Integrating SAM2 with Bayesian Convolutional Neural Networks: Predicting Diseases in Medical Imaging

Richard Chin

Abstract

Bayesian Convolutional Neural Networks (BCNNs) offer a valuable advantage in medical diagnostics by quantifying diagnostic uncertainty with probabilistic predictions. By integrating segmentation techniques, BCNNs can capture more intricate details in medical images, further enhancing their predictive performance. The Segment Anything Model 2 (SAM2), known for its strong zero-shot segmentation capabilities in natural and video imagery, presents a promising opportunity for medical image segmentation. Although SAM2's application in medical imaging, particularly under zero-shot conditions, warrants further exploration, its segmentation outputs can be leveraged to improve BCNN training. By facilitating the learning of specific region boundaries and characteristics, SAM2 can help BCNNs achieve more accurate and reliable diagnostic outcomes, despite its limitations in boundary detection.

In this study, we present a novel framework that combines SAM2 with BCNNs to address two primary challenges in medical imaging: achieving precise anatomical segmentation and accurately estimating prediction uncertainty. Within this framework, SAM2 is responsible for segmenting anatomical regions, while BCNNs enhance abnormality detection and provide measures of diagnostic confidence, improving interpretability and reliability.

Our evaluations on a subset of the NIH Chest X-ray dataset demonstrate slight improvements in both diagnostic accuracy and uncertainty estimation using this hybrid approach compared to the baseline approach without segmentation. However, in the segmentation process itself, SAM2 exhibited limitations and failed to accurately delineate all the significant medical components in these lower-contrast imaging modalities, where under-segmentation and over-segmentation issues persisted under zero-shot conditions. To mitigate these challenges, we applied bounding boxes and edge detection techniques, which led to **slight** performance gains in predictions and uncertainty scores for this dataset.

1 Introduction

The accurate diagnosis of diseases and anomalies from medical images is crucial in clinical practice. However, interpreting these images with precision and confidence poses significant challenges due to the inherent complexity and variability of medical data. Traditional diagnostic models often lack mechanisms to express the uncertainty associated with predictions, which is critical in ambiguous or complex cases.

The Segment Anything Model (SAM2) is a high-performing segmentation tool that can delineate anatomical regions within images with impressive zero-shot segmentation capabilities [1]. A key innovation in SAM2 is its ability to extend segmentation to video data, making it a versatile choice for dynamic imaging environments compared to its predecessor, SAM. Although some research indicates that SAM2 performance is comparable to that of SAM[2], our decision to use SAM2 is based on its overall enhanced boundary detection compared to its counterpart and potential for refinement. To address continued segmentation struggles in certain lower-contrast imaging modalities, such as CT and ultrasound [3]. , we have implemented additional techniques such as bounding boxes and edge detection, which have shown to mitigate these issues more noticeably.

While SAM2 can perform segmentation on images, it doesn't inherently detect abnormalities or diseases. Clinical diagnostics also often require additional insights, particularly in estimating the uncertainty associated with disease detection. Bayesian Convolutional Neural Networks (BCNNs) excel in this domain, as they not only classify abnormalities but also provide probabilistic uncertainty estimates. This feature is essential in clinical settings, where diagnostic confidence greatly impacts decision-making. BCNNs enhance the interpretability of predictions by providing confidence intervals, offering an advantage over traditional neural networks in high-stakes applications.

While models like UNet and DeepLabV3+ have demonstrated superior performance in medical image segmentation [4], our study focuses on leveraging SAM2 to explore its unique capabilities in capturing uncertainty. UNet, with its well-established U-shaped architecture, has been extensively validated and widely adopted in medical image segmentation tasks. However, SAM2 offers a distinct advantage with its ability to incorporate various prompts (e.g., points, boxes, masks). This flexibility allows us to investigate how these prompts can enhance segmentation performance in complex medical imaging scenarios. By using SAM2, we aim to uncover new insights into segmentation techniques and improve the model's ability to handle uncertainty in medical images.

In this study, we propose an integrated SAM2 and BCNN model, aiming to leverage SAM2's segmentation capabilities with BCNN's uncertainty-aware classification, potentially leading to more accurate and reliable disease detection in medical images.

