

FINE TUNING - SAM

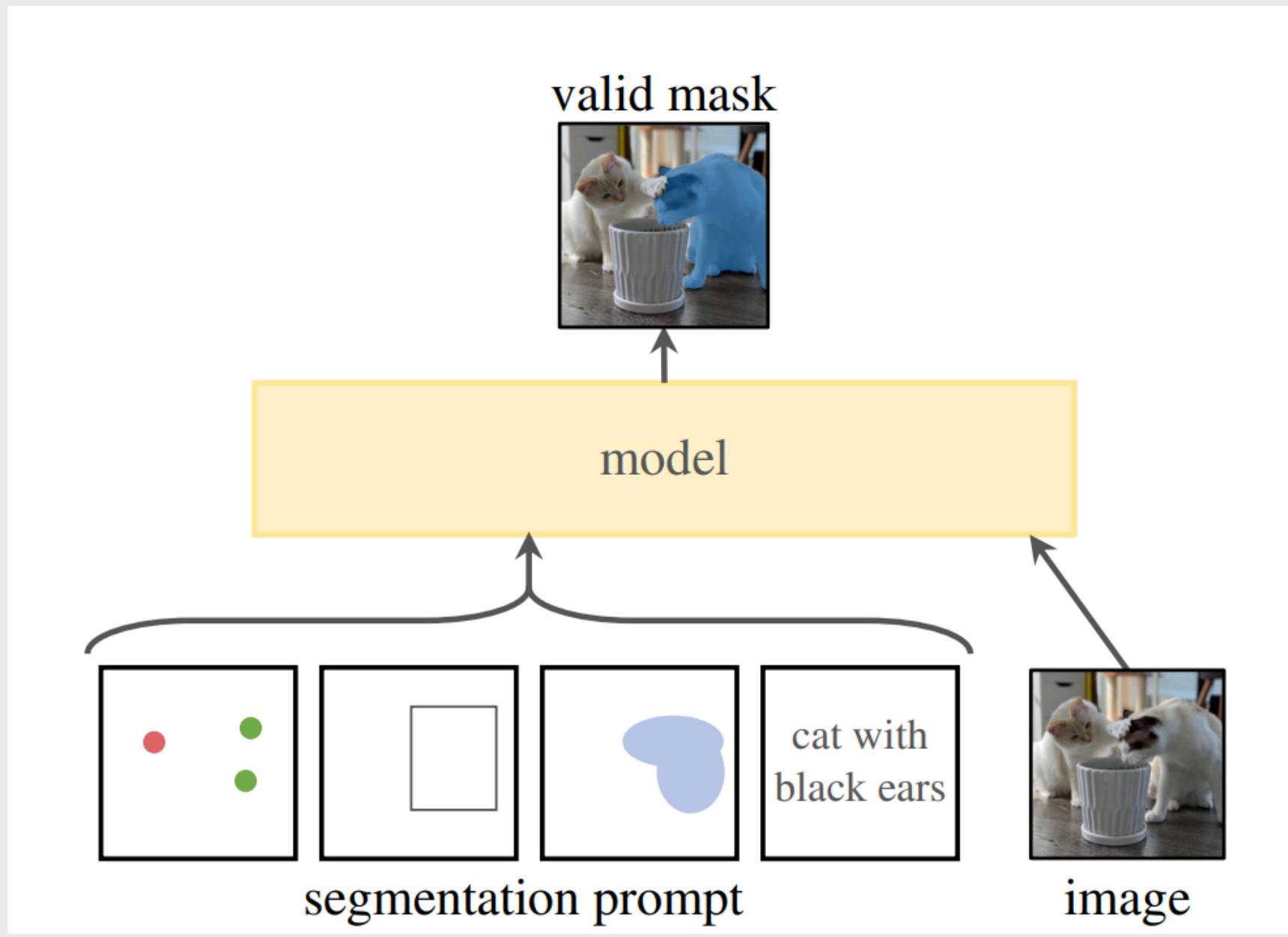
Anderson Falcão
Leonardo Alexandre
Ramyro Corrêa

SEGMENT ANYTHING MODEL

O SAM é um Foundation Model da Meta AI para segmentação de imagens.

