SAM
STREET HOLE

Henrique Borges
Mariana
Paula

SAM

A inteligência artificial mais avançada para detectar, editar e experimentar com imagens e vídeos.

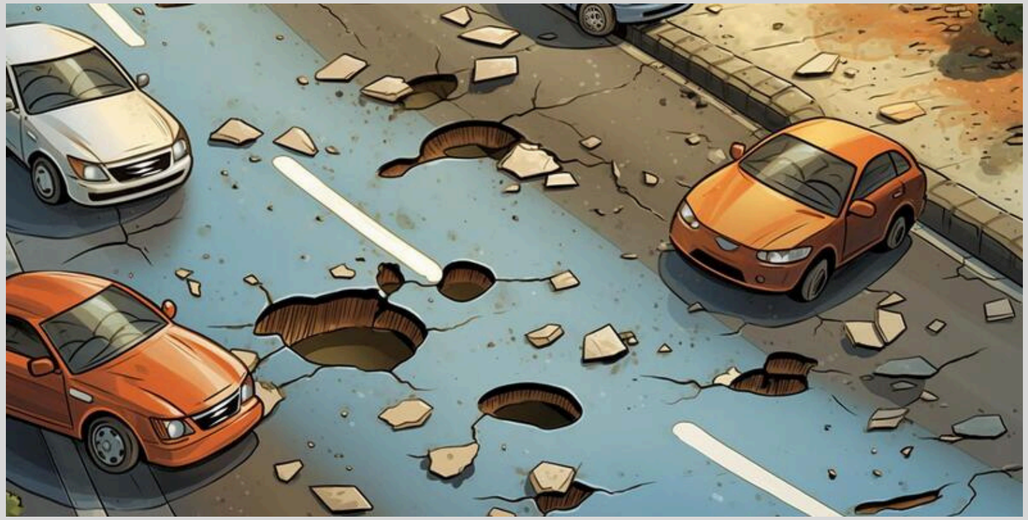
- Modelo de segmentação genérico.
- Produz máscaras automáticas a partir de imagens.
- Usa prompts simples: ponto, caixa ou nenhum.
- Treinado em grande volume de dados variados.
- Aplica o mesmo método em diferentes tipos de cena.

SAM

- Encoder de imagem (ViT) extrai embeddings.
- Prompts (pontos/caixas/texto) são convertidos em embeddings de prompt.
- Um decodificador leve combina embeddings da imagem + embeddings do prompt.
- O decodificador gera a máscara final.

kaggle

POTHOLE IMAGE SEGMENTATION DATASET



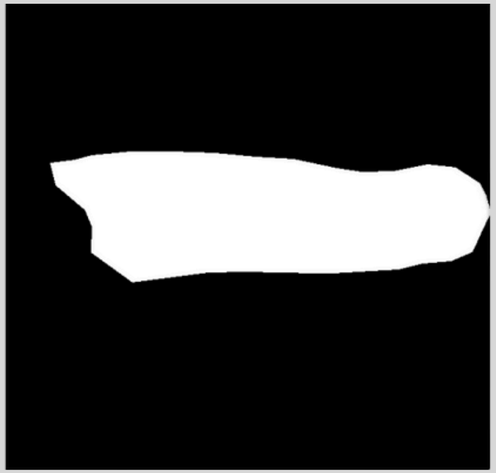
IMPORTÂNCIA DA DETECÇÃO DE BURACOS NA PAVIMENTAÇÃO:

Detecção de buracos na pavimentação é crucial pois contribui para a segurança, manutenção e eficiência nas vias públicas e estradas. Identificando-os e prontamente os consertando, nós podemos prevenir acidentes, economizar em grandes reparos,e garantir um melhor fluxo de trânsito.

DETALHES DO DATASET

Esse dataset foi disponibilizado via roboflow.com e contém no total 780 imagens anotadas no formato YOLOv8 para a detecção e segmentação de buracos na pavimentação. As imagens foram submetidas a pré-processamento e augmentation para garantir que estão consitentes para treinar modelos robustos

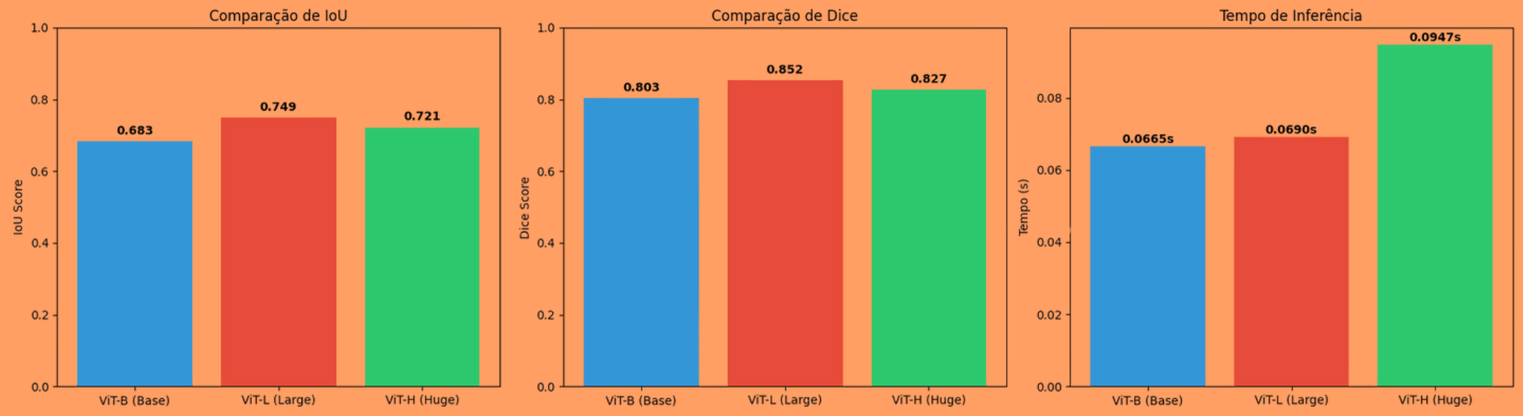
IMAGENS EXEMPLOS



MODELOS PRÉ-TREINADOS

	IOU	DICE	TIME
VIT-B (BASE)	0.6831	0.8032	0.0665
VIT-L (LARGE)	0.7489	0.8520	0.0690
VIT-H (HUGE)	0.7214	0.8270	0.0947



1 **VIT-B (BASE)**

- Porte pequeno
- 12 camadas de Transformer, 768 de dimensão
- Mais rápido e com menor custo