2 Methods

2.1 Segmentation with SAM2

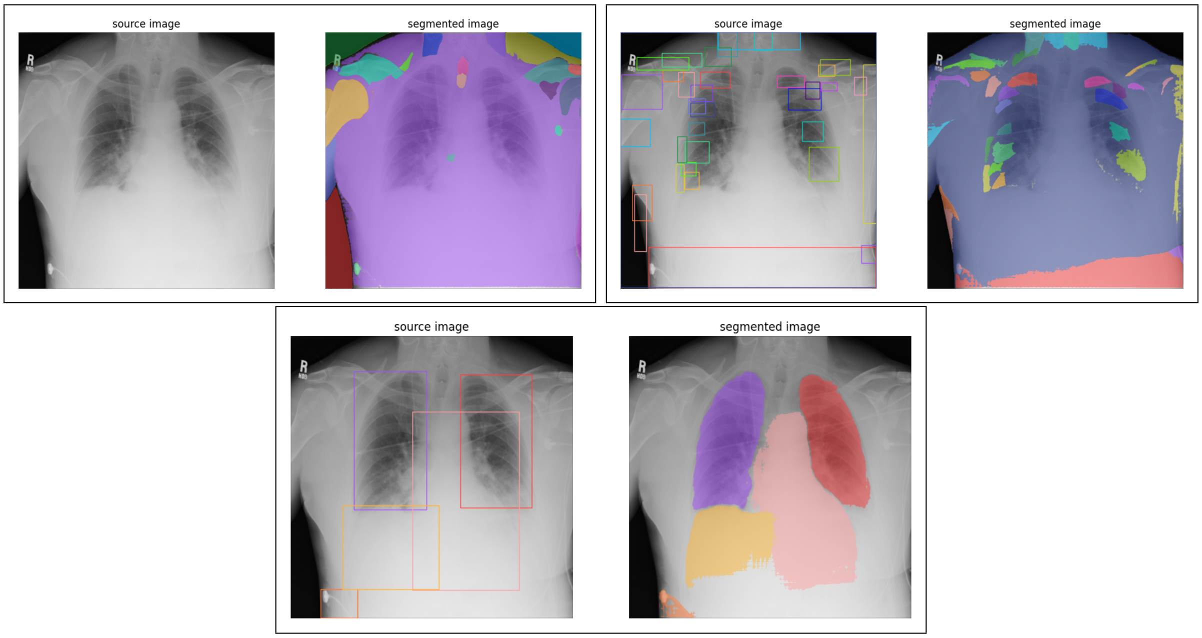

Figure 1: On the left-hand side, an example of SAM2 segmentation performance is shown after applying the SAM2 automatic masker. On the right-hand side, SAM2 segmentation performance is displayed using boundary boxes created based on contours identified through edge detection. At the bottom, SAM2 manual segmentation was performed, annotating the most important areas of the X-rays for analysis (e.g. heart, lungs, pleura, ribs)

Segmentation Workflow: Our experiment consisted of utilizing different strategies for SAM2 segmentation on a subset of 200 images from the NIH Chest X-ray14 dataset, including leveraging automatic zero-shot with SAM2 masks, creating custom boundary box masks using contours, and manually making the boundary boxes to isolate key anatomical regions from the images. The boundary boxes were used as prompts for SAM2 inference to generate annotations.

Refinement for Bounding Boxes: The bounding boxes were generated via a specific pipeline to improve segmentation accuracy during inference. The process involved pre-processing techniques such as applying Gaussian blur to reduce noise, followed by Canny edge detection to identify precise boundaries. Contours were then detected and bounding boxes were created around these contours. Small bounding boxes were combined into larger ones to focus on significant regions. After using the combined bounding boxes to prompt the SAM2 inference predictor and generate annotations, the annotated images were prepared as input for our custom-built Bayesian Convolutional Neural Network (BCNN).

Fine-tuning the SAM2 model on the chest x-ray dataset was attempted, but it proved to be too computationally intensive on a T4 GPU. Training for 10 epochs, at 3 minutes per epoch, only resulted in an approximate 0.02 increase in the IOU score. Given that running thousands of epochs is recommended to obtain satisfactory fine-tuning results for the SAM2 model [5], this approach was not feasible given our limited resources.