- Zero-Shot: Capaz de segmentar objetos nunca vistos.
- Promptable: Aceita pontos, caixas (bounding boxes) ou texto como entrada para guiar a segmentação.
- Treinado com mais de 10 milhões de imagens e mais de 1 bilhão de máscaras.

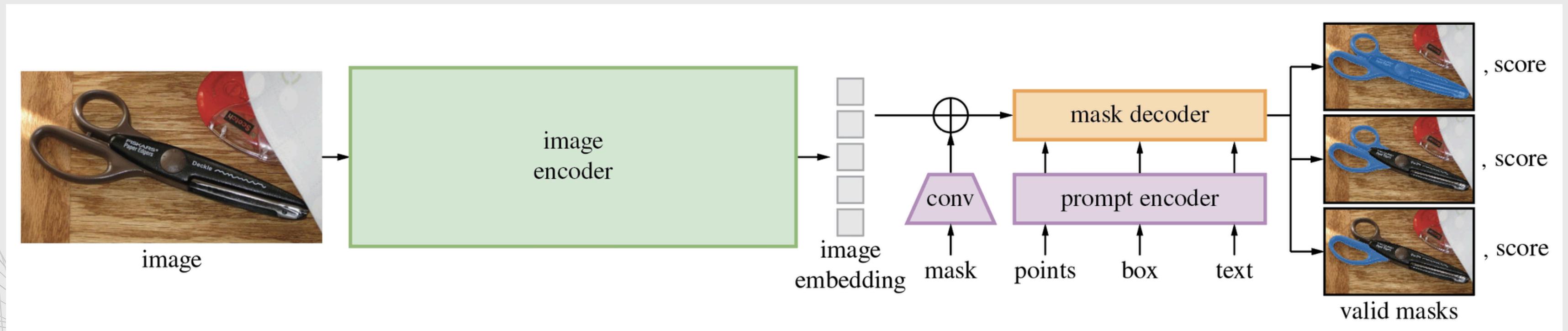




02

ARQUITETURA

- Image Encoder: Vision Transformer (ViT-Base).
- Prompt Encoder: Codifica a Bounding Box.
- Mask Decoder: Gera a máscara final.

