ML4Seismic Partners Meeting 2023

Finding What You Want: Prompting Based Segmentation in Foundation Models

Prithwijit Chowdhury, Mohit Prabhushankar and Ghassan AlRegib

OLIVES @GeorgiaTech

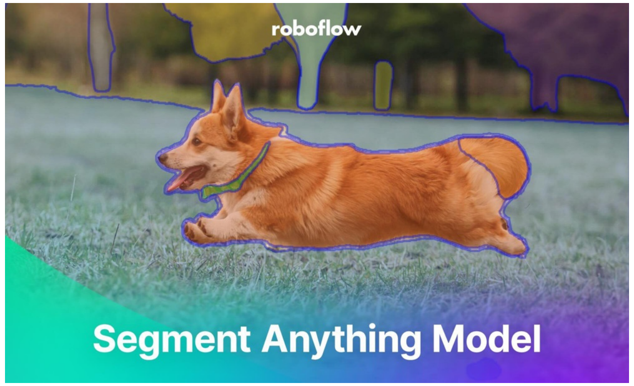
Georgia Tech

Foundation Models
Generality, Adaptability & Scalability: One model to solve them all

Foundation models are multi (several billion) parameter AI models that are trained on a massive dataset of text and code and image

PaLM 2

ChatGPT



•Generalization: Perform a wide range of tasks, without the need to be trained on each task.

•Adaptation: Adapt to new tasks and domains by fine-tuning on a small amount of data.

•Scaling: Create and annotate data for its own training purposes.

Generality, Adaptability & Scalability: One model to solve them all

Creating new data and scaling them for adaptation has never been easier.

Foundation models can help users create data for a new task and allow its annotation or labelling

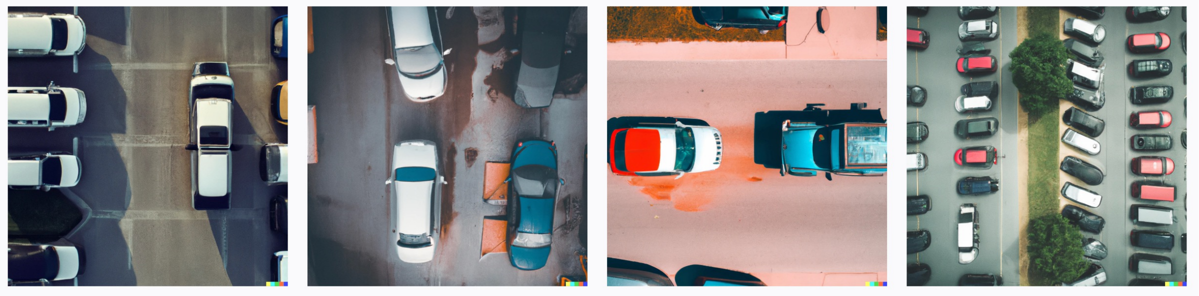
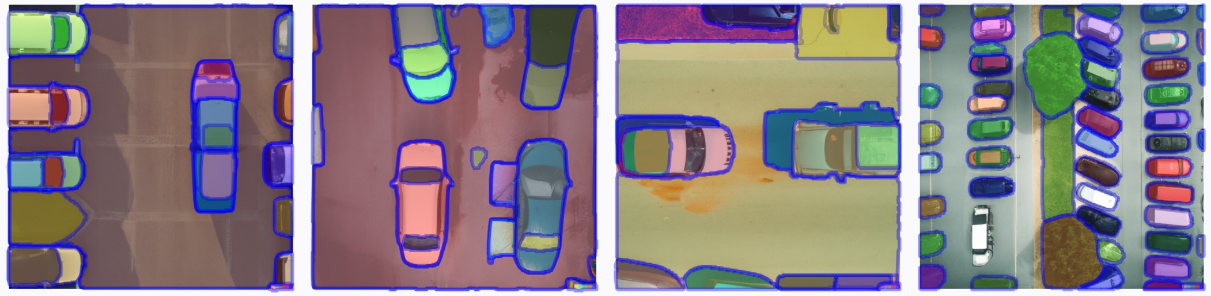


Image generation using DALLE-2



Everything segmentation in SAM

A new annotated dataset for overhead car images was generated in minutes using DALLE and SAM

Generalizability of Foundation Models

Large Foundation Models claim to allow zero shot generalization to unseen downstream tasks

However, most of these foundation models fail to generalize to seismic tasks

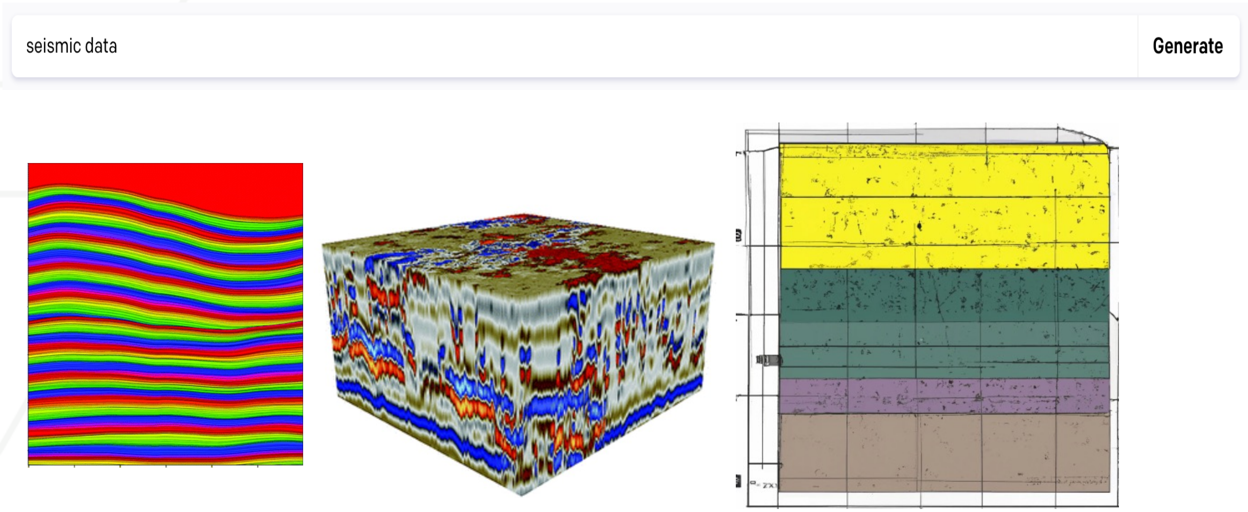
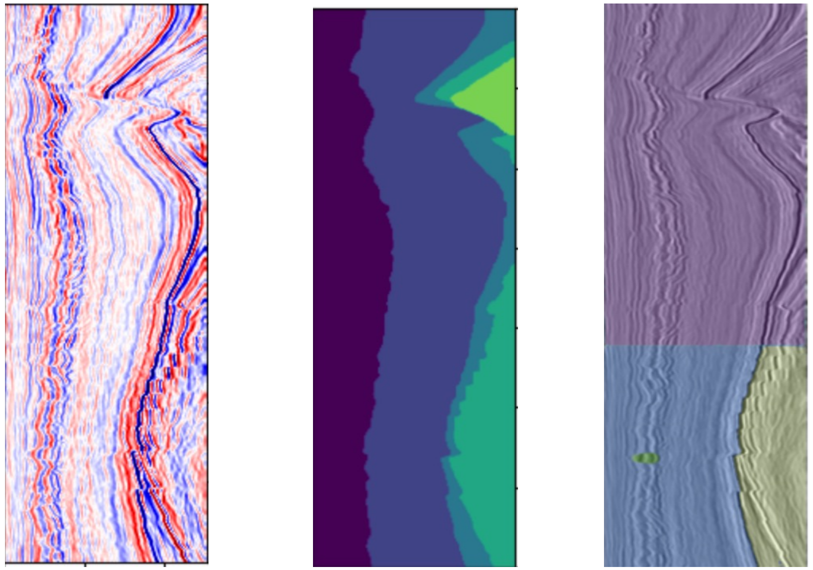


