6.8301 Project Proposal

Brian Rapanan, John Readlinger, James Moore

Problem/question and relevant readings:

Our group will be experimenting with ControlNet to see if Stable Diffusion can be controlled to generate high quality X-Ray images from a sketch consisting of orientation lines and a prompt. In other words, the sketch consists of only lines (similar to the output of the openpose library) to signal the orientation of the body part of interest. The prompt then lists out the body part of interest as well as a condition which is either 'healthy' or 'fractured,' allowing our ControlNet to generate X-Ray imagery from this input. This application is specifically interesting because of the many downstream use cases in preoperative planning, remote diagnostics, and medical education. Relevant resources we found most useful in designing our approach were the FracAtlas dataset, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", "Deep Learning in Medical Image Analysis" and "Diffusion models in medical imaging: A comprehensive survey."

Data and methods/algorithms:

For our dataset, we intend to use the FracAtlas dataset. This dataset contains over 4,000 manually labeled x-rays of different body regions and it is the only dataset of its kind to focus only on fractures and normal X-rays. While large, one limitation of this dataset is the inability to extend our model to other medical problems diagnosable through X-ray. However, for the scope of this project FracAtlas suffices. To transform these X-rays into controlmaps similar to those generated by openpose, we intend to adapt their methods for object segmentation to X-rays, as their current methods use keypoint markers like external body parts which are not available in X-rays. In Openpose, keypoints at the ankles, knees, hips, and other joints are used to map humans in the image to two dimensional poses. For our uses, keypoints would include relevant joints, distinguishing bone features, and common fracture points which we will identify through a CNN architecture. To develop sketches for our ground truth data we can apply the Canny edge detection algorithm which both smooths out the original X-ray and contains hyperparameters which can be tuned to make sketches more rough, detailed, etc. In addition, we will create synthetic data from our original data by rotating and scaling the existing images so that our model can generalize across a wide range of X-ray images, orientations, and fracture types.

Computing resources:

We recognize that training a ControlNet for stable diffusion and handling, preprocessing, and postprocessing images requires significant computing resources. We intend to use Colab or Colab Pro to rely on the GPU access to boost training time and memory.

Evaluating results:

We intend to split our dataset into standard train, validation, and test splits and will follow a standard training process with a large emphasis on performance on the test set. Qualitatively, we expect our ControlNet to be able to output a wide array of X-Rays of various body parts in many different orientations. To assess this, we can employ noise tests on the input image/prompt to test the model's robustness and overall examine the quality of generated images from the ControlNet. To test this quantitatively, we can use metrics like Inception Score (IS) that assesses the diversity and clarity of the model's outputs, and Frechet Inception Distance (FID) to compare the distribution of the generated images to the ground truth data.

Contribution:

If a robust and accurate model is built, this model can be useful in a myriad of applications. Doctors would be able to generate example X-Rays for patients or for comparison without having to use another patient's X-Ray, potentially resulting in a HIPAA violation. Moreover, medical students and professionals could use the model for general studying purposes, prompting the model to generate various X-Rays. An extension of this work could include the generation of other images such as CT scans, ultrasounds or MRIs.

References:

Abedeen, I., Rahman, M. A., Prottyasha, F. Z., et al. "FracAtlas: A Dataset for Fracture Classification, Localization and Segmentation of Musculoskeletal Radiographs." Scientific Data, vol. 10, 2023, p. 521. *Springer Nature*, https://doi.org/10.1038/s41597-023-02432-4.

Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. https://doi.org/10.48550/arXiv.1812.08008

Chan, Heang-Ping et al. “Deep Learning in Medical Image Analysis.” *Advances in experimental medicine and biology* vol. 1213 (2020): 3-21. doi:10.1007/978-3-030-33128-3_1

Kazerouni, Amirhossein, et al. "Diffusion models in medical imaging: A comprehensive survey." *Medical Image Analysis* (2023): 102846. https://doi.org/10.48550/arXiv.2211.07804