2.2 BCNN for Disease/Abnormality Classification

Model Architecture: The SAM2/MedSAM-segmented images were used as input to a BCNN designed for disease detection and uncertainty estimation. Given the complexity of developing a model from scratch, we leveraged a pre-trained DenseNet-121 model as the backbone for feature extraction. DenseNet-121 is known for its dense connectivity pattern, where each layer receives input from all preceding layers, promoting feature reuse and efficient gradient flow. To integrate the DenseNet model into a BCNN implementation, the classifier layer of DenseNet-121 was replaced with an identity layer to retain the extracted features.

The extracted features from the DenseNet backbone are then passed through a series of connected layers. First, a dropout layer with a dropout rate of 0.5 is applied, randomly setting 50 percent of the input units to zero during training to aid in regularization and prevent overfitting. Following this, the features are fed into a fully connected layer (fc1) with 512 neurons, utilizing a ReLU activation function to introduce non-linearity. Afterwards, the output from fc1 is passed to a second fully connected layer (fc2) with 15 neurons, corresponding to the 15 classes in the multi-label classification task, outputting the raw logits for each class.

Uncertainty Estimation: During inference, we applied Monte Carlo Dropout to the BCNN model to estimate the uncertainty in the classification of segmented regions. By performing multiple forward passes with dropout enabled, we generated an ensemble of slightly varied models. The variances in the resulting predictions served as the measure of uncertainty, providing insights into the model's confidence in its predictions.

3 Results

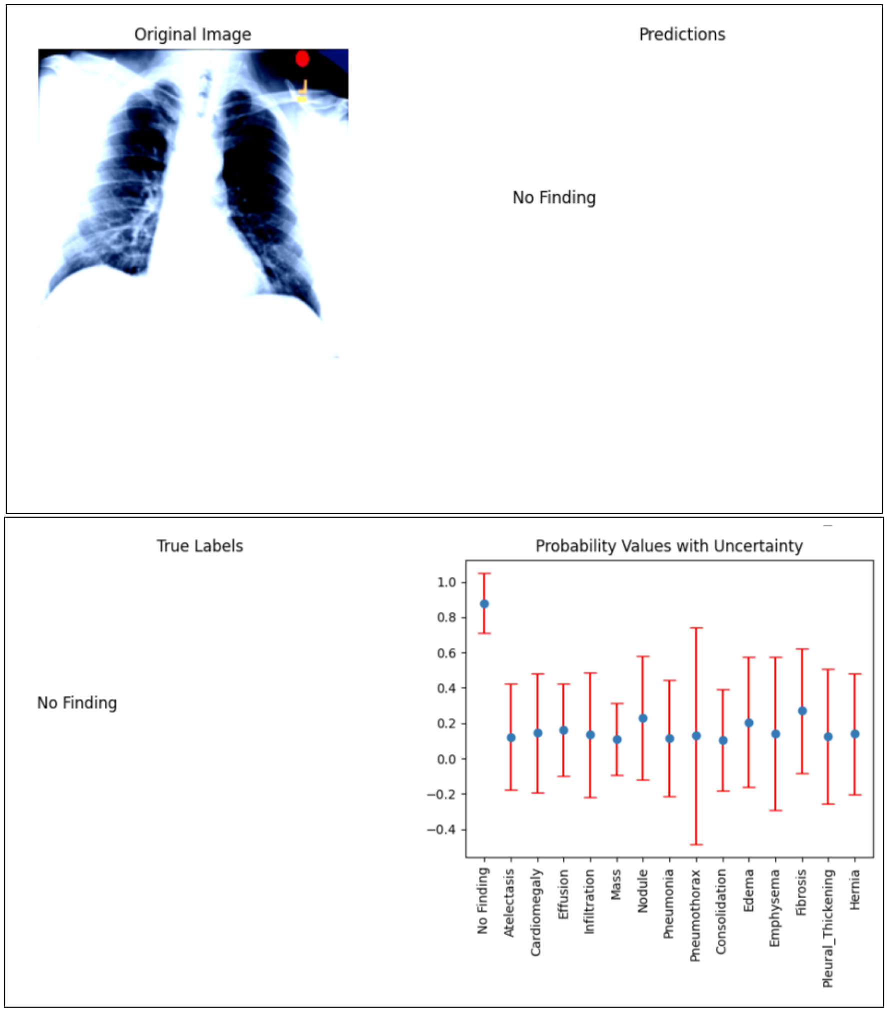

Figure 2: An example of BCNN prediction performance is shown after applying annotations on the image using SAM2 bounding box segmentation. In the bottom right graph (error bar graph), the blue dots represent the mean probability of the predicted class, while the red lines depict the spread of the probability distribution, illustrating uncertainty as confidence intervals.