2 **VIT-L (LARGE)**

- Médio porte.
- 24 camadas Transformer, 1024 de dimensão.
- Custo maior, melhores performances

3 **VIT-H (HUGE)**

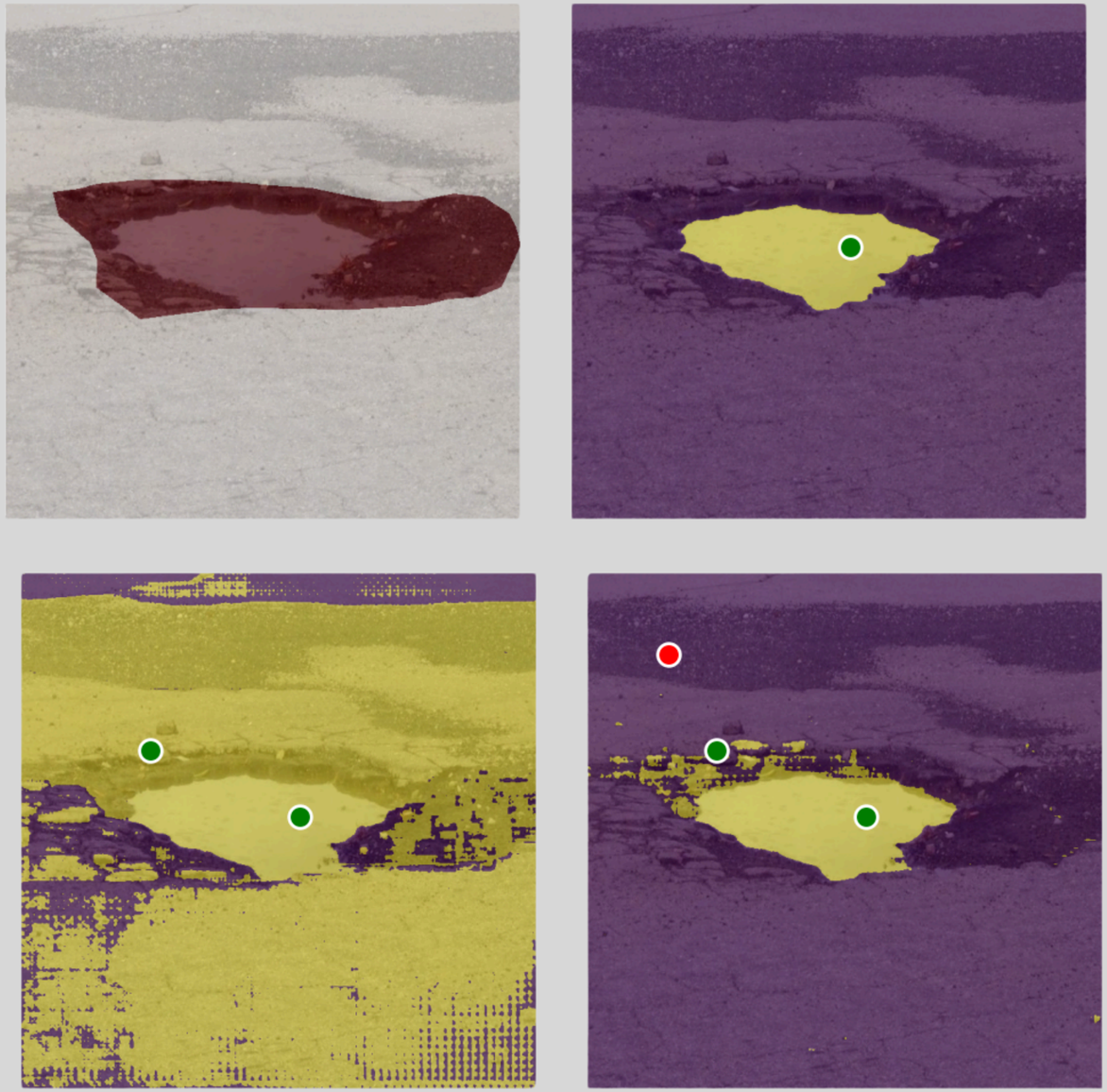
- Maior.
- 32 camadas Transformer, 1280 de dimensão.
- Desempenho superior, uso pesado de memória.