Image generation using DALLE-2



Seismic Image | Ground Truth | SAM Output

Everything segmentation in SAM

Adaptation of Foundation Models: Prompting

Prompt-able Foundation Models can be adapted to unseen downstream tasks

Adding prompts to allow models to generate outputs that are more informative based on user's needs

Prompt:

a seismic volume showing different facies classification like the F3 block | **Generate**

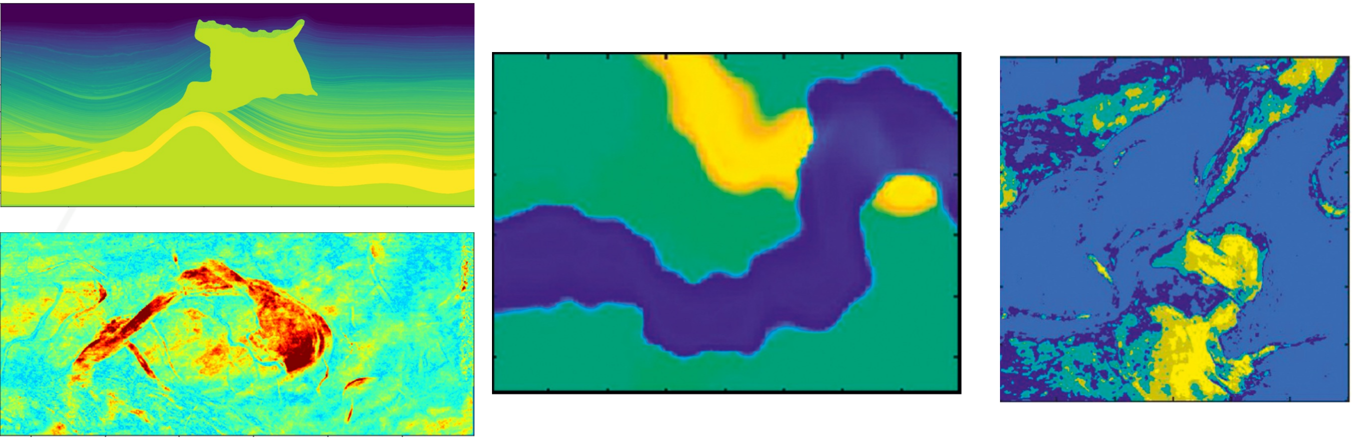
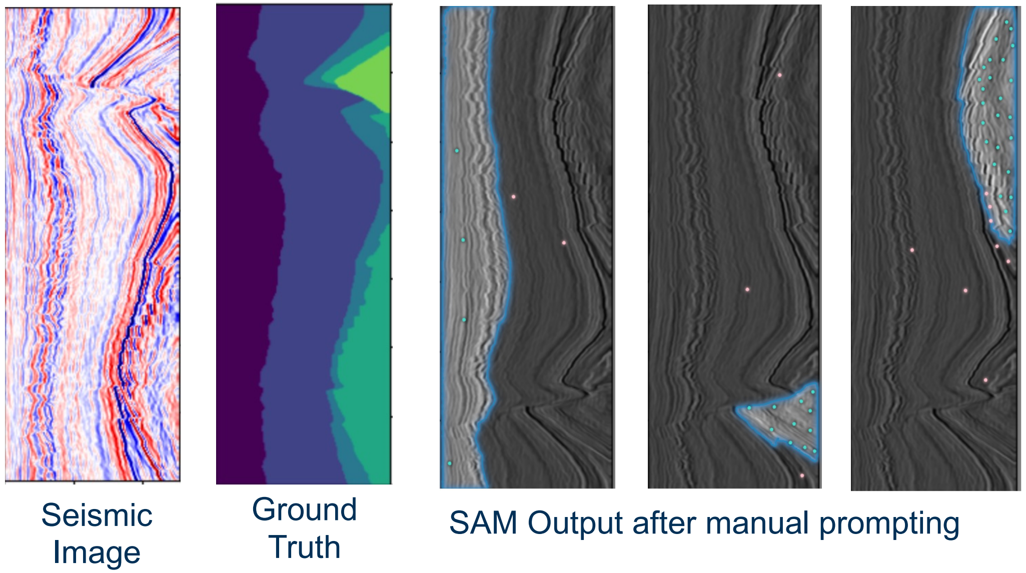


Image generation using DALLE-2



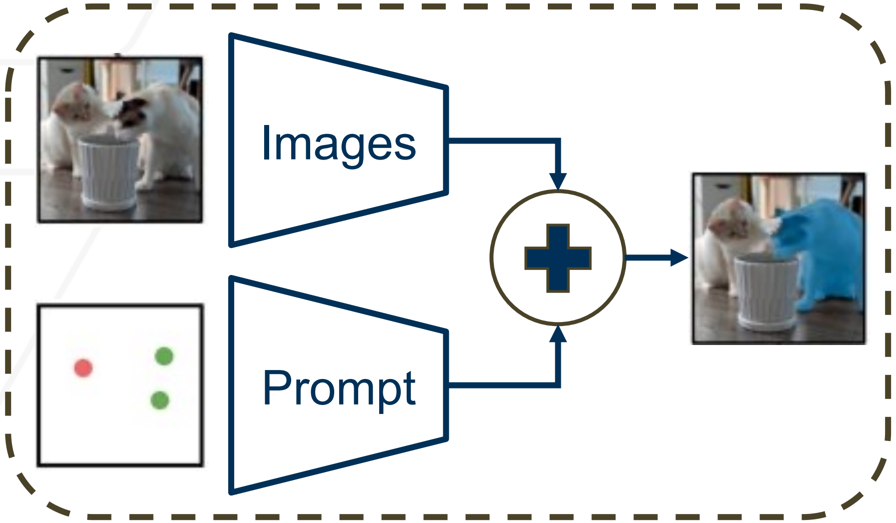
Prompted segmentation in SAM

Still the output is either not accurate or very costly (time/labor)

Segment Anything (SAM)

A prompt-able segmentation system

SAM is a capable of zero-shot generalization to unfamiliar objects and images, without the need for additional training.



