4th International Workshop of Advances in Simplifying Medical UltraSound (ASMUS)
October 8, 2023

Synthetic Boost: Leveraging Synthetic Data for Enhanced Vision-Language Segmentation in Echocardiography

Rabin Adhikari*, Manish Dhakal*, **Safal Thapaliya***, Kanchan Poudel, Prasiddha Bhandari, Bishesh Khanal

Nepal Applied Mathematics and Informatics Institute for research (NAAMII)

**equal contribution*

Outline

- Echocardiography and Segmentation Models
- Synthetic Data and Prompt Engineering
- Experiments and Results

Outline

- **Echocardiography and Segmentation Models**
- Synthetic Data and Prompt Engineering
- Experiments and Results

Echocardiography

- Cheap, portable and gives HD images
- Requires strong segmentation algorithms

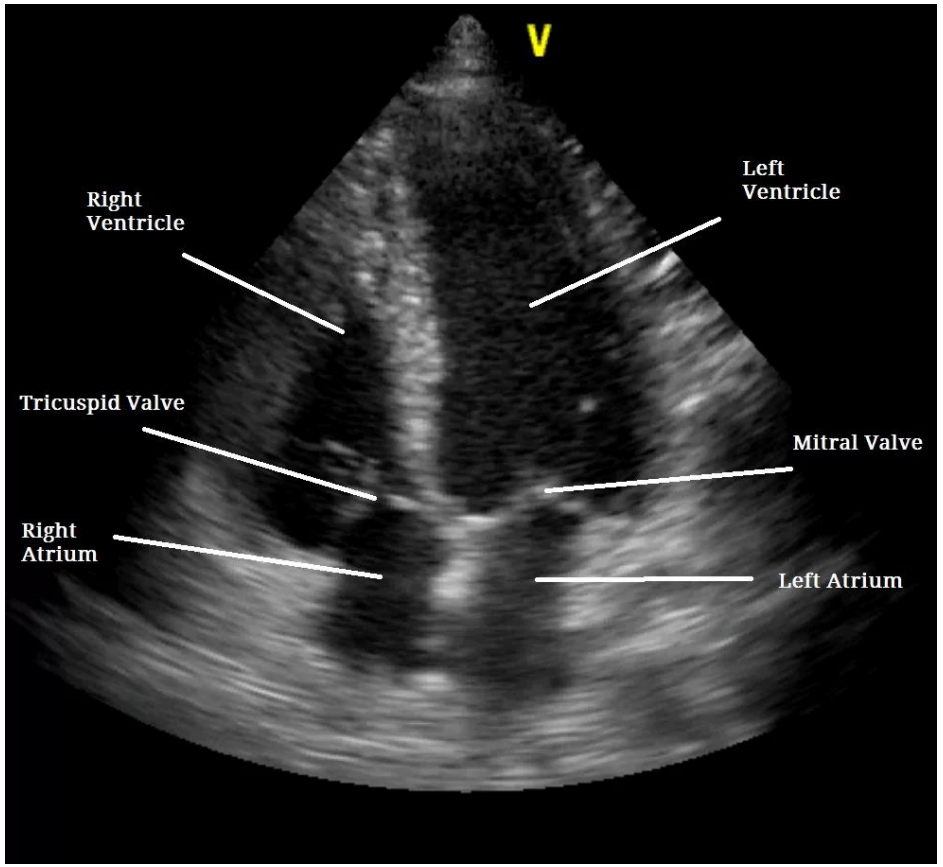


Image source: https://heartsense.in/echocardiogram/

How hard is it to accurately segment different parts in an echocardiography?