VIT-L

50 AMOSTRAS COM PROMPT=BBOX
IOU: 0.7201 | DICE: 0.8275

50 AMOSTRAS COM PROMPT=POINT
IOU: 0.4918 | DICE: 0.6006

20 AMOSTRAS COM PROMPT=AUTOMATIC
IOU: 0.4151 | DICE: 0.5075

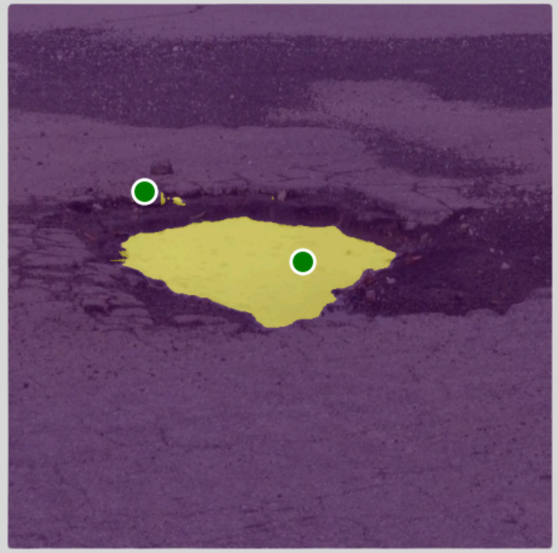


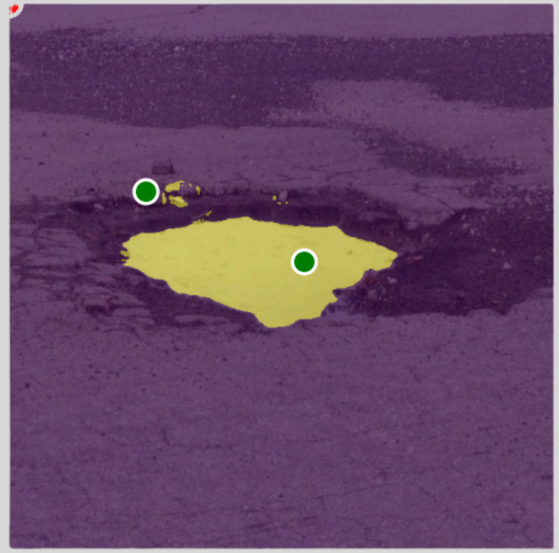
VIT-H

50 AMOSTRAS COM PROMPT=BBOX
IOU: 0.7133 | DICE: 0.8238