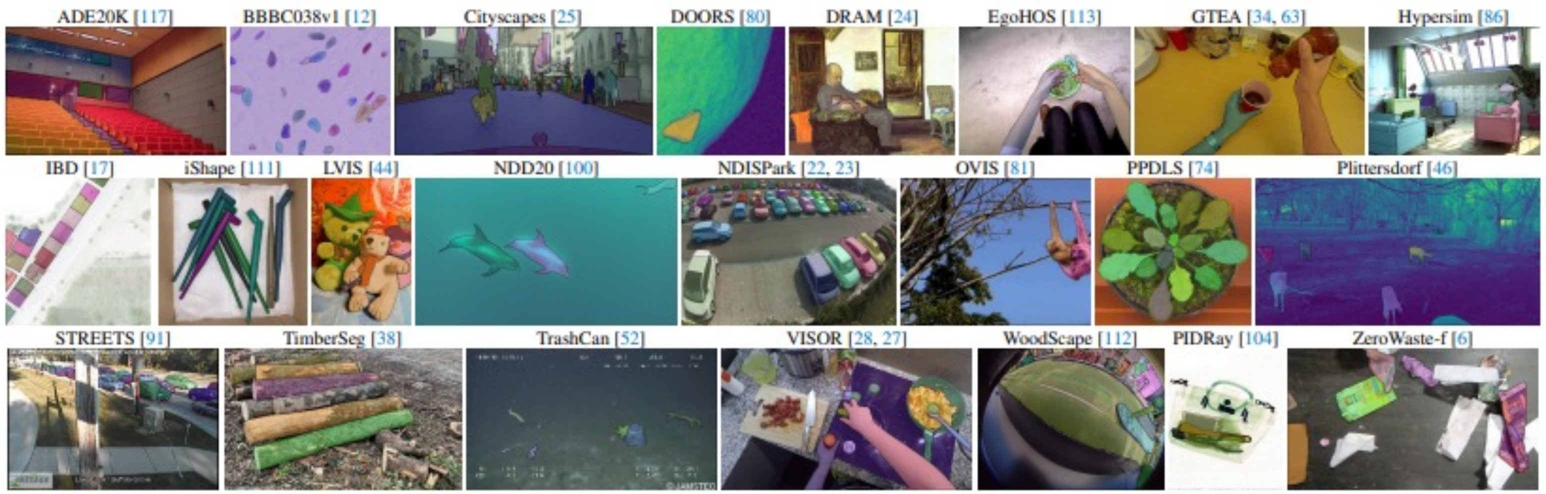
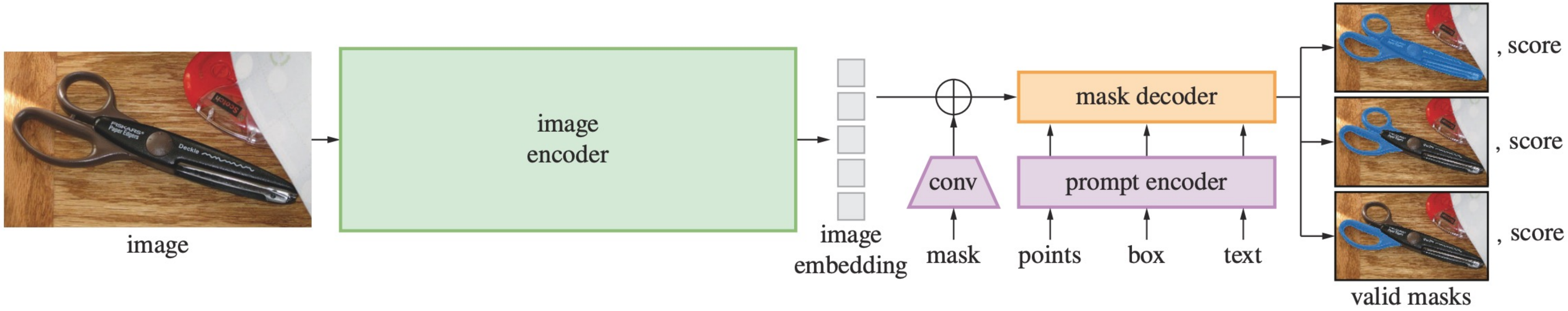


Figure 8: Samples from the 23 diverse segmentation datasets used to evaluate SAM's zero-shot transfer capabilities.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. *arXiv preprint arXiv:2304.02643*

Segment Anything (SAM)
Model Architecture

Consists to three main blocks: Image Encoder, Prompt Encoder and Mask Decoder

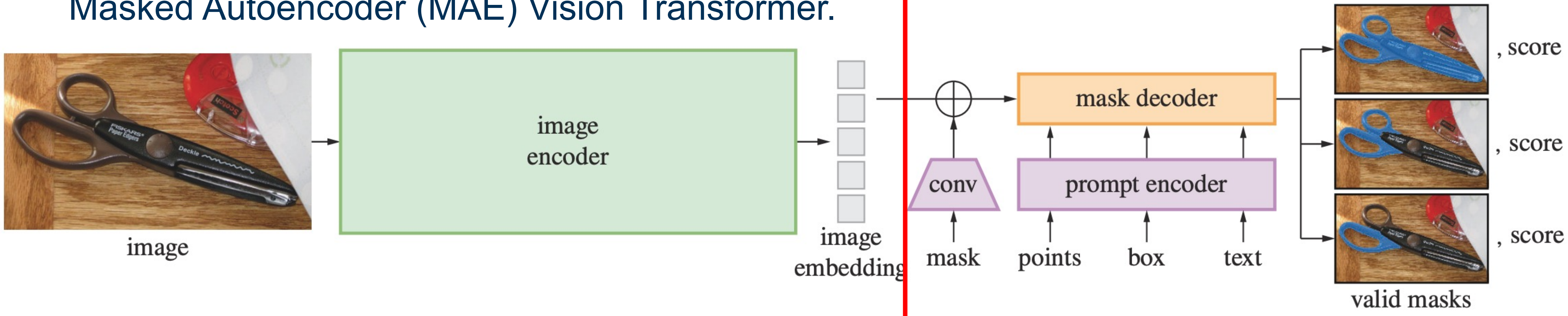


Kirillov, Alexander, et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).

Segment Anything (SAM)
Model Architecture

Transformer based image encoder downscales an image to a 1 channel embedding

Masked Autoencoder (MAE) Vision Transformer.