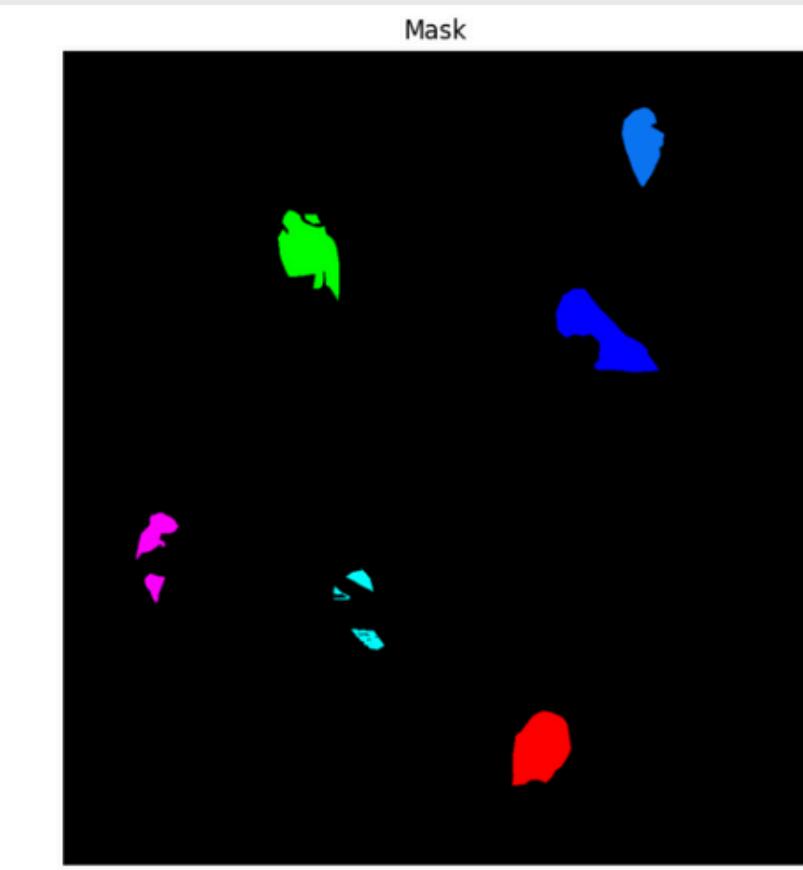
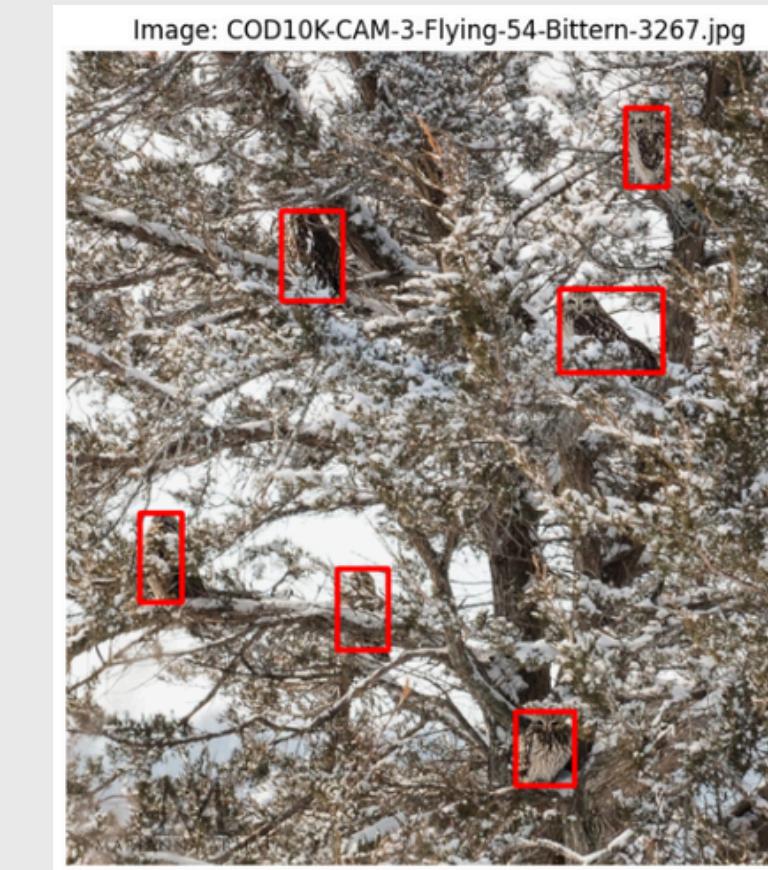


TAREFA: DETECAÇÃO DE CAMUFLAGEM

O Desafio: Segmentar objetos camuflados.

Dataset: Camouflaged/Concealed Object Detection (**COD10K-V3**)

- Foco: Objetos Camuflados (Camouflaged Object Detection).
- Máscaras de segmentação de instância, bounding boxes
- Diversidade de objetos e contextos.