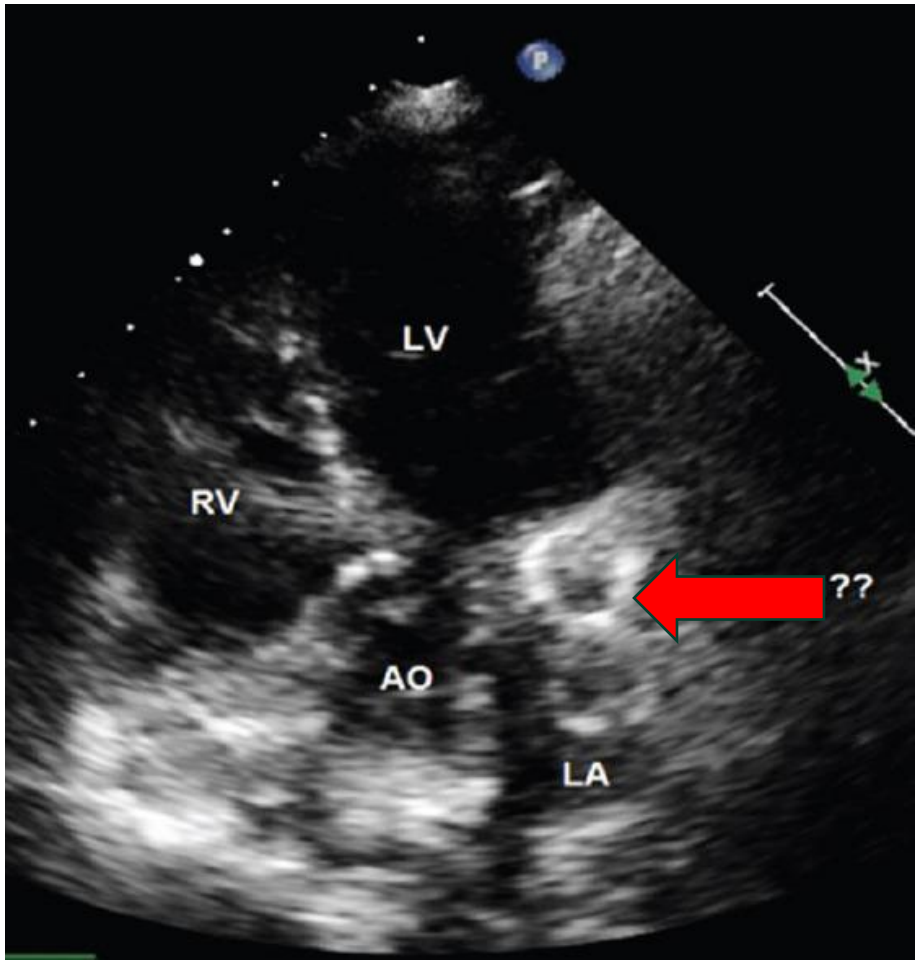
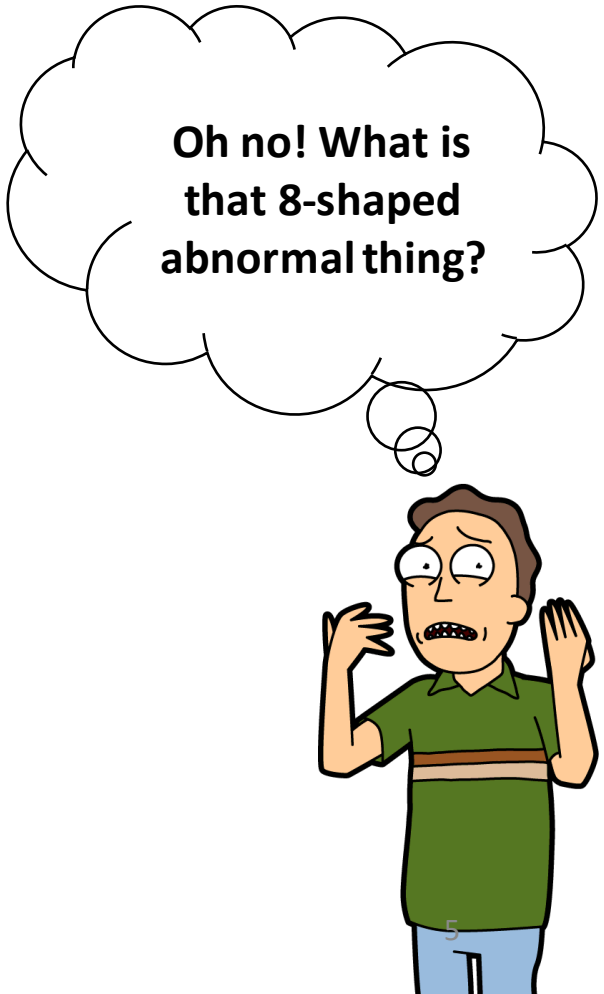
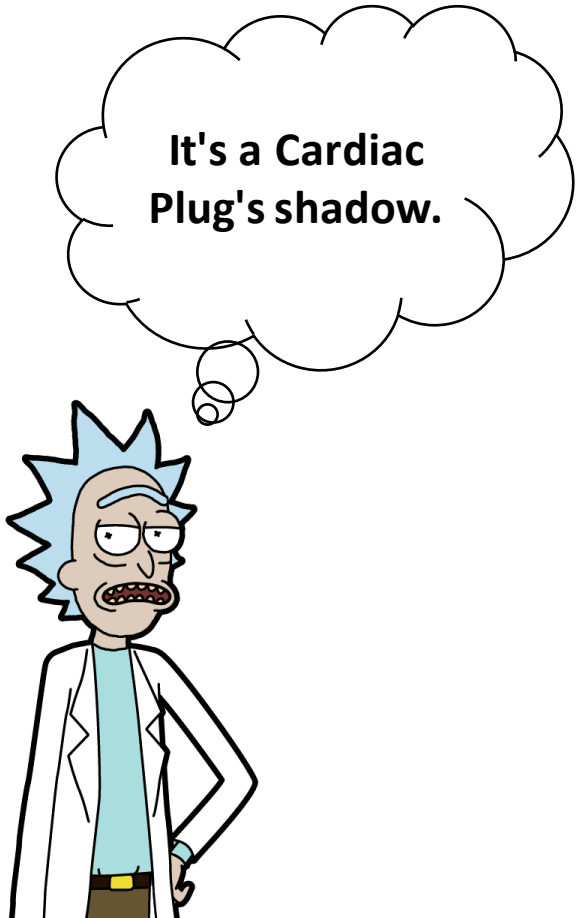


Image source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7799067/



How hard is it to accurately segment different parts in an echocardiography?