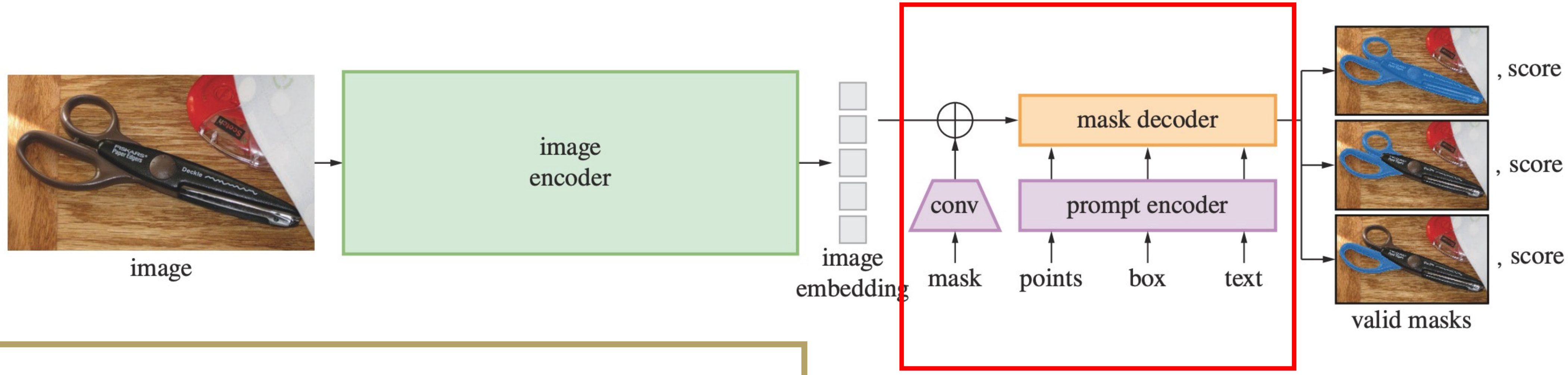


- Can learn from unlabeled data.
- Robust to images of various resolutions without sacrificing performance.

Kirillov, Alexander, et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).

Segment Anything (SAM)
Model Architecture

Prompt encoder embeds the positional information with the image channel



Prompt encoder takes Pixel based points, bounding boxes, or masks.

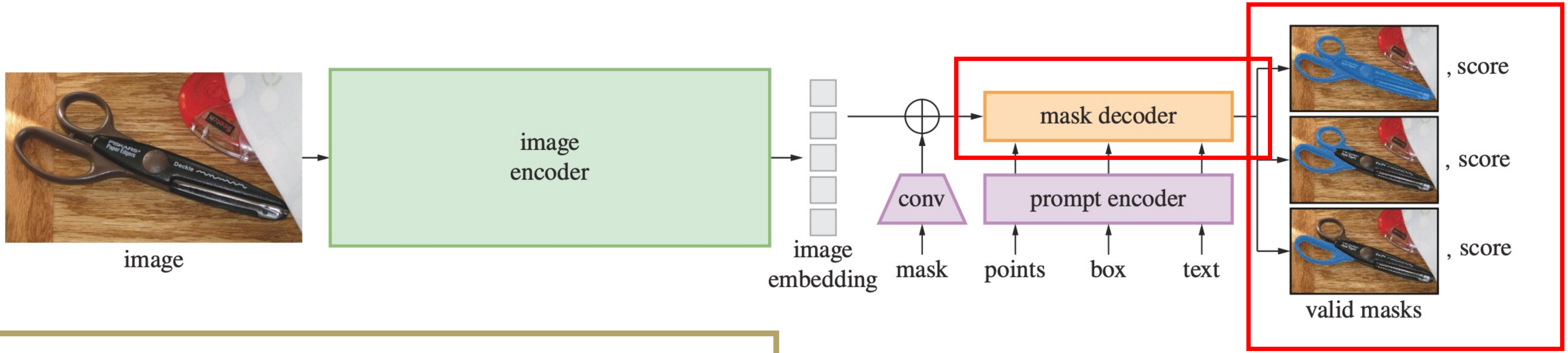
Texting prompt encoder takes text-image encoder using CLIP.

Encodes the positional information to the image embedding

Kirillov, Alexander, et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).

Segment Anything (SAM)
Model Architecture

Prompt encoder embeds the positional information with the image channel



SAM outputs are all **class agnostic**: objects are **segmented without distinguishing between different instance**s or class labels.

It can cut out any object by not classify it.

Mask decoder maps the image and prompt embeddings to an output mask

Kirillov, Alexander, et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).

Segment Anything (SAM) - Inferencing

Out of the box Inferencing allows two mode: Everything detection and prompt prediction

Manual prompt-based detection only allowed segmentation of a single object/ region of interest.

Everything detection

Prompt Prediction

Point prompts generated every 4X4 pixels

All objects segmented

Manual prompting selects only one segment

Kirillov, Alexander, et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).

SAM – Inferencing (Out of the Box)

Different types of prompt prediction based on points and bounding boxes.

Selection and exclusion can be done based on the type of prompt chosen.

