Decision Making in Orthopedic Surgery Through Hyper Low Dose Images
Previously "Low Dose Fluoroscopy for Orthopedic Surgery"
Final Paper

Members: Mariya Kazachkova, Michael Mudgett
Mentors: Mathias Unberath PhD, Nico Zaech

Background

In internal pelvic fracture reconstruction surgery, the surgeon inserts a k-wire into the pelvis to align the fractured bones. In order to know where the wire is and what trajectory it is following, the surgeon must use x-ray images. Typically, over 100 digital radiographs are taken during the insertion of one k-wire. In order to get a clear view of the instrumentation and pelvic bones, the surgeon images with high dose digital radiographs. The dose of each image adds up over the course of the surgery and can have negative effects on the patient. More importantly, the surgeon will accumulate a small radiation dose due to the photon scattering. Over many procedures, this dose will accumulate quickly. Finally, the discrete nature of digital radiographs causes the procedure to stop every time an image is taken. Our aim is to address these issues by enabling the use of hyper-low dose images. In the long term, we would like to shorten the overall surgery by applying hyper-low dose fluoroscopy.

Problem

The obvious solutions for the high dose problem are to either take less images or to take lower dose images. Because the surgeon needs to stay updated on the k-wire trajectory, the second choice seems best. However, low-dose images lose much of their information content and are noisy. It can be very difficult to ascertain the bone structures and tool location. Likewise, surgeons do not want to risk misjudging the trajectory from a poor image. Thus, the problem is finding a method to extract what little information there is in a low-dose image and then display it in a way the surgeon can interpret. The surgeon only needs to verify the location of the tool relative to the bone, so this serves as a minimum functionality test for the solution.

Approach

We took a two-step approach to this problem. The first approach was to lower the dose by physically affecting the filtration on the x-ray emitter. It is known that copper and aluminum filtration filter low-energy photons so that a larger proportion of relatively high dose photons reach the subject and detector. Because the attenuation difference between bone and soft tissue increases at higher energies, this suggests that added filtration would make the bone in an image more clear. Although the higher proportion of high energy photons means that for the same number of photons, the dose will be larger, we noted that if the image quality is improved significantly we could lower the dose while still maintaining the same image quality. In order to investigate this, we used, MC-GPU, a software for generating x-ray images based on monte carlo simulations [2]. Along with this, we generated custom x-ray spectra using Spektr [6]. By projecting x-ray images at various spectra and recording the simulated dose measurements, we evaluated image quality and dose with changing filtration parameters.

The second step, which would allow us to lower the dose by lowering the mAs of the x-ray machine, was a deep learning network which could resolve noisy images. Many denoising networks perform very well and many have been applied to x-ray images but there are some limitations here [5]. For example, denoising networks perform very well with Gaussian noise; it is not difficult to learn noise that is normally distributed. Digital radiographs, however, exhibit Poisson noise rather than Gaussian, and the noise at each pixel does not come from the same Poisson distribution. In short, this means that as the number of photons increases in a region, so does the variance of the noise. Because the noise variance is not the same throughout the image, there is much more for a network to learn. Thus, performance is not as strong as would

normally be expected. Our deep learning pipeline addresses this in a key step, the Generalized Anscombe Transform, which is discussed in more detail in the methods section.

With these two steps, we have investigated two different methods for effectively lowering the emitted dose in the operating room. The *in silico* analysis is detailed below but testing with an actual scanner would be required for full validation. This is clearly a future step for the project.