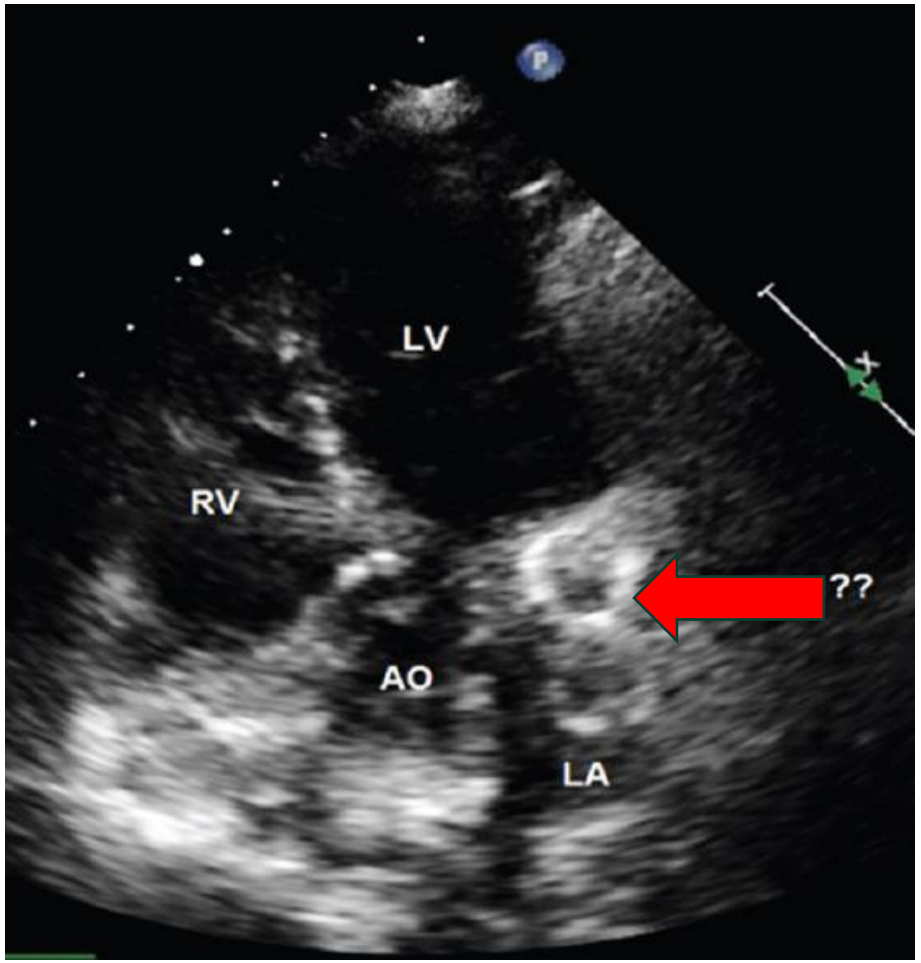
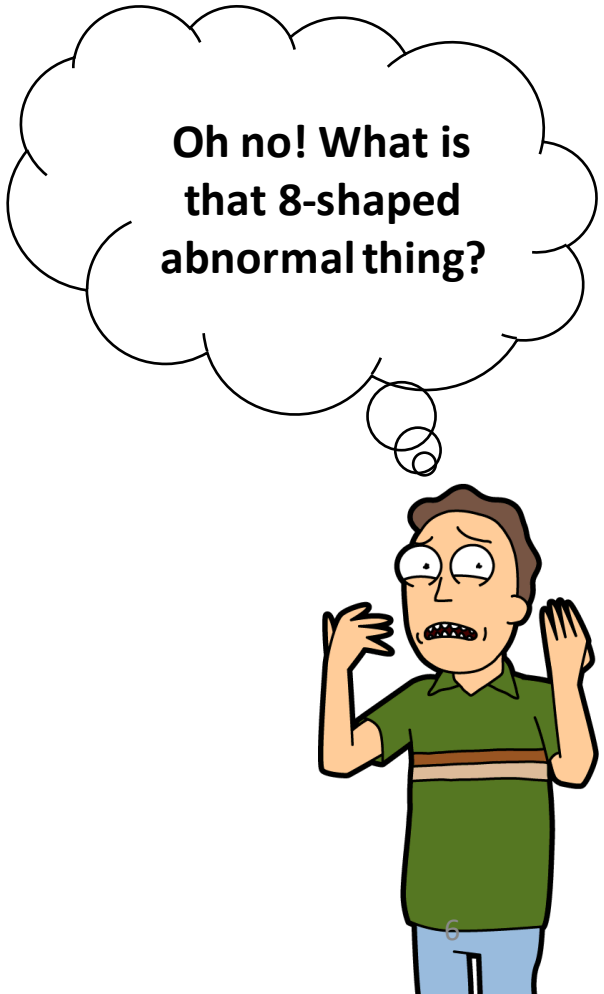
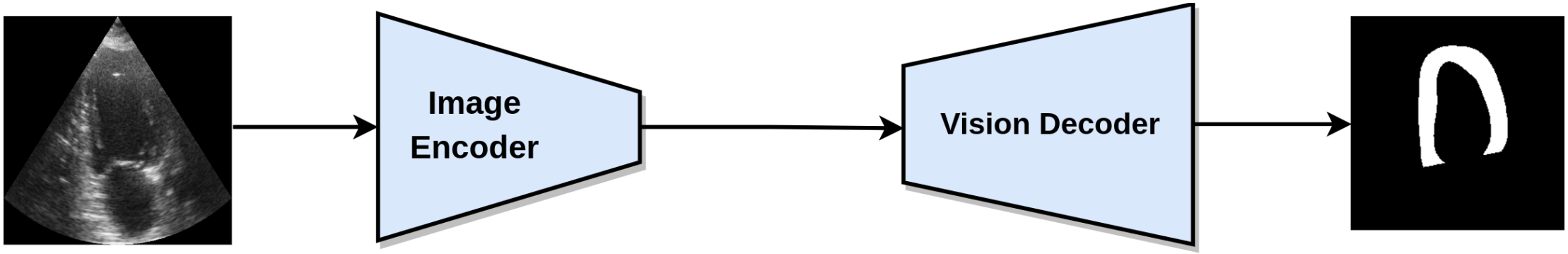


