposito, V., Seaberg, M. H., Yoon, C. H., & Turner, J. J. (2021). Absolute contrast estimation for soft X-ray photon fluctuation spectroscopy using a variational droplet model. *Scientific Reports*, *11*(1), 1-9.**
2. **Sun, Y., Montana-Lopez, J., Fuoss, P., Sutton, M., & Zhu, D. (2020). Accurate contrast determination for X-ray speckle visibility spectroscopy. *Journal of Synchrotron Radiation*, *27*(4), 999-1007.**
3. **Sun, Y., Esposito, V., Hart, P. A., Hansson, C., Li, H., Nakahara, K., ... & Zhu, D. (2021). A Contrast Calibration Protocol for X-ray Speckle Visibility Spectroscopy. *Applied Sciences*, *11*(21), 10041.**

Proposed Work

Here, I list a series of tasks that I hope to achieve in the project (in order of difficulty):

1. Develop a probabilistic model for the read noise by looking at only measurements without any sample (i.e. “dark shots”). Add this noise to the data from the simulated forward model and experiment with denoising algorithms (and validate on simulated data). In particular, I will try DnCNN as well as removing the bias terms in order to generalize to unknown noise levels.
2. Develop a more accurate forward model of the data. For example, this includes a more accurate noise model as well as incorporating a probability distribution over charge cloud sizes. Finally, I also plan to implement an intensity simulator based on the observed scattering profile. The idea will be to train more U-nets based on the improved model and experiment on published real data which has been previously analyzed using the Droplet algorithm. In addition, I will also ‘check’ by using a cycle-consistent pass through the forward model – i.e. take my denoised photon map and put it back through the simulator and compare the MSE between the simulated detector and the real detector.
3. If I get time, I would like to try a cycleGAN model for unpaired image to image translation. In this case, I can generate realistic ground truth photon maps and I have a number of experimental noisy photon maps. However, I do not have any paired examples.