TREINAMENTO

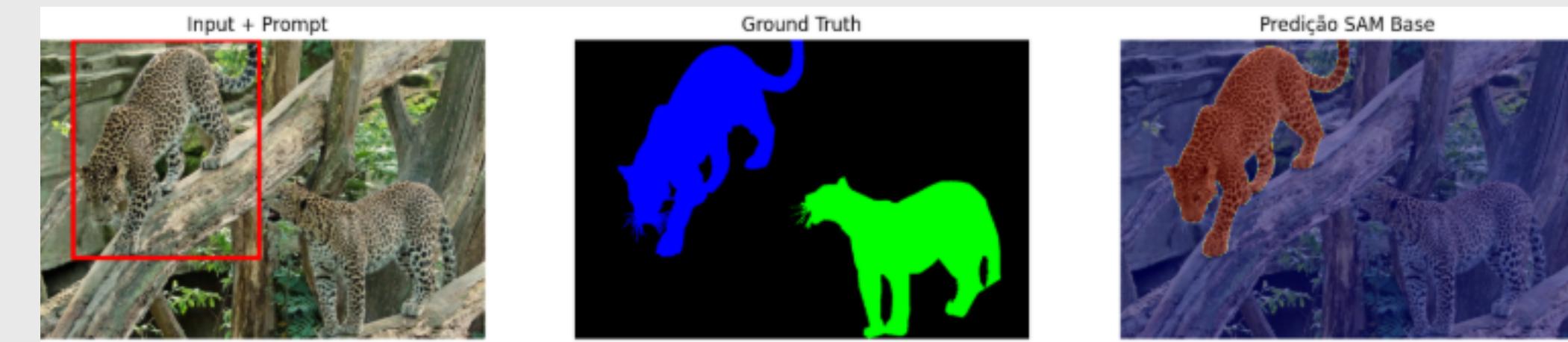
Para adaptar o SAM à tarefa de camuflagem, realizamos o fine-tuning apenas do decoder.

Configuração

- Pesos Congelados: Image Encoder e Prompt Encoder.
- Pesos Treináveis: Mask Decoder.
- Otimizador: AdamW ($\text{lr}=1\text{e-}4$).
- Loss Function: Combinação de Dice Loss e Cross Entropy.

Processo

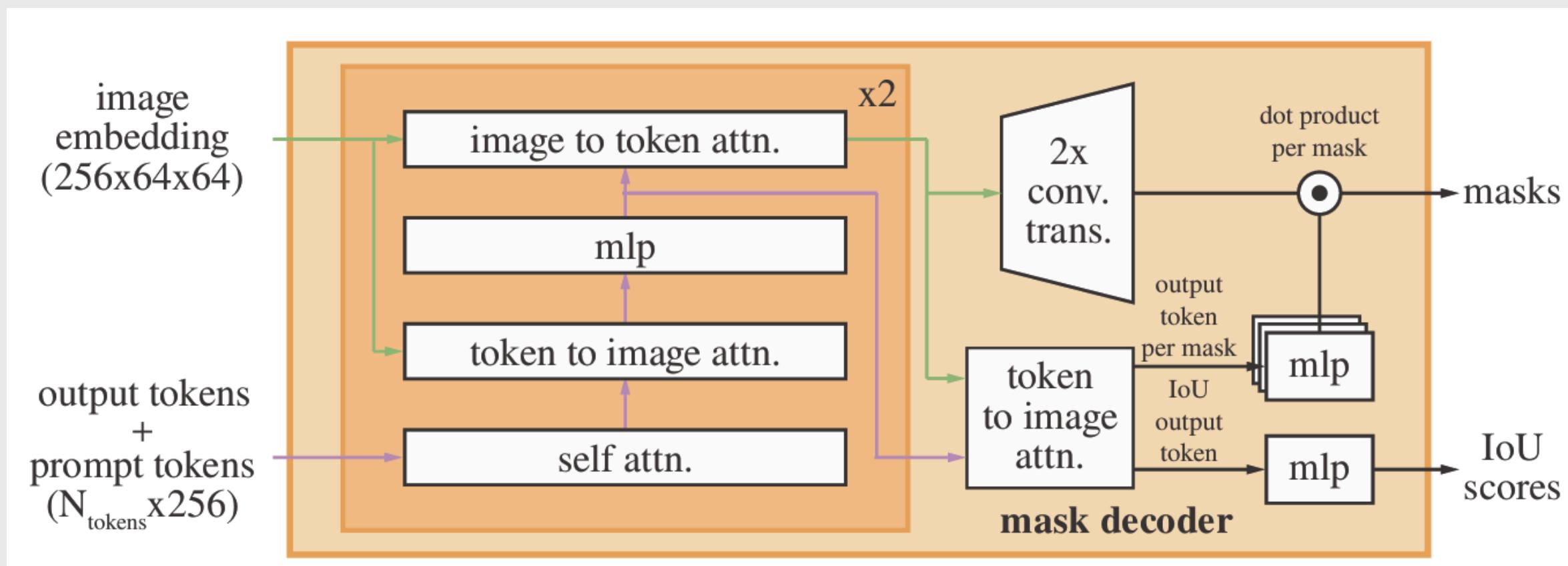
- Input: Imagem Original + Bounding Box (Ground Truth).
- Target: Máscara Binária.
- Épocas: 15.