We compared the performance of the hybrid SAM2 and BCNN model based on the utilization of different segmentation methods with the SAM2 model.

The SAM2 segmentation with BCNN models demonstrated comparable, but slightly better performance in terms of accuracy, precision, recall, and AUC-ROC and compared to the baseline BCNN model on the test dataset. Specifically, the BCNN models zero-shot SAM2 segmentation achieved an accuracy of 0.8770 and 0.8800, precision of 0.2857 and 0.3333, recall of 0.1471 and 0.1912, and AUC-ROC of 0.1942 and 0.2430 respectively. These performances are comparable to the BCNN with no previous segmentation on the image, which achieved an accuracy of 0.8681, precision of 0.2308, recall of 0.1324, and AUC-ROC of 0.1682.

The lower scores in precision, recall, and F-score in the test results are most likely due to some classes not having any true positives given the 40 images used as the test dataset, which were calculated as 0 when producing overall performance metrics. To address this, larger training and testing datasets could be used to ensure a more comprehensive evaluation and potentially improve these metrics by providing true positives for all classes.

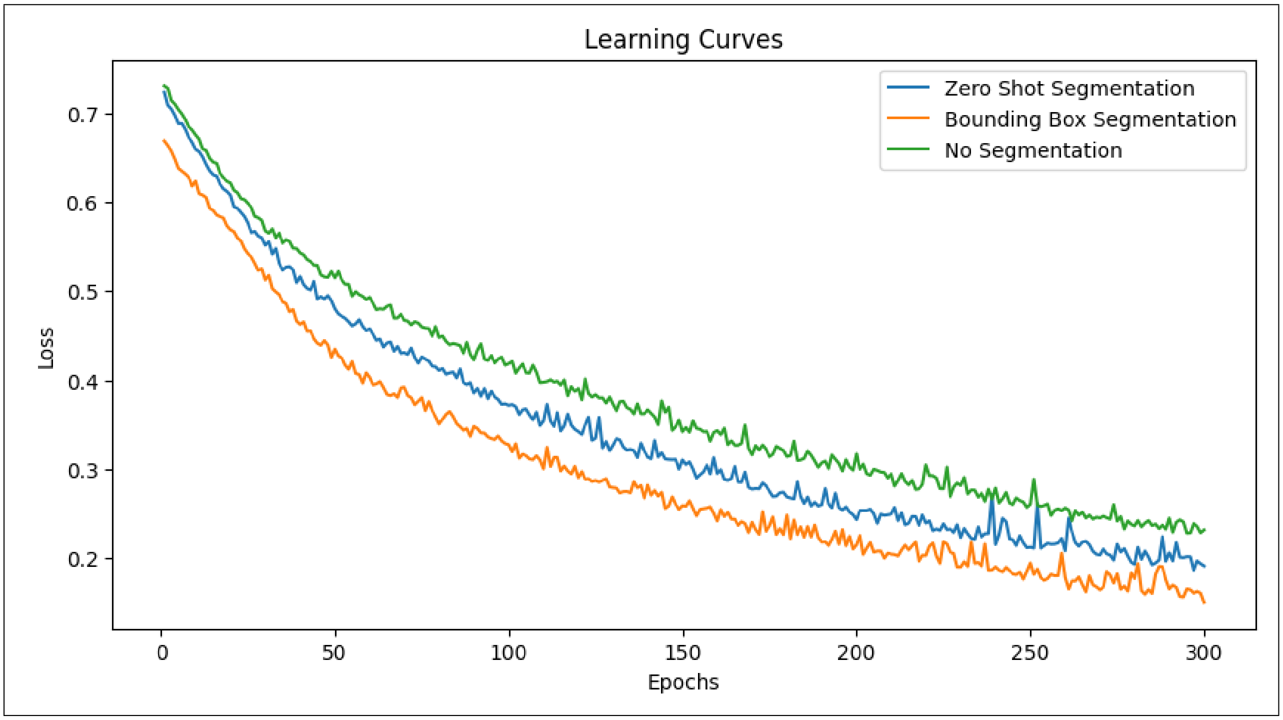

Figure 3: Training Curves of the BCNN model given different segmentation implementations