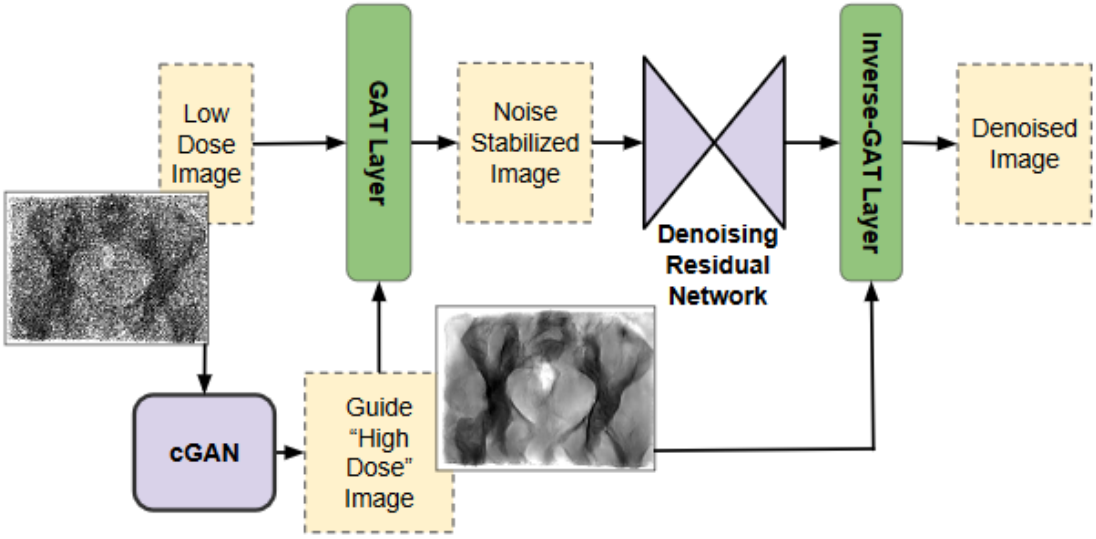

Fig 1. Denoising Pipeline Structure

Methods

For the spectrum-dose analysis portion of the project, we downloaded MC-GPU for monte carlo image simulation and set it up running on a GPU. As stated above, energy spectra were generated with Spektr. Given a voxel file, MC-GPU was able to interface with the energy spectrum files and simulated an x-ray image. We changed the source code slightly to compute the overall dose rather than the dose per photon and printed this to a separate text file. This way, if we generated multiple images, we could store the overall dose for later analysis. A user can alter the dose in MC-GPU by changing the number of photons, and then dose scales linearly with number of photons. We simulated radiographs for a set of spectra at different photon levels and then scaled the photon levels accordingly so that we could generate images with equivalent soft tissue dose across multiple spectra. These equidose images allowed us to study the change in image quality cross spectra when dose is held constant. From a literature search, we found that image quality improves with increasing copper filtration at lower kVp values and that this improvement levels off at around 0.3mm of copper. Our initial findings suggested that more copper will continuously improve the image quality but this is not realistic. With more filtration, the x-ray source must be on for more time and with more than 0.3mm of copper, the load becomes very large and the tube can overheat. This is not something that would be apparent in a computer simulation and thus it is important that the work done here be validated with an actual imaging system. Furthermore, there is no concrete method of measuring the quality of an image. SSIM (Structural Similarity Index) is a common method but it does not work very well on noisy monte carlo images because although images look similar, the individual pixel values can vary by a lot. In conclusion, we verified that as filtration levels increase, the the SNR (signal to noise ratio) will improve when dose is held constant). The next step would be to hold the SNR or some other quality measure constant and then find the minimum dose required to achieve that quality for each spectrum. This would take much more computational time because the image quality does not linearly scale with dose and we would have to test a range of dose levels. Moving forward with this would be very dependent on how

different the dose is to create a “good” quality image in no filtration and high filtration. Once this value is known, we would continue with testing on an actual system and selecting a best dose profile for this procedure.