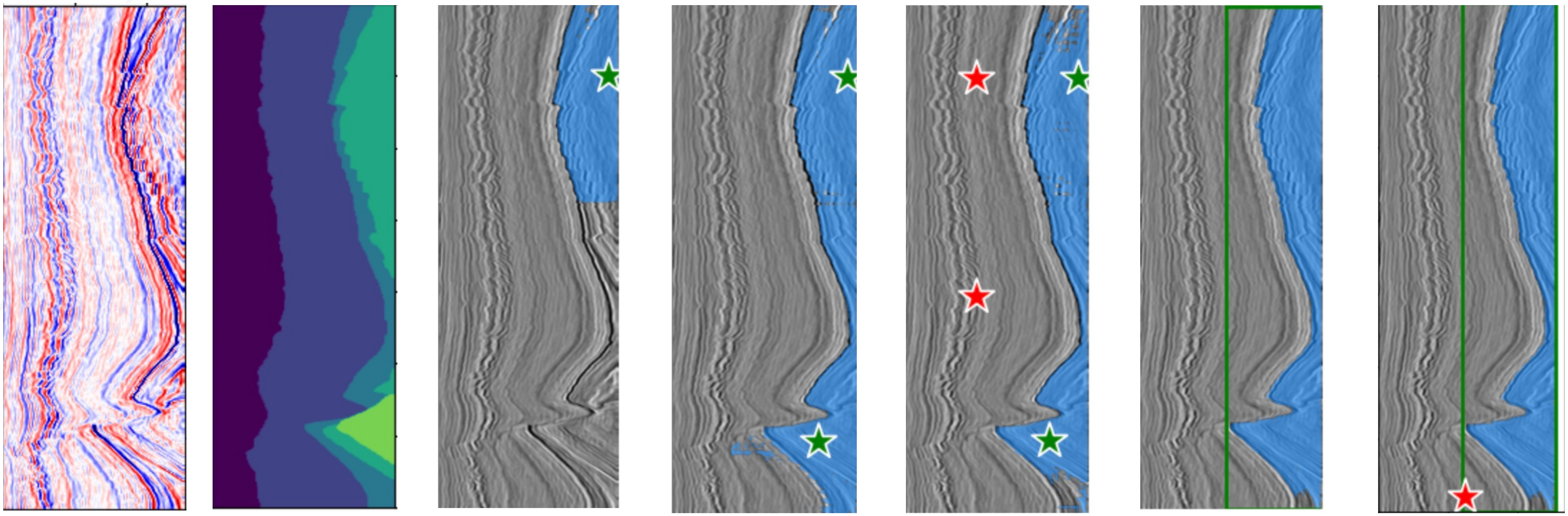
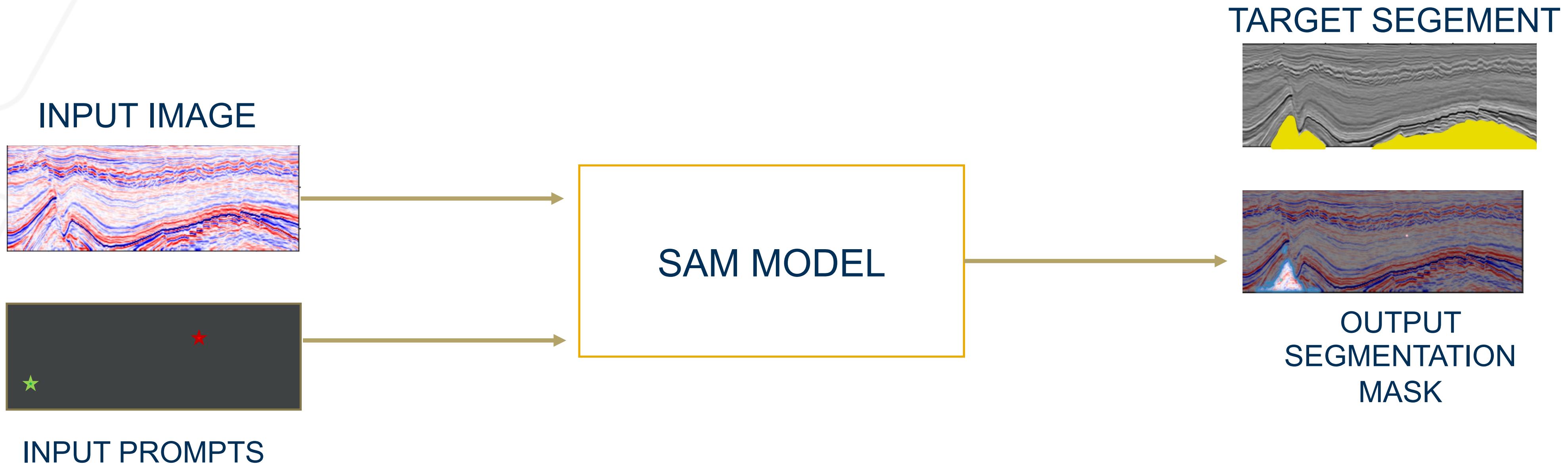


Image | Ground truth | Single point prompt | Double Point prompt | Exclusion Point prompt | Bounding box prompt | Mixed prompts

SAM – Inferencing (Out of the Box)

Re-prompting: Adding or removing prompts based on the model response.

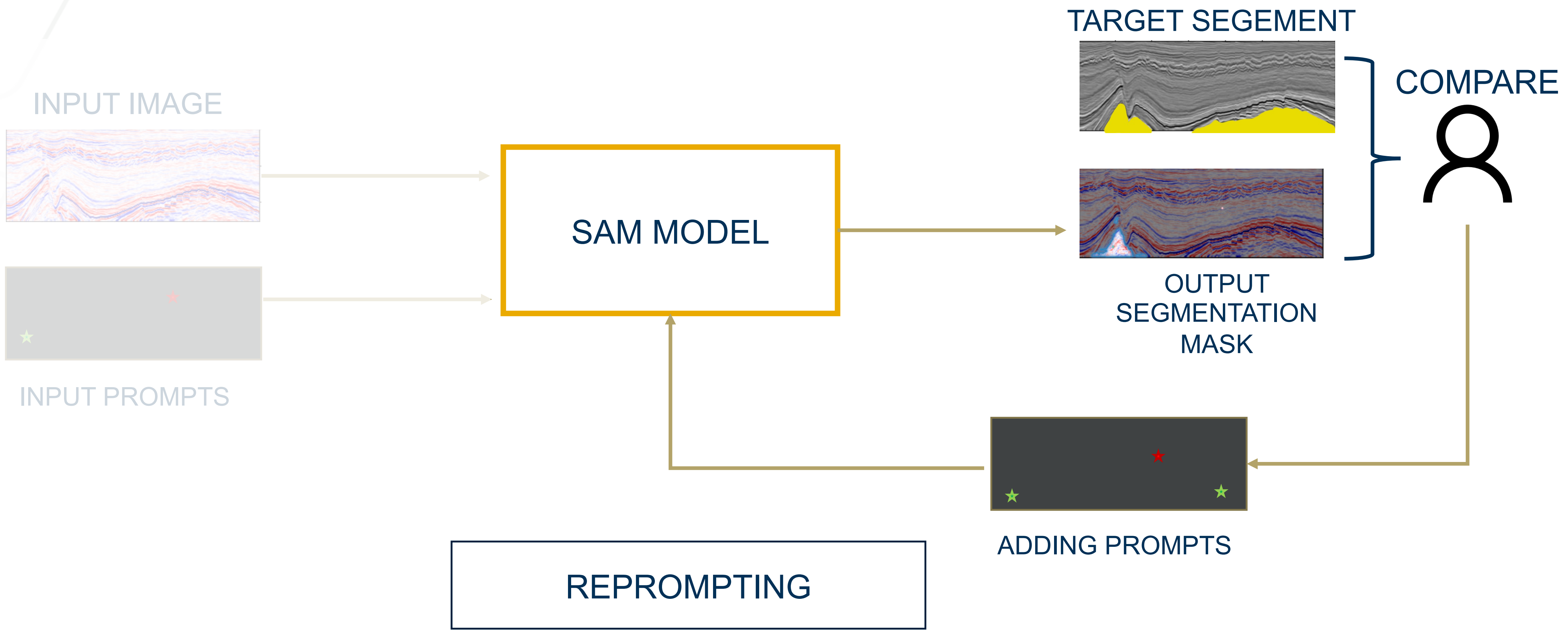
Re-prompting allows to finetune the response without re-training the model.



SAM – Inferencing (Out of the Box)

Re-prompting: Adding or removing prompts based on the model response.