Segmentation Method to BCNN	Dataset	Overall Accuracy	Overall Precision	Overall Recall	Overall F1-Score
Automatic Segmentation	Validation	0.9467	0.6667	0.5455	0.6000
	Test	0.8770	0.2857	0.1471	0.1942
Bounding Box Segmentation	Validation	0.9520	0.7021	0.6000	0.6471
	Test	0.8800	0.3333	0.1912	0.2430
No Segmentation	Validation	0.9613	0.7955	0.6364	0.7071
	Test	0.8681	0.2308	0.1324	0.1682

Figure 4: Performance Scores of the BCNN model given different segmentation implementations

The integration of SAM2 with BCNN also provided comparable uncertainty estimates, allowing for a more transparent understanding of the model's predictions and performances.

Despite the BCNN with bounding box segmentation for SAM2 inference displaying the best performance among the three models on the test dataset, it also exhibited the greatest uncertainty and variance in uncertainty values. Specifically, the mean uncertainty was 0.4357 with a standard deviation of 0.1321 (see Figure 5). This observation underscores the trade-off between performance and uncertainty, indicating that while the model achieves high accuracy, its confidence in predictions varies significantly. The larger variance in uncertainty values further suggests that the model may be more susceptible to overfitting or encountering difficulties with harder-to-classify instances, resulting in inconsistent confidence across different classes.

In comparison, the zero-shot SAM2 segmentation inference model with BCNN showed a mean uncertainty of 0.3978, with a standard deviation of 0.1057. Despite its slightly lower performance to the other hybrid model, this model implementation demonstrates more consistent confidence in its predictions while still performing better than the baseline BCNN model.

By analyzing this measure of uncertainty, we can pinpoint areas of data where the model's confidence was lower, highlighting regions that may need closer examination or further validation. This capability is particularly valuable in medical diagnostics, where identifying areas of uncertainty can guide further investigation and improve diagnostic accuracy.

Overall, the SAM2 and BCNN models, when applying segmentation, show slightly different performances in terms of accuracy and uncertainty estimates, suggesting that this approach could serve as a valuable tool for future clinical applications by helping determine which segmentation techniques would make the model more confident while maintaining accuracy.