The deep learning network was designed to denoise a noisy low dose image and restore the image quality to a level where it could be used by a surgeon. Note that this does not necessarily mean it must be restored to the level of a high dose image. If the surgeon only has to make a decision on whether or not the tool is within the bone, only a small area of the image needs strong structural definition. As mentioned above, applying an existing denoising network to an x-ray image is difficult because the noise is drawn from a Poisson distribution. This is why we chose to employ the GAT (Generalized Anscombe Transform), which transforms a random variable with a Possion distribution into the Gaussian distribution space [7]. An issue that arises here, however, is that the transform cannot be applied directly to the noisy image. The transform assumes that the means which the Poisson distributions are drawn from are known and this is not true for the noisy image. In this case, we would need an example high dose image to know the actual value (the means of the Poisson distributions) of an image and then stabilize the noise with this information. Since we are trying to eliminate the need for a high dose image, this posed a problem. We decided to use a cGAN (conditional generative adversarial network) to generate a guide “high dose” image which can supply those means given only the low dose image [3]. The structure of the overall pipeline is as follows:

The initial low dose image is fed into the cGAN. We transform the image to the attenuation domain to reduce dynamic range in order to coax the cGAN to learn the more detailed structure of the pelvis rather than the background (this transformation is undone before saving the output images). Because the cGAN was trained to generate high dose images from low dose images, it outputs a “high dose” image with smooth values close to those of the high dose image. This image cannot be used for surgical procedures because the cGAN does not necessarily recreate the structure perfectly. The output of this network is now called the guide image. The guide image and the original low dose image are fed into the first GAT layer of the network. Here, the guide image is transformed from the intensity space into the photon space (multiplication by a scalar) and then the GAT of this image is computed. To stabilize the variance in the low dose image, we linearize the transform for each pixel around the value provided by the guide image. We use a first order Taylor expansion to accomplish this:

$$\text{GAT}(\text{lowdose}) \quad \text{GAT}(\text{guide}) + \text{GAT}'(\text{guide})(\text{lowdose}-\text{guide})$$

This transformation returns a low dose (noisy) image with an approximately constant standard deviation in local regions throughout the image. The stabilized low dose image then goes into a denoising network. The network we chose applies residual learning to learn the noise of input images [4]. Thus, the output of the network is just noise (loss is ultimately computed between the predicted noise and actual noise). Next, the noise is subtracted from the noisy image to produce a stabilized denoised image. The last step is to perform the inverse of the original GAT layer and return the denoised image to the Poisson domain. The inverse of the above equation is used, where the output is *lowdose* rather than *GAT(lowdose)*. After this step, we have our denoised image in its original domain.

For this pipeline, the cGAN and denoising network (DnCNN) were implemented from established repositories and edited to fit our problem. For example, we trained on TIFF images to maintain precision. The code for the GAT equation was downloaded but we developed the GAT and inverse-GAT layers on our own. They were implemented as PyTorch layers and thus required forward() and backward() methods to allow for correct backpropagation through our network. The pipeline can train and test end-to-end with low and high dose training and test images supplied.

To generate the data for the network, DeepDRR was used to generate simulated x-ray images [1]. Because DeepDRR takes in photon count (photons per detector pixel) as a parameter, we could not easily quantify the patient dose. For the dataset, the clean images

were saved at a step in the pipeline before any noise is introduced. This image resembles a very high dose image with no noise and serves as a strong reference image for the network. DeepDRR then applies Poisson noise based on the photon count. We looked at image metrics for a variety of photon counts and chose the lowest count which maintained most of the clean image properties to be “100% dose.” This value was then divided by 10 for “10% dose” and these were deemed hyper-low dose. We trained the network with the 10% dose images.

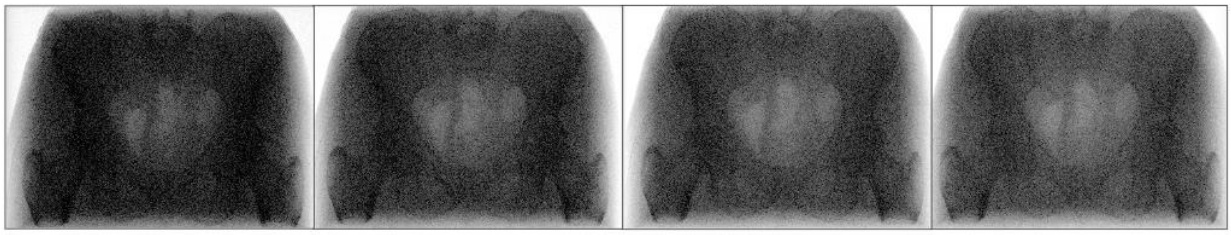

Fig 2. Monte Carlo simulated radiographs generated at 60 kVp and constant soft tissue dose. From left to right, the spectrum was filtered with 0.0mm Cu, 0.1mm Cu, 0.2mm Cu, 0.3mm Cu. The copper filters out low-energy photons, increasing the contrast between bone and the background.