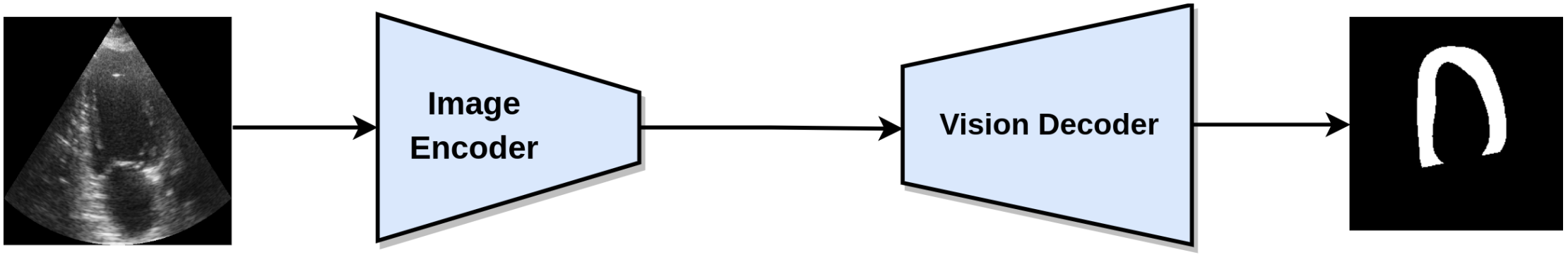
Image source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7799067/



Existing Segmentation Models

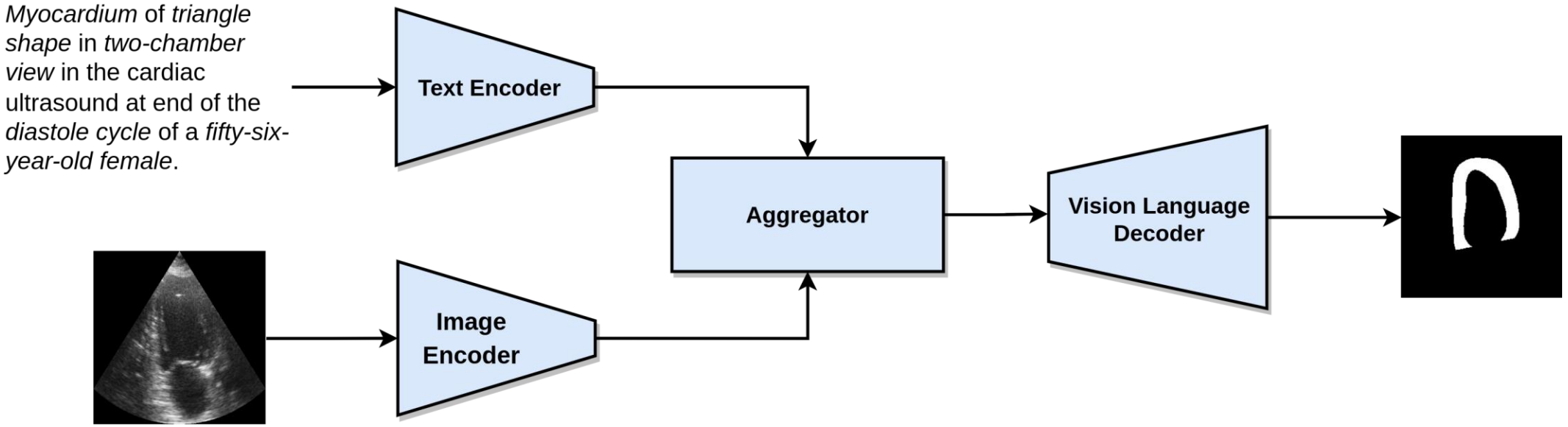


Existing Segmentation Models

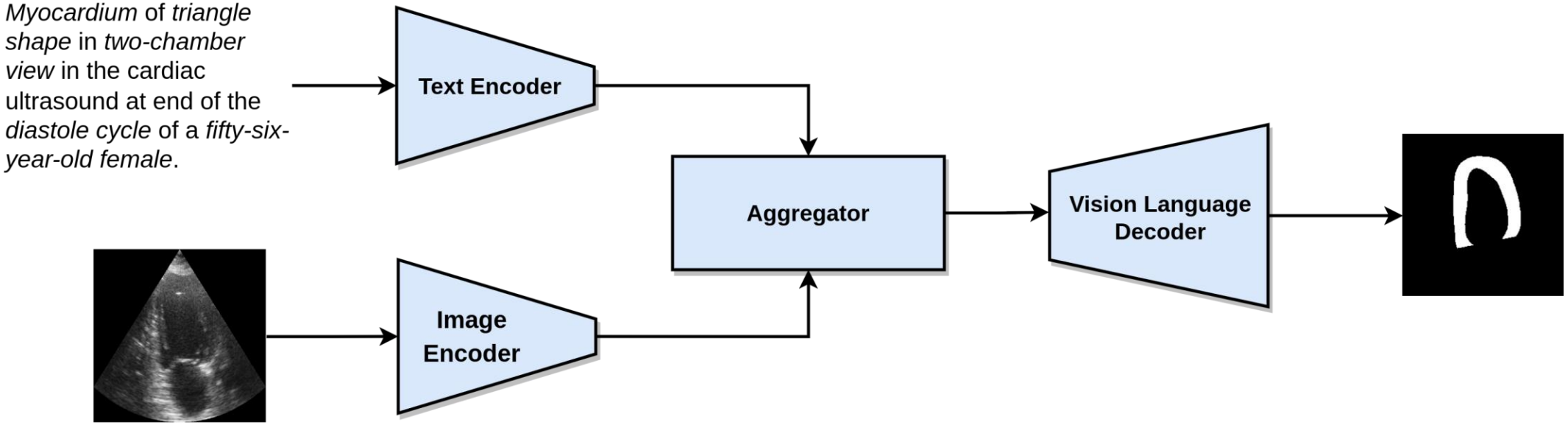


- **Rely on large amount of annotated data for supervised training**
- **Lack explainability**
- **Require retraining when new classes are introduced**
- **Not resilient to distribution shifts**