Total de Parâmetros: 93,735,472

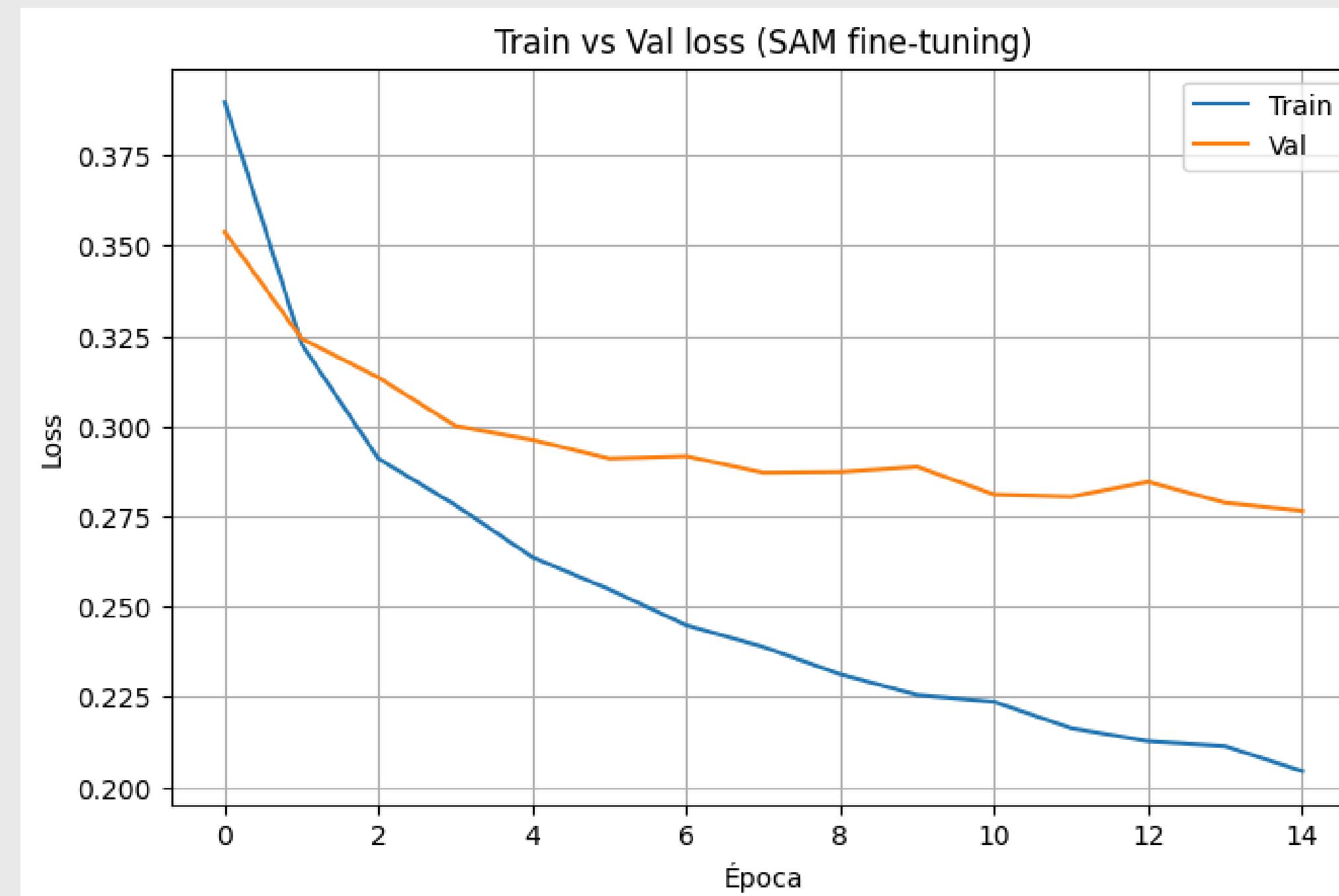
Parâmetros Treináveis (Mask Decoder): 4,058,340