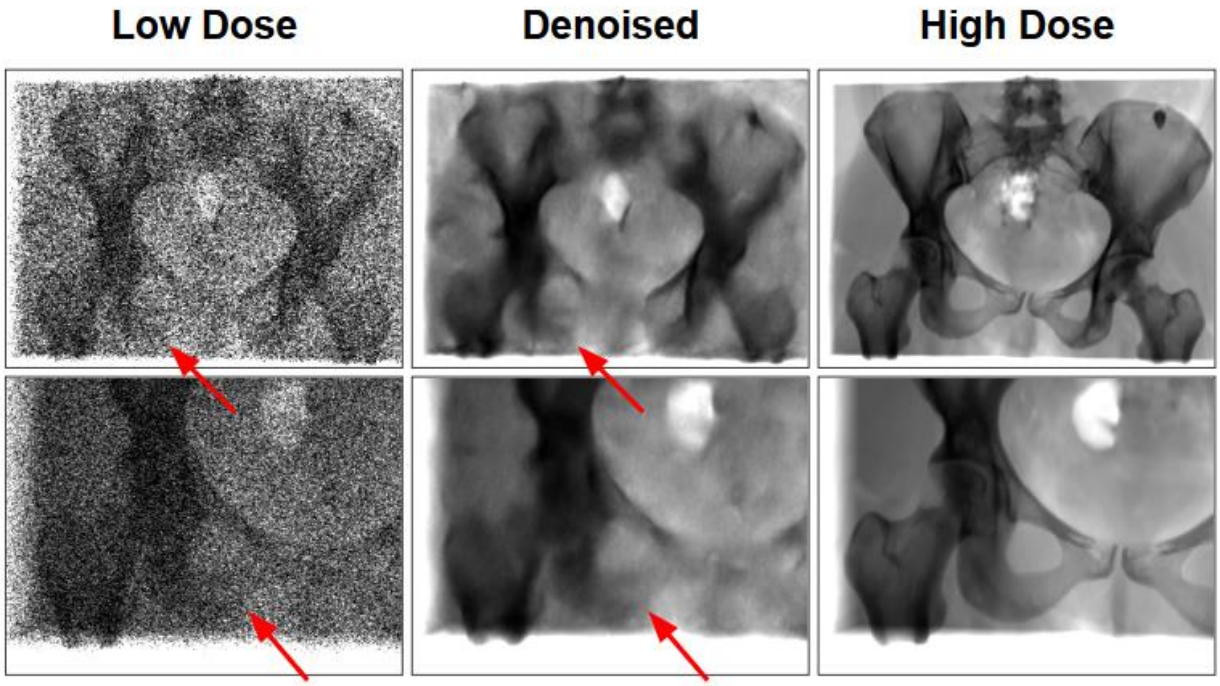

Fig 3. Results of the denoising pipeline. Each row contains a low dose, denoised, and high dose reference image. The rows show two different images from the test set. Red arrows indicate structures recovered by the deep learning pipeline