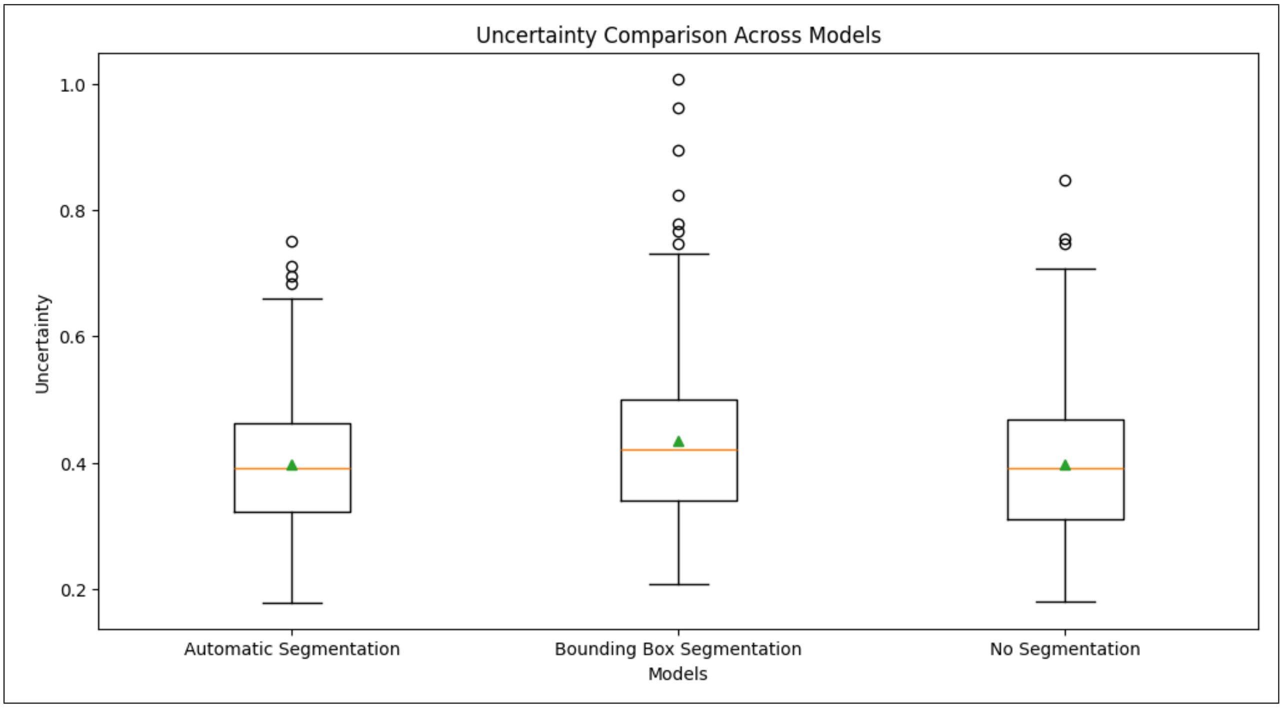

Figure 5: A box plot comparing the distribution of uncertainty values across the three different models on the test dataset (40 images)

4 Conclusion

The integration of the Segment Anything Model 2 (SAM2) with Bayesian Convolutional Neural Networks (BCNNs) presents a promising approach to enhancing multi-disease prediction from medical images. By combining the unique segmentation capabilities of SAM2 with the uncertainty quantification strengths of BCNNs, this hybrid model can potentially offer more reliable and confident diagnostic predictions.

By using transformations and edge detection to improve segmentation results, this approach not only improves the accuracy of boundary delineation but also addresses the challenges of under-segmentation in lower-contrast imaging modalities such as those found in the NIH Chest X-ray dataset. In turn, this improvement in segmentation allows the neural network to focus on uncertainty quantification in the most critical areas, leading to more informed and confident clinical decision-making.

Future work should focus on further validating this integrated approach across a larger scale of chest X-ray data and diverse datasets beyond chest x-rays, utilizing various data and feature processing techniques, and/or exploring its applicability to other medical imaging tasks and models. By leveraging the complementary strengths of segmentation models (e.g. SAM2) and uncertainty-aware models (e.g. BCNNs), it may be possible to achieve more accurate and trustworthy medical image analysis on a general scale, ultimately benefiting patient care and outcomes, as well as enhancing professional analytical efficiency.

References

[1] Alexander Kirillov, Kaiming He, Ross Girshick, and Piotr Dollár, "SAM 2: Segment Anything in Images and Videos." 2024.

[2] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, Christoph Feichtenhofer, "SAM 2: Segment Anything in Images and Videos." arXiv preprint arXiv:2408.00714v1, 2024. Available: `https://arxiv.org/html/2408.00714v1`

[3] Sourya Sengupta, Satrajit Chakrabarty, and Ravi Soni, "Is SAM 2 Better than SAM in Medical Image Segmentation?" arXiv preprint arXiv:2408.04212v2, 2024. Available: `https://arxiv.org/pdf/2408.04212v2`

[4] Carvalho, G. B., & Almeida, J. (2024). Exploiting the Segment Anything Model (SAM) for Lung Segmentation in Chest X-ray Images. *2024 27th Iberoamerican Congress on Pattern Recognition (CIARP)*, 1-14. Available: https://arxiv.org/html/2411.03064v1#bib.bib5

[5] DataCamp. (2024). Fine-Tuning SAM 2 on a Custom Dataset: Tutorial. Available: https://www.datacamp.com/tutorial/sam2-fine-tuning

[6] Ma, J., Kim, S., Li, F., Baharoon, M., Asakereh, R., Lyu, H., & Wang, B. (2024). Segment Anything in Medical Images and Videos: Benchmark and Deployment. *arXiv preprint arXiv:2408.03322v1*. Available: https://arxiv.org/html/2408.03322v1

[7] Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R. K., Antani, S., Thoma, G., Wang, Y. X., Lu, P. X., & McDonald, C. J. (2014). Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging, 33*(2), 233-245. https://doi.org/10.1109/TMI.2013.2284099

[8] Candemir, S., Jaeger, S., Palaniappan, K., Musco, J. P., Singh, R. K., Xue, Z., Karargyris, A., Antani, S., Thoma, G., & McDonald, C. J. (2014). Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Transactions on Medical Imaging, 33*(2), 577-590. https://doi.org/10.1109/TMI.2013.2290491

[9] Gaggion, N., Mosquera, C., Aineseder, M., Mansilla, L., Milone, D., & Ferrante, E. (2024). CheXmask Database: a large-scale dataset of anatomical segmentation masks for chest x-ray images (version 0.4). *PhysioNet*. https://doi.org/10.13026/pgag-by42

[10] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray14: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases [Data set]. National Institutes of Health Clinical Center. https://nihcc.app.box.com/v/ChestXray-NIHCC/file/220660789610

[11] Bargagna, F., De Santi, L. A., Martini, N., Genovesi, D., Favilli, B., Vergaro, G., Emdin, M., Giorgetti, A., Positano, V., & Santarelli, M. F. (2023). Bayesian convolutional neural networks in medical imaging classification: A promising solution for deep learning limits in data scarcity scenarios. *Journal of Digital Imaging, 36*(6), 2567-2577. https://doi.org/10.1007/s10278-023-00897-8

5 Appendix

A Data

We used the chest X-ray 14 dataset for training and testing, sampling on a small size of 200 images. The dataset was divided into training, validation, and test sets in a 60/20/20 ratio to ensure a balanced distribution. To prevent data leakage and bias, we ensured that images from the same patient were only included in one of the training, validation, or test sets.

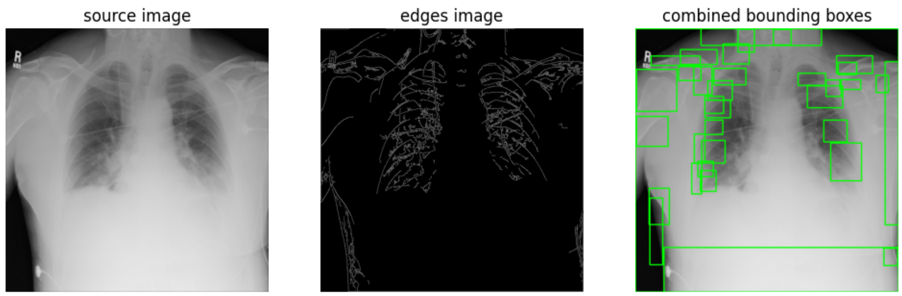

Figure 6: Side-by-side comparision between of an chest X-ray original, edge detection, boundary box images

As a method to enhance the segmentation quality by the SAM2 model, images were also preprocessed using Gaussian Blur and Canny Edge Detection. To convert the corresponding true labels (disease detections) of the images into a more usable format for the Bayesian convolutional neural network and improve model performance, we performed one-hot encoding and normalization.

B SAM2 Model Configuration

The SAM2 model was loaded with the `sam2_hiera_small.pt` checkpoint. Various parameters for the zero-shot segmentation generator, `SAMAutomaticMaskGenerator`, included `points_per_side` (64), `pred_iou_thresh` (0.8), and `stability_score_thresh` (0.92).

C Training and Evaluation

The BCNN was trained for 300 epochs using a T4 GPU and its performance was assessed on the validation and test sets. Uncertainty estimation was performed on the test set using Monte Carlo Dropout by obtaining the mean predictions and their corresponding variances. These metrics helped provide insights into the model's confidence in its predictions.

D Code Implementation

The code used to produce these results is available at the following link: GitHub Repository (`https://github.com/DumplingCodeEater/SAM-with-BCNN`).