Vision Language Segmentation Models



Vision Language Segmentation Models



- **Extract rich information from image and language prompt pairs**
- **Aid in accurate and explainable segmentation**
- **Requires large image-prompt pairs for good finetuning performance**

Outline

- Echocardiography and Segmentation Models
- **Synthetic Data and Prompt Engineering**
- Experiments and Results

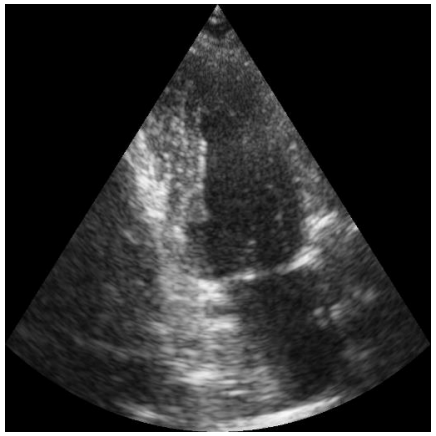
Real and synthetic echocardiography

CAMUS [1]

- Real cardiac segmentation dataset
- 2D apical two-chamber and four-chamber views at end-diastole (ED) and end-systole (ES) cycles
- *Train-val-test split:* 600-400-200

Synthetic Echocardiography [2]

- Synthetic echocardiography images generated using SDMs [3]
- Takes perturbed anatomical masks as conditioning information to denoise the noisy images and generates echocardiographic images
- *Train-val split:* 8000-1000



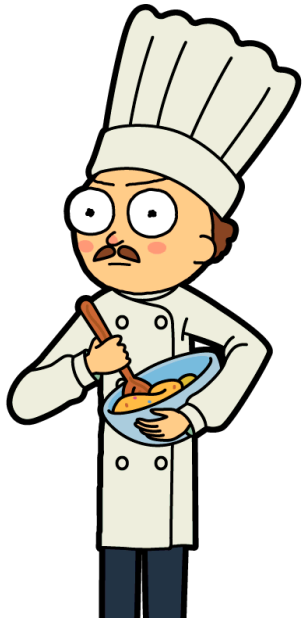
9x more training samples

[1] Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., ... & Bernard, O. (2019). Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE transactions on medical imaging, 38*(9), 2198-2210.

[2] Stojanovski, D., Hermida, U., Lamata, P., Beqiri, A., & Gomez, A. (2023). Echo from noise: synthetic ultrasound image generation using diffusion models for real image segmentation. In *International Workshop on Advances in Simplifying Medical Ultrasound* (pp. 34-43). Cham: Springer Nature Switzerland.

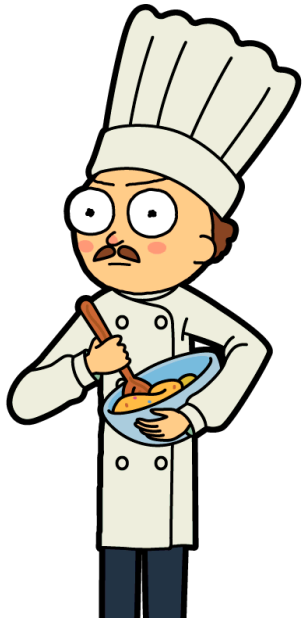
[3] Wang, W., Bao, J., Zhou, W., Chen, D., Chen, D., Yuan, L., & Li, H. (2022). Semantic image synthesis via diffusion models. *arXiv preprint arXiv:2207.00050*.

But we don't have prompts for CAMUS. How to get them?



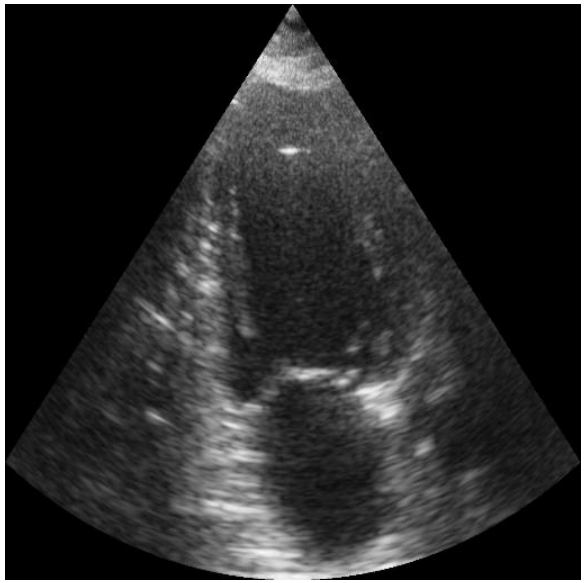
We'll create our own prompts.

But we don't have prompts for CAMUS. How to get them?



Prompt Engineering

Following our previous work [1], we created 7 different prompts.



Target Structure — Structure's Shape — Apical View

Left ventricular cavity of oval shape in two-chamber view in the cardiac ultrasound at the end of the diastole cycle of a seventy-three-year-old male with good image quality.