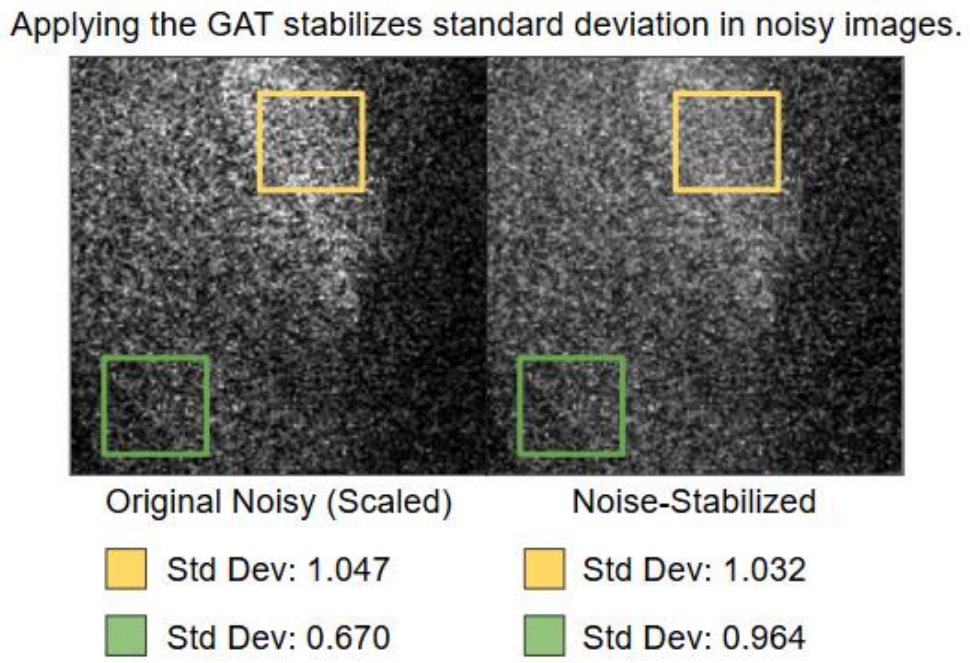

Fig 4. Two patches used to train the denoising network. The left image shows the scaled noisy patch with inconsistent standard deviation. The noise-stabilized image exhibits much more consistent standard deviation values throughout the image.

Results + Discussion

The results from our work with dose/spectrum/image quality analysis show that in simulation there is an improvement in image quality as filtration increases. However, though the images at similar dose look different at the same window/level, these can be altered and the images will look very similar to the naked eye. As discussed in the methods section, more work would need to be done in this area to fully evaluate and identify the best spectrum. For this type of problem, it is difficult to relate the simulation data to real-world images and this step is imperative for discovering something that can be applied in a surgical setting.

We trained the pipeline with a training set of 3,780 simulated x-ray images. These were taken on 5 different CT volumes. Each volume was imaged at a variety of angles (rotations around the sphere) and from different source-detector distances (zoomed in vs zoomed out). As mentioned above, the input hyper-low dose images were deemed “10% dose.” Because these were not generated with a real imaging system, this value is heuristic and might not reflect the actual change in dose to a patient. The test set consisted of images taken from similar camera positions but on a completely separate volume. This component was very important for assessing the performance of our network. All generated images were 320 x 240 px to prevent memory overflows during training. The denoising network trained on patches taken at a small stride to increase the amount of training data. 100 x 100 px patches snipped from the training set were used to train the network. Both the cGAN and denoising network were trained for 100 epochs, at which point the models appeared to converge.

Looking at the data, it is visually apparent that the denoised image is substantially more clear than the noisy image. Some of the harder to discern structures are recovered by the network and there is a strong resemblance between the denoised and clean image. It is not clear that the image would be ready for use in a surgery. Furthermore, the images with a synthetic tool were tested on the network. The results showed that the k-wire was almost entirely removed, but this should be expected. The pipeline was not trained on any images with instrumentation so when it is given an image with a tool, it would not “treat” it correctly. We hypothesize that if a

large dataset of images with instrumentation was used for training, the network would easily recover the tool placement. Likewise, to improve the contrast around the pubic arch, we could add more training images which highlight only this location. Nonetheless, the CT volumes we used were cropped, giving much brighter regions around the edges. This brightness was caused by the lack of soft tissue and air in the volume. In a full CT, this effect would not occur. Moving forward, a training set that does not have this effect would be ideal.