Percentual Treinável: 4.33%



EVOLUÇÃO DO TREINAMENTO

O modelo demonstrou convergência rápida, reduzindo a loss tanto no treino quanto na validação.



RESULTADOS

SAM BASE:

==== RESULTADOS (SAM PRÉ-TREINADO) ====

Test loss médio: 0.6270

Test IoU médio: 0.1392

Test Dice médio: 0.2082

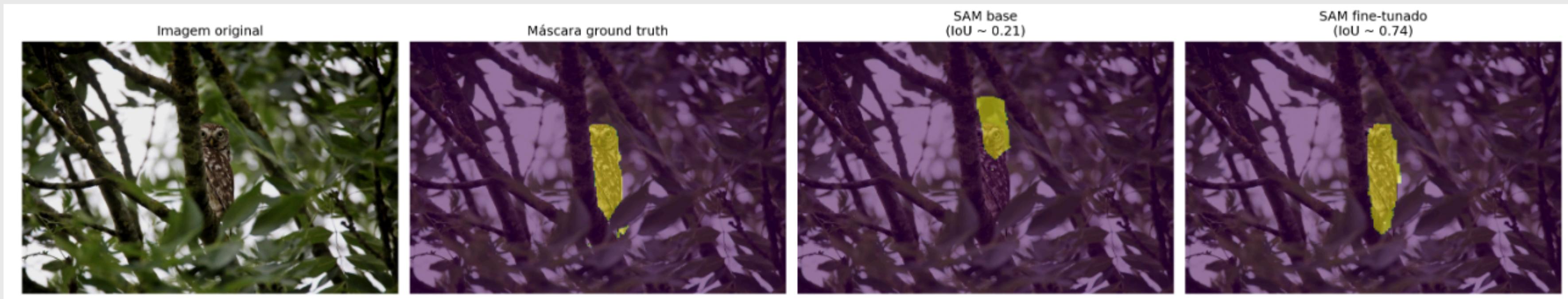
SAM COM FINE TUNING:

==== RESULTADOS NO TESTE =====

Test loss médio: 0.2594

Test IoU médio: 0.5112

Test Dice médio: 0.6466



RESULTADOS



CONCLUSÃO E PRÓXIMOS PASSOS

Conclusão

- O fine-tuning do decoder foi suficiente para adaptar o SAM ao problema de camuflagem.
- O uso de Bounding Boxes como prompt garantiu alta precisão na localização do objeto.
- A loss de validação caiu de ~0.35 para ~0.27, indicando aprendizado efetivo.

Próximos passos

- Testar segmentação automática (sem box prompt) para avaliar a capacidade de detecção pura.
- Comparar com modelos específicos de camuflagem (ex: SInet).

Obrigado!