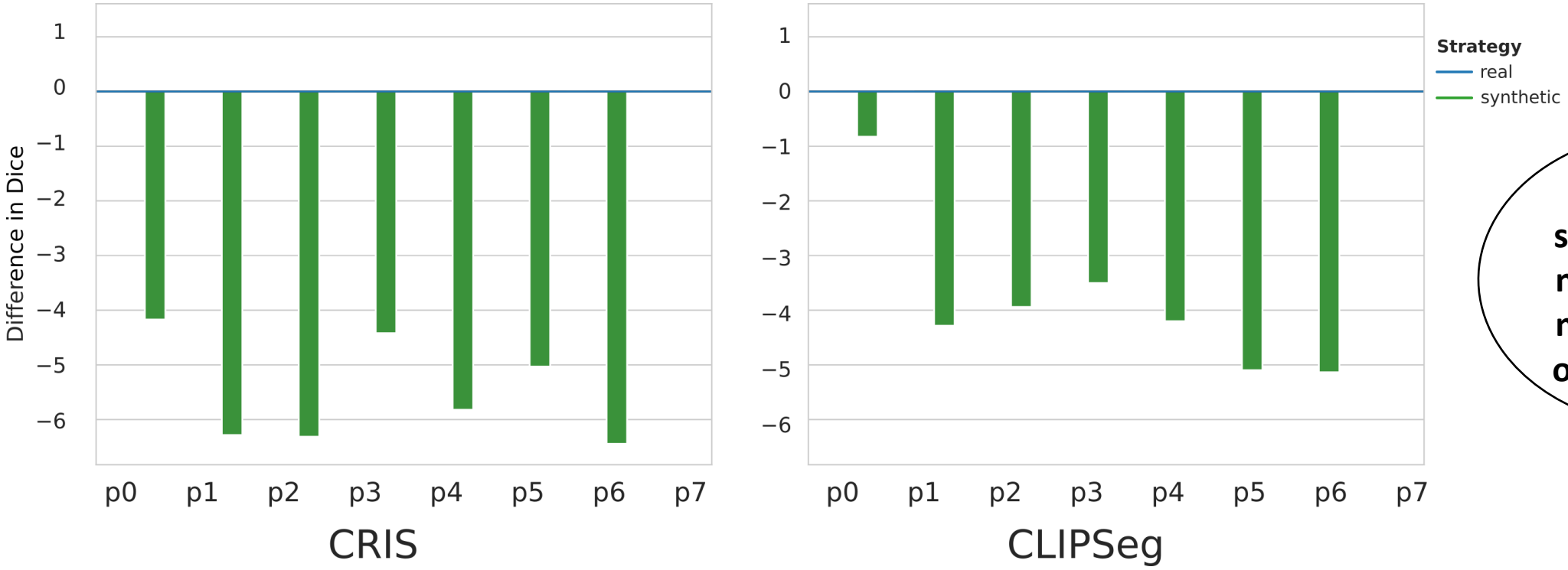
Cardiac Cycle — Patient's Age — Patient's Sex — Image Quality

[1] Poudel, K., Dhakal, M., Bhandari, P., Adhikari, R., Thapaliya, S., & Khanal, B. (2023). Exploring transfer learning in medical image segmentation using vision-language models. *arXiv preprint arXiv:2308.07706*.

Outline

- Echocardiography and Segmentation Models
- Synthetic Data and Prompt Engineering
- **Experiments and Results**

How does synthetic data help in finetuning?

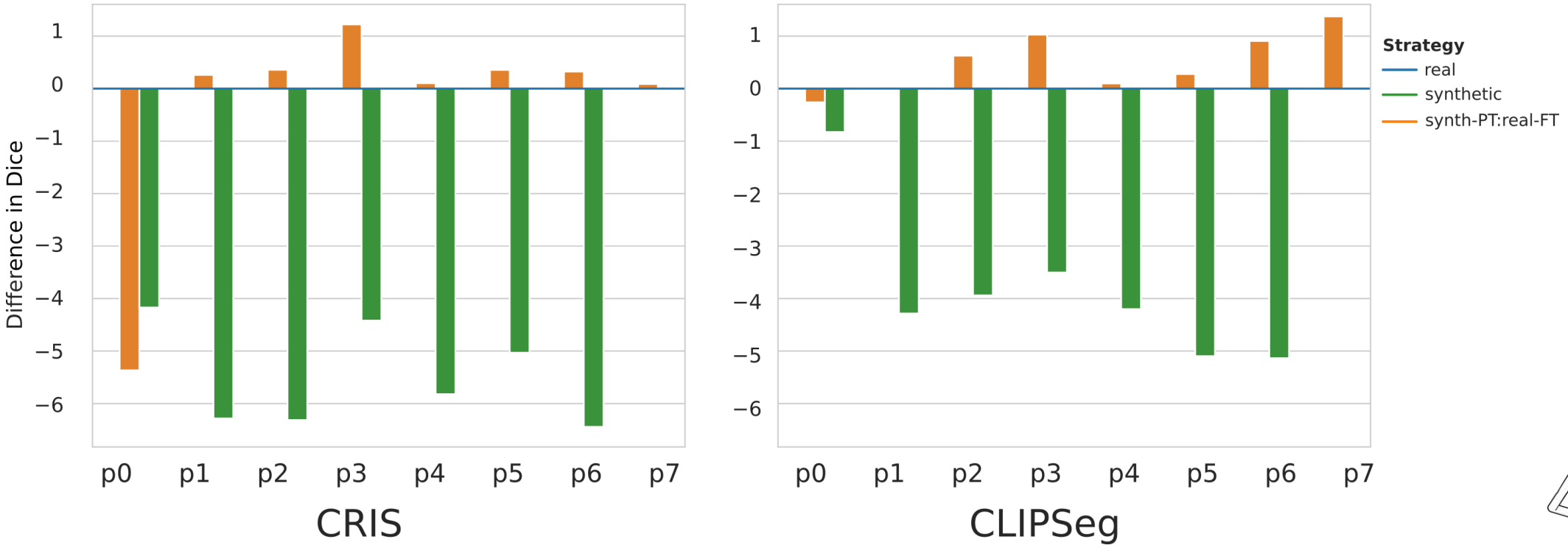


Using just the synthetic data is not helping...we need to look for other strategies.