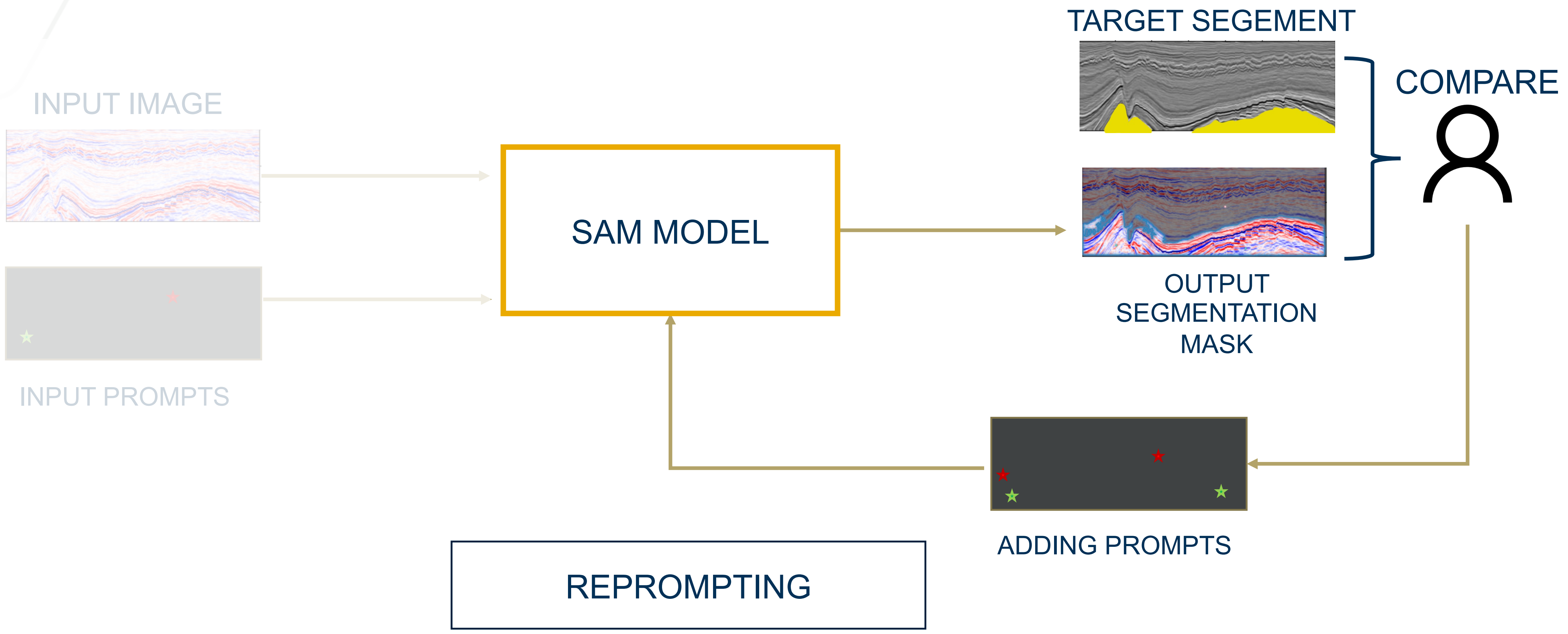
Re-prompting allows to finetune the response without re-training the model.



SAM – Inferencing (Out of the Box)

Re-prompting: Adding or removing prompts based on the model response.

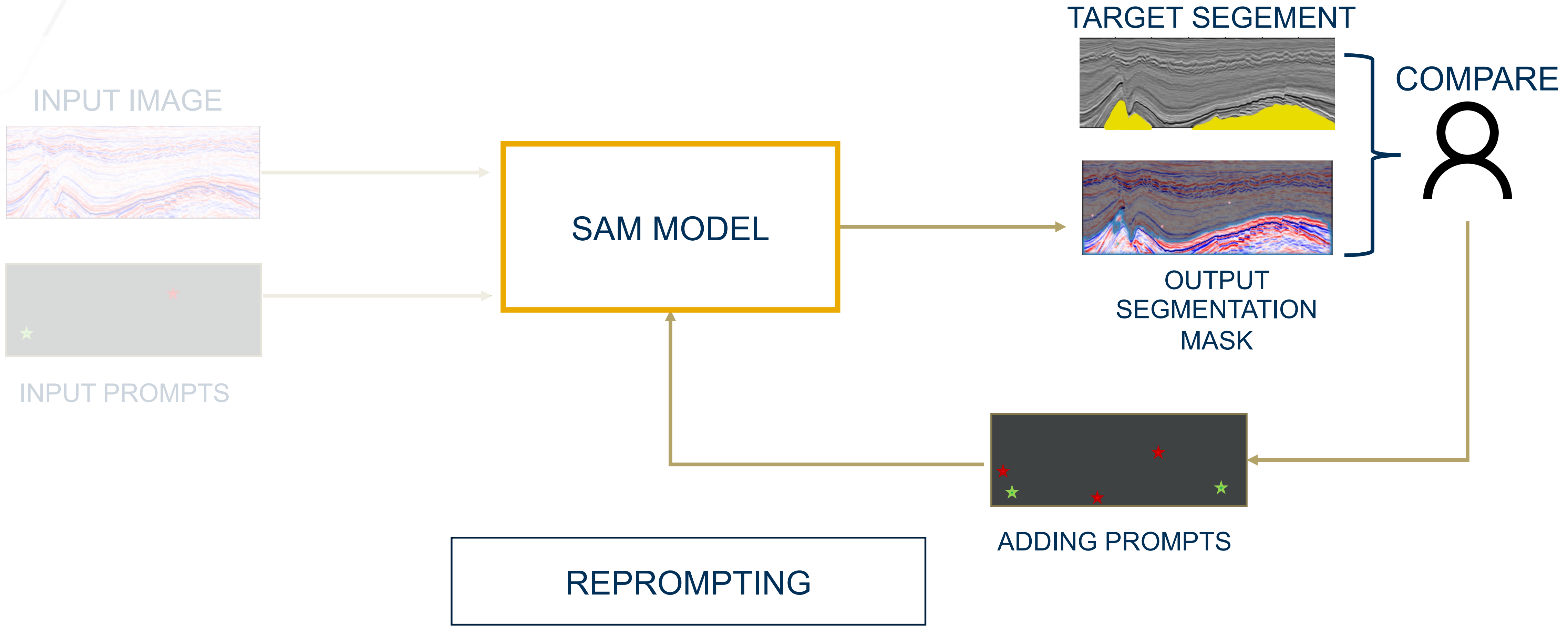
Re-prompting allows to finetune the response without re-training the model.



SAM – Inferencing (Out of the Box)

Re-prompting: Adding or removing prompts based on the model response.

Re-prompting allows to finetune the response without re-training the model.