Looking specifically at the effects of the GAT layer, we see that when a noisy image is stabilized by the GAT layer, the local standard deviation in a region is far more constant across a patch than in the original noisy image. This step greatly improved the performance of the denoising network.

Significance + Future Work

This work is significant because it pushes the limit of how far one can reduce the radiation dose while still denoising a useful image. With more time, better training data, and real-world experiments, the lower limit for dose can be characterized. Likewise, the model is fast and could be implemented in the operating room for real-time image improvement.

In the future, the goal is to apply this model to live fluoroscopic video. Our plan is to use an LSTM (long short-term memory) to retain information about past images and improve the quality of low-dose video. If implemented, the pipeline would address the start-stop nature of the current procedure, as discussed above. Before this however, an improved and expanded training set and preliminary testing on a physical imaging system would be necessary.

Conclusion

We have developed and implemented a deep learning pipeline to improve the quality of hyper-low dose digital radiographs. We have trained the model and assessed test output but there are still many steps before it is ready for surgical use. Nonetheless, the pipeline structure is in place and we believe that a more comprehensive training set would vastly improve the output of this network.

Management Summary

Who did what

Mariya's main task was implementing the deep learning pipeline. Michael generated all images, both for dose/spectrum analysis and for deep learning training/testing. When troubleshooting network performance and debugging (there was a lot of debugging), we made use of partnered programming.

What was accomplished vs what was planned

Our original maximum deliverable was to have a functional LSTM network which could denoise continuous fluoroscopic video. We were not able to achieve this goal, but that is because our expected deliverable (the pipeline) became much more complex as the semester went on. Originally, our plan was to implement a denoising network and then move directly to LSTM implementation. However, the issue with Poisson noise (discussed above) became apparent and our mentors explained the importance of the GAT stabilization step. We experimented with multiple network structures (i.e. using a variational autoencoder in place of a cGAN) before settling on a final design. We only accomplished our expected deliverable, but we were able to dive much deeper into the topic and develop a complex model that addresses the inherent problems in denoising x-ray images. Overall, we believe it was better to focus more on one task and generate strong results with that rather than overlook the details and have a sub-par LSTM.

Next steps

As was previously mentioned, the next steps in this project would be to modify the pipeline to be applicable for fluoroscopic data. This would likely mean using an LSTM network to retain relevant information throughout timesteps. Prior to doing this, it would be necessary to create a larger training set in order to better generalize to testing data. Additionally, testing the pipeline in a mock-OR would be crucial in verifying its success on real images rather than generated ones.

What we learned

We learned that when it comes to software, setup takes much longer than expected and should not be overlooked! We set aside much less time for setup than it actually took. Likewise, we better understand the value of partnered programming, as having two people looking at the code means we are twice as likely to catch mistakes.

On the subject matter, we learned a great deal about x-ray imaging and deep learning.

Technical Appendix
See course website for link to code and documentation.

References
[1] Unberath, M., Zaech, J. N., Lee, S. C., Bier, B., Fotouhi, J., Armand, M., & Navab, N. (2018). DeepDRR--A Catalyst for Machine Learning in Fluoroscopy-guided Procedures. Proc. MICCAI
[2] Badal, A., & Badano, A. (2009). Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit. Medical physics, 36(11), 4878-4880.
[3] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. arXiv preprint.
[4] Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing, 26(7), 3142-3155.
[5] Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., & Wang, G. (2017, April). Low-dose CT denoising with convolutional neural network. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on* (pp. 143-146). IEEE.
[6] Siewerdsen, J. H., Waese, A. M., Moseley, D. J., Richard, S., & Jaffray, D. A. (2004). Spektr: A computational tool for x-ray spectral analysis and imaging system optimization. *Medical physics*, *31*(11), 3057-3067.
[7] Makitalo, M., & Foi, A. (2011). Optimal inversion of the Anscombe transformation in low-count Poisson image denoising. *IEEE transactions on Image Processing*, *20*(1), 99-109.