Synthetic data better than no data.

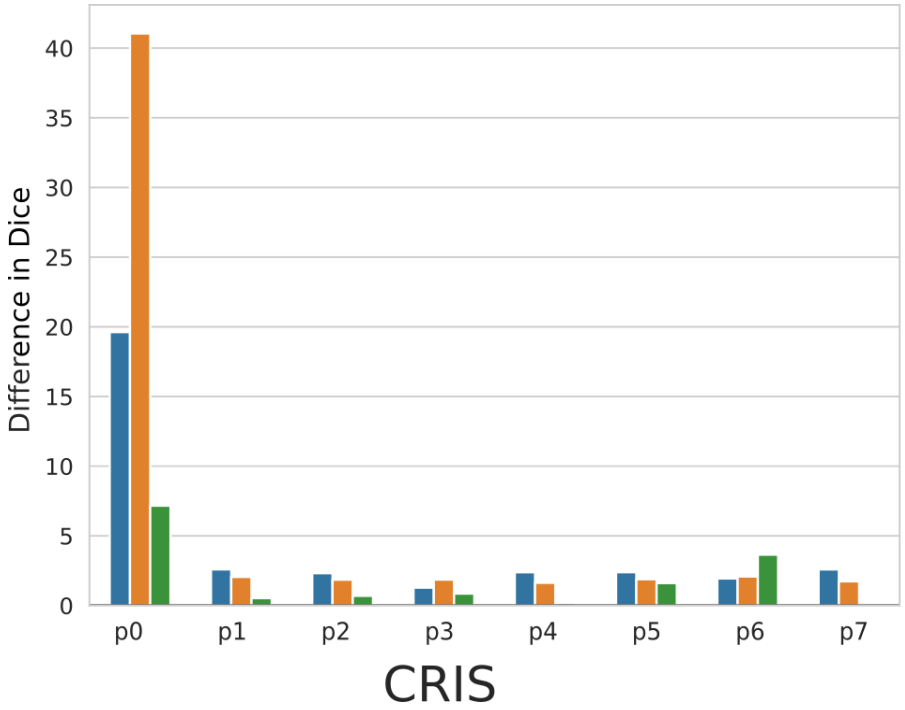
Quality > Quantity

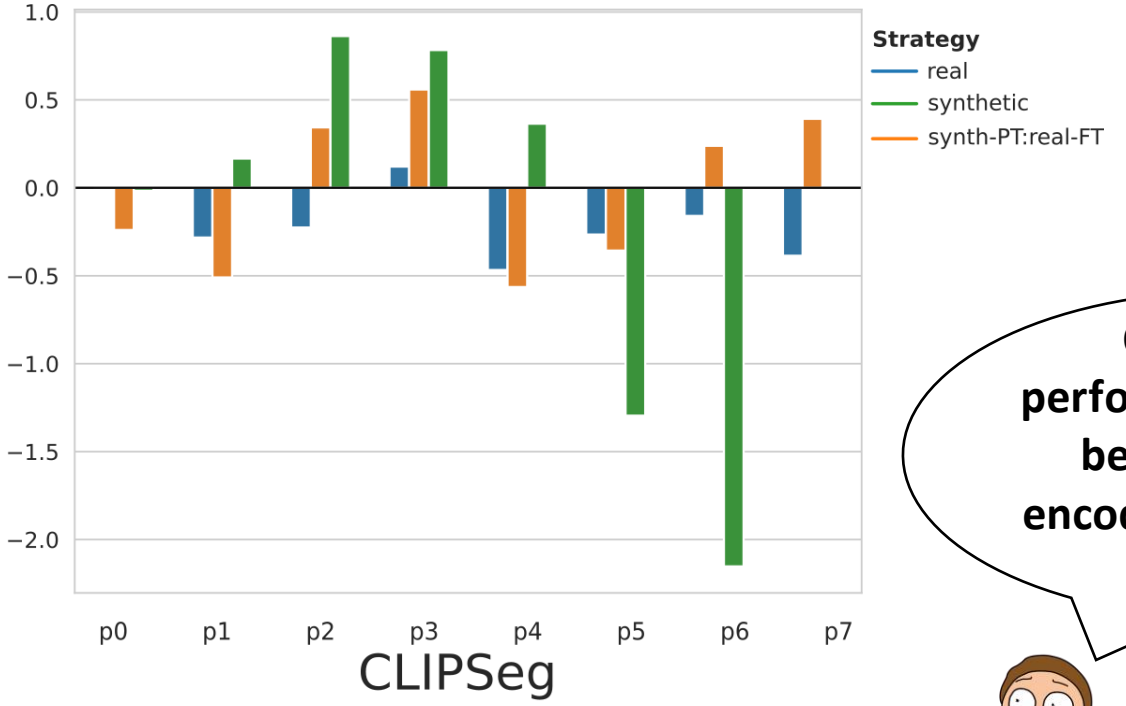
How does synthetic data help in finetuning?



