COMP 691- Research Proposal

Translating X-ray images to RGB images to perform CLIP based analysis

Contrastive vision- language pre- training (CLIP) is a machine learning model developed by OpenAI and introduced in January 2021. CLIP is based on contrastive learning which involves training the model to differentiate between positive and negative pairs. It is trained using contrastive learning between images and unstructured texts to predict which images are associated with which texts. It enables the model to create meaningful features from visual and textual information and perform matching by encoding them to vectors. CLIP can perform image classification, object detection and zero shot learning, where it is possible to make predictions about the classes that it hasn't been trained with.

Before performing CLIP analysis on x-ray images, we apply a pixel-to-pixel image translation method to convert them to RGB images using cycle-consistent adversarial networks. Following this, the dataset undergoes zero shot learning where the model is trained to recognize and perform tasks for which it has no experience during the training phase. The model learns to predict which image belongs to which text and outputs cosine similarities between the embedded texts and images. This enables the model to generalize tasks it has not encountered before.

To perform CLIP based model on x-ray images for object detection will allow more robust performances in the detection of security threats that can be useful in object detection and x-ray security imaging in real time environment.

Resources:

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. & Sutskever, I.. (2021). Learning Transferable Visual Models From Natural Language Supervision. Proceedings of the 38th International Conference on Machine Learning, in Proceedings of Machine Learning Research 139:8748-8763 Available from https://proceedings.mlr.press/v139/radford21a.html.

Paul A, Shen TC, Lee S, Balachandar N, Peng Y, Lu Z, Summers RM. Generalized Zero-Shot Chest X-Ray Diagnosis Through Trait-Guided Multi-View Semantic Embedding With Self-Training. IEEE Trans Med Imaging. 2021 Oct;40(10):2642-2655. doi: 10.1109/TMI.2021.3054817. Epub 2021 Sep 30. PMID: 33523805; PMCID: PMC8591713.

Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.

R. Zhang et al., "PointCLIP: Point Cloud Understanding by CLIP," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 8542-8552, doi: 10.1109/CVPR52688.2022.00836. keywords: