Paper Review: Medical SAM2: Segment Medical Images as Video via Segment Anything Model 2

By: Lam Nguyen

1. SUMMARY

The Medical Segment Anything Model 2 (SAM 2) has the ability to segment both 2D images and 3D models.

Segmentation put simply is using a machine learning model to divide input data into distinct zones/ classes that are useful.

To implement this, SAM 2 uses a self-sorting memory bank mechanism that selects informative embeddings based on confidence and dissimilarity. The memory bank mechanism allows segmentation across multiple images using a single prompt.

To evaluate the performance of medical SAM 2, experiments were performed on 5 2D tasks and 9 3D tasks: Spleen, Left and Right Kidney, Gallbladder, Esophagus, Liver, Stomach, Aorta, IVC, Veins, Pancreas, and Adrenal Glands.

The other part of the experiment was to get users to input a prompt to segment images. The experiment included a mixture of prompts and no prompts to segment the images and 3d models.

Experiments show that MedSAM-2 was consistently outperforming the most state-of-the-art 2d and 3d segmentation models while reducing the need for constant human interaction.

2. STRENGTHS

- The MedSAM-2 model when set up is easy for the end user to use as it is mostly automated and can be prompted with a single text input as well as the input images.
- Outperforms other Segmentation Models in both 2D and 3D image tasks using one model.
- Can handle both 2D image and 3D Model inputs which make it extremely versatile for medical imaging tasks.

3. WEAKNESSES

- Prompting and Vision Language Models require a very large investment in hardware for both training and inference. This model used 64 Nvidia A100 GPUs for training. And although inference will require less hardware computation, a significant investment in either hardware or cloud computing is required to run this model. Hospitals without the willingness, technical knowledge and/or capital to invest in hardware or cloud computing requirements won't be able to use this model… Especially considering that a human most likely will have to confirm the results of the model anyways… Is MedSAM-2 even feasible when medical care in the United States already is extremely expensive?
- To effectively prompt the model, a user has to be trained to be able to ask the MedSAM-2 the right prompts in the right manner in order to get an accurate and consistent result. When lives are on the line, any sort of deviation in prompting might cause hallucinations or differences. This requires a trained expert to check and confirm outputs.

4. TECHNICAL EXTENSIONS

- The researchers could create a quantized model that requires much less resources to do inference on or to train with. This will allow more hospitals and/or researchers to access this model for training and inference and use. In turn this might actually allow better refinement of the model since more data and testing is available.
- A list or application of premade prompts might be useful to have in order to standardize prompt inputs across multiple individual and hospitals in order to reduce the chances of hallucination. When it comes to human lives, more care should be taken in

order to be able to use this technology in the real world.

5. OVERALL REVIEW

The MedSAM-2 model is able to outperform many of the state of the art medical segmentation models in both 2D and 3D visual data. It can be interacted with using a single prompt which can make it easy to work with for the medical practitioner.

The downside of this model is that it requires a significant investment in computing power which can limit its use in hospital settings…. Especially where liability is concerned and a trained human needs to be there to confirm the results anyways.

To reduce cost and increase consistency in output, a quantized model could be created as well as an Application to standardize the prompts might be useful.