50 AMOSTRAS COM PROMPT=POINT
IOU: 0.4661 | DICE: 0.5803

20 AMOSTRAS COM PROMPT=AUTOMATIC
IOU: 0.4878 | DICE: 0.5853

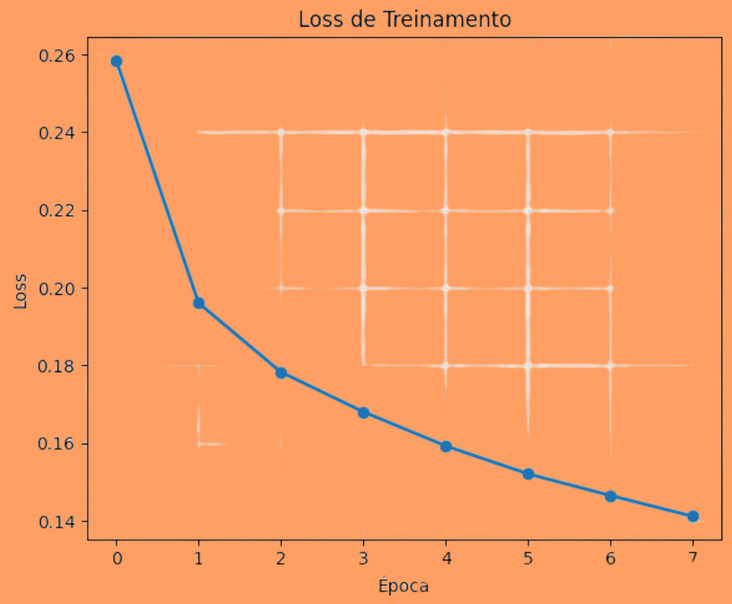


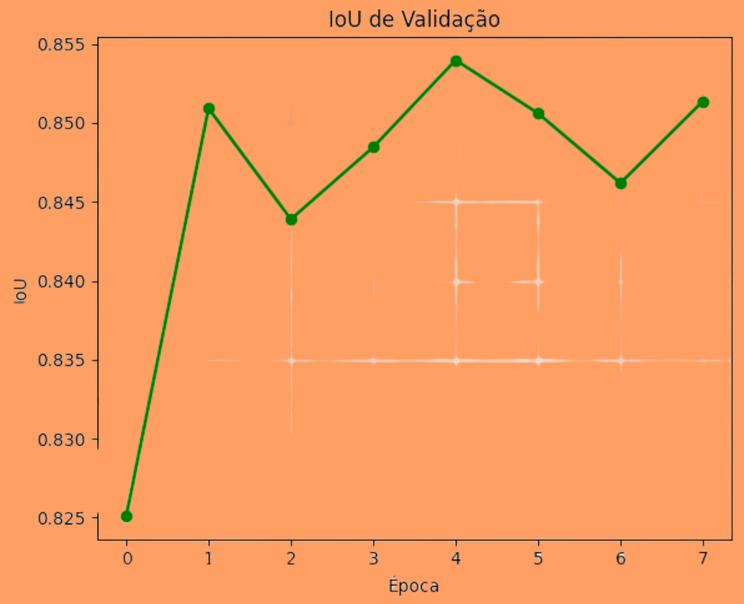


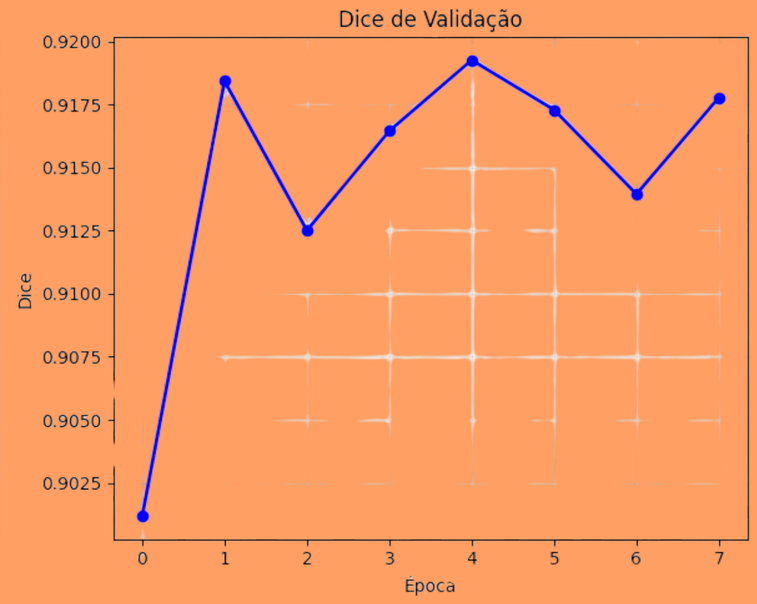
FINE TUNE - SAM

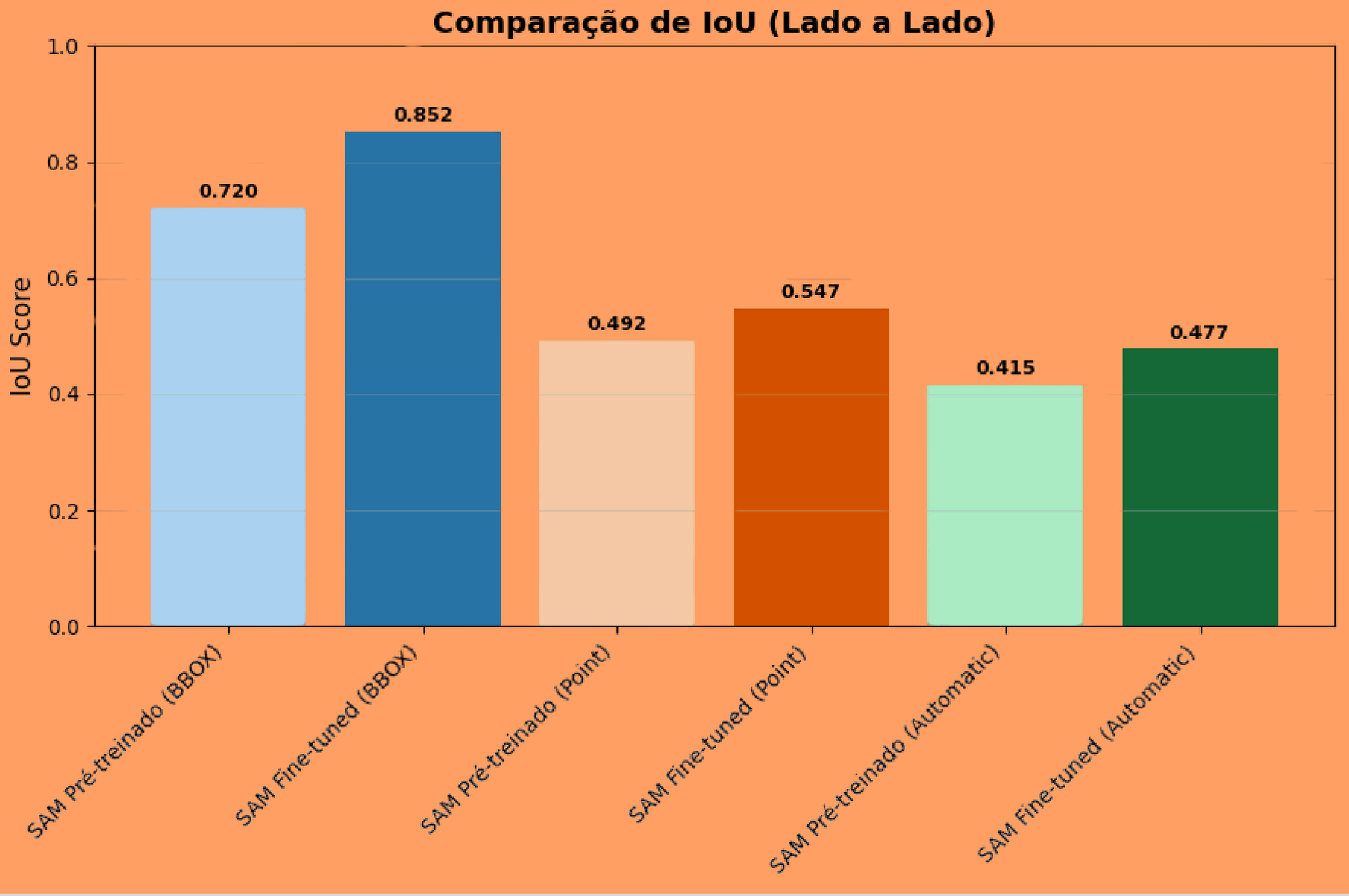
FINE TUNE - SAM

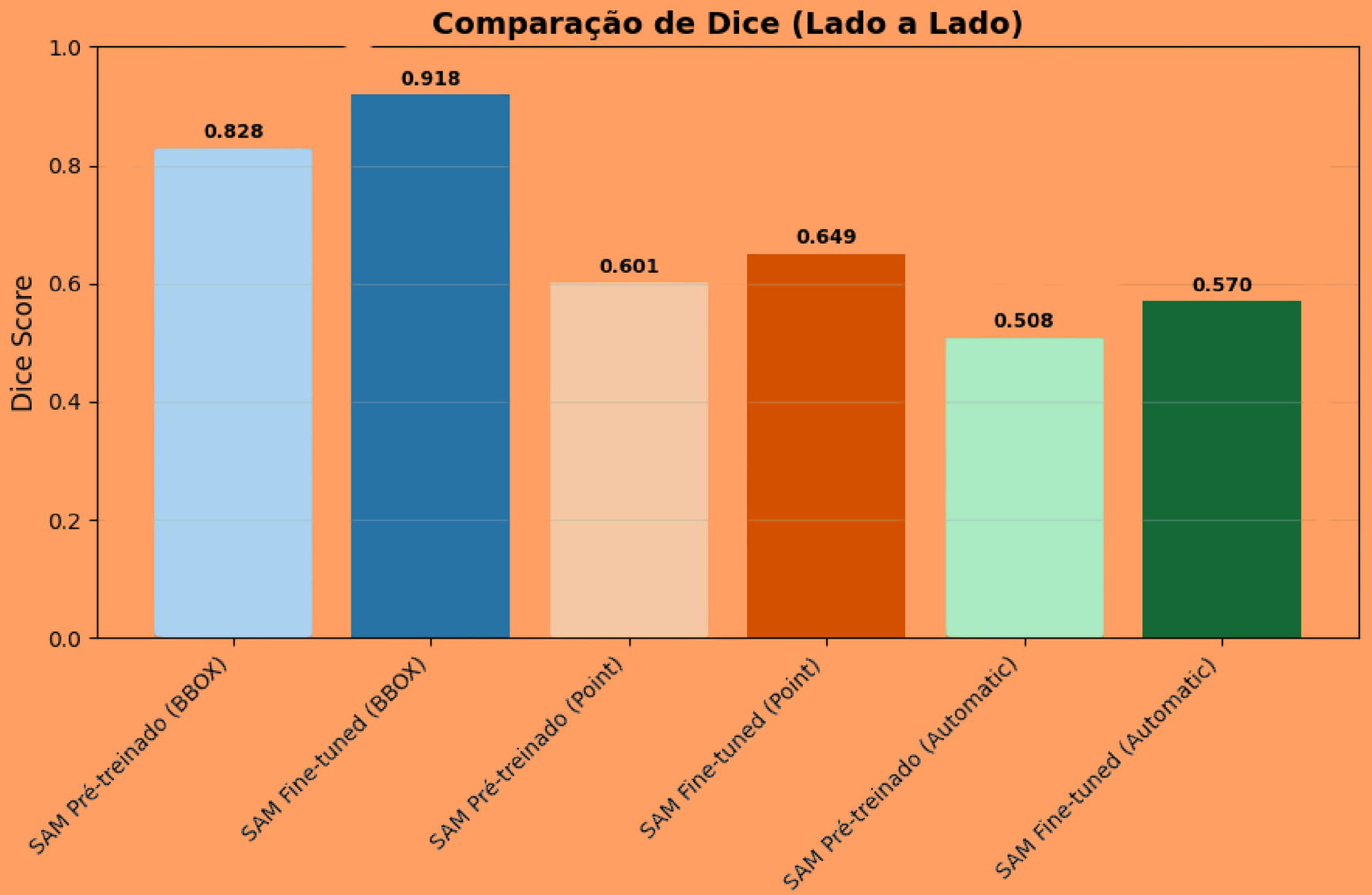