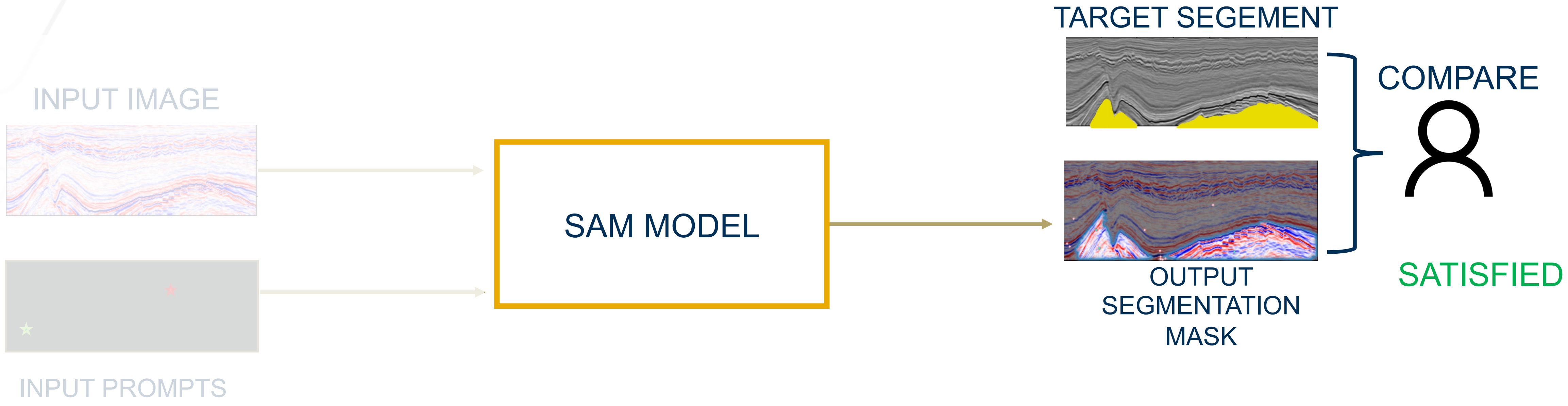
OLIVES @GeorgiaTech

Georgia Tech

SAM – Inferencing (Out of the Box)

Re-prompting: Adding or removing prompts based on the model response.

Re-prompting allows to finetune the response without re-training the model.

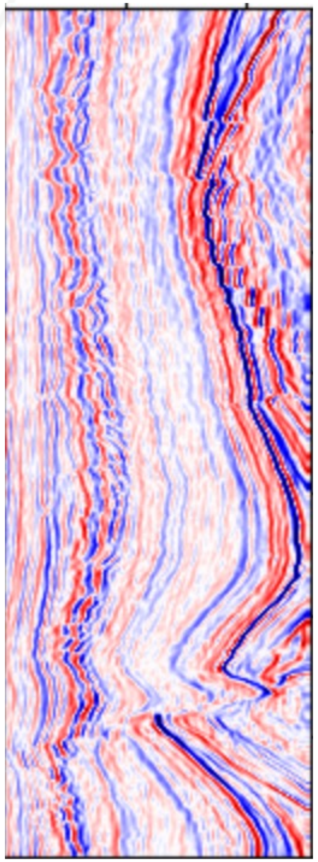


REPROMPTING

SAM – Inferencing (Out of the Box)

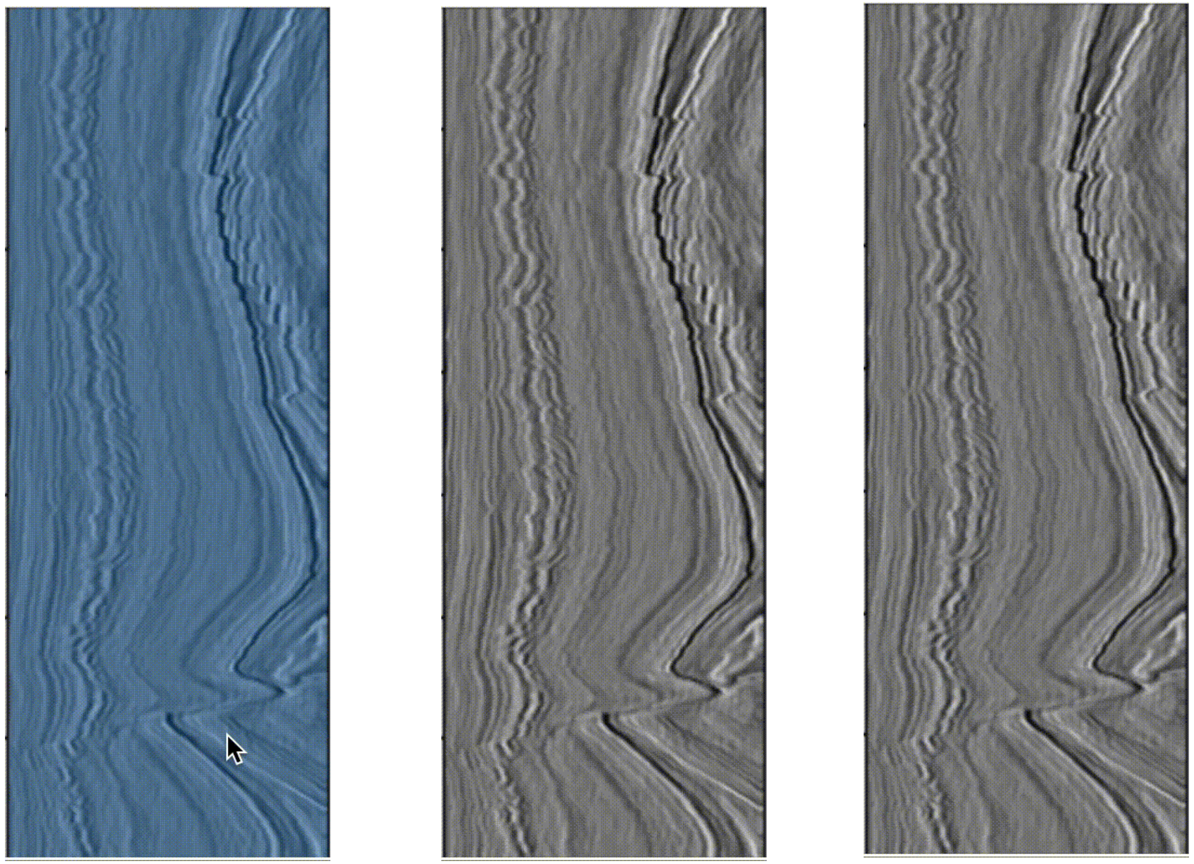
Re-prompting: Adding or removing prompts based on the model response.

Re-prompting allows to finetune the response without re-training the model.