Pretraining on synthetic

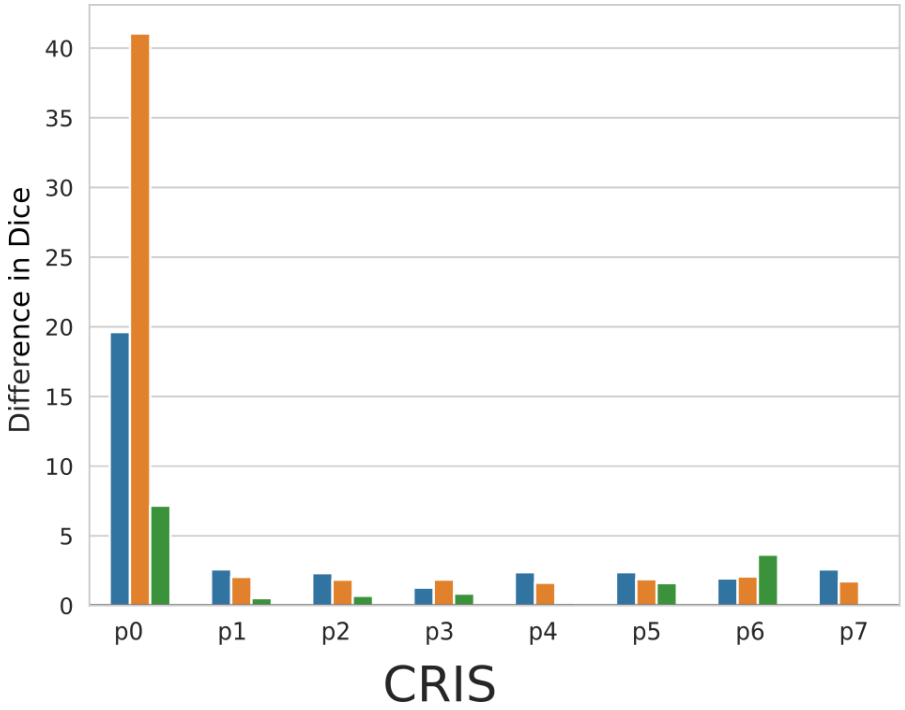
Does training the encoders during finetuning give different results?

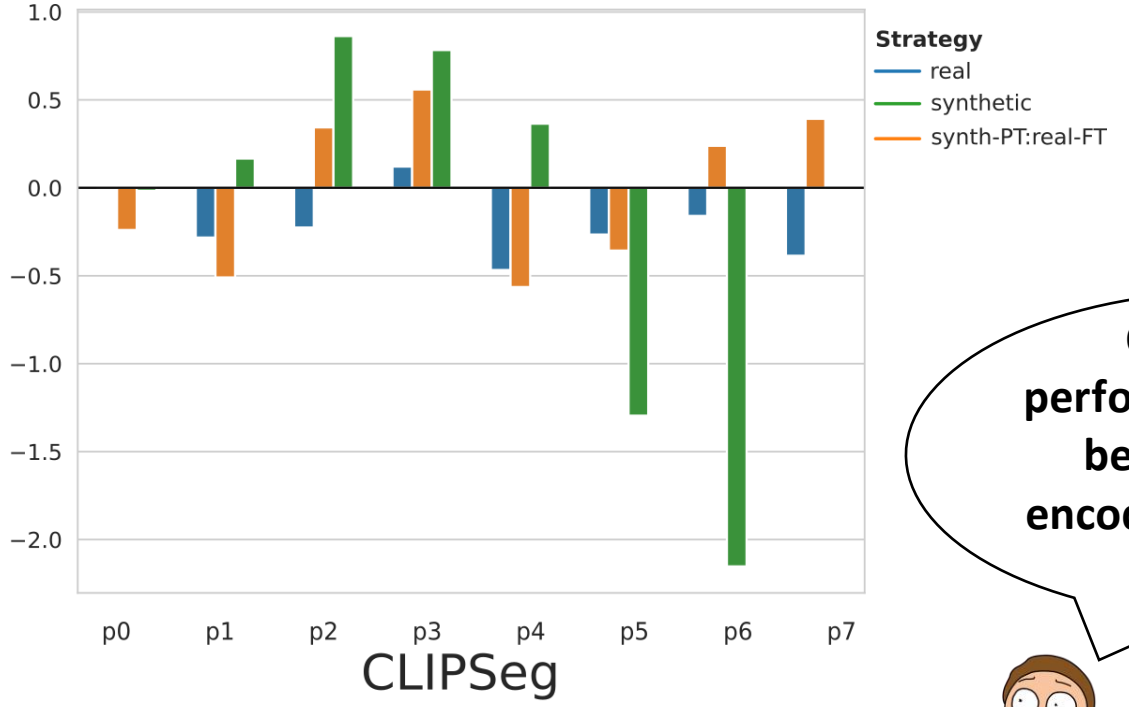




Only CRIS's performance seem to be better when encoders are trained. But....why?

Does training the encoders during finetuning give different results?





Only CRIS's performance seem to be better when encoders are trained. But....why?

CRIS finetuned encoders 🥲

CLIPSeg did not 😳

Future Directions

- Train using both real and synthetic data and describe whether image is real or synthetic in the language prompt.
- Find ways to generate synthetic triplets of image-language-mask at scale without annotated image-mask pairs.

Thank You

Scan this to know more about our work