FOI REALIZADO UM FINE TUNING COM 8 EPOCHS VOLTADO PRA ABORDAGEM DE BOUND BOXES





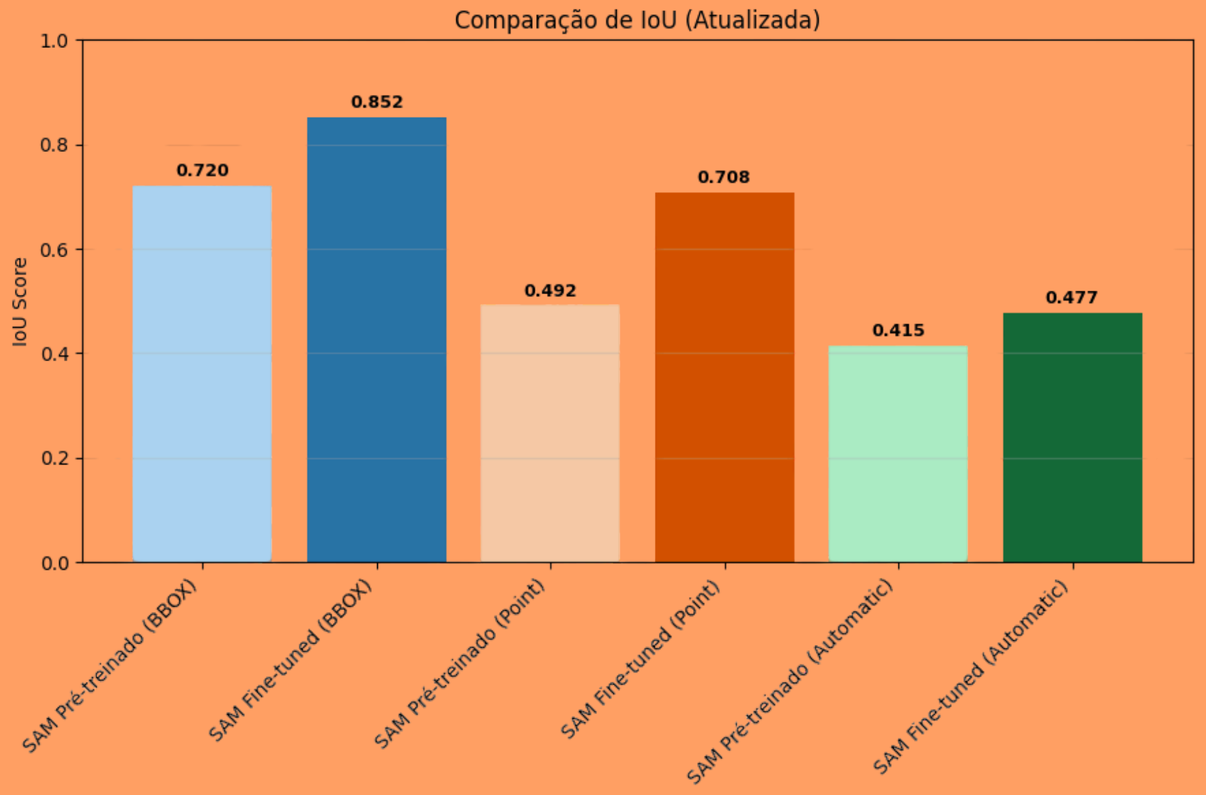


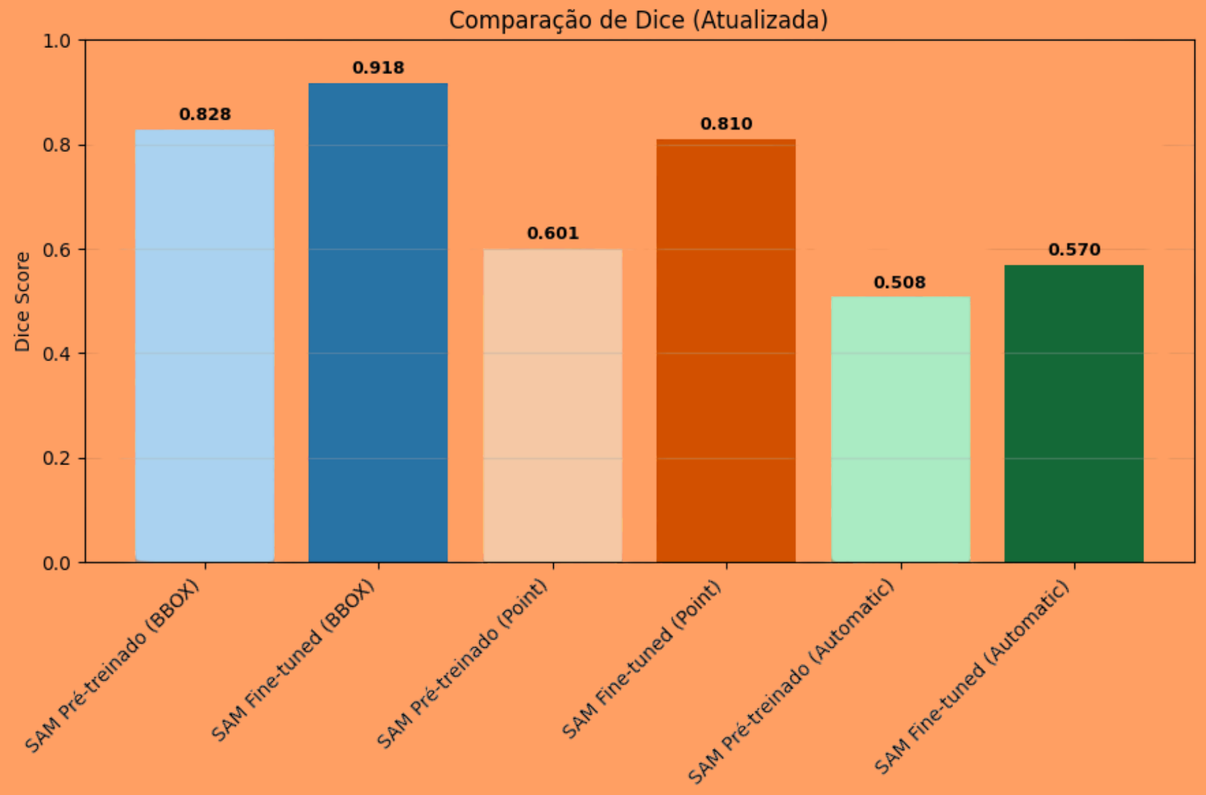

Comparação de IoU (Lado a Lado)
IoU Score
1.0
0.8
0.6
0.4
0.2
0.0
0.720
0.852
0.492
0.547
0.415
0.477
SAM Pré-treinado (BBOX)
SAM Fine-tuned (BBOX)
SAM Pré-treinado (Point)
SAM Fine-tuned (Point)
SAM Pré-treinado (Automatic)
SAM Fine-tuned (Automatic)


Comparação de Dice (Lado a Lado)
Dice Score
1.0
0.8
0.6
0.4
0.2
0.0
0.828
0.918
0.601
0.649
0.508
0.570
SAM Pré-treinado (BBOX)
SAM Fine-tuned (BBOX)
SAM Pré-treinado (Point)
SAM Fine-tuned (Point)
SAM Pré-treinado (Automatic)
SAM Fine-tuned (Automatic)