Image

Ground truth

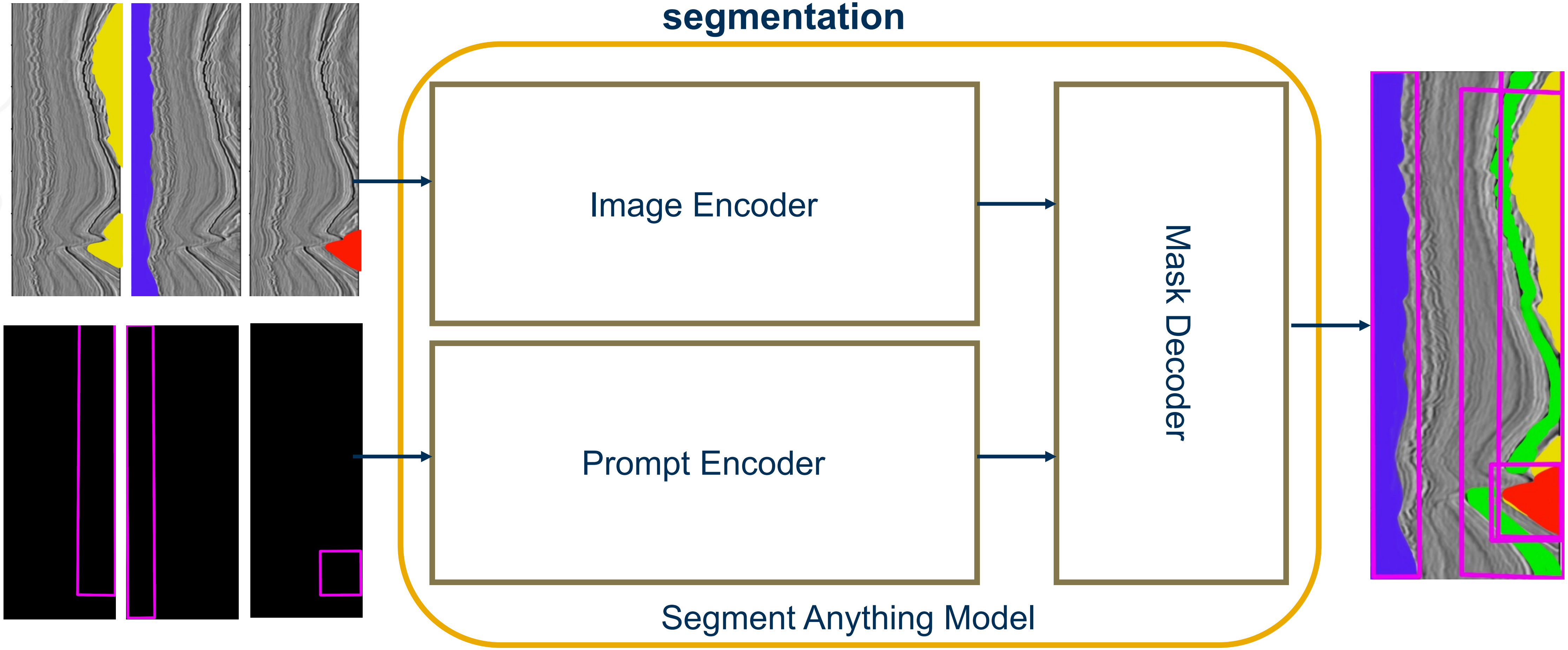


Some examples of re-prompting for various target segmentation

SAM – Encoder Finetuning

The Vision Transformer of SAM can be finetuned like any other transformer based finetuning

Since SAM is class-agnostic all labels must be sent as individual masks for facies segmentation

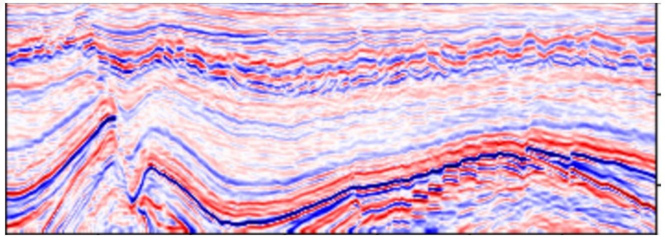


SAM – Encoder Finetuning

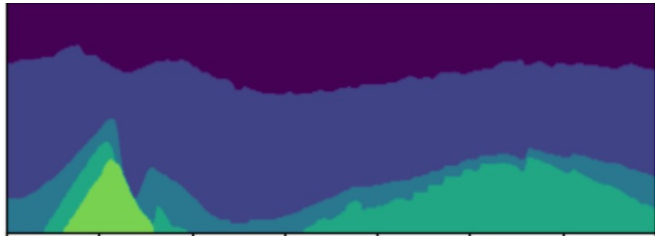
Fine tuning SAM can help reduce the cost of prompting

Both inclusion and exclusion points halved from base-SAM to finetuned-SAM

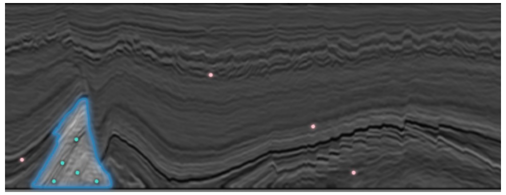
Seismic Image



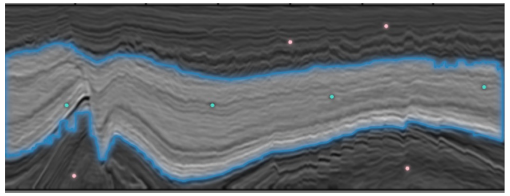
Ground Truth



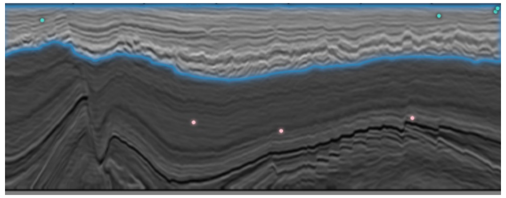
Point prompts: 9



Point prompts: 8

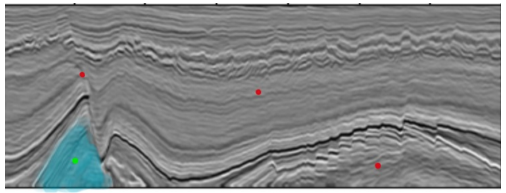


Point prompts: 7

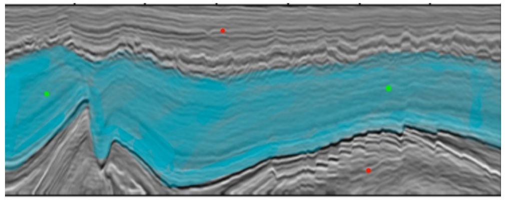


Base SAM

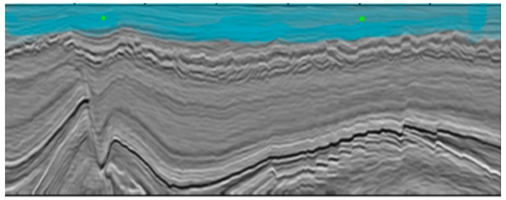
Point prompts: 4



Point prompts: 4



Point prompts: 2



finetuned SAM

However, the cost of finetuning such a large model severely outweighs the benefits

Conclusion

- **Large Foundation Models**, claiming to tackle the problem of zero-shot learning, **fails miserably in generalizing and adapting seismic tasks.**
    - Finetuning can help. But is too costly.
- **Prompting:**
    - It can **help generate more informative or accurate inferences** without need to finetune the model.
    - Several types and strategies of prompting (like re-prompting) opens the scope for various type of experimental designs and case studies of these large models.

For more OLIVES content, please visit:

GitHub

SCAN ME

Publications

SCAN ME