FINE TUNE - SAM

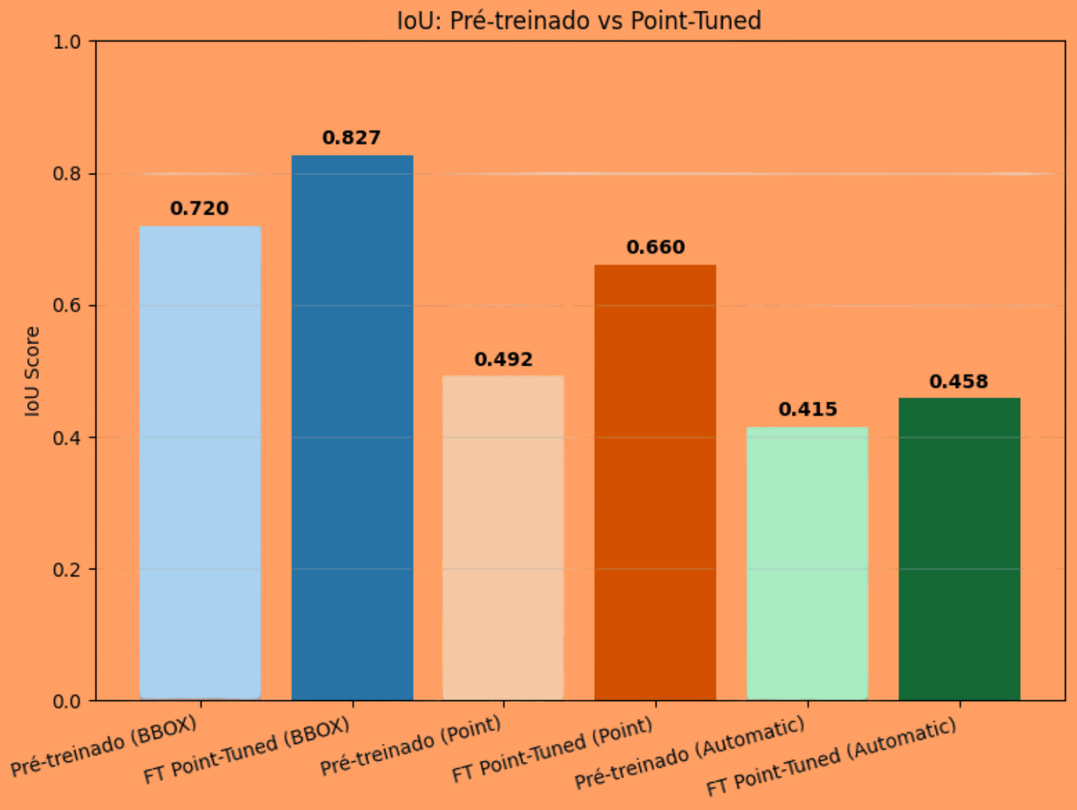
FINE TUNE DE 5 EPOCHS ADICIONAL PARA SEGMENTAÇÃO COM POINT COMO ENTRADA

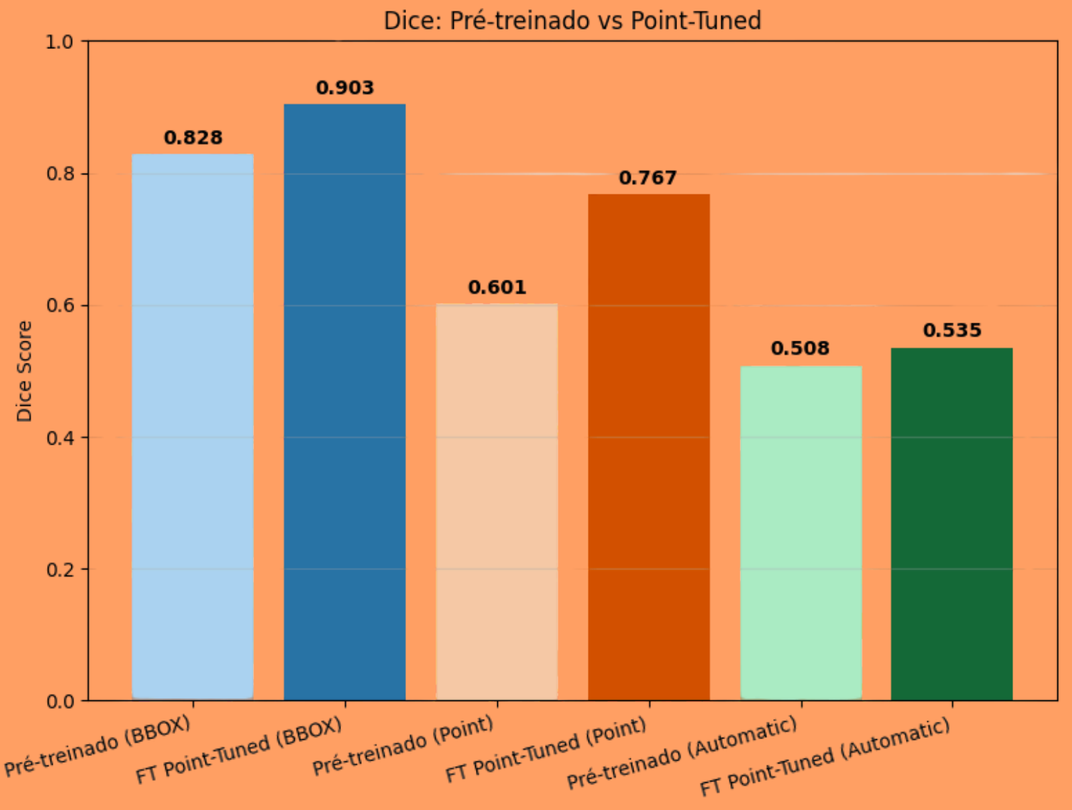




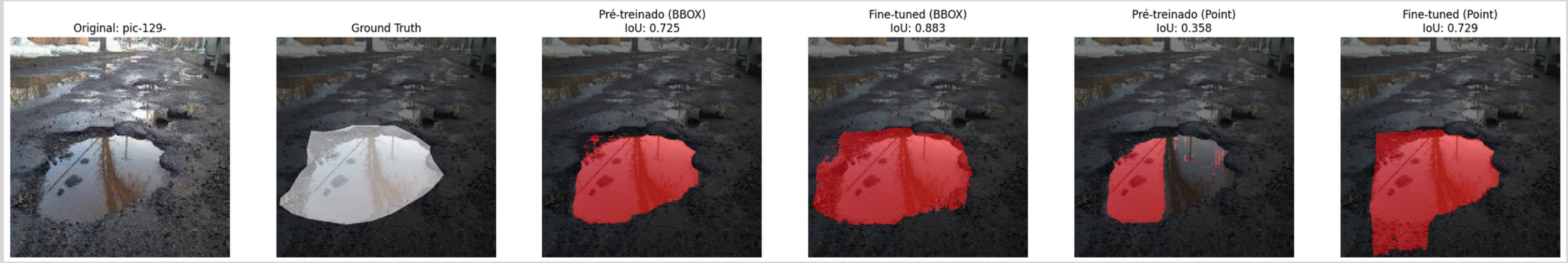
FINE TUNE - SAM

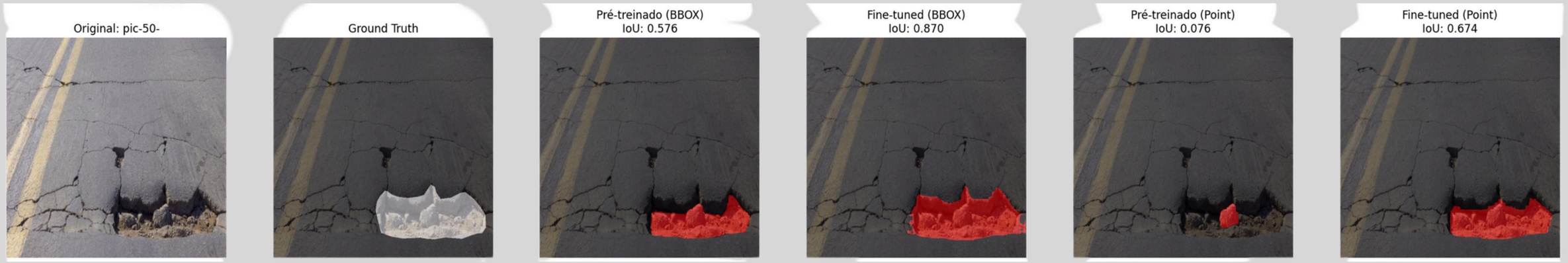
FINE TUNE DE 8 EPOCHS PARA SEGMENTAÇÃO COM POINT COMO ENTRADA





FINE TUNE - SAM





GROUNDED-SAM: SEGMENTAÇÃO COM TEXT PROMPTS

=== TESTANDO PROMPT: 'POTHOLE' ===
AVALIANDO GROUNDED-SAM COM TEXTO: 'POTHOLE'
IOU: 0.4908 | DICE: 0.5806

=== TESTANDO PROMPT: 'HOLE' ===
AVALIANDO GROUNDED-SAM COM TEXTO: 'HOLE'
IOU: 0.4170 | DICE: 0.4976

=== TESTANDO PROMPT: 'ROAD DAMAGE' ===
AVALIANDO GROUNDED-SAM COM TEXTO: 'ROAD DAMAGE'
IOU: 0.1444 | DICE: 0.2181

OWL-VIT + SAM

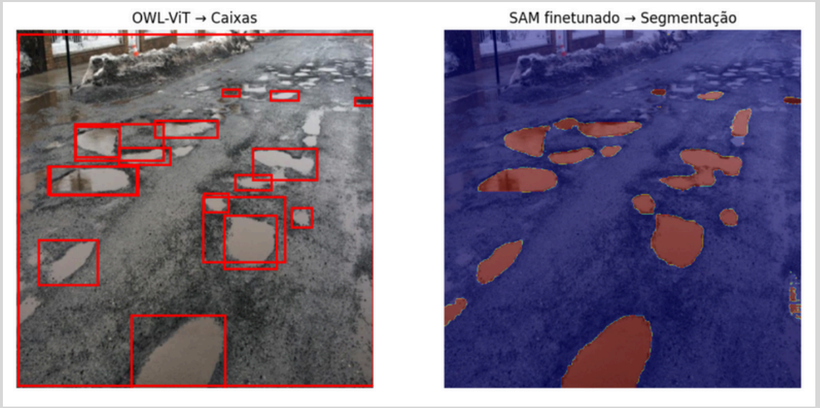
SAM não processa prompts textuais. Ele foi projetado para segmentação baseada em prompts visuais. Pensando nisso, unimos ele a um outro modelo que consegue entender prompts e retornar boundingbox.

OWL-ViT (Open-World Localization Vision Transformer) é um modelo de detecção a partir de prompts textuais ou classes arbitrárias, mesmo que o modelo não tenha visto essas classes durante o treinamento.
Saída:

- Bounding boxes para os objetos correspondentes ao texto.
- Scores de confiança para cada detecção.

Pipeline comum SAM + OWL-ViT

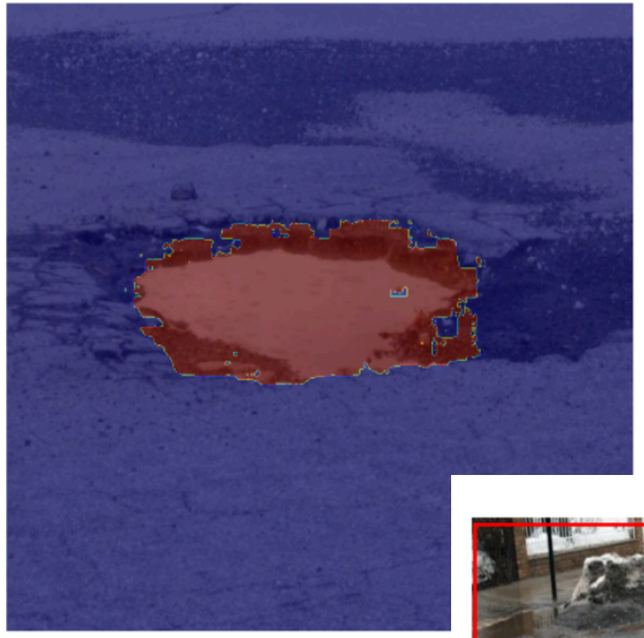
1. Usuário fornece texto: "buraco na rua".
2. OWL-ViT retorna caixas onde provavelmente há buracos.
3. Cada caixa é passada como prompt visual para o SAM.
4. SAM gera máscara segmentada de alta precisão para cada objeto detectado.



OWL-VIT + SAM

OWL-ViT → Caixas detectadas

SAM finetunado (segmentação final)



OWL-ViT → Caixas

SAM finetunado → Segmentação

Segmentação com VIT-H com fine tune

Segmentação com DINO + SAM

Demo oficial SAM

